UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Solvatochromic dyes for the discovery and manipulation of Escherichia coli protein-protein interactions

Permalink

https://escholarship.org/uc/item/67m2d0h4

Author

Charov, Katherine

Publication Date

2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Solvatochromic dyes for the discovery and manipulation of *Escherichia coli* protein-protein interactions

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

in

Chemistry

By

Katherine Charov

Committee in charge:

Professor Michael D. Burkart, Chair
Professor Kamil Godula
Professor Joseph O'Connor
Professor Stanley Opella
Professor Larissa Podust

2020

This dissertation of Katherine Charov is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

To my teachers:

Susan Kenny

Julia Sheldon

Roger Cantello

Tony Davis

Nik Charov

Alex Charov

And last, but certainly not least,

Helen Charov, my wonderful mother

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Fatty acid synthase ..... FAS
Polyketide synthase ..... PKS
Non-ribosomal peptide synthetase ..... NRPS
Carrier protein ..... CP
Acyl carrier protein ..... ACP
E. coli fatty acid synthase acyl carrier protein ..... EcACP
4’ Phosphopantetheinyl transferase ..... PPTase
Surfactin phosphopantetheinyl transferase ..... Sfp
Acyl carrier protein hydrolase ..... AcpH
Holo-ACP synthase ..... AcpS
Coenzyme A ..... CoA
E. coli Pantothenate Kinase ..... CoaA
E. coli Phosphopantetheine adenylyl transferase ..... CoaD
E. coli Dephosphocoenzyme A kinase ..... CoaE
Adenosine triphosphate ..... ATP
Protein-protein interaction..... PPI
4-dimethylnaphthalene..... 4-DMN

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank the Burkart lab members for making my time in graduate school a little less painful than it otherwise might have been. Mike for being an understanding and approachable advisor, who very generously let me take a few months off to do an internship. Jim La Clair for the coffee runs, company during late night microscopy sessions and general advice. Tony Davis the post-doc mentor who set the extremely high bar for mentorship by always being available as a sounding board, advice giver, listener, problem solver, foodie and, overall, a great friend. Warren “Wardog” Chan, my lab brother, for a friendship born from a mutual love of food and dislike of everything else. Jeff Mindrebo for teaching me all the biochemistries when I was a first and second year but, more importantly, for teaching me that having a dance party in the lab during working hours is not a good way to be productive but is a good way to keep the stoke high.

To my family: Ma, you made this all possible by setting me up for success in high school and college. It’s been a while since I had to call you in tears to help me write a history paper, and I haven’t gotten any care packages with dried peaches in a while, but I know that support is always there. My older broski, Nik, who is partly responsible for my penchant for pranks and mostly responsible for me passing high school physics. My Pa, who told me that I was “becoming an interesting person” just as I started liking science but then passed away before we could really start talking about the cool stuff; I know you’d be proud. Mikey Betts: my not-really-related-to-me-but-still-counts Uncle who never stopped gifting me Legos and toys that required batteries and building and someone who always talked to me like an adult when it came to science and math. Uncle Gene, my

actual uncle, who used to take us kids to the coast of Maine to look at the rings of Saturn or the Moon up close and who also never stopped treating me like a scientist, even when we kicked up too much dust into your fancy telescope and ruined your night vision with our flashlights. And of course, the rest of my Labor Day family.

My wonderful partner for life (probably???), Michael, deserves his own paragraph for being a) wonderful, b) handsome, c) caring and d) extremely patient with me while I learned how to code in Python. This last point deserves its own reward. You brighten every room you enter; you make everyone you meet feel special and welcome; you're the most wonderful person I know (tied with my Ma). Your love and support throughout the years has made it all possible, whether we're getting Thai food takeout for the third time in one week, or you learning to make delicious coffee (sorry for getting you addicted to caffeine), or just giving me a nice warm hug. You make my jokes better, and while I'll never admit it, you're pretty funny.

Chapter 2, in full, is a reprint of the material as it appears: Charov, Katherine; Burkart, Michael D. "A Single Tool to Monitor Multiple Protein-Protein Interactions of the *Escherichia coli* Acyl Carrier Protein." *ACS Infectious Diseases* 5, (9) 2019: 1518-1523. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication and is currently under revision after peer review of the material as it may appear: Charov, Katherine; Burkart, Michael D. "*In silico* identification and *in vitro* evaluation of a protein-protein interaction inhibitor of *Escherichia coli* fatty acid biosynthesis." *Chemical Biology & Drug Design.* The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is currently being prepared for submission for publication as: Charov, Katherine; Burkart, Michael D. “Development of a rapid, fluorescent assay for assessing the effects of partner enzyme surface mutations on the Escherichia coli acyl carrier chain flipping mechanism.” The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, was resubmitted after revision for publication as it appears: Woojoo E. Kim, Katherine Charov, Mária Džunková, Eric D. Becraft, Julia Brown, Frederik Schulz, Tanja Woyke, James J. La Clair, Ramunas Stepanauskas, Michael D. Burkart “Synthase-selective exploration of a tunicate microbiome by activity-guided single-cell genomics.” The dissertation author is a primary co-author of this manuscript along with Dr. Woojoo Eunice Kim, Dr. Mária Džunková, Dr. Eric D. Becraft, and Dr. Julia Brown.

VITA

2020 Doctor of Philosophy, Chemistry, University of California San Diego

2017 Master of Science, University of California San Diego

2014 Bachelor of Arts, Biophysics, Johns Hopkins University

PUBLICATIONS

Woojoo E. Kim*, **Katherine Charov***, Mária Džunková*, Eric D. Becraft*, Julia Brown*, Frederik Schulz, Tanja Woyke, James J. La Clair, Ramunas Stepanauskas, and Michael D. Burkart. "Synthase-selective exploration of a tunicate microbiome by activity-guided single-cell genomics." *Submitted to Nature Chemical Biology.*

Katherine Charov and Michael D. Burkart. "In silico identification and in vitro evaluation of a protein-protein interaction inhibitor of *Escherichia coli* fatty acid biosynthesis." *Submitted to Chemical Biology & Drug Design.*

Katherine Charov and Michael D. Burkart. "Quantifying protein-protein interactions of the acyl carrier protein with solvatochromic probes." *Methods in Enzymology.* 2020 638, 321-340. DOI: https://doi.org/10.1016/bs.mie.2020.03.016

Katherine Charov and Michael D. Burkart. "A Single Tool to Monitor Multiple Protein–Protein Interactions of the Escherichia coli Acyl Carrier Protein." *ACS Infectious Diseases* 2019 5 (9), 1518-1523. DOI: 10.1021/acsinfecdis.9b00150

ABSTRACT OF THE DISSERTATION

Solvatochromic dyes for the discovery and manipulation of

Escherichia coli protein-protein interactions

By

Katherine Charov

Doctor of Philosophy in Chemistry

University of California San Diego, 2020

Professor Michael D. Burkart, Chair

Protein-protein interactions (PPIs) are ubiquitous in Nature and found in all domains of life. They mediate signal transduction, muscle contraction, electron transport and metabolic pathway processing, among many other processes, and the molecular underpinnings of PPIs remain a rich source of scientific exploration. One model system used to study PPIs is fatty acid biosynthesis (FAS) from *Escherichia coli.* In *E. coli* FAS each enzyme of the pathway is a discrete unit to which cargo is delivered by a central acyl carrier protein (EcACP). EcACP tethers the growing acyl chain, sequesters the chain

within its alpha helical core to protect it from hydrolysis and cross reactivity, finds the correct partner enzyme, delivers the cargo to the correct partner enzyme and then finds the next partner enzyme in the cycle, all while navigating the crowded milieu of the cytoplasm. During *E. coli* FAS, EcACP engages in nine different PPIs, and has been implicated in at least 23 additional PPIs with enzymes and proteins from other metabolic pathways in *E. coli* alone. This wide range of PPIs somehow allow EcACP to deliver substrates to partner enzymes with exquisite selectivity and specificity.

While numerous biophysical and structural tools such as isothermal titration calorimetry or X-ray crystallography exist to characterize PPIs, progress in the field has been hindered by a lack of universal, or rapid screen. This dissertation focuses on the development of four different fluorescent-based assays for the rapid visualization of PPIs between EcACP and its numerous partner enzymes, and for the discovery of new carrier protein interactions in previously uncultured marine bacteria.

First, the utility of a 4-dimethylnaphthalene (4-DMN) pantetheine probe to monitor thirteen different EcACP PPIs was expanded. 4-DMN-pantetheine was appended to EcACP to generate 4-DMN-EcACP and it was shown that when a partner enzyme interacted with 4-DMN-EcACP, the 4-DMN cargo from EcACP was chain flipped into the active site of the partner enzyme, which led to an increase in fluorescence intensity. 4-DMN-EcACP was used to visualize PPIs between EcACP and enzymes from FAS, biotin and Lipid A biosynthesis, iron-sulfur cluster formation and chromosome trafficking to demonstrate the wide applicability of this fluorescent tool. It was also shown that this fluorescent tool could be used to detect the presence of active site inhibitors in partner enzymes.

Next, *in silico* screening methods were used to identify a potential inhibitor of the PPI between EcACP and a dehydratase from *E. coli* FAS, FabA. Following identification of a lead compound, suramin, the inhibitory effect of suramin was tested *in vitro* using the previously developed solvatochromic assay, and the inhibition of suramin was confirmed using a secondary, crosslinking based assay. This work provided the basis for a major R01 grant to the NIH and demonstrated how detailed structural knowledge of these complexes can guide antibiotic discovery efforts and the development of a high-throughput screening program to be undertaken with collaborators at Sanford-Burnham-Prebys.

The utility of 4-DMN-EcACP was also demonstrated for mapping residues on the surface of two partner enzymes from *E. coli* FAS, FabF and FabI. Previous structural methods had identified residues on the surface of FabF and FabI as being critical to the EcACP interaction with each enzyme, and each of these residues were systematically mutated to either alanine or a charge swap mutation. The effect of these mutations on the PPIs between EcACP and FabF and EcACP and FabI was studied using an adapted 4-DMN-EcACP assay. It was found that some point mutations could effectively destroy the PPI, while others left the PPI relatively unchanged. This work is significant in future PPI engineering efforts in being able to rapidly identify which residues on the surface of EcACP partner enzymes are most critical to the PPI.

Finally, the same 4-DMN-pantetheine fluorescent tool was used to label individual bacterium *in vivo* from the microbiome of a tunicate. Sorting of these fluorescently labeled bacteria, followed by single-cell genomics and genome mining revealed that bacteria that up took the 4-DMN-pantetheine dye had a larger number of active secondary metabolite

pathways compared to untreated samples. This study provides a workflow for the future discovery of secondary metabolite pathways from uncultured bacteria that might otherwise be overlooked by more traditional genome mining methods.

Chapter 1. Introduction to the study of protein-protein interactions

1.1 Overview of protein-protein interactions

Interest in understanding the molecular underpinnings of these interactions has been driven by research in structure-aided drug design of PPI inhibitors. To design effective PPI inhibitors, it is imperative to map the surface resides most significant to the PPI in question to tailor drug scaffolds to target these residues, instead of attempting to block the entire PPI surface area which can be many hundreds of Angstroms in size.[1,2] PPIs have also been researched for protein engineering of biosynthetic pathways in which PPIs play a critical role, such as fatty acid (FAS), polyketide (PKS) and non-ribosomal (NRPS) synthases.[3,4] Much work has been dedicated to engineering new biosynthetic pathways by swapping enzymatic domains from different FAS, PKS or NRPS pathways to produce new bioactive compounds. However, any attempt to stitch together chimeric enzymes must be accompanied by addressing new PPIs and the challenge of rapidly identifying which residues are critical for the PPIs in question remains a major bottleneck.

One of the best characterized systems for studying PPIs is *Escherichia coli* fatty acid biosynthesis.[5–8] Fatty acids are biosynthesized in an iterative fashion: in both eukaryotes and prokaryotes, each turn of the fatty acid synthase (FAS) cycle extends the growing acyl chain by two carbons. Plants and bacteria use a type II system (FAS-II, Figure 1), in which each enzyme is a discrete entity, while eukaryotes use a megasynthase (FAS-I). Central to both FAS-I and FAS-II is the acyl carrier protein (ACP), which serves as a scaffold and shuttle for the growing fatty acyl chain as it is acted upon by each of the enzymes in the pathway. ACP sequesters its cargo within its three alpha helical core to protect the cargo from hydrolysis and cross reactivity until the correct partner enzyme is

encountered, after which ACP presents its cargo into the active site of the partner enzyme using a universal chain flipping mechanism. In FAS-II alone, ACP engages in eight different PPIs. These PPIs are responsible for the correct delivery of cargo by ACP to each enzyme in the pathway and are critical to recognition and catalysis. This dissertation will focus on the development of fluorescent-based assays to rapidly evaluation PPIs between the *Escherichia coli* acyl carrier protein and its various partner enzymes.

1.2 Overview of fatty acid biosynthesis

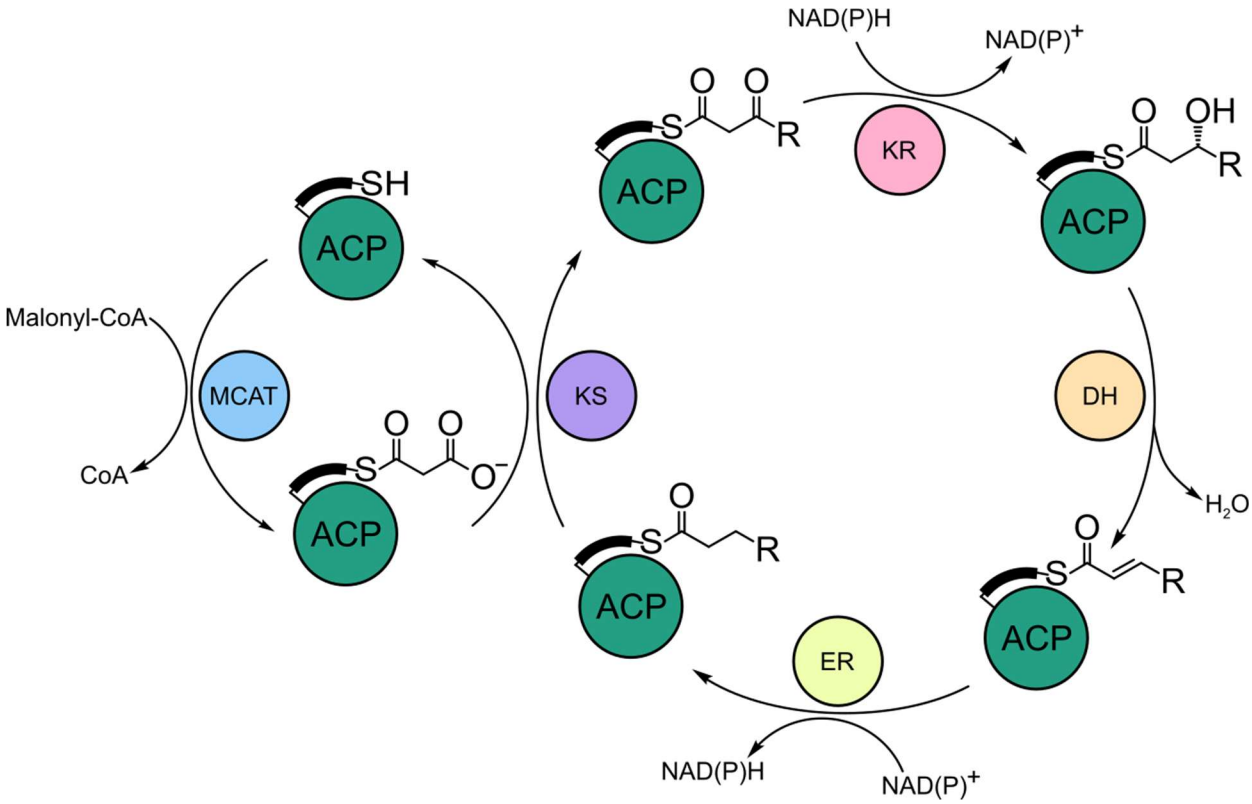


Figure 1.1: Cartoon representation of type II fatty acid biosynthesis.

One of the best characterized systems for studying PPIs is *Escherichia coli* fatty acid biosynthesis. Fatty acids are biosynthesized in an iterative fashion: in both eukaryotes and prokaryotes, each turn of the fatty acid synthase (FAS) cycle extends the growing acyl chain by two carbons. Plants and bacteria use a type II system (FAS-II,

Figure 1.1), in which each enzyme is a discrete entity, while eukaryotes use a megasynthase (FAS-I). Central to both FAS-I and FAS-II is the acyl carrier protein (ACP), which serves as a scaffold and shuttle for the growing fatty acyl chain as it is elongated and tailored by ACP partner enzymes in the FAS cycle.

ACP sequesters its cargo within its three alpha helical core to protect the cargo from hydrolysis and cross reactivity until the correct partner enzyme is encountered, after which ACP presents its cargo into the active site of the partner enzyme using a universal chain flipping mechanism.[16] In *E. coli* FAS-II alone, ACP engages in eight different PPIs. hese PPIs are responsible for the correct delivery of cargo by ACP to each enzyme in the pathway and are critical to recognition and catalysis.

1.3 Overview of chemical tools to study *E. coli* FAS

The Burkart laboratory has become a leader in usually chemical biology tools to probe PPIs in *E. coli* FAS.[17–23] EcACP is expressed in an inactive, "*apo-",* form and postranslationally modified by action of a phosphopantetheinyl transferase (PPTase) which appends a 4'-phosphopantetheine (PPant) arm to a conserved serine residue on the surface of EcACP (Figure 1.2). This yields the active, "*holo-*," form of EcACP. The PPant arm terminates in a thiol which *holo*-EcACP uses to tether acyl chain intermediates to itself via thioester bond to form *acyl*-EcACP.

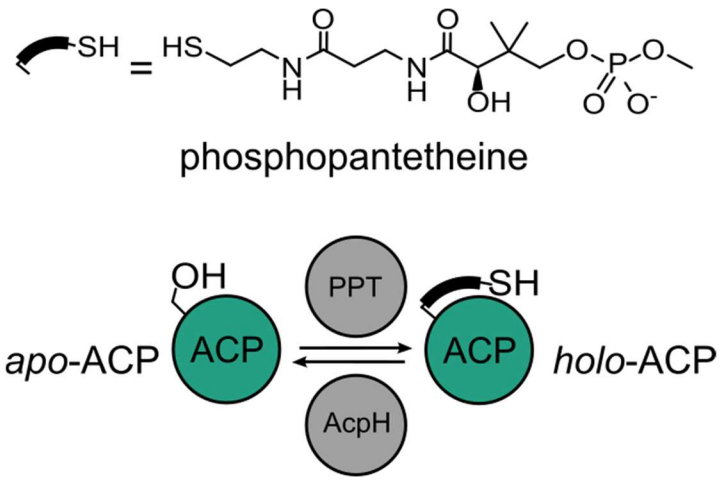


Figure 1.2: Structure of phosphopantetheine arm and cartoon representation of the post-translational modification event to generate holo-ACP from the inactive, apo-ACP, form. The phosphopantetheine arm can be removed by action of a phosphodiesterase, AcpH.

The core of the Burkart laboratory technology relies upon exploitation of this natural post-translational modification of EcACP. Early work in the Burkart laboratory established the synthesis of pantetheine analogues and a chemoenzymatic method to attach these synthetically prepared pantetheines to EcACP *in vitro,* in a one-pot reaction.[24] To accomplish this, EcACP is expressed and purified recombinantly from *E. coli* cultures, yielding a mixture of both *holo*- and *apo*-EcACP. Removal of the PPant arm with the action of a separately expressed and purified phosphodiesterase, AcpH, yields 100% *apo*-EcACP. *Apo*-EcACP can then be chemoenzymatically modified by action of three Coenzyme A preparative enzymes, CoaA, CoaD and CoaE, plus another phosphopantetheinyl transferase, Sfp (Figure 1.3). This labeling generates a *crypto*-EcACP form. The Burkart laboratory has demonstrated that many different pantetheine analogues can be appended to EcACP in this manner. The pantetheine analogue discussed in this dissertation bears a terminal fluorophore, but other analogues with various warheads and crosslinking moieties have been used for other studies in the lab.

1.4 Introduction to solvatochromic tools to study ACP-partner enzyme interactions

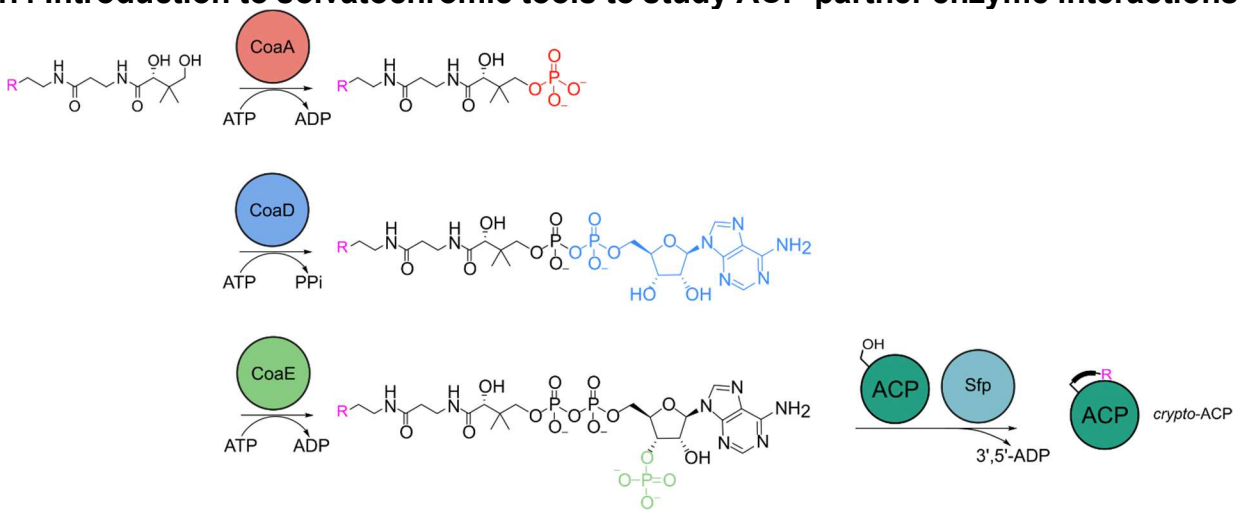


Figure 1.3: Cartoon representation of the chemoenzymatic "one-pot" method to append Coenzyme A analogues onto apo-ACP to form crypto-ACP. R groups vary from fluorophores to crosslinkers.

The fluorescent dye, 4-dimethylnaphthalene (4-DMN), used for the work discussed in this dissertation was first synthesized by the Imperioli group at the Massachusetts Institute for Technology in 2008.[25] 4-DMN is a fused ring cycle bearing an anhydride portion and a dimethyl group positioned on one of the fused aromatic rings. Imperioli et al. showed that 4-DMN is a solvatochromic fluorophore, meaning that the fluorescence intensity emitted by the dye is strongly dependent on the hydrophobicity of its environment.

In 2014, Beld et al. demonstrated the 4-DMN could be synthesized into a pantetheine analogue to generate 4-DMN-pantetheine, which could then undergo chemoenzymatic transfer onto *apo*-EcACP using the same one-pot method as described above, to generate 4-DMN-EcACP (Figure 1.4).[26] EcACP is known to sequester its cargo within the core of its alpha helical bundle, and it does so with the 4-DMN cargo. This sequestration leads to an increase in fluorescence intensity of the dye relative to the dye free in solution because of the hydrophobic nature of the interior of EcACP.

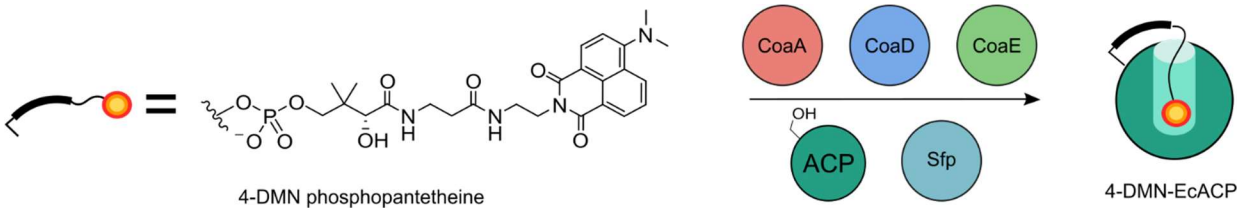


Figure 1.4: structure of 4-dimethylnaphthalene phosphopantetheine and one-pot reaction to generate 4-DMN-EcACP.

EcACP uses a universal chain flipping mechanism to transfer cargo from the interior of its alpha helical core into the active site of its various partner enzymes.[16] Beld et al. demonstrated that 4-DMN-EcACP completes this chain flipping mechanism with FabF, a ketosynthase from *E. coli* FAS, which leads to a large increase in fluorescence intensity when measured with a fluorimeter. This is due to the change in hydrophobicity encountered by the 4-DMN dye as it is transferred into the relatively greasier active site of FabF. This study confirmed that the 4-DMN cargo is being chain flipped into the active site of FabF, and not a cryptic pocket on the surface of the enzyme instead, by use a second pantetheine probe, this time with the 4-DMN dye bearing a crosslinking warhead. The 4-DMN-crosslinker chain flips into the active site, again leading to a large increase in fluorescence intensity, but then reacts with the FabF active site cysteine to form a covalent bond. This effectively traps the 4-DMN-EcACP·FabF interaction and can be visualized by an SDS-PAGE assay. Further experiments with protein NMR titrations and single molecule fluorescence microscopy confirmed that 4-DMN-EcACP chain flips the 4-DMN cargo into the active site of FabF and that this method could be used to visualize the chain flipping mechanism between these two proteins.

1.5 Solvatochromic dye assay expansion (Chapters 2 – 4)

Following the work by Beld et al. demonstrating the use of the 4-DMN fluorophore for monitoring EcACP · FabF interaction, we turned our attention to expanding the scope of this dye.[27] We discovered that many EcACP · partner enzyme interactions could be

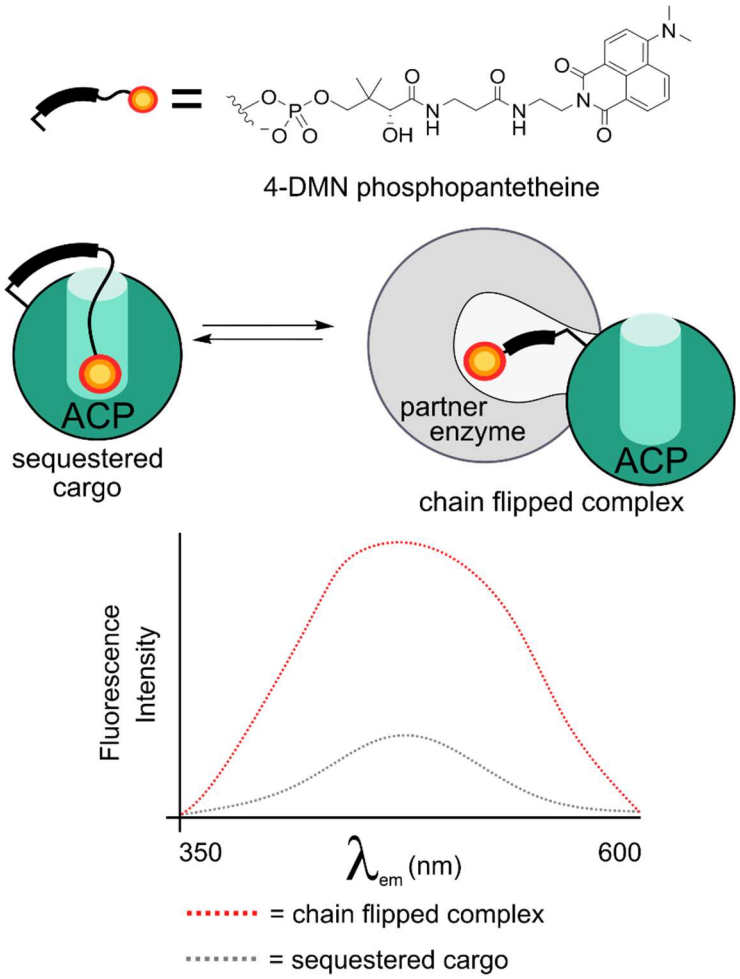


Figure 1.5: ACP sequesters the 4-DMN cargo within its alpha helical bundle, leading to mild fluorescence output. When interacting with a partner enzyme, the 4-DMN cargo is transferred into the active site, leading to a large increase in fluorescence intensity.

monitored with this single method and expanded the utility of this method for the detection of both active site and PPI inhibitors (Figure 1.5). We also determined that perturbations to EcACP · partner enzyme interactions resulting from mutations to surface residues could

be monitored by adapting this assay into a rapid screen for alanine scanning. The motivation of this work was to determine if the 4-DMN dye could be useful in a high-throughput screen searching for inhibitors of EcACP · partner enzyme interactions, or inhibitors targeting partner enzyme active sites. With the hot spot mapping work, we also showed that 4-DMN-EcACP assays can provide complementary data to kinetic and crystallography studies, to decouple the effect of surface residue mutations on either catalysis or the PPI itself.

A single tool to monitor multiple EcACP · partner enzyme interactions (Chapter 2)

EcACP has been implicated in 30 different PPIs in *E. coli* metabolism to date. Each of these interactions is critical to cell survival and many are orthogonal to those found in eukaryotes, making these numerous PPIs an attractive, if not challenging, drug target. Previously, each EcACP · partner enzyme interactions PPI might be characterized by painstaking X-ray crystallography or protein NMR studies, but we sought out to create a method to monitor a multitude of these interactions in one single experiment. The ability to monitor many different interactions increases the number of targets available for high-throughput screening approaches. This first work demonstrates that thirteen different partner enzymes have a solvatochromic response when titrated against 4-DMN-EcACP. These enzymes represent EcACP partner enzymes from FAS, biotin and Lipid A biosynthesis, iron-sulfur cluster formation and chromosome trafficking, showing the breadth of EcACP · partner enzyme interactions able to be monitored with a single 4-DMN-EcACP species. We showed that this method is sensitive enough to monitor subtle conformational changes in partner enzymes that engage cofactors such as NADPH, meaning that if adapted to a high-throughput screen, using 4-DMN-EcACP to detect

inhibitors might mean that allosteric inhibitors are also found. Titrations of each partner enzyme studied against 4-DMN-EcACP allowed for the determination of binding constants for some of the interactions, although we noted that because the 4-DMN cargo is much different from the natural acyl chain substrate of EcACP's that the K_d's calculated using this method should only be used to compare relative K_d's and not as an absolute value of binding affinity. This work also showed that known active site inhibitors such as phenylmethylsulfonyl fluoride (PMSF, a known hydrolase inhibitor) and cerulenin (a known inhibitor of the *E. coli* FAS enzyme FabF) could be detected with 4-DMN-EcACP as a dose dependent decrease in fluorescence intensity relative to uninhibited enzymes.

Computational and *in vitro* discovery of an EcACP · FabA interaction inhibitor (Chapter 3)

Following the validation of 4-DMN-EcACP based assays in the detection of active site inhibitors (Chapter 2), we next asked if this method could be used to find inhibitors of the EcACP · partner enzyme interaction itself. The work presented in this chapter started as a class project for an Introduction to Computational Chemistry course the dissertation author completed during the 2017 – 2018 academic year. The class project used the previously solved EcACP-FabA crosslinked crystal structure to virtually screen for compounds making contacts at the interface of these two proteins. Residues on the surface were chosen as targets for docking to a small panel of 200 commercially available and well-known drugs. The screen identified five compounds with potential inhibitory action, and one, known as suramin, was purchased due to its commercial availability. Suramin was studied *in vitro* using the previously published solvatochromic assay to access the disruption to the 4-DMN-EcACP · FabA interaction and was found to a have

dose dependent inhibitory effect on this PPI. A secondary, crosslinking based assay was used to confirm the inhibitory action of suramin on the interaction as well. Finally, the *in vivo* activity of suramin was accessed by a minimum inhibitory concentration assay (MIC). No antibiotic activity of suramin was detected in this assay, likely due to the highly charged nature of the molecule rendering it unable to cross the cell membrane. This work forms the foundation of a major assay development grant in collaboration with the high-throughput screening center at Sanford-Burnham-Prebys.

Solvatochromic dyes for hot-spot mapping (Chapter 4)

This solvatochromic assay for EcACP chain flipping with partner enzymes was also used to probe the perturbations to EcACP · partner enzyme interactions due to point mutations on the surface of each protein. X-ray crystallography and protein NMR studies on the EcACP · FabF interaction had shown that there are three regions on the surface of FabF that are involved in both EcACP recognition and catalysis and we wondered if this same solvatochromic chain flipping assay could be used to rapidly identify which of these residues is most important to EcACP · FabF chain flipping, and which are more critical to the FabF catalysis itself.[18] By systematically mutating hot-spot residues on the surface of FabF to either alanine or the charge swap mutation and testing the solvatochromic activity of each of these FabF variants, we were able to map which residues are critical to the EcACP · FabF chain flipping interaction. This work complements other research produced by our group earlier in 2020 studying the same FabF variants using both enzyme kinetics and *in vivo* complementation work.[13] These results taken all together suggest that some residues on the surface of FabF may play

only a critical role in EcACP · FabF recognition, while others are only necessary for FabF catalysis rather than the PPI itself.

Activity-based discovery of proteins using solvatochromic dyes (Chapter 5)

The 4-DMN pantetheine dye was also used to discover new synthase domains from previously uncultured marine bacteria. Pantetheine analogues readily cross the bacterial cell wall and can label carrier proteins *in vivo* provided a promiscuous phosphopantetheinyl transferase (PPTase).[28] We reasoned that marine bacteria harboring secondary metabolite pathways would have such enzymatic machinery for both carrier proteins and active PPTases and set out to create an activity guided program for the discovery of these pathways from the microbiome of marine tunicates. In this study, we first labeled a small subset of a tunicate microbiome with the 4-DMN pantetheine probe and sorted individual bacterium using fluorescently-activated cell sorting (FACS). Following FACS, bacteria genomes were sequenced using single-cell genomics methods and then the genome mined for active secondary metabolite pathways using the computational tool, antiSMASH.[29] AntiSMASH annotates genomic data according to homology with known carrier-protein based pathways such as non-ribosomal peptide, polyketide and fatty acid synthases. Each of these pathways is known to use a PPTase to append the carrier protein with the natural phosphopantetheine arm and so it follows that bacteria with highly active secondary metabolite pathways would be labeled *in vivo* by our 4-DMN pantetheine analogue. Genes from a representative organism were synthesized and inserted into a vector for recombinant expression and purification. Following successful purification, the *in vitro* labeling activity of the new carrier protein-PPTase pair was tested and found to be active. This workflow demonstrated the utility of

fluorescently labeled CoA analogues for activity-guided discovery of new carrier protein dependent secondary metabolite pathways in otherwise uncultured or undiscovered bacteria.

Acknowledgements

Chapter 1, in part, uses figures and text from the manuscript titled “Quantifying protein-protein interactions of the acyl carrier protein with solvatochromic probes” by Katherine Charov and Michael D. Burkart, which was published in Methods in Enzymology 638, 2020, 321-340. The dissertation author is the primary author of this manuscript. The rest of the work presented in this chapter is unpublished.

1.6 References

1. Modell, A. E., Blosser, S. L. & Arora, P. S. Systematic Targeting of Protein-Protein Interactions. *Trends Pharmacol. Sci.* **37**, 702–713 (2016).

2. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* **21**, 1102–1114 (2014).

3. Azzarito, V., Long, K., Murphy, N. S. & Wilson, A. J. Inhibition of α-helix-mediated protein-protein interactions using designed molecules. *Nat. Chem.* **5**, 161–173 (2013).

4. Klaus, M., D’Souza, A. D., Nivina, A., Khosla, C. & Grininger, M. Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chem. Biol.* **14**, 426–433 (2019).

5. Newman, A. G., Vagstad, A. L., Storm, P. A. & Townsend, C. A. Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J. Am. Chem. Soc.* **136**, 7348–7362 (2014).

6. Yuzawa, S., Deng, K., Wang, G., Baidoo, E.E.K., Northen, T.R., Adams, P.D., Katz, L. & Keasling, J.D. Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol.* **6**, 139–147 (2017).

7. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).

8. Kortemme, T., Kim, D. E. & Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE Signal Transduct. Knowl. Environ.* **2004**, pl2 (2004).

9. Douzi, B. Protein-Protein Interactions: Surface Plasmon Resonance. *Methods Mol. Biol. Clifton NJ* **1615**, 257–275 (2017).

10. Pierce, M. M., Raman, C. S. & Nall, B. T. Isothermal titration calorimetry of protein-protein interactions. *Methods San Diego Calif* **19**, 213–221 (1999).

11. Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Martin, J.L., Marfori, M., & Forwood, J.K. Crystallography and protein-protein interactions: biological interfaces and crystal contacts. *Biochem. Soc. Trans.* **36**, 1438–1441 (2008).

12. Zuiderweg, E. R. P. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1–7 (2002).

13. 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* **59**, 3626–3638 (2020).

14. 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. U. S. A.* **117**, 24224–24233 (2020).

15. Misson, L. E., Mindrebo, J.T., Davis, T.D., Patel, A., McCammon, J.A., Noel, J.P., & Burkart, M.D. Structure and Dynamic Basis of Molecular Recognition Between Acyltransferase and Carrier Protein in *E. coli* Fatty Acid Synthesis. http://biorxiv.org/lookup/doi/10.1101/2020.05.15.098798 (2020) doi:10.1101/2020.05.15.098798.

16. Cronan, J. E. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. *Biochem. J.* **460**, 157–163 (2014).

17. Nguyen, C., Haushalter, R., 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., & Burkart, M.D. Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505**, 427–431 (2014).

18. 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.* **11**, 1727 (2020).

19. 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.* **37**, 355–379 (2020).

20. 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.* **1**, 8–12 (2020).

21. 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.* **116**, 6775–6783 (2019).

22. Tallorin, L., Finzel, K., Nguyen, Q.G., Beld, J., La Clair, J.J., & Burkart, M.D. Trapping of the Enoyl-Acyl Carrier Protein Reductase-Acyl Carrier Protein Interaction. *J. Am. Chem. Soc.* **138**, 3962–3965 (2016).

23. Moynié, L., Hope, A.G., Finzel, K., Schmidberger, J., Leckie, S.M., Schneider, G., Burkart, M.D., Smith, A.D., Gray, D.W., & Naismith, J.H. A Substrate Mimic Allows High-Throughput Assay of the FabA Protein and Consequently the Identification of a Novel Inhibitor of Pseudomonas aeruginosa FabA. *J. Mol. Biol.* **428**, 108–120 (2016).

24. Worthington, A. S., Porter, D. F. & Burkart, M. D. Mechanism-based crosslinking as a gauge for functional interaction of modular synthases. *Org. Biomol. Chem.* **8**, 1769–1772 (2010).

25. Loving, G. & Imperiali, B. A versatile amino acid analogue of the solvatochromic fluorophore 4-N,N-dimethylamino-1,8-naphthalimide: a powerful tool for the study of dynamic protein interactions. *J. Am. Chem. Soc.* **130**, 13630–13638 (2008).

26. Beld, J., Cang, H. & Burkart, M. D. Visualizing the chain-flipping mechanism in fatty-acid biosynthesis. *Angew. Chem. Int. Ed Engl.* **53**, 14456–14461 (2014).

27. Charov, K. & Burkart, M. D. A Single Tool to Monitor Multiple Protein-Protein Interactions of the Escherichia coli Acyl Carrier Protein. *ACS Infect. Dis.* **5**, 1518–1523 (2019).

28. Clarke, K. M., Mercer, A. C., La Clair, J. J. & Burkart, M. D. In vivo reporter labeling of proteins via metabolic delivery of coenzyme A analogues. *J. Am. Chem. Soc.* **127**, 11234–11235 (2005).

29. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., & Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

Chapter 2. A single tool to monitor multiple protein-protein interactions of the *Escherichia coli* acyl carrier protein

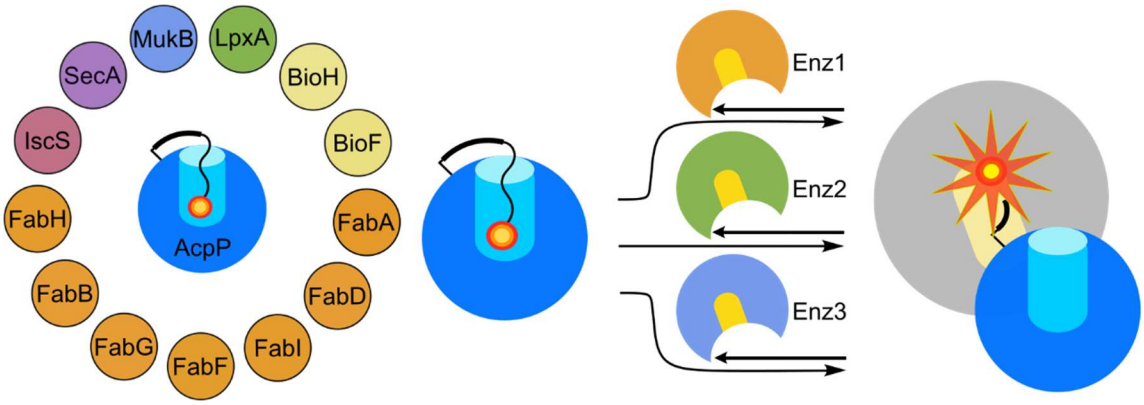


Figure 2.1: Graphical abstract of chapter 2.

Protein-protein interactions are ubiquitous to all domains of life and have gained recent interest as drug targets. However, many current methods to study protein-protein interactions can be costly and are low-throughput. Here, we demonstrate a solvatochromic tool based on the natural post-translational modification of the *Escherichia coli* acyl carrier protein (EcACP) used to visualize protein-protein interactions between EcACP and thirteen different partner enzymes from several biosynthetic pathways. We use this tool to confirm proposed interactions between EcACP and both catalytic and regulatory proteins. We also show the utility of this method towards detecting allosteric changes to partner enzyme structure and the validation of active site inhibitors. We anticipate the future adaptation of this assay into a high-throughput screen for antibiotic discovery.

Protein-protein interactions (PPIs) play critical roles in stabilizing structure, integrating signal transduction networks, and guiding protein delivery. While originally viewed as undruggable, the clinical approval of the integrin antagonist, tirofaban, followed by the taxanes, rapamycin and cyclosporine in the late 1990s, confirmed PPIs as

therapeutic targets.[1] PPIs have also gained importance as targets for infectious disease playing an increasing role in anti-infectives and antibiotics.[2–4] Our work and others' have highlighted the role of PPIs in bacterial fatty acid biosynthesis. In these systems, bacteria synthesize fatty acids using a type II fatty acid synthase (FAS) in which the acyl carrier protein (ACP) shuttles the growing acyl chain to a series of discrete partner enzymes, engaging in a minimum of nine PPIs to produce fatty acids of varying chain lengths *de novo* (Figure 2.2). Early studies of the thiolactomycin antibiotics in the late 1980s validated bacterial fatty biosynthesis as a therapeutic target, and the clinical approval of triclosan and isoniazid offer further support.[5,6] Bacterial type II FAS is distinct from the type I FAS megasynthases of mammalian cells, making PPIs between type II enzymes and the central FAS ACP a potentially specific avenue for drug discovery and design.

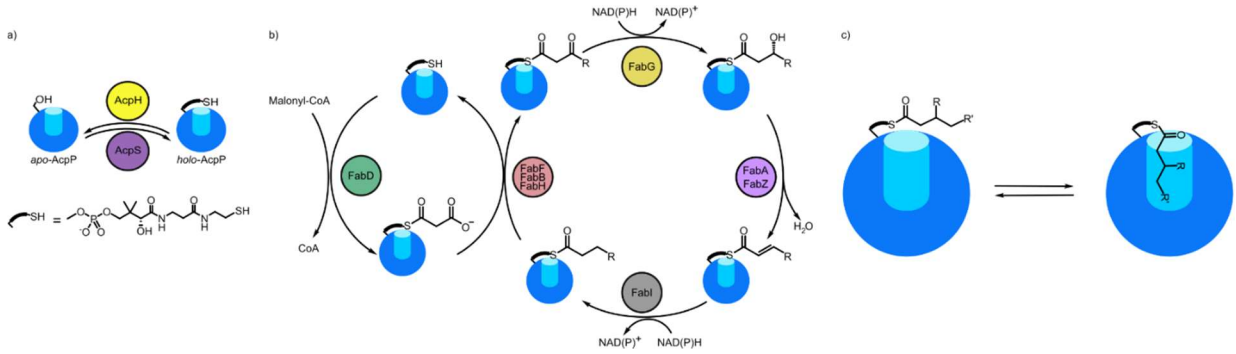


Figure 2.2: Cartoon representation of E. coli type II FAS. a) apo-EcACP is first post-translationally modified by a phosphopantetheinyl transferase enzyme to append a phosphopantetheine arm to form holo-EcACP. B) Following post-translational modification, the malonyl-CoA acyl transferase, FabD, transfers a malonyl unit to the phosphopantetheine arm of ACP to form malonyl-ACP, which is then condensed with another unit of malonyl-ACP by the first ketosynthase, FabH, to extend the chain by two carbons. Next, the ketoreductase, FabG, reduces the beta-keto group to a beta-hydroxy moiety, which is then acted upon by the dehydratase, FabA, to form an alpha-beta unsaturated unit. The enoyl reductase, FabI, forms the fully saturated acyl chain, which can then re-enter the FAS cycle to condense with another unit of malonyl-ACP by the second or third ketosynthases, FabF or FabB. When the proper chain length has been reached, the acyl chain is removed from ACP by a thioesterase or acyl transferase enzyme. c) sequestration of cargo within the alpha helical bundle of EcACP.

As a bacterial protein of central metabolism, ACP has been implicated in 18 additional interactions in biosynthetic pathways beyond FAS. As part of the hydrocarbon factory of the cell, ACP shuttles acyl precursors to generate fatty acids, which are used in the biosynthesis of lipid A, hemolysin, biotin and lipoic acid.[7–10] Bacterial ACP plays further roles in regulation of the stress response and iron-cluster formation, and is hypothesized to participate in chromosome segregation.[11–13] Each of these ACP·partner enzyme PPIs are critical to cell survival and represent a potential target for new antibiotic discovery.

Protein-protein interactions of ACP are by necessity transient and weak and require special methods to study them fully. We and others have previously used solution NMR and X-ray crystallography to characterize the surface residues that engage between ACP and a small set of its partner enzymes.[14] While these methods reveal significant detail, they are low-throughput, can be costly, and are limited by technical challenges like solubility and size. Other methods used to assess the thermodynamics of PPIs, such as isothermal titration calorimetry and surface plasmon resonance, often require placing interacting proteins in exotic, non-native environments. The lack of a universal screen that enables the parallel analysis of ACP interactivity has hindered the high-throughput discovery process. Herein we describe the application of a solvatochromic probe for visualizing ACP·partner enzyme interactions. In 2014, we reported a new method to rapidly visualize the chain flipping mechanism between *Escherichia coli* ACP (EcACP) and the fatty acid partner enzyme ketosynthase II (FabF) using a solvatochromic dye, 4-dimethylnaphthalene-phosphopantetheine (4-DMN-ppant, Figure 2.3a).[15]

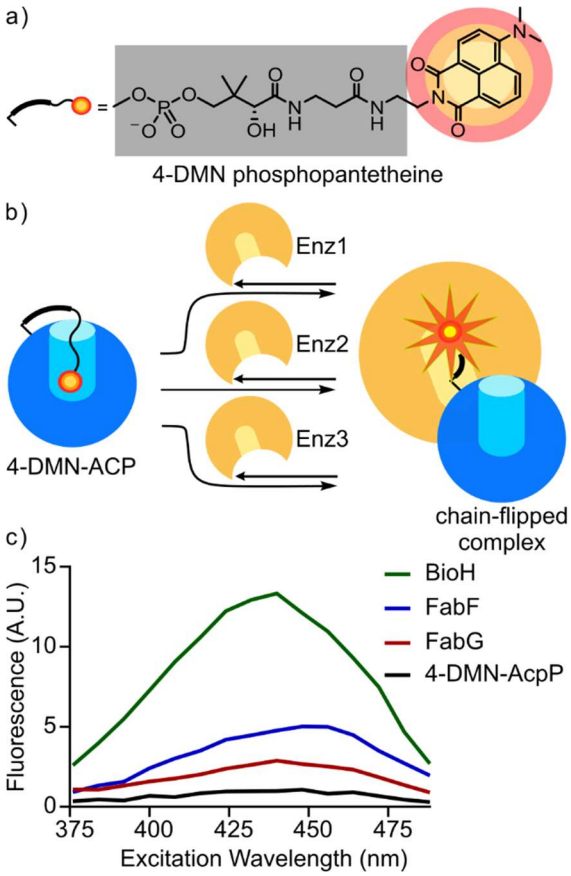


Figure 2.3: a) structure of solvatochromic dye, 4-DMN-phosphopantetheine, b) Labeled ACP, 4-DMN-EcACP, flips the 4-DMN cargo into the active site of a partner enzyme upon proper PPI, c) fluorescent excitation spectrum (λem = 512 nm) for 4-DMN-EcACP interacting with three different partner enzymes

When EcACP is labeled with 4-DMN-ppant (herein referred to as 4-DMN-EcACP), the 4-DMN fluorophore is sequestered within the ACP helical bundle. Upon binding to FabF, the solvatochromic cargo chain flips into the active site of FabF, resulting in a measurable increase in fluorescent intensity due to the change in hydrophobic environment (Figure 2.3b, c). The chain flipping mechanism is dependent on correct PPI and is universal in ACP·partner enzymes interactions.[16] A fluorescence intensity increase was not observed when 4-DMN-EcACP was titrated against bovine serum albumin, which is known to not interact with ACP, indicating that this method is specific to ACP·partner

enzyme interactions. This solvatochromic method can be rapidly monitored using a plate reader or fluorimeter, allowing for real time monitoring of 4-DMN-EcACP·FabF

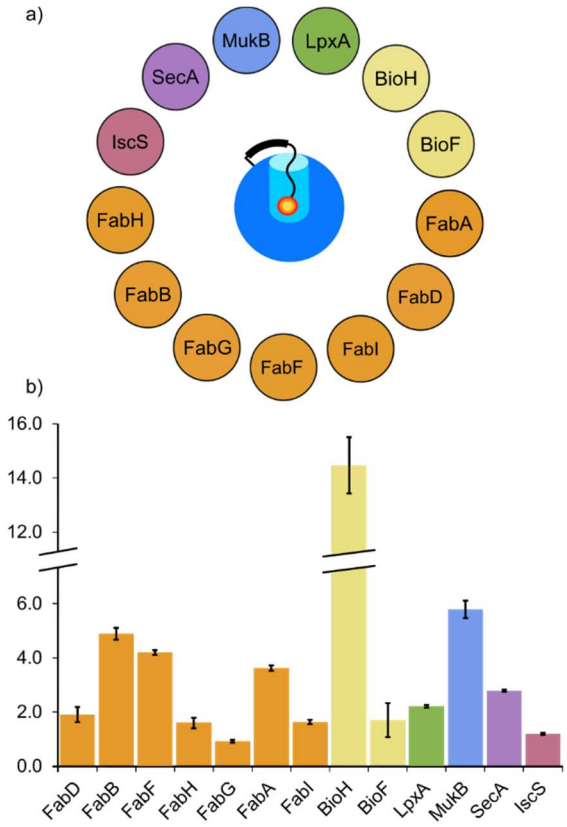


Figure 2.4: a) EcACP-partner enzyme interactions studied by this assay grouped by metabolic pathway. Previously discussed fatty acid biosynthetic enzymes are depicted in orange, while chromosome partitioning and secretory enzymes (MukB, SecA, respectively) are in blue and purple. In addition to enzymes discussed in the text, we also evaluated the biotin biosynthetic BioH and BioF (yellow) which catalyze the hydrolysis of pimeloyl-EcACP methyl ester and the decarboxylative condensation of pimeloyl-ACP and alanine, respectively. Lipid A biosynthesis involves at least four transfers of acyl chains from EcACP to lipid A precursors by four different enzymes, of which LpxA (green) is the first, transferring a 3-OH-C14 chain from EcACP to uridine diphosphate N-acetylglucosamine.The iron cluster assembly enzyme, IscS, which removes a sulfur from cysteine and transfers it to the terminal thiol of the phosphopantetheine arm of EcACP was also evaluated. b) Solvatochromic response of each enzyme. Response roughly correlates with active site polarity and size: enzymes with larger and more hydrophobic pockets can accommodate the large 4-DMN fluorophore and lead to a larger increase in fluorescence intensity. Data were completed in triplicate on three separate occasions for a total of 9 titrations for each enzyme; error bars represent 95% confidence intervals. BioF contains a pyridoxal 5'-phosphate cofactor, which fluoresces in the same wavelength as the 4-DMN solvatochromic dye, making analysis of the protein difficult and prone to error.

interactions. Here, we demonstrate the capability of 4-DMN-EcACP to monitor PPIs between *E. coli* EcACP and 13 known partner enzymes, enabling the monitoring of a wide breadth of PPIs with a single tool (Figure 2.4b). This work has long-range implications for the design of high-throughput screens for drug discovery against bacterial ACP

interactions and can be adapted to any bacterial ACP·partner enzyme interaction network. To detect interactions between 4-DMN-EcACP and partner enzymes, individual enzymes were titrated into a solution of 4-DMN-EcACP with a fluorescence reading taken at each point to generate titration curves for each interaction (SI Figsure S2.1-S2.13). We first used this tool to monitor PPIs between 4-DMN-EcACP and well characterized partner enzymes from the fatty acid, biotin, lipid A and iron-sulfur cluster assembly biosynthetic pathways. The end points of each titration are summarized in Figure 2.4b. These titrations revealed that it is possible to visualize the interaction between EcACP and a wide variety of biosynthetic partner enzymes.

However, while each ACP·partner enzyme interaction studied resulted in an increase in fluorescence intensity, the change in intensity was unique to the partner enzyme assayed. For instance, the ketosynthase enzymes FabB and FabF perform similarly in this assay, resulting in a 4.9 and 4.3-fold fluorescence increase, respectively; however, the ketosynthase FabH shows a 1.7-fold fluorescence increase under the same assay conditions. FabH is responsible for the first FAS condensation of two malonyl-EcACP units, while FabB and FabF are responsible for later condensations of elongated acyl chains in FAS (Figure 2.2). While all three enzymes perform the same basic reaction, FabB and FabF exhibit nearly the same chain length specificity, while FabH accepts only short acyl chain lengths.[17] The active site of FabH is smaller and likely cannot accommodate the bulky 4-DMN fluorophore resulting in a smaller fluorescence intensity increase relative to the solvatochromic effect observed with FabF and FabB.. This trend is also observed with the malonyl-CoA acyl transferase, FabD, which appends a malonyl unit onto the phosphopantetheine arm of EcACP at the beginning of the FAS cycle. FabD

has a small active site which results in a low fluorescence increase when titrated with 4-DMN-EcACP. 4-DMN is sensitive to the relative hydrophobicity of each active site as well. The active site of FabD is characterized by relatively hydrophilic residues (glutamine, arginine and serine, SI Fig. 18); when 4-DMN is chain flipped into the active site of FabD it encounters an environment that is more hydrophobic than the alpha helical bundle of EcACP, which results in the 1.9-fold fluorescence intensity increase. However, the hydrophobicity of the FabD active site is lower when compared to the active sites of the ketosynthases FabF and FabB (hydrophobic heat map Fig. S18). This is reflected in the relative fluorescence intensity increases: FabF and FabB showed a fluorescence intensity increase of 4.3 and 5.1, respectively, while the relatively more hydrophilic FabD resulted in a 1.9-fold fluorescence intensity increase. We also present a case in which the utility of this assay is limited. The active site of the biotin enzyme 8-amino-7-oxononanoate synthase, BioF contains the highly fluorescent pyridoxal 5'-phosphate cofactor. While the background fluorescence of BioF was corrected for, the readings of the 4-DMN-EcACP·BioF interaction are intrinsically noisy due to the fluorescent cofactor and resulted in a larger error than the other partner enzymes studies.

We next investigated ACP·partner enzyme interactions that have been less well characterized. Affinity purification studies have suggested that the chromosome partitioning enzyme MukB and the protein translocase subunit SecA might form a complex with EcACP, but it remains unclear whether EcACP copurifies with these enzymes due to its high acidity and intracellular concentration, or because of a functional role.[12,13] MukB is responsible for chromosome organization along with SecA, which helps in the translocation of membrane proteins and chromosome segregation. Our assay

shows that EcACP and both these enzymes engage in selective interaction (Figure 2.4b). We are encouraged that this method will prove useful for quickly accessing interactions between EcACP and new partner enzymes.

Having demonstrated that the 4-DMN probe can monitor a wide breadth of ACP·partner enzyme interactions, we next asked if this method could be used to detect allosteric effects of cofactors. To study the effects of cofactors, we used the FAS II ketoreductase, FabG, which forms an active tetramer with up to four NADPH and four EcACP binding sites. In 2001, White et al. showed that FabG undergoes a conformational change upon binding of $NADP^+$, producing a defined tunnel for the acyl-EcACP substrate, providing a regulatory role for PPIs.[18] We observed a difference in fluorescent response between 4-DMN-EcACP interacting with FabG in the absence of $NADP^+$ versus in the presence of 0.25 equivalents of the cofactor, confirming that $NADP^+$ induces changes to the active site of FabG and binding site of EcACP (SI Fig. 14). We propose that 4-DMN chain flips into the tunnel arising from $NADP^+$ binding, resulting in the fluorescence intensity increase, relative to *apo*-FabG. Allosteric modulators represent another mode of antibiotic drug discovery and these results demonstrate the potential utility of this assay in detecting a wider scope of inhibitors.

Finally, we demonstrated the utility of our method towards extrapolating dissociation constants and evaluating the binding of active site inhibitors. We calculated the K_d of the 4-DMN-EcACP·FabF, 4-DMN-EcACP·BioH and 4-DMN-EcACP·FabA interaction to be 0.89 (0.35, 1.6)[a] μM (literature value 4.1 μM), 1.13 (0, 5.053)[a] μM (literature value 3.1 μM) and 2.85 (2.3, 3.6)[a] μM, respectively.[19, 20] It is important to note that the installation of the 4-DMN probe on EcACP could alter the binding affinity of

EcACP to partner enzymes. It is hypothesized that EcACP confers information about its sequestered cargo to partner enzymes by subtly changing the conformation of its helices.[21] It is likely that the sequestration of the 4-DMN dye causes such a change, resulting in an altered K_d as extracted from the fluorescent data presented here.

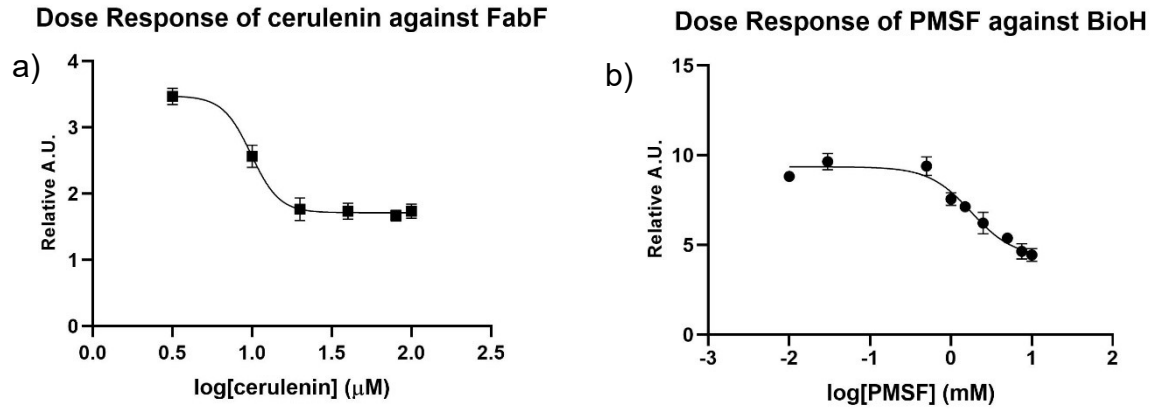


Figure 2.5: Dose response curves of a) cerulenin and b) PMSF against FabF and BioH, respectively. Calculated IC_{50} values are 9.87 (8.52, 10.82)[b] μM for cerulenin and 1.78 (1.32, 3.30)[b] mM for PMSF. High concentrations of PMSF were limited by solubility of the compound. Data were completed in triplicate and IC_{50} values calculated using GraphPad Prism 8.

We next determined the IC_{50} of known inhibitors, cerulenin and phenyl methyl sulfonyl fluoride (PMSF) (Figure 2.5). Cerulenin targets the FAS ketosynthases FabF and FabB by reacting with the active site cysteine to form a covalent adduct.[22] PMSF is a general serine protease inhibitor, forming a covalent bond in the active site following reaction with an active site serine residue.[23] Our assay determined the IC_{50} of cerulenin to be 9.87 (8.52, 10.82)[b] μM against FabF (literature value of IC_{50} = 0.4 - 9 μM) and the IC_{50} of PMSF against BioH to be 1.78 (1.32, 3.30)[b] mM.[24]

Based on these findings, we conclude that this fluorescence assay provides a platform from which to develop a high-throughput screen of both PPIs and active site inhibitors. PPIs are challenging to target with small molecules due to their large, diffuse surfaces, and do not benefit from known natural ligands or substrates to guide drug discovery. Moreover, high-throughput screening does not typically identify inhibitors of these interactions.[25,26] However, PPIs are quickly being recognized as not only druggable targets but ones that carry the potential for high specificity against bacteria and other pathogens.[27] The sensitivity of this method to conformation, as demonstrated with our evaluation of FabG-$NADP^+$ solvatochromism, is particularly notable in the context of drug discovery, wherein this method could be used to detect both active site inhibitors and compounds that act allosterically to alter protein-protein interactions.

Recently, other published studies have used fluorescent and infrared probes to study carrier protein·partner enzyme complexes, including the conversion of the terminal phosphopantetheine thiol into a thiocyanate and the use of a different solvatochromic probe.[28, 29] While each of these methods provide applications for specific enzymes of protein-protein interactions, our present study demonstrates the unique ability to probe ACP binding events with over a dozen enzymes, offering distinct advantages.

We have developed a single fluorescent tool that can monitor multiple protein-protein interactions between the ACP and multiple partner enzymes. We have also demonstrated the utility of this assay towards the detection of conformational changes and the presence of active site inhibitors in partner enzymes of EcACP. We anticipate developing this assay into a high-throughput screen to detect not only active site inhibitors

and inhibitors of protein-protein interaction of multiple EcACP-partner enzyme complexes.

Methods

Protein purification: Plasmids encoding protein of interest were transformed into *E. coli* BL21 cells and grown in 1-liter cultures of lysogeny broth at 37°C on a shaker. When an OD_{600} of 0.6 was reached, cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside or 0.1% L-arabinose and shaken at 37°C for another 3 hours and then harvested at 2000 rpm and frozen at -20°C. Frozen cell pellets harboring the expressed His-tagged protein of interest were thawed and resuspended in lysis buffer (25 mM HEPES, 150 mM NaCl, 15% glycerol, 25 mM imidazole, pH 7.4) with 10 mg of lysozyme and 10 mg of deoxyribonuclease. After thawing, cell pellets were placed on ice and sonicated for 10 minutes to lyse cells. Cell lysate was clarified by centrifugation at 10000 rpm for 45 minutes at 4°C. The clarified supernatant was then transferred to a 50 mL falcon tube with 1 mL of nickel-agarose resin and allowed to batch bind at 4°C for 20 minutes. After 20 minutes, the nickel resin was placed in a 20 mL column and washed with 50 mL of lysis buffer. To elute protein, the nickel resin was washed with 10 mL of elution buffer (25 mM HEPES, 150 mM NaCl, 15% glycerol, 250 mM imidazole, pH 7.4) and collected in 1 mL fractions. 12% SDS-PAGE was used to determine fractions containing protein of interest, which were then collected and concentrated using a spin concentrator. Imidazole was removed by dialyzing the concentrated fractions overnight into lysis buffer. Concentrated protein was flash frozen at -78°C and stored at -20°C until further use. For LpxA protein, which lacks a His-tag, 1 mL of Reactive Red resin was used in place of nickel-agarose and the protein was eluted with 400 mM NaCl.

4-DMN-EcACP preparation: E. coli acyl carrier protein (EcACP) was expressed and purified from *E. coli* as described above, yielding a mixture of *apo*- and *holo*-EcACP. To remove the phosphopantetheine arm, the mixture of *apo*- and *holo*-EcACP was incubated at 37°C overnight with an equal volume of previously purified phosphodiesterase from *Pseudomonas aeruginosa* (AcpH), to yield 100% *apo*-EcACP as previously described.[30] The success of the removal of the phosphopantetheine arm was determined by partially denaturing 20% urea-PAGE. To append the 4-DMN-pantetheine solvatochromic dye to *apo*-EcACP, the protein was incubated overnight at 37°C with 0.1 mg/ml each of previously purified CoA-biosynthetic enzymes from *E. coli* (CoaA, CoaD and CoaE), and the phosphopantetheinyl transferase from *Bacillus subtilis* (Sfp), 12 mM MgCl2, 8 mM ATP, 0.1% Triton-X and 2 mM 4-DMN pantetheine dissolved in dimethylsulfoxide, as previously described.[15] 4-DMN-pantetheine was synthesized as previously described.[15] The success of the labeling of EcACP with 4-DMN-phosphopantetheine was determined by 20% urea-PAGE. 4-DMN-EcACP was purified using an S75 Superdex size exclusion column with an ÄKTA Pure FPLC in 50mM phosphate buffer, pH 8.0.

Partner enzyme titration: Readings were completed using a black COSTAR 96-well plate containing 50 microliters per well of 120 μM 4-DMN-EcACP into which partner proteins (40 μM stock concentration) were individually titrated. Fluorescence was measured using a ThermoFisher VarioLux plate reader at λ_{ex} = 408 nm and λ_{em} = 512 nm over the course of 3 minutes for each titration point to allow the system to equilibrate. The value of fluorescence for each titration point was averaged over the final two and half minutes of each reading to examine only fluorescent values at equilibrium. Partner proteins were also titrated into a well of 50 μL phosphate buffer without 4-DMN-EcACP to provide a

value for baseline background fluorescence of each partner enzyme. Each titration was completed in triplicate on three separate occasions. K_d values for FabF, BioH and FabA were calculated by fitting to a one site total binding equation:

$$Fluorescence\ Increase = NS * X + \frac{B_{max} * X}{(K_d + X)} \qquad \text{(Equation 1)}$$

Where NS is the slope of nonspecific binding, B_{max} is the maximum specific binding, and X is the concentration of partner enzyme added. [a]Curves were fit using GraphPad Prism 8 and K_d reported with 95% confidence intervals.

Titration data analysis: The average of each titration point was taken for t > 30s. Each titration point was then corrected for background fluorescence by subtracting the fluorescence intensity of the partner protein in phosphate buffer without 4-DMN-EcACP from the fluorescence intensity of the partner protein with 4-DMN-EcACP. Each titration point was then normalized by dividing the fluorescence intensity of the protein plus 4-DMN-EcACP by the fluorescence intensity of the same volume of 4-DMN-EcACP alone.

Inhibitor studies: Inhibitors were added to 200 µL of their respective partner proteins at varying concentrations and incubated at room temperature with light shaking for twenty minutes. 50 µL of partner protein with inhibitor was added to a black 96-well COSTAR plate containing 50 µL of 4-DMN-EcACP. Fluorescence was measured at λ_{ex} = 408 nm and λ_{em} = 512 nm using a ThermoFisher VarioLux plate reader for twenty minutes to allow the system to equilibrate. Each data point was collected in triplicate. [b]IC_{50} values were calculated using GraphPad Prism 8 and data reported with 95% confidence intervals.

Supporting Information

SI Figures 1-13: Titration curves for each 4-DMN-EcACP·partner enzyme interaction studied. SI Figure 14: Fluorescence emission scan of 4-DMN-EcACP interacting with *apo*-FabG and FabG-$NADP^+$. SI Figure 15. Urea-PAGE analysis of *apo*-EcACP conversion to 4-DMN-EcACP. SI Figure 16. LC-EST-TOFMS analysis of *apo*-EcACP. Figure S17. LC-ESI-TOFMS analysis of 4-DMN-EcACP. Figure S18. Comparison of hydrophobicity of FabD, FabF and FabB active sites. Figure S19. SDS-PAGE of Nickel column fractions of SecA and MukB purification. Figure S20. Final preparations of purified BioH, BioF, LpxA, IscS, FabD, FabH, FabB, FabF, FabG, FabA, FabI and SecA. Figure S21. Titration of LpxA and FabD into solution of 4-DMN-phosphopantetheine.

Acknowledgements

Research reported in this publication was supported by NIH grant RO1 GM095970. K.C. was supported by the National Science Foundation Graduate Assistance in Areas of National Need (NSF GAANN) Fellowship. We also thank E. Bouveret, PhD., V. Rybenkov, PhD., P. Zhou, PhD., and D. Oliver, PhD. for providing plasmids for the expression of IscS, MukB, LpxA and SecA. We thank J.T. Mindrebo for protein purification discussion, T.D. Davis, PhD. for synthetic chemistry advice, M.H. Ostertag for assistance with data analysis, and the UCSD Molecular Mass Spectrometry Facility for LC-ESI-TOFMS analysis

References

(1) Arkin, M. R.; Tang, Y.; Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chem. Biol.* **2014**, *21* (9), 1102–1114. https://doi.org/10.1016/j.chembiol.2014.09.001.

(2) Yang, X.; Lennard, K. R.; He, C.; Walker, M. C.; Ball, A. T.; Doigneaux, C.; Tavassoli, A.; van der Donk, W. A. A Lanthipeptide Library Used to Identify a Protein-Protein Interaction Inhibitor. *Nat. Chem. Biol.* **2018**, *14* (4), 375–380. https://doi.org/10.1038/s41589-018-0008-5.

(3) Male, A. L.; Forafonov, F.; Cuda, F.; Zhang, G.; Zheng, S.; Oyston, P. C. F.; Chen, P. R.; Williamson, E. D.; Tavassoli, A. Targeting Bacillus Anthracis Toxicity with a Genetically Selected Inhibitor of the PA/CMG2 Protein-Protein Interaction. *Sci. Rep.* **2017**, *7* (1), 3104. https://doi.org/10.1038/s41598-017-03253-3.

(4) Boshoff, H. I. Caught between Two Proteins: A Mycobacterial Inhibitor Challenges the Mold. *Mol. Microbiol.* **2017**, *103* (1), 2–6. https://doi.org/10.1111/mmi.13570.

(5) Nishida, I.; Kawaguchi, A.; Yamada, M. Effect of Thiolactomycin on the Individual Enzymes of the Fatty Acid Synthase System in Escherichia Coli. *J. Biochem. (Tokyo)* **1986**, *99* (5), 1447–1454. https://doi.org/10.1093/oxfordjournals.jbchem.a135614.

(6) Hayashi, T.; Yamamoto, O.; Sasaki, H.; Kawaguchi, A.; Okazaki, H. Mechanism of Action of the Antibiotic Thiolactomycin Inhibition of Fatty Acid Synthesis of Escherichia Coli. *Biochem. Biophys. Res. Commun.* **1983**, *115* (3), 1108–1113. https://doi.org/10.1016/s0006-291x(83)80050-3.

(7) Lin, S.; Hanson, R. E.; Cronan, J. E. Biotin Synthesis Begins by Hijacking the Fatty Acid Synthetic Pathway. *Nat. Chem. Biol.* **2010**, *6* (9), 682–688. https://doi.org/10.1038/nchembio.420.

(8) Jordan, S. W.; Cronan, J. E. A New Metabolic Link. The Acyl Carrier Protein of Lipid Synthesis Donates Lipoic Acid to the Pyruvate Dehydrogenase Complex in Escherichia Coli and Mitochondria. *J. Biol. Chem.* **1997**, *272* (29), 17903–17906. https://doi.org/10.1074/jbc.272.29.17903.

(9) Issartel, J. P.; Koronakis, V.; Hughes, C. Activation of Escherichia Coli Prohaemolysin to the Mature Toxin by Acyl Carrier Protein-Dependent Fatty Acylation. *Nature* **1991**, *351* (6329), 759–761. https://doi.org/10.1038/351759a0.

(10) Brozek, K. A.; Raetz, C. R. Biosynthesis of Lipid A in Escherichia Coli. Acyl Carrier Protein-Dependent Incorporation of Laurate and Myristate. *J. Biol. Chem.* **1990**, *265* (26), 15410–15417.

(11) Battesti, A.; Bouveret, E. Acyl Carrier Protein/SpoT Interaction, the Switch Linking SpoT-Dependent Stress Response to Fatty Acid Metabolism. *Mol. Microbiol.* **2006**, *62* (4), 1048–1063. https://doi.org/10.1111/j.1365-2958.2006.05442.x.

(12) Gully, D.; Moinier, D.; Loiseau, L.; Bouveret, E. New Partners of Acyl Carrier Protein Detected in Escherichia Coli by Tandem Affinity Purification. *FEBS Lett.* **2003**, *548* (1–3), 90–96.

(13) Adachi, S.; Murakawa, Y.; Hiraga, S. Dynamic Nature of SecA and Its Associated Proteins in Escherichia Coli. *Front. Microbiol.* **2015**, *6*, 75. https://doi.org/10.3389/fmicb.2015.00075.

(14) 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.; et al. Trapping the Dynamic Acyl Carrier Protein in Fatty Acid Biosynthesis. *Nature* **2014**, *505* (7483), 427–431. https://doi.org/10.1038/nature12810.

(15) Beld, J.; Cang, H.; Burkart, M. D. Visualizing the Chain-Flipping Mechanism in Fatty-Acid Biosynthesis. *Angew. Chem. Int. Ed Engl.* **2014**, *53* (52), 14456–14461. https://doi.org/10.1002/anie.201408576.

(16) Cronan, J. E. The Chain-Flipping Mechanism of ACP (Acyl Carrier Protein)-Dependent Enzymes Appears Universal. *Biochem. J.* **2014**, *460* (2), 157–163. https://doi.org/10.1042/BJ20140239.

(17) Heath, R. J.; Rock, C. O. Inhibition of Beta-Ketoacyl-Acyl Carrier Protein Synthase III (FabH) by Acyl-Acyl Carrier Protein in Escherichia Coli. *J. Biol. Chem.* **1996**, *271* (18), 10996–11000. https://doi.org/10.1074/jbc.271.18.10996.

(18) Price, A. C.; Zhang, Y.-M.; Rock, C. O.; White, S. W. Cofactor-Induced Conformational Rearrangements Establish a Catalytically Competent Active Site and a Proton Relay Conduit in FabG. *Struct. Lond. Engl. 1993* **2004**, *12* (3), 417–428. https://doi.org/10.1016/j.str.2004.02.008.

(19) Agarwal, V.; Lin, S.; Lukk, T.; Nair, S. K.; Cronan, J. E. Structure of the Enzyme-Acyl Carrier Protein (ACP) Substrate Gatekeeper Complex Required for Biotin Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (43), 17406–17411. https://doi.org/10.1073/pnas.1207028109.

(20) Worthington, A. S.; Porter, D. F.; Burkart, M. D. Mechanism-Based Crosslinking as a Gauge for Functional Interaction of Modular Synthases. *Org. Biomol. Chem.* **2010**, *8* (8), 1769–1772. https://doi.org/10.1039/b925966j.

(21) Evans, S. E.; Williams, C.; Arthur, C. J.; Burston, S. G.; Simpson, T. J.; Crosby, J.; Crump, M. P. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *Chembiochem Eur. J. Chem. Biol.* **2008**, *9* (15), 2424–2432. https://doi.org/10.1002/cbic.200800180.

(22) Price, A. C.; Choi, K. H.; Heath, R. J.; Li, Z.; White, S. W.; Rock, C. O. Inhibition of Beta-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin. Structure and Mechanism. *J. Biol. Chem.* **2001**, *276* (9), 6551–6559. https://doi.org/10.1074/jbc.M007101200.

(23) Lundblad, R. L. A Rapid Method for the Purification of Bovine Thrombin and the Inhibition of the Purified Enzyme Wtih Phenylmethylsulfonyl Fluoride. *Biochemistry* **1971**, *10* (13), 2501–2506. https://doi.org/10.1021/bi00789a012.

(24) Kodali, S.; Galgoci, A.; Young, K.; Painter, R.; Silver, L. L.; Herath, K. B.; Singh, S. B.; Cully, D.; Barrett, J. F.; Schmatz, D.; et al. Determination of Selectivity and Efficacy of Fatty Acid Synthesis Inhibitors. *J. Biol. Chem.* **2005**, *280* (2), 1669–1677. https://doi.org/10.1074/jbc.M406848200.

(25) Snowden, M.; Green, D. V. The Impact of Diversity-Based, High-Throughput Screening on Drug Discovery: "Chance Favours the Prepared Mind." *Curr. Opin. Drug Discov. Devel.* **2008**, *11* (4), 553–558.

(26) Cochran, A. G. Antagonists of Protein-Protein Interactions. *Chem. Biol.* **2000**, *7* (4), R85-94.

(27) Carro, L. Protein-Protein Interactions in Bacteria: A Promising and Challenging Avenue towards the Discovery of New Antibiotics. *Beilstein J. Org. Chem.* **2018**, *14*, 2881–2896. https://doi.org/10.3762/bjoc.14.267.

(28) Thiele, G. A. R.; Friedman, C. P.; Tsai, K. J. S.; Beld, J.; Londergan, C. H.; Charkoudian, L. K. Acyl Carrier Protein Cyanylation Delivers a Ketoacyl Synthase-Carrier Protein Cross-Link. *Biochemistry* **2017**, *56* (20), 2533–2536. https://doi.org/10.1021/acs.biochem.7b00219.

(29) Biswas, R.; Singh, B. K.; Dutta, D.; Das, P. K.; Maiti, M. K.; Basak, A.; Das, A. K. Decrypting the Oscillating Nature of the 4'-Phosphopantetheine Arm in Acyl Carrier Protein AcpM of Mycobacterium Tuberculosis. *FEBS Lett.* **2019**, *593* (6), 622–633. https://doi.org/10.1002/1873-3468.13339.

(30) Kosa, N. M.; Haushalter, R. W.; Smith, A. R.; Burkart, M. D. Reversible Labeling of Native and Fusion-Protein Motifs. *Nat. Methods* **2012**, *9* (10), 981–984. https://doi.org/10.1038/nmeth.2175.

General experimental procedures:

Partner enzyme titration: Readings were completed using a ThermoFisher VarioLux plate reader. For each titration, 50 µL of 120 µM 4-DMN-EcACP containing 0.1% Triton-X was first dispensed into a black COSTAR 96-well plate and fluorescence was measured at λ_{ex} = 408 nm and λ_{em} = 512 nm over the course of 3 minutes. 3 µL of a partner enzyme (stock concentration 40 µM containing 0.1% Triton-X) was then added to three separate wells

of 4-DMN-EcACP and a fluorescence reading at λ_{ex} = 408 nm and λ_{em} = 512 nm taken every 3 seconds for a total of 3 minutes to allow the system to come to equilibrium. In between readings the plate was lightly shaken at 600 rpm. This process was repeated until a total of 50 µL of the partner enzyme was added to the well of 4-DMN-EcACP. Partner proteins were also titrated into a well of 50 µL phosphate buffer without 4-DMN-EcACP to provide a value for baseline background fluorescence of each partner enzyme.

To analyze the data for each partner enzyme titration into 4-DMN-EcACP, the average of each titration point was first taken for t > 30s. Then, each titration point was corrected for intrinsic fluorescence of partner enzyme by subtracting the fluorescence intensity of the same volume of partner enzyme in 50 µL of phosphate buffer. Each titration point was then normalized by dividing the fluorescence intensity of the well containing the partner enzyme plus 50 µL of 4-DMN-EcACP by the fluorescence intensity of the just 50 µL of 4-DMN-EcACP to generate the “relative A.U.” value for each titration point. Titration plots were generated by plotting the relative fluorescence intensity A.U. of each titration point vs. the concentration of partner enzyme at that point.

Supporting Information

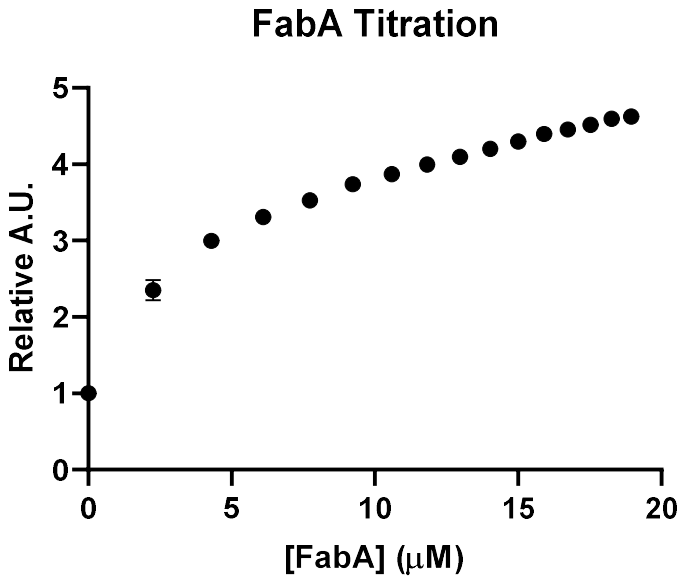


Figure S2.1: Titration of FabA against 4-DMN-EcACP.

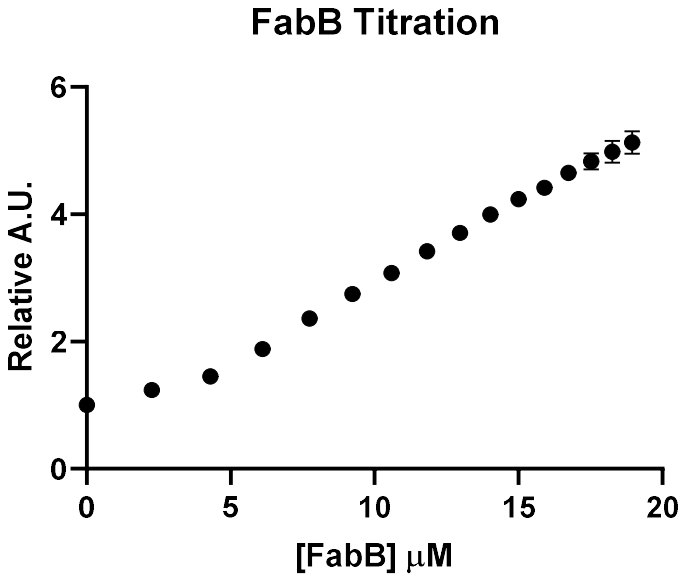
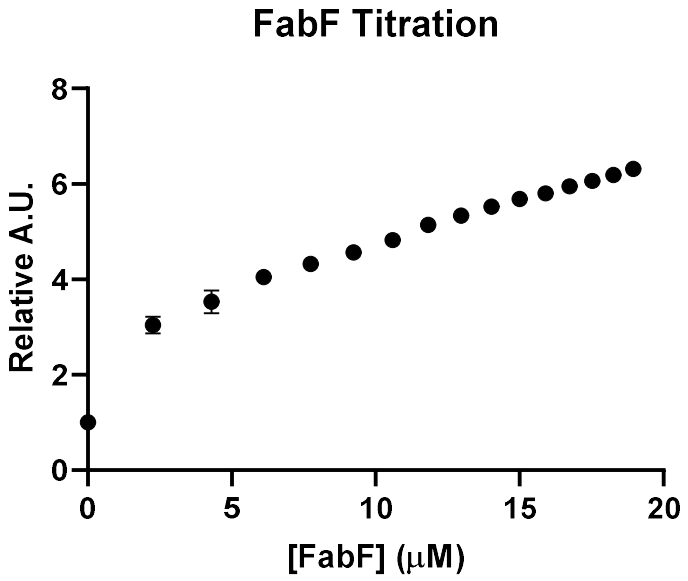


Figure S2.2: Titration of FabB against 4-DMN-EcACP.



S

Figure S2.3: Titration of FabF against 4-DMN-EcACP

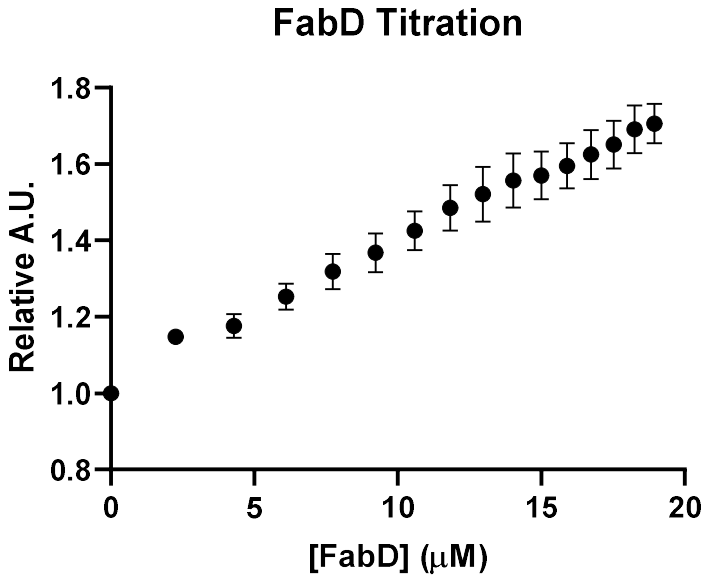


Figure S2.4: Titration of FabD against 4-DMN-EcACP.

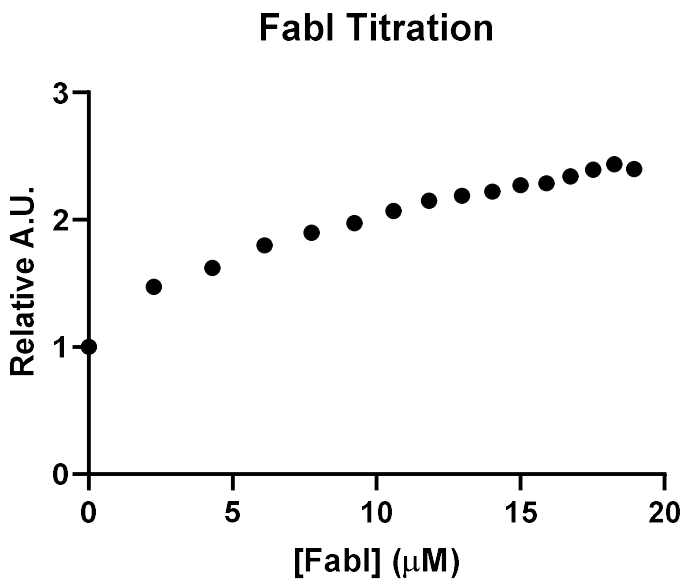


Figure S2.5: Titration of FabI against 4-DMN-EcACP.

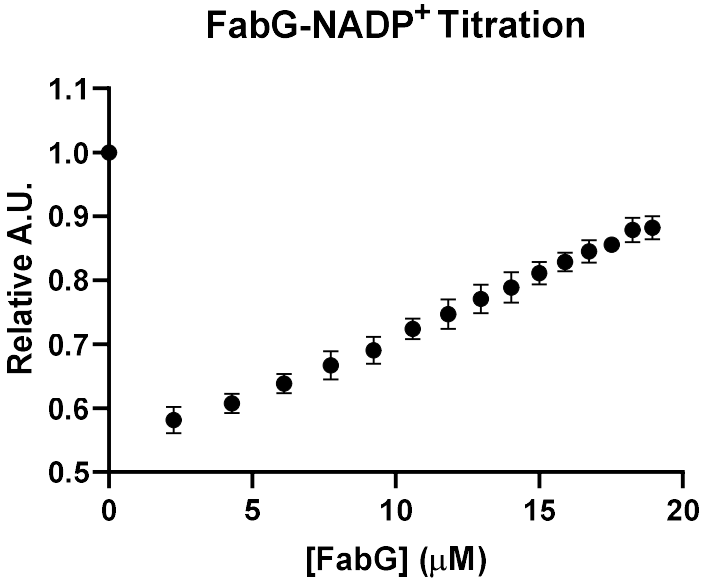


Figure S2.6: Titration of FabG against 4-DMN-EcACP. FabG contains 1.0 equivalent of $NADP^+$.

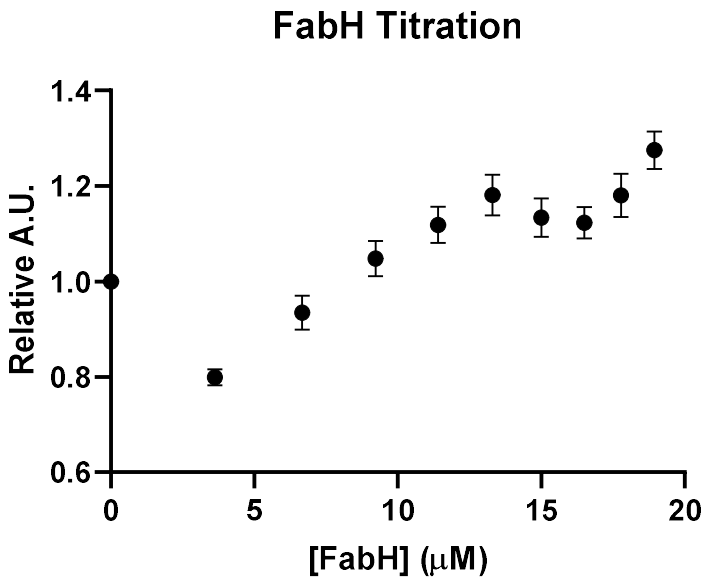


Figure S2.7: Titration of FabH against 4-DMN-EcACP.

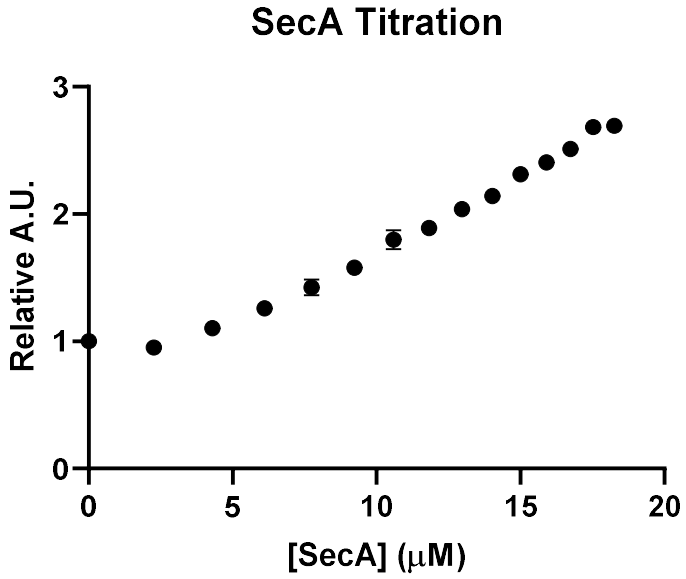


Figure S2.8: Titration of SecA against 4-DMN-EcACP.

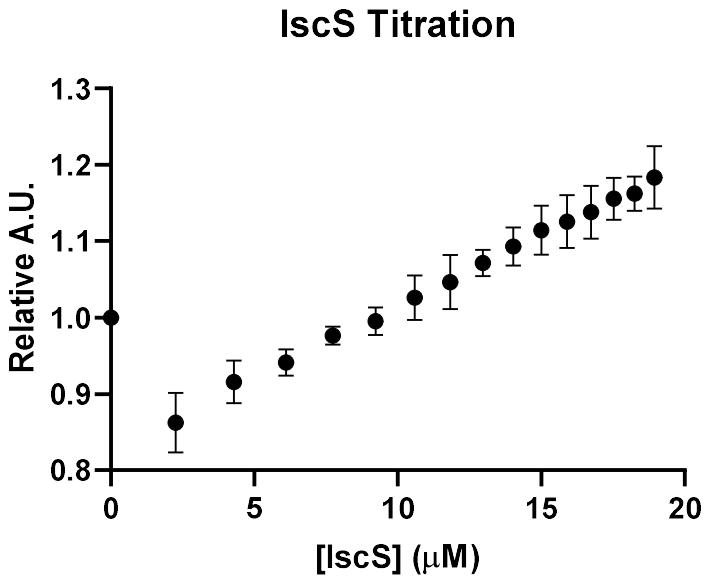


Figure S2.9: Titration of IscS against 4-DMN-EcACP.

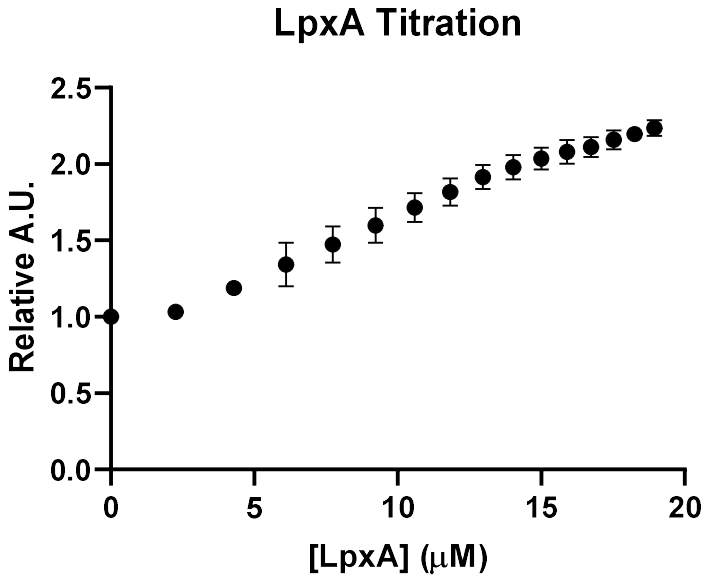


Figure S2.10: Titration of LpxA against 4-DMN-EcACP.

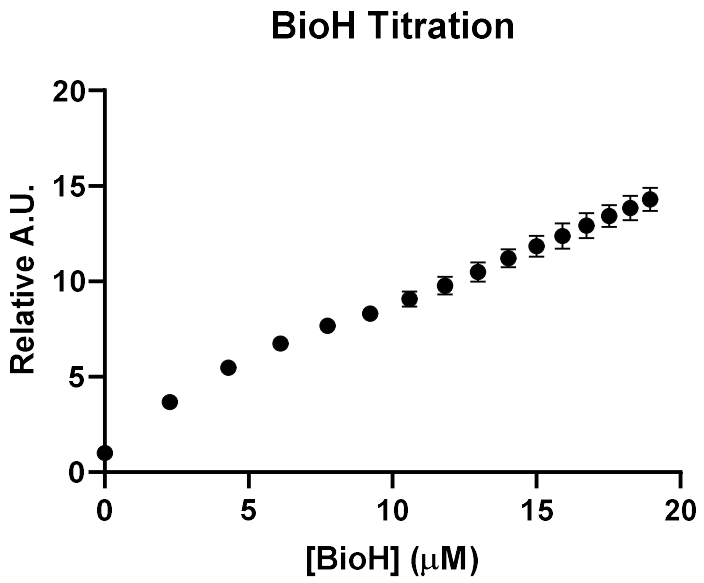


Figure S2.11: Titration of BioH against 4-DMN-EcACP.

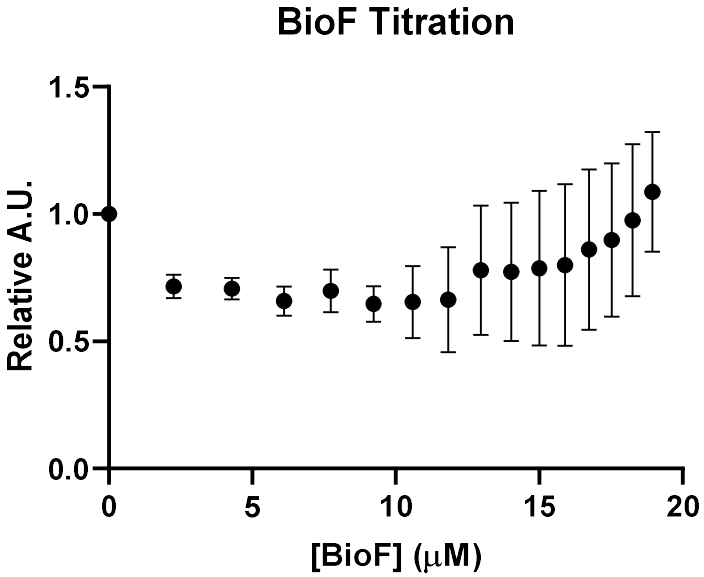


Figure S2.12: Titration of BioF against 4-DMN-EcACP. BioF contains a pyridoxal phosphate cofactor that strongly fluoresces under the conditions of this assay leading to a difficulty distinguishing solvatochromic response of 4-DMN vs. intrinsic fluorescence of BioF.

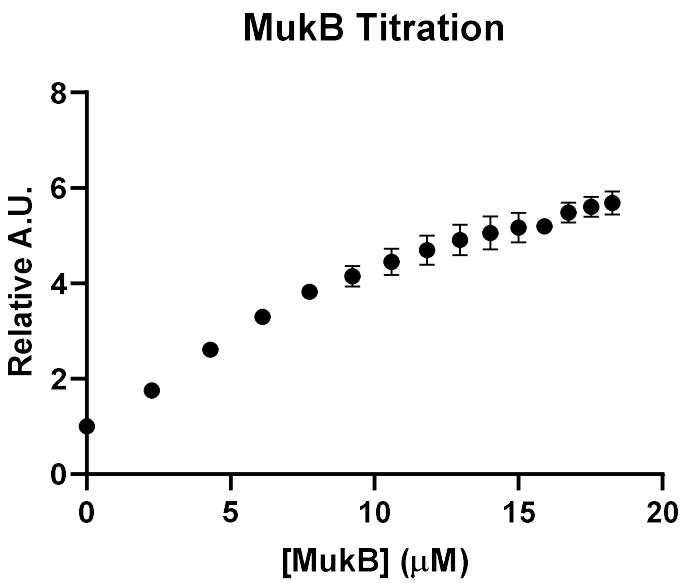


Figure S2.13: Titration of MukB against 4-DMN-EcACP.

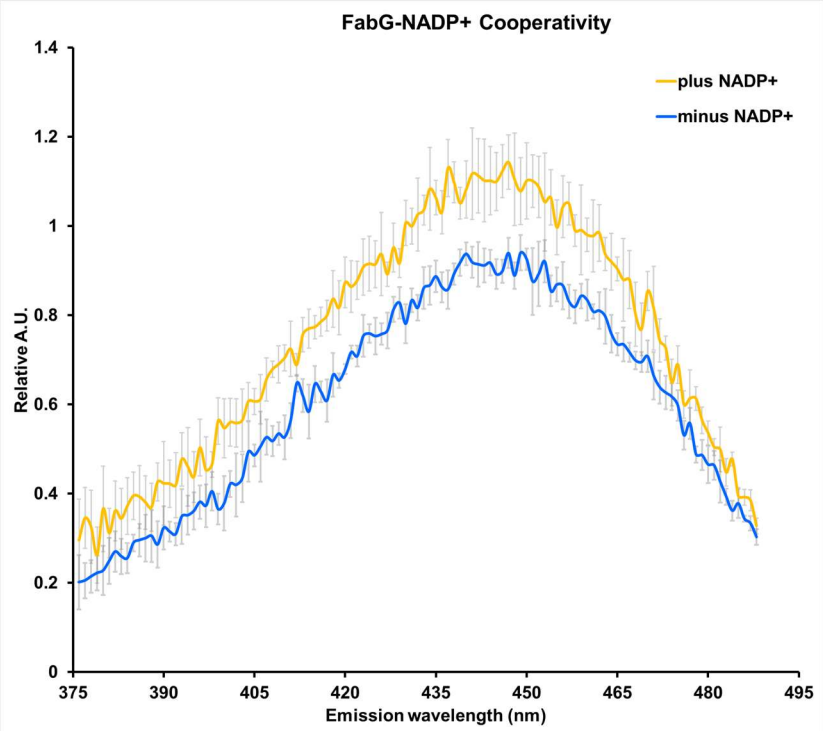


Figure S2.14: Effect of $NADP^+$ cofactor binding on solvatochromic response between 4-DMN-EcACP and FabG. The presence of $NADP^+$ causes a conformational change in FabG, creating a new tunnel for the ACP cargo, which places the 4-DMN cargo in a more hydrophobic environment leading to a larger increase in fluorescence intensity. 50 µM of FabG was incubated with 50 µM of $NADP^+$ for ten minutes at 0°C and then added to 30 µL of 45 µM 4-DMN-EcACP. FabG at 50 µM without $NADP^+$ was added to a separate aliquot of 30 µL of 45 µM 4-DMN-EcACP. The fluorescence at λ_{ex} = 408 nm and λ_{em} = 512 nm was measured over the time course of 4 minutes to allow the system to come to equilibrium. After equilibration, the emission spectrum of the 4-DMN-EcACP-FabG complex in both the presence and absence of $NADP^+$ was measured (λ_{ex} = 408 nm). Data were collected in triplicate. The emission spectrum at λ_{ex} = 408 nm was also taken of FabG (without 4-DMN-EcACP) in the presence and absence of $NADP^+$. The spectrum of FabG containing $NADP^+$ (without 4-DMN-EcACP) was subtracted from the spectrum of each 4-DMN-EcACP·FabG containing $NADP^+$ replicate, which was then normalized to the max fluorescence A.U. of the samples in the absence of $NADP^+$. The same data processing was completed for samples in the absence of $NADP^+$.

Figure S2.15: 20% acrylamide urea-PAGE run at 150V for 2 hours and15 minutes stained with Coomassie blue showing conversion of *apo*-EcACP (right lane) to 4-DMN-EcACP (left lane). The installation of the 4-DMN probe onto EcACP results in a conformational shift to a more compact species, which causes the 4-DMN-EcACP to run further in a urea-PAGE experiment.

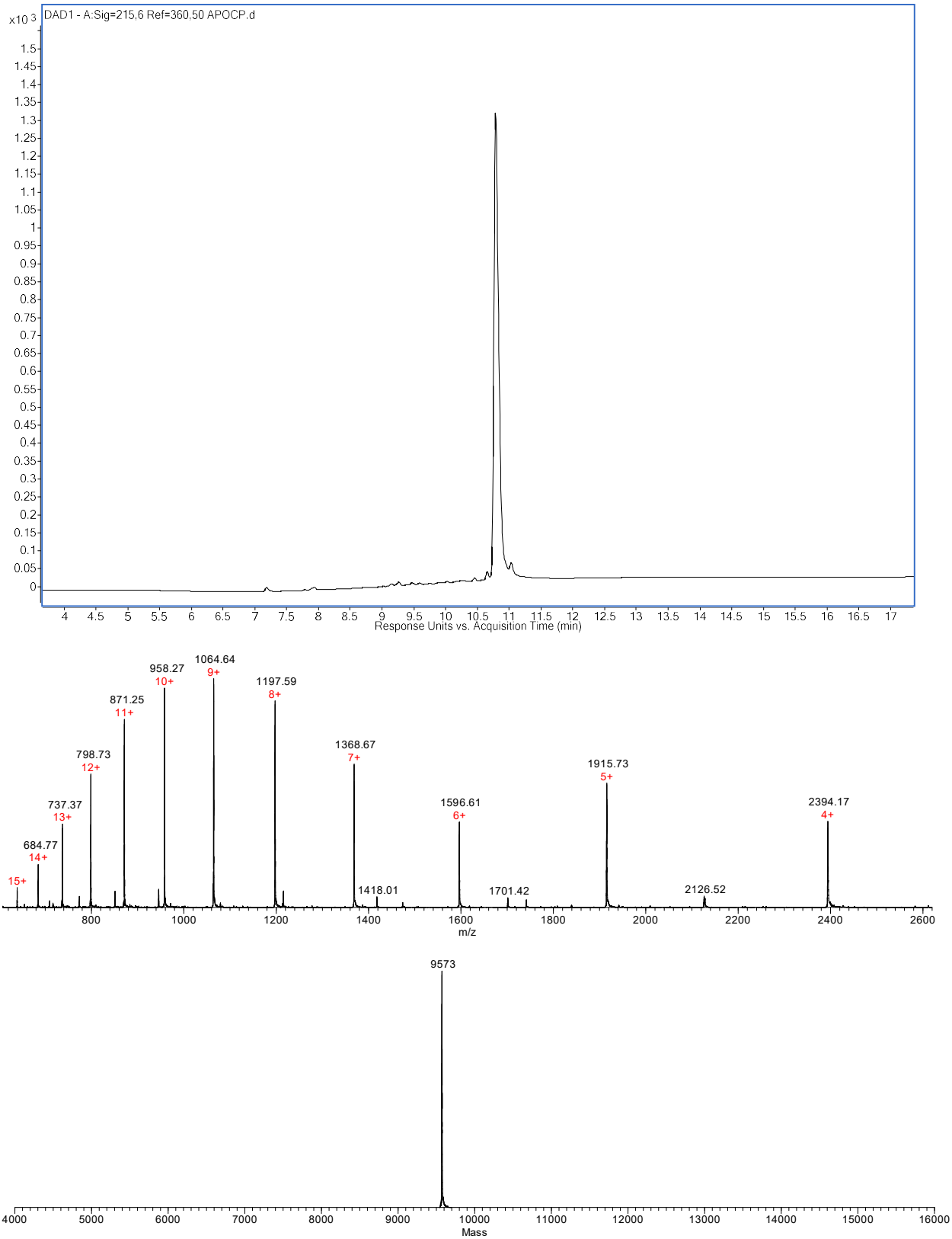
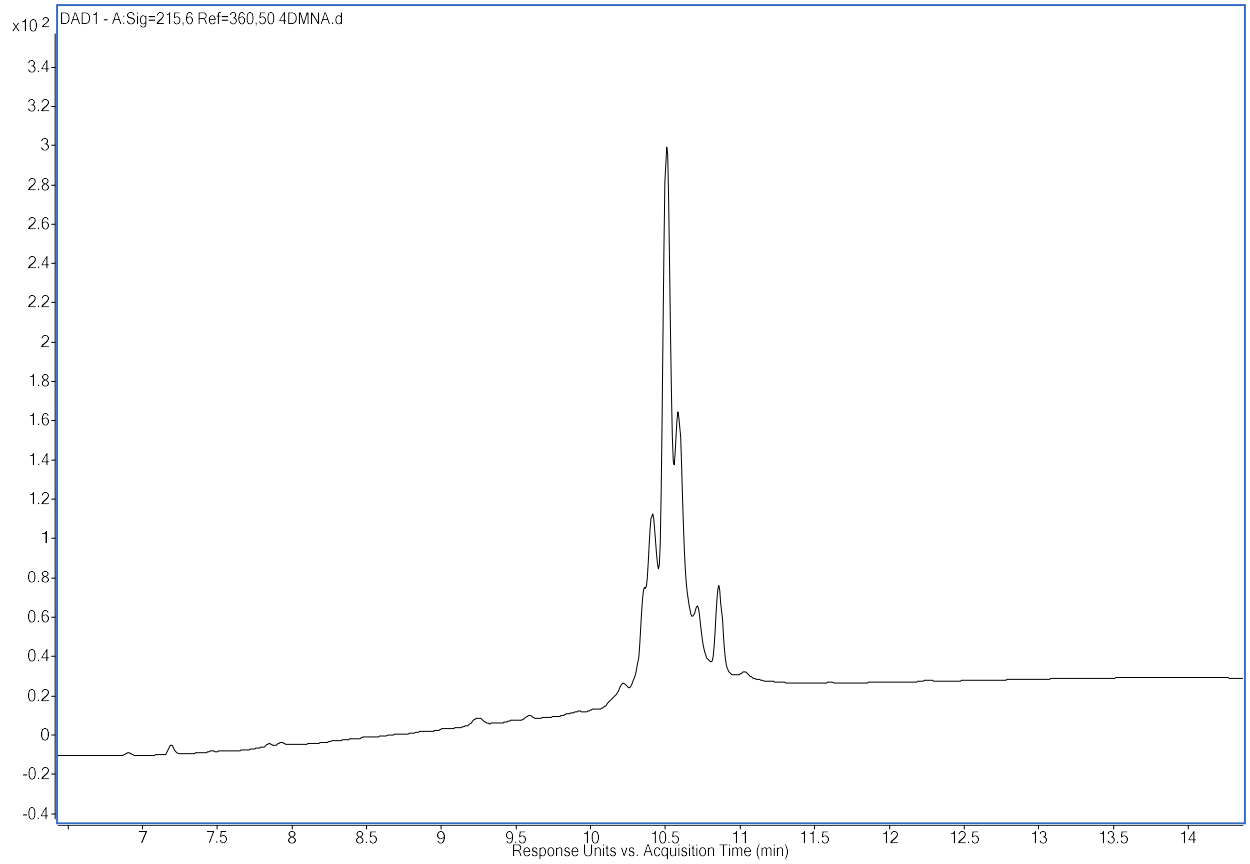
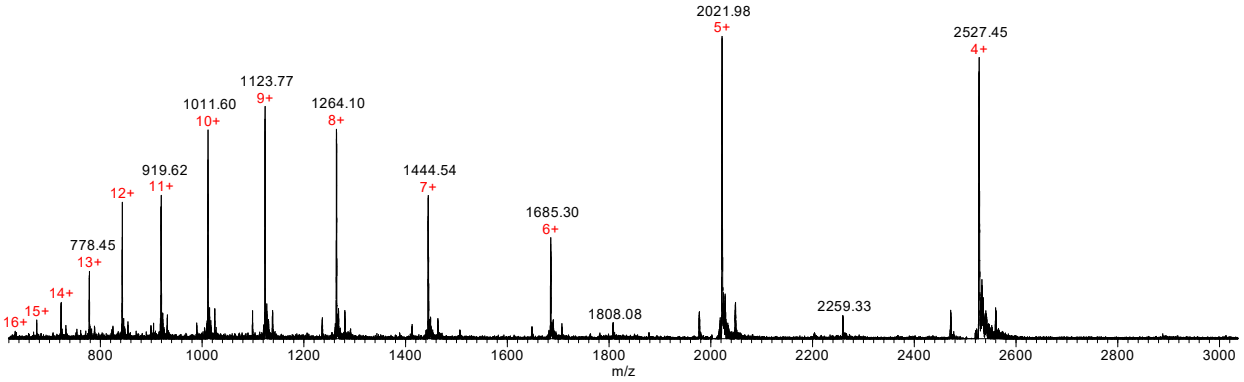


Figure S2.16: LC trace, ESI-TOFMS spectrum and deconvoluted mass spectrum of *apo*-EcACP. An Agilent 6230 time-of-flight mass spectrometer (TOFMS) coupled with an

Agilent 1260 HPLC was used for LC-ESI-TOFMS analysis. The chromatographic separation was performed under room temperature on a Phenomenex Aeris Widepore XB-C18 column (50 mm length, 2.1 mm ID, and 3.6 um particle size). Mobile phase A was HPLC-grade water with 0.1% TFA, and HPLC grade Acetonitrile with 0.1% TFA was used as mobile phase B. The mobile phase was delivered at a rate of 0.3 ml/min under gradient conditions as follows: held at 5% mobile phase B for two minutes, then increased to 90% mobile phase B in 10 minutes, held at 90% mobile phase B for 2 minutes, returned to 5% mobile phase B in 1 minute, and equilibrated with 5% mobile phase A for 7 minutes. Agilent MassHunter software was used for data acquisition and processing, and MagTran software[1] was used for MS data deconvolution.


DAD1 - A:Sig=215,6 Ref=360,50 4DMNA.d
x10 2
Response Units vs. Acquisition Time (min)


2021.98
5+
2527.45
4+
1123.77
9+
1011.60
10+
1264.10
8+
919.62
11+
12+
1444.54
7+
1685.30
6+
778.45
13+
14+
15+
16+
1808.08
2259.33
m/z

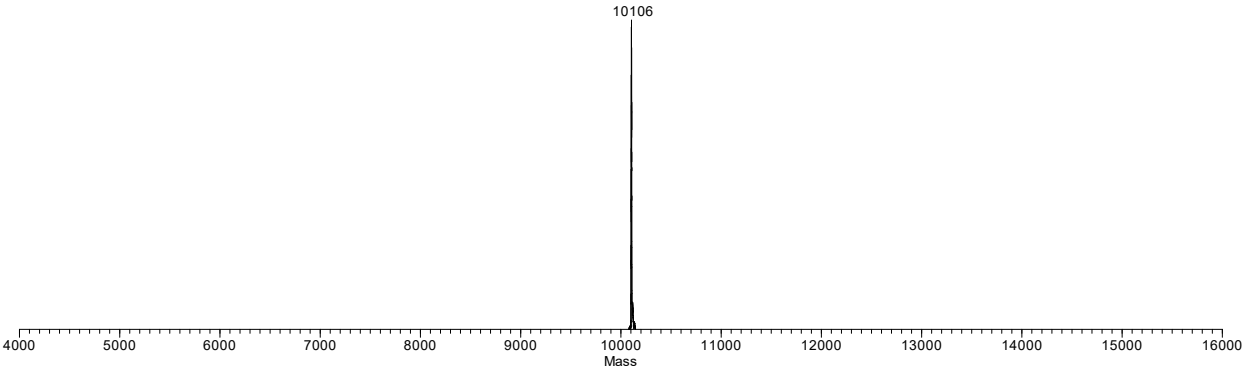

10106
Mass

Figure S2.17: LC trace, ESI-TOFMS spectrum and deconvoluted mass spectrum of 4-DMN-EcACP. An Agilent 6230 time-of-flight mass spectrometer (TOFMS) coupled with an Agilent 1260 HPLC was used for LC-ESI-TOFMS analysis. The chromatographic separation was performed under room temperature on a Phenomenex Aeris Widepore XB-C18 column (50 mm length, 2.1 mm ID, and 3.6 um particle size). Mobile phase A was HPLC-grade water with 0.1% TFA, and HPLC grade Acetonitrile with 0.1% TFA was used as mobile phase B. The mobile phase was delivered at a rate of 0.3 ml/min under gradient conditions as follows: held at 5% mobile phase B for two minutes, then increased to 90% mobile phase B in 10 minutes, held at 90% mobile phase B for 2 minutes, returned to 5% mobile phase B in 1 minute, and equilibrated with 5% mobile phase A for 7 minutes. Agilent MassHunter software was used for data acquisition and processing, and MagTran software[1] was used for MS data deconvolution.

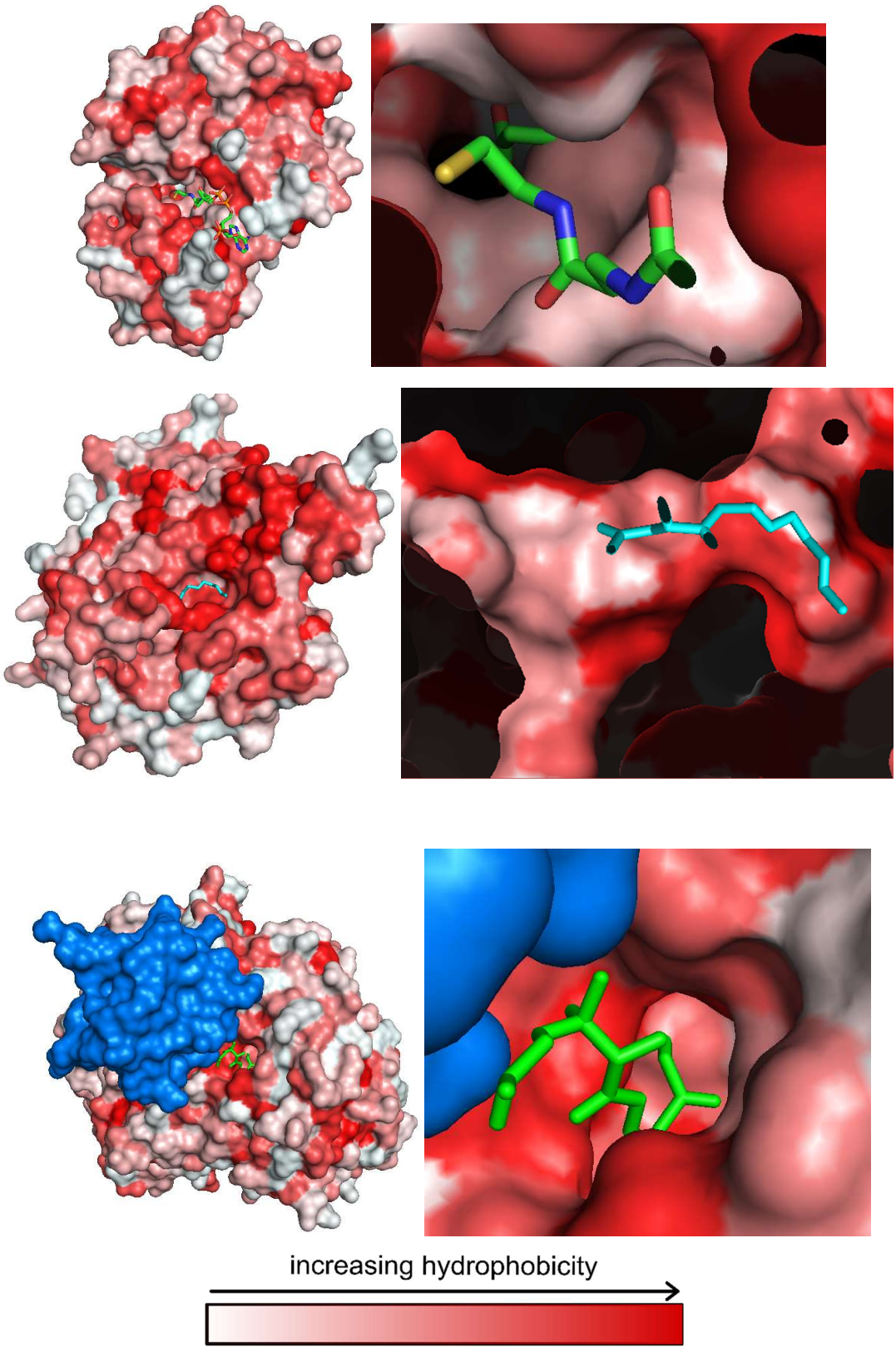


Figure S2.18. Heat map of EcACP partner enzymes and zoomed in images of the active sites of the malonyl-CoA acyl transferase in complex with malonyl-CoA (FabD, PDB:

2G2Z), ketosynthase II (FabF, PDB: 1B3N) in complex with active site inhibitor cerulenin, and ketosynthase I (FabB, PDB: 5KOF) in complex with acyl carrier protein (blue) and crosslinker (green) colored by hydrophobicity using PyMol color_h script. Ligands in the active site of each protein are shown for clarity of active site location. While the active site of FabD is more hydrophobic than the helical bundle of 4-DMN-EcACP, as evidenced by the increase in fluorescence intensity when 4-DMN-EcACP chain flips the 4-DMN cargo into the active site of FabD, titration with FabB or FabF results in an even larger fluorescence, suggesting that FabB and FabF feature active sites that are even more hydrophobic than FabD. Coloration of each active site by hydrophobicity confirms this trend: mildly hydrophobic residues, colored light pink, line the active site of FabD, while deep red coloration in the active sites of FabB and FabF reveal a more hydrophobic environment.

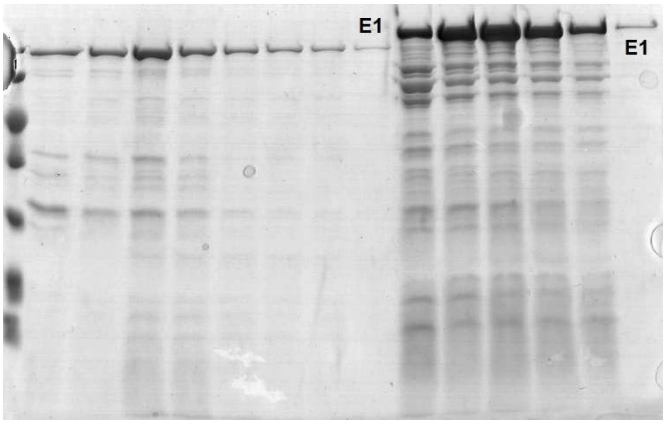


Figure S2.19. 12% SDS-PAGE of Nickel column fractions of SecA (left, 102 kDa) and MukB (right, 170 kDa). The gel was run at 160V for 65m and stained using Coomassie blue stain. Fractions labeled E1 were concentrated and used without further purification.

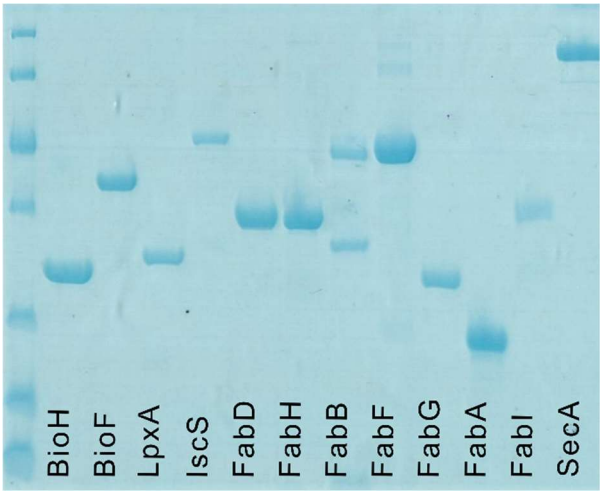


Figure S2.20: 12% SDS-PAGE of FPLC purified fractions of BioF, BioH, IscS, LpxA, FabD, FabH, FabF, FabG, FabA, FabI and SecA. The gel was run at 160V for 65m and stained using Coomassie blue stain.

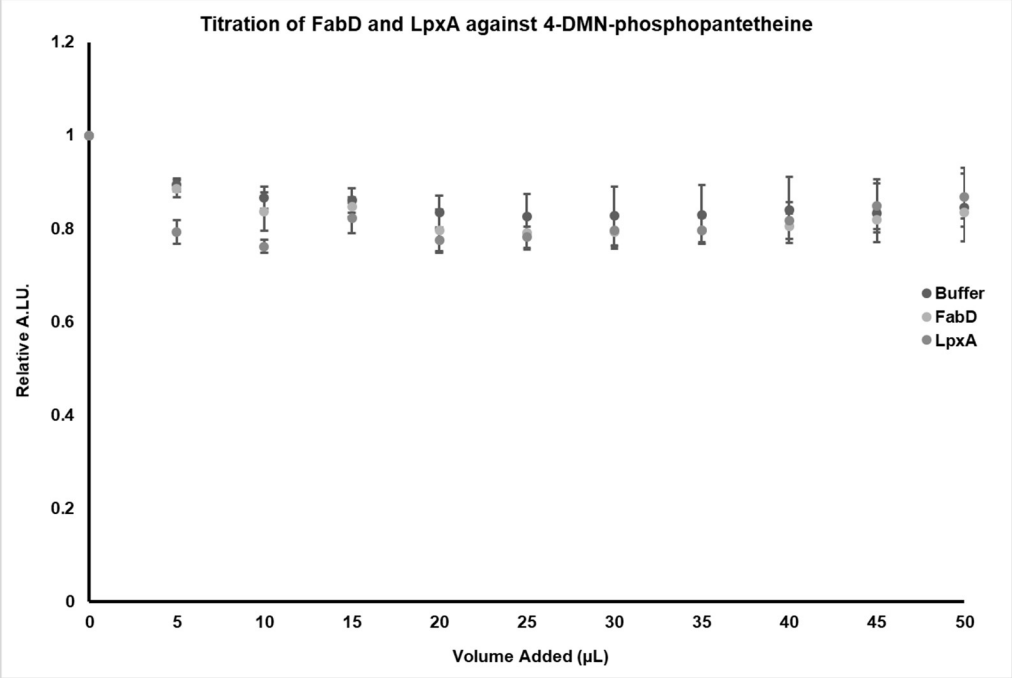


Figure S2.21. Titrations of LpxA and FabD with 4-DMN-phosphopantetheine were completed identically to titrations with 4-DMN-EcACP. Readings were completed using a ThermoFisher VarioLux plate reader. For each titration, 50 µL of 4-DMN-phosphopantetheine, previously purified by FPLC, dissolved in 50 mM phosphate buffer, pH 8.0 was first dispensed into a black COSTAR 96-well plate and fluorescence was measured at λ_{ex} = 408 nm and λ_{em} = 512 nm over the course of 3 minutes. Next, 5 µL of a partner enzyme (stock concentration 40 µM containing 0.1% Triton-X), or just buffer, was then added to three separate wells of 4-DMN-phosphopantetheine and a fluorescence reading at λ_{ex} = 408 nm and λ_{em} = 512 nm taken every 3 seconds for a total of 3 minutes to allow the system to come to equilibrium. In between readings the plate was lightly shaken at 600 rpm. This process was repeated until a total of 50 µL of the partner enzyme was added to the well of 4-DMN-phosphopantetheine.

To analyze the data for each partner enzyme titration into 4-DMN-phosphopantetheine, the average of each titration point was first taken for t > 30s. Then, each titration point was corrected for intrinsic fluorescence of partner enzyme by subtracting the fluorescence

intensity of the same volume of partner enzyme in 50 µL of phosphate buffer. Each titration point was then normalized by dividing the fluorescence intensity of the well containing the partner enzyme plus 50 µL of 4-DMN-phosphopantetheine by the fluorescence intensity of the just 50 µL of 4-DMN-phosphopantetheine to generate the "relative A.U." value for each titration point. Titration plots were generated by plotting the relative fluorescence intensity A.U. of each titration point vs. the concentration of partner enzyme at that point.

Supplementary References:

1. **Z. Zhang and A.G. Marshall, *J. Am. Soc. Mass Spectrom*., 1998, 9, 225-233.**

Acknowledgements:

Chapter 2, in full, is a reprint of the material as it appears: Charov, K., Burkart, M.D. "A Single Tool to Monitor Multiple Protein–Protein Interactions of the Escherichia coli Acyl Carrier Protein." *ACS Infectious Diseases* (2019): 5, 9, 1518-1523. The dissertation author is a contributing author of this manuscript.

Chapter 3. *In silico* identification and *in vitro* evaluation of a protein-protein interaction inhibitor of *Escherichia coli* fatty acid biosynthesis

To combat the rise in antibiotic resistance, new targets must be identified and probes against them developed. Protein-protein interactions (PPI) of bacterial type II fatty acid biosynthesis (FAS-II) represent an untapped, yet rich area for new antibiotic discovery. Here, we present a computational and in vitro workflow for the discovery of new inhibitors of PPI in Escherichia coli FAS-II. As part of this study, we identified suramin, an existing treatment for African sleeping sickness, to effectively block the interaction of E. coli dehydratase FabA and the acyl carrier protein EcACP, with an IC_{50} = 85 μM. This finding validates a workflow that combines in silico screening with in vitro PPI assays to identify probes appropriate for further optimization.

Antibiotic resistance is a critical challenge in drug discovery today. The US Centers for Disease Control and Prevention estimates that each year, 2.8 million people are infected with multi-drug resistant (MDR) bacteria, resulting in over 35,000 deaths.[1] The continuing evolution and spread of MDR bacteria make the development of new antibiotic agents an urgent necessity.

Protein-protein interactions (PPIs) are emerging as a new target for antibiotic development.[2] PPIs are often highly species specific making them attractive for the development of narrow spectrum antibiotics. Although PPIs were once thought of as undruggable due to their large, diffuse surfaces and lack of distinct binding pockets, the recent successes of selective PPI inhibitors have spurred renewed interest in the area. The interaction between the *E. coli* cell-division protein ZipA and FtsZ has been the target of numerous antibiotic discovery studies.[3–5] Inhibitors of the bacterial sliding clamp have

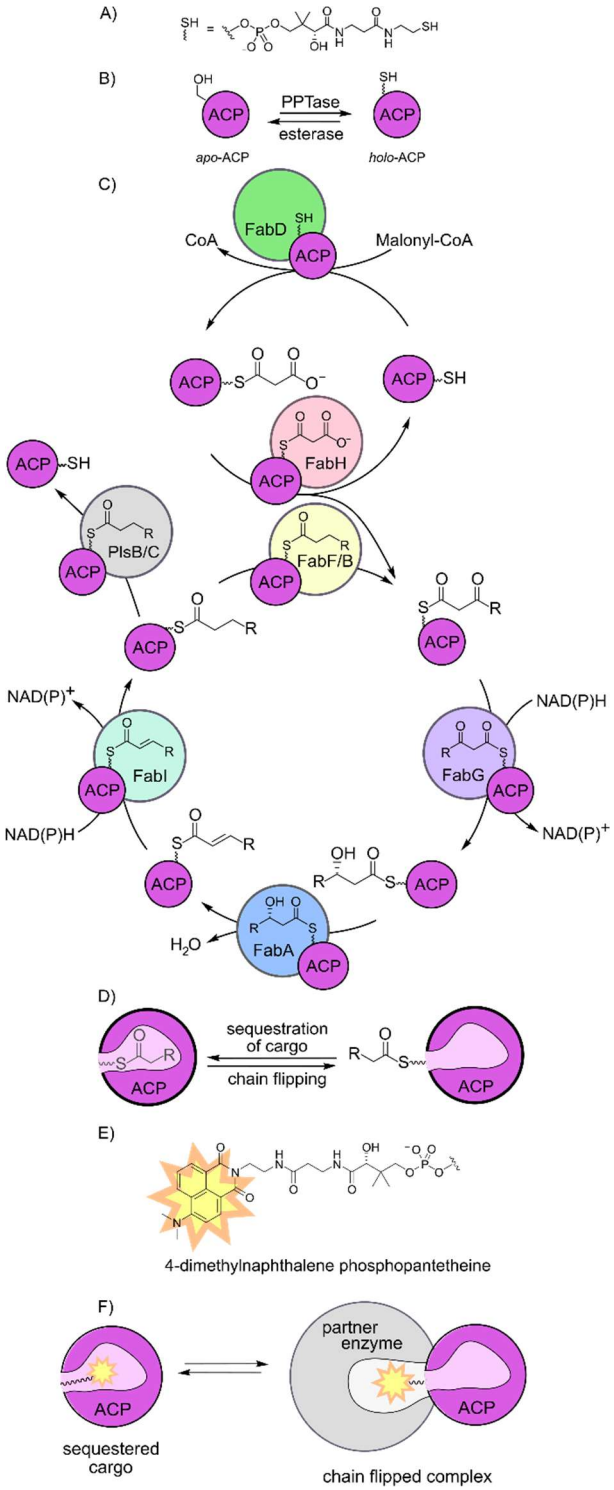


Figure 3.1. Schematic representation of E. coli FAS-II. A) Structure of phosphopantetheine (PPant) arm. B) EcACP is expressed in an inactive apo form and is post-translationally modified by a phosphopantetheinyl transferase (PPTase) to form holo-EcACP. A phosphodiesterase cleaves off the PPant arm. The terminal thiol of holo-EcACP forms a thioester bond to acyl substrates during C) the cycle of type II fatty acid biosynthesis. D) EcACP sequesters cargo within its alpha helical bundle. E) Structure of 4-dimethylnaphthalene-phosphopantetheine probe used in this study. F) Schematic of the chain flipping mechanism between 4-DMN-ACP and partner enzymes.

also proven that targeting PPIs in bacteria is a rich source for new antibiotic development.[6–8]

PPIs play a significant role in bacterial fatty acid biosynthesis. Most bacteria biosynthesize fatty acids via a type II fatty acid synthase (FAS-II) in which each enzyme of the pathway is a discrete unit, in contrast to the eukaryotic type I system (FAS-I) in which the entire fatty acid synthase is comprised as one or two large proteins. This difference has already made bacterial FAS-II a prime target for antibiotics, as evidenced by the discovery and application of triclosan, cerulenin and platensimycin, among many others.[9,10] Furthermore, the PPIs of FAS-II are by necessity orthogonal to those of FAS-I used by eukaryotes because of the large difference in enzymatic architecture (discrete units of FAS-II versus a single polypeptide of FAS-I). As such, FAS-II PPI inhibitors are less likely to interfere with the PPIs of mammalian FAS-I, which has not always been the case for active site FAS inhibitors such as cerulenin which is known to inhibit both FAS-II and FAS-I, rendering it clinically irrelevant. However, while FAS-II has been the subject of many drug discovery efforts, no known PPI inhibitors of FAS-II have been validated or brought to market. The work presented here develops the foundations of a FAS-II PPI inhibitor discovery pipeline.

Central to FAS-II is the acyl carrier protein (ACP), which shuttles the growing acyl chain to each enzyme of the pathway, engaging in a minimum of nine PPIs along the de novo FAS-II pathway (Fig. 1C). ACP has also been implicated in an additional 18 PPIs, involved in iron-sulfur cluster biosynthesis, Lipid A and biotin biosynthesis and chromosome organization.[11–17] Each of these PPIs represents a potential druggable interaction, as each is critical to bacterial survival.

Aiding in the discovery and design of inhibitors, the past decade has seen a large advance in the structural knowledge of the FAS-II system. Our group and others have succeeded in characterizing many *E. coli* ACP (EcACP) protein-protein interactions using X-ray crystallography, protein NMR and other biophysical techniques.[18–26] In addition to the wealth of structural knowledge of EcACP·partner enzyme interactions, we also recently published a fluorescent-based in vitro assay to rapidly monitor PPIs between EcACP and thirteen different partner enzymes.[27,28]

We reasoned that structural data from previous studies along with our in vitro assay could be leveraged to search for new inhibitors of the PPIs between EcACP and its partner enzymes. In the present study, we used a structure-based virtual screen to identify the African sleeping sickness treatment, suramin, as a protein-protein interaction inhibitor of the EcACP · 3-hydroxydecanoyl-[acyl-carrier protein] dehydratase (FabA) interaction. Suramin was then evaluated for PPI inhibition using a recently developed in vitro assay and followed with a secondary crosslinking assay. Our results represent the power of using highly detailed structural information to guide the discovery of new PPI inhibitors and provide a platform from which to develop a high-throughput screening workflow.

Results and discussion

Structural analysis of EcACP-FabA complex. We began our study by identifying hot-spot residues on EcACP and FabA through analysis of previously solved X-ray crystallography and protein nuclear magnetic resonance structures of the EcACP-FabA complex.[23] Hot-spot residues are defined as residues that when mutated to alanine result in a change in the free energy of PPI binding by at least 1.5 kcal/mol.[29] These residues

are typically a small subset of the total interacting residues on the surface of each protein and derive from either salt bridges or hydrophobic interactions. Knowledge of hot-spot residues is important in the context of PPI inhibitor design. Of the many successful PPI inhibitors reported, the majority of these compounds make contact with hot-spot residues and inhibit the complex formation by disrupting these critical binding sites.[30]

The X-ray crystal structure of EcACP crosslinked with FabA (PDB: 4KEH) showed that acidic residues on the surface of EcACP interact with basic residues on FabA to form salt bridges, as well as hydrophobic interactions between the two proteins. Heteronuclear single quantum coherence (HSQC) NMR experiments further confirmed the importance of these residues to binding of EcACP to FabA.[23] We anticipated that a computational search for compounds that interact with these residues could identify inhibitors of the EcACP · FabA interaction.

Virtual screening. We next turned to structure-based virtual screening to identify lead compounds for inhibition of the EcACP · FabA PPI using the Molecular Operating Environment (MOE) pharmacophore query.[31] The MOE pharmacophore query allows the user to define binding sites and specific interactions between docked ligands and a target protein. We aimed to virtually dock a panel of compounds on the surface of FabA or EcACP. Furthermore, each compound was docked by MOE making key interactions with the hot spot residues responsible for EcACP · FabA PPI. By docking compounds with hot-spot residues of FabA or EcACP, we aimed to identify compounds that may disrupt the binding of FabA and EcACP. R132, R136 and K161 residues on the surface of FabA, and E41 and E47 on the surface of EcACP were chosen as preferred sites for docking of small molecules. A panel of 200 natural products was then docked and evaluated for

interaction in this defined space. The screen identified five compounds (Fig. S1) that participate in contacts on the surface of EcACP and FabA. Of the five, the African sleeping sickness treatment, suramin (Fig. 2a), was selected for further evaluation due to commercial availability.

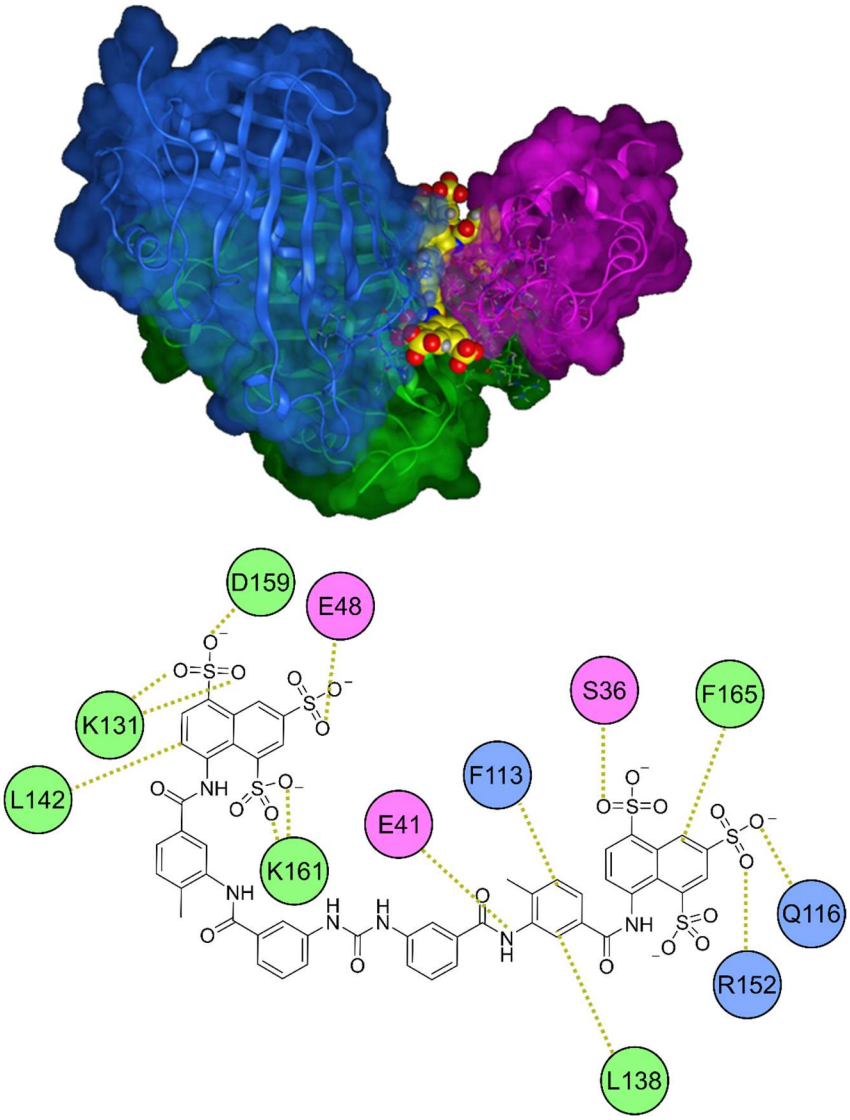


Figure 3.2. a) 3D docked structure of suramin (spheres) on the surface of FabA dimer (blue, green) and EcACP (magenta). EcACP contacts both FabA dimers during chain flipping. b) Representation of surface residues of FabA (blue, green) and ACP (magenta) interacting with suramin predicted by virtual screening with MOE. Contacts made are majority electrostatic and mediated by the poly-sulfonate groups present on suramin. Hydrophobic residues on the surface of FabA also participate in binding by interacting with the aromatic rings of suramin. K131, L138, L142, D159, K161, F165 belong to one FabA monomer, while F113, Q116 and R152 are residues from the other FabA monomer. K161 of FabA and E41 of ACP were previously identified by X-ray crystallography and NMR studies as important to PPI between the proteins.

Evaluation of suramin docking. Suramin is a poly-sulfonated urea that makes multiple contacts on the surface of both EcACP and FabA (Fig. 2a). On the surface of the FabA homodimer suramin engages residues K131, D159 and K161 of one monomer and R152, Q116 of the second monomer. Notably, K161 of FabA forms a salt bridge with E47 of EcACP, but engaged suramin via a salt bridge, thereby disrupting a critical binding mode with EcACP. Similarly, another targeted hot-spot residue of EcACP, E41, interacts with an amide nitrogen of suramin rather than forming a salt bridge with R132 of FabA. Suramin also makes contacts with other surface residues that were not part of the original pharmacophore query: E48 and S36 of EcACP and K131, D159, R152 and Q116 of FabA bind suramin through salt brides with the multiple other sulfonate groups.

In vitro evaluation. We recently reported a fluorescent-based assay for visualizing PPIs between the EcACP and thirteen of its partner enzymes (Fig. 1e-f).[27,28] We utilized this assay to further evaluate the inhibitory effect of suramin on the EcACP · FabA interaction. Briefly, EcACP is appended with a solvatochromic dye, 4-dimethylnaphthalene (4-DMN, Fig. 1e), which is responsive the hydrophobicity of its environment. This new species is termed 4-DMN-EcACP. When FabA, a partner enzyme of EcACP, interacts with 4-DMN-EcACP, the 4-DMN cargo is transferred int the active site of FabA, via a universal chain flipping mechanism used by EcACP for transferring cargo to partner enzymes (Fig. 1f).[34] The active site of FabA is more hydrophobic than the alpha helical core of EcACP and so this transfer of 4-DMN into the FabA active site causes a large increase in fluorescence intensity that is easily monitored with a plate reader or fluorimeter.

To study inhibition of the 4-DMN-EcACP · FabA by suramin, FabA was first incubated with a range of concentrations of suramin and then added to 4-DMN-EcACP. The disruption of PPIs between EcACP and FabA by suramin lead to a decrease in fluorescence intensity relative to the interaction in the absence of the inhibitor. We calculated an IC50 of 78.73 - 91.66 μM (95% confidence intervals) for suramin against the 4-DMN-EcACP · FabA interaction (Fig. 3a).

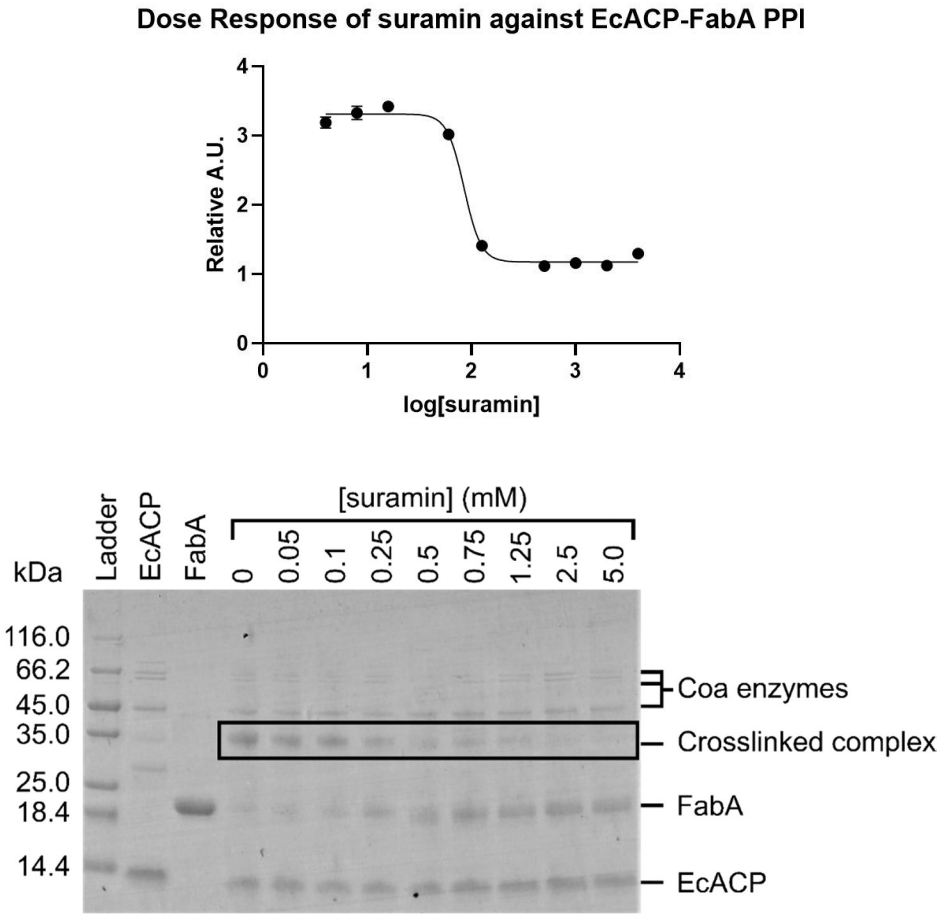


Figure 3.3. A) A dose response curve of suramin inhibitory effect against the EcACP·FabA interaction determined the IC_{50} of suramin to be 85.00 (78.73, 91.66) μM. Data were collected in triplicate and reported with 95% confidence intervals. B) A secondary crosslinking assay confirms mode of action. EcACP is loaded with a reactive warhead that, in the absence of PPI inhibitor, forms a covalent bond in the active site of FabA thereby crosslinking EcACP and FabA (lane 4). ACP is added in excess at a concentration of 60 μM and FabA at a concentration of 10 μM, with a total reaction volume of 10 μL. As the concentration of suramin is increased, crosslinking decreases with 1.25 mM of suramin was sufficient to block crosslinking. Crosslinking experiments were performed in triplicate. A dose response curve generated from crosslinking experiments is shown in Figure S10.

Secondary assay confirmation. For secondary validation, we next turned to our previous work using a crosslinking based assay to orthogonally confirm the PPI inhibiting activity of suramin (Fig. 3b).[32] Here, EcACP is loaded with a phosphopantetheine analogue which bears a mechanism-based reactive warhead, referred to as sulfonyl alkyne-EcACP, that forms a covalent bond with the active site H70 on FabA resulting in a crosslinked complex, EcACP=FabA (Fig. S6). We have previously demonstrated the utility of this crosslinking assay to confirm active site hits from a high-throughput screen.[32] Because successful crosslinking between EcACP and FabA is dependent on the proteins engaging in correct PPI, we hypothesized that suramin would block crosslinking by inhibiting the EcACP · FabA interaction as a PPI inhibitor, which could easily be detected by a gel-based assay.

To test the action of suramin against crosslinking, we incubated sulfonyl alkyne-EcACP with FabA and a range of suramin concentrations overnight at 37 °C and then analyzed the resulting crosslinking reactions by 12% SDS-PAGE (Fig. 3b). Experiments were completed in triplicate, additional gels are included in Supplementary Information Figure S8. We found that a 1.25 mM concentration of suramin was sufficient to block 100% of crosslinking. To compare the results of the inhibition of crosslinking to the inhibition of chain flipping as detected by our solvatochromic assay, we determined the IC50 of suramin against EcACP-FabA crosslinking (Fig. S10) to be 118.0 to 333.7 μM (data reported with 95% confidence intervals), which is higher than the IC_{50} as determined by the solvatochromic assay. This is likely due to difference in probe attached to EcACP. While the 4-DMN probe allows for rapid evaluation of PPI, the fluorophore is a large, fused ring moiety that does not resemble the natural fatty acyl substrate usually carried by

EcACP. Unpublished studies from our laboratory have shown that mutations to surface residues on partner enzymes of EcACP can diminish chain flipping activity, as evidenced by a lower fluorescence response compared to wild type enzymes, and for some examples, completely block the transfer of the 4-DMN dye from EcACP into the active site of a partner enzyme. This suggests that chain flipping of an irregular substrate is more tightly controlled than that of substrates which more closely resemble the fatty acyl chain. The C10 chain of the sulfonyl alkyne probe mimics the natural substrate of FabA. Therefore, chain flipping into the active site of FabA is less sensitive to disruptions to PPI, resulting in a lower concentration of suramin necessary to disrupt crosslinking as observed in our secondary assay.

We were unable to determine a minimum inhibitory concentration (MIC) of suramin against Escherichia coli (K12 strain) (Fig. S7). This finding was not surprising, as the highly charged nature of suramin, blocks cell penetration. This study was designed to demonstrate proof of concept, and as such showed that PPI inhibitors can be identified using this workflow. Further studies are ongoing to identify new protein-protein interaction inhibitors through this powerful combination of in silico screening and in vitro validation.

Conclusions

We have demonstrated a workflow for using highly detailed structural studies to complete virtual screening combined with a fluorescent-based PPI assay to identify an inhibitor of the EcACP · FabA complex. A crosslinking assay is used as a secondary in vitro assay to confirm inhibitory activity. While suramin does not exhibit any in vivo activity, we show that the wealth of structural data of carrier protein-partner enzyme interactions can be used to guide antibiotic drug discovery. This method can be easily adapted to

screen for compounds against FAS-II PPI in other pathogens such as Mycobacterium tuberculosis for which there are detailed structures available to guide virtual screening programs followed by in vitro evaluation.

Materials and methods

Virtual screening. The crosslinked EcACP=FabA X-ray crystallography structure (PDB: 4KEH) was loaded into the MOE software suite under an AMBER forcefield with a gas-phase dielectric constant of 80. The MOE QuickPrep tool with Protonate 3D was used to find the overall lowest potential energy configuration for different states of terminal amides, hydroxyls, thiols, histidines and titratable groups. Residues on the surface of FabA (R132, R136 and K161) and EcACP (E41, E47) were selected using the pharmacophore query tool. A panel of 200 compounds was docked individually into the site and receptor as specified by the selected residues.

Solvatochromic dye assay. EcACP was expressed and purified and loaded with a 4-dimethylnaphthalene pantetheine analogue as previously described, and then further purified again by FPLC using an S75 Sephadex column in 50 mM phosphate buffer, pH 8.0 to generate 4-DMN-EcACP (Fig. S4) at a concentration of 116 μM.[33] FabA was expressed from E. coli BL21 cells as previously described and purified using Ni-agarose resin followed by FPLC purification using an S200 Sephadex column in 50 mM phosphate buffer, pH 8.0.[23] To access the inhibition of PPI by suramin, 200 μL aliquots of 40 μM FabA were incubated with a range of suramin concentrations (0–4 M) for twenty minutes at room temperature with light shaking. Following incubation, 50 μL aliquots of FabA were placed in wells of a black COSTAR 96-well plate and an initial fluorescence reading at 408/512 nm (excitation/emission) taken using a ThermoFisher VarioLux plate reader. 4-

DMN-EcACP was then added to each well of FabA and the fluorescence at 408/512 nm (ex/em) read over thirty minutes to allow the system to come to equilibrium. An IC50 curve for suramin against the 4-DMN-EcACP · FabA interaction was generated using GraphPad Prism by plotting the fluorescence value at the end of 4-DMN-EcACP · FabA incubation against the log concentration of suramin. Data were collected in triplicate and IC50 values reported with 95% confidence intervals.

Crosslinking assay. EcACP was expressed and purified and loaded with a sulfonyl alkyne pantetheine analogue as previously described to generate sulfonyl-alkyne-EcACP. This ACP desalted to remove excess sulfonyl alkyne probe (Fig. S5).[23] Sulfonyl-alkyne-EcACP was concentrated to a final concentration of 116 μM. To access the inhibition of crosslinking by suramin, FabA, sulfonyl alkyne-EcACP and various concentrations of suramin (0–5 mM) were incubated at 37 °C overnight with light shaking. The final concentration of sulfonyl-alkyne-EcACP in the crosslinking mixture was 60 μM and the final concentration of FabA in the crosslinking mixture was 10 μM. After incubation, 2 μL samples of each crosslinking reaction were analyzed by 12% SDS-PAGE run for 65 minutes at 160 Volts. The gel was stained with Coomassie blue. To determine the IC50 of suramin against EcACP-FabA crosslinking, crosslinking percentages were determined by scanning gels and quantifying the density of gel bands using ImageJ (Fig. S9). Crosslinking percentages were normalized to the band density of crosslinking in the absence of suramin. GraphPad Prism software was used to generate the IC50 curve and analysis. The dose-response curve is presented in Figure S10.

MIC determination. 5 mL of sterile Luria-Bertani broth was inoculated with E. coli K12 cells and incubated overnight at 37 °C. The overnight culture was diluted 1:15000

into fresh sterile Luria-Bertani broth and aliquoted into a 96-well plate. Suramin was added in increasing concentration (0-4M range) to each well of the 96-well plate and then incubated overnight at 37 °C. An optical density reading at 600 nm was taken 16 hours later to access growth in the presence of suramin (Fig. S4). Ampicillin was used as a positive control.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Prof. Michael Gilson for training and discussion on virtual screening, Prof. Kara Jaremko for synthesis of the sulfonyl-alkyne probe and Prof. Tony Davis for advice. This work was supported by NIH grant RO1 GM095970. K.C. was supported by the National Science Foundation Graduate Assistance in Areas of National Need (NSF GAANN) Fellowship.

References

1. About Antibiotic Resistance. https://www.cdc.gov/drugresistance/about.html (2020).

2. Cossar, P. J., Lewis, P. J. & McCluskey, A. Protein-protein interactions as antibiotic targets: A medicinal chemistry perspective. Med. Res. Rev. 40, 469–494 (2020).

3. Tsao, D. H. H., Sutherland, A.G., Jennings, L.D., Li, Y., Rush, T.S., Alvarez, J.C., Ding, W., Dushin, E.G., Dushin, R.G., Haney, S.A., Kenny, C.H., Malakian, A.K., Nilakantan, R., & Mosyak, L. Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design. Bioorg. Med. Chem. 14, 7953–7961 (2006).

4. Jennings, L. D., Foreman, K.W., Rush, T.S., Tsao, D.H.H., Mosyak, L., Yuanhong, L., Sukhdeo, M.N., Ding, W., Dushin, E.G., Kenny, C.H., Moghazeh, S.L., Petersen, P.J., Ruzin, A.V., Tuckman, M., & Sutherland, A. Design and synthesis of indolo[2,3-

a]quinolizin-7-one inhibitors of the ZipA-FtsZ interaction. Bioorg. Med. Chem. Lett. 14, 1427–1431 (2004).

5. Sutherland, A. G., Alvarez, J., Ding, W., Foreman, K.W., Kenny, C.H., Labthavikul, P., Mosyak, L., Petersen, P.J., Rush, T.S., Ruzin, A., Tsao, D.H.H., & Wheless, K.L. Structure-based design of carboxybiphenylindole inhibitors of the ZipA-FtsZ interaction. Org. Biomol. Chem. 1, 4138–4140 (2003).

6. Yin, Z., Whittell, L.R., Wang, Y., Jergic, S., Liu, M., Harry, E.J., Dixon, N.E., Beck, J.L., Kelso, M.J., & Oakley, A.J. Discovery of lead compounds targeting the bacterial sliding clamp using a fragment-based approach. J. Med. Chem. 57, 2799–2806 (2014).

7. Wijffels, G., Dalrymple, B.P., Prosselkov, P., Kongsuwan, K., Epa, V.C., Lilley, P.E., Jergic, S., Buchardt, J., Brown, S.E., Alewood, P.F., Jennings, P.A., & Dixon, N.E. Inhibition of protein interactions with the beta 2 sliding clamp of *Escherichia coli* DNA polymerase III by peptides from beta 2-binding proteins. Biochemistry 43, 5661–5671 (2004).

8. Georgescu, R. E., Yurieva, O., Kim, S., Kuriyan, J., Kong, X.P., & O'Donnell, M. Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp. Proc. Natl. Acad. Sci. U. S. A. 105, 11116–11121 (2008).

9. Hayashi, T., Yamamoto, O., Sasaki, H., Kawaguchi, A. & Okazaki, H. Mechanism of action of the antibiotic thiolactomycin inhibition of fatty acid synthesis of Escherichia coli. Biochem. Biophys. Res. Commun. 115, 1108–1113 (1983).

10. Kodali, S., Galgoci, A., Young, K., Painter, R., Silver, L.L., Herath, K.B., Singh, S.B., Cully, D., Barrett, J.F., Schmatz, D., & Wang, J. Determination of selectivity and efficacy of fatty acid synthesis inhibitors. J. Biol. Chem. 280, 1669–1677 (2005).

11. Lin, S., Hanson, R. E. & Cronan, J. E. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. Nat. Chem. Biol. 6, 682–688 (2010).

12. Jordan, S. W. & Cronan, J. E. A new metabolic link. The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in Escherichia coli and mitochondria. J. Biol. Chem. 272, 17903–17906 (1997).

13. Issartel, J. P., Koronakis, V. & Hughes, C. Activation of Escherichia coli prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation. Nature 351, 759–761 (1991).

14. Brozek, K. A. & Raetz, C. R. Biosynthesis of lipid A in Escherichia coli. Acyl carrier protein-dependent incorporation of laurate and myristate. J. Biol. Chem. 265, 15410–15417 (1990).

15. Battesti, A. & Bouveret, E. Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. Mol. Microbiol. 62, 1048–1063 (2006).

16. Gully, D., Moinier, D., Loiseau, L. & Bouveret, E. New partners of acyl carrier protein detected in Escherichia coli by tandem affinity purification. FEBS Lett. 548, 90–96 (2003).

17. Adachi, S., Murakawa, Y. & Hiraga, S. Dynamic nature of SecA and its associated proteins in Escherichia coli. Front. Microbiol. 6, 75 (2015).

18. 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. 11, 1727 (2020).

19. Misson, L. E., Mindrebo, J.T., Davis, T.D., Patel, A., McCammon, J.A., Noel, J.P. & Burkart, M.D. Structure and Dynamic Basis of Molecular Recognition Between Acyltransferase and Carrier Protein in E. coli Fatty Acid Synthesis. http://biorxiv.org/lookup/doi/10.1101/2020.05.15.098798 (2020) doi:10.1101/2020.05.15.098798.

20. Du, D., Katsuyama, Y., Horiuchi, M., Fushinobu, S., Chen, A., Davis, T.D., Burkart, M.D., & Ohnishi, Y. Structural basis for selectivity in a highly reducing type II polyketide synthase. Nat. Chem. Biol. (2020) doi:10.1038/s41589-020-0530-0.

21. 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. 1, 8–12 (2020).

22. 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. 116, 6775–6783 (2019).

23. 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 505, 427–431 (2014).

24. Jain, N. U., Wyckoff, T. J. O., Raetz, C. R. H. & Prestegard, J. H. Rapid analysis of large protein-protein complexes using NMR-derived orientational constraints: the 95 kDa complex of LpxA with acyl carrier protein. J. Mol. Biol. 343, 1379–1389 (2004).

25. Zhang, Y.-M., Wu, B., Zheng, J. & Rock, C. O. Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. J. Biol. Chem. 278, 52935–52943 (2003).

26. Tallorin, L., Finzel, K., Nguyen, Q.G., Beld, J., La Clair, J.J., & Burkart, M.D. Trapping of the Enoyl-Acyl Carrier Protein Reductase-Acyl Carrier Protein Interaction. J. Am. Chem. Soc. 138, 3962–3965 (2016).

27. Charov, K. & Burkart, M. D. A Single Tool to Monitor Multiple Protein-Protein Interactions of the Escherichia coli Acyl Carrier Protein. ACS Infect. Dis. 5, 1518–1523 (2019).

28. Charov, K. & Burkart, M. D. Quantifying protein-protein interactions of the acyl carrier protein with solvatochromic probes. Methods Enzymol. 638, 321–340 (2020).

29. Li, L., Zhao, B., Cui, Z., Gan, J., Sakharkar, M.K., & Kangueane, P. Identification of hot spot residues at protein-protein interface. Bioinformation 1, 121–126 (2006).

30. Modell, A. E., Blosser, S. L. & Arora, P. S. Systematic Targeting of Protein-Protein Interactions. Trends Pharmacol. Sci. 37, 702–713 (2016).

31. Chemical Computing Group. Molecular Operating Environment (MOE). (2013).

32. Moynié, L, Hope, A.G., Finzel, K., Schmidberger, J., Leckie, S.M., Schneider, G., Burkart, M.D., Smith, A.D., Gray, D.W., & Naismith, J.H. A Substrate Mimic Allows High-Throughput Assay of the FabA Protein and Consequently the Identification of a Novel Inhibitor of Pseudomonas aeruginosa FabA. J. Mol. Biol. 428, 108–120 (2016).

33. Beld, J., Cang, H. & Burkart, M. D. Visualizing the chain-flipping mechanism in fatty-acid biosynthesis. Angew. Chem. Int. Ed Engl. 53, 14456–14461 (2014).

34. Cronan JE. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. Biochem J. 2014 Jun 1;460(2):157-63.

Supporting information

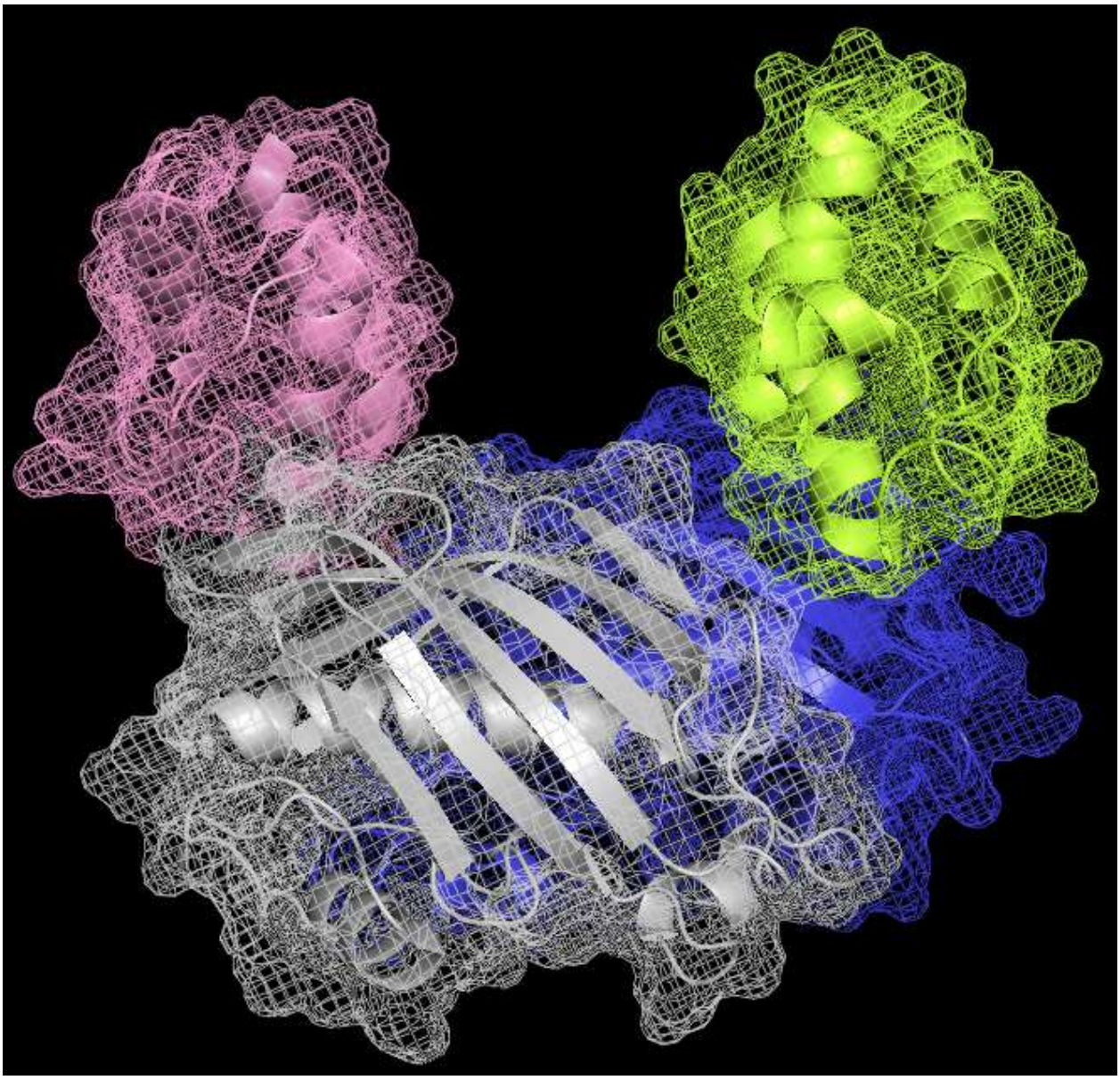
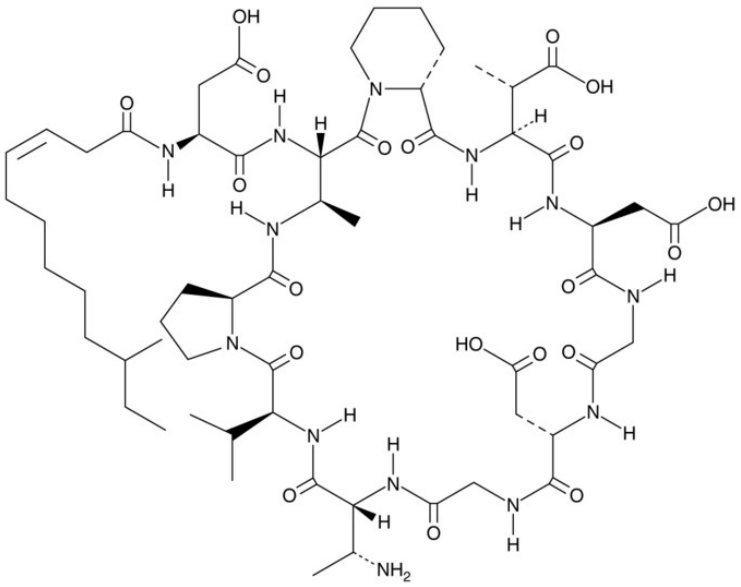
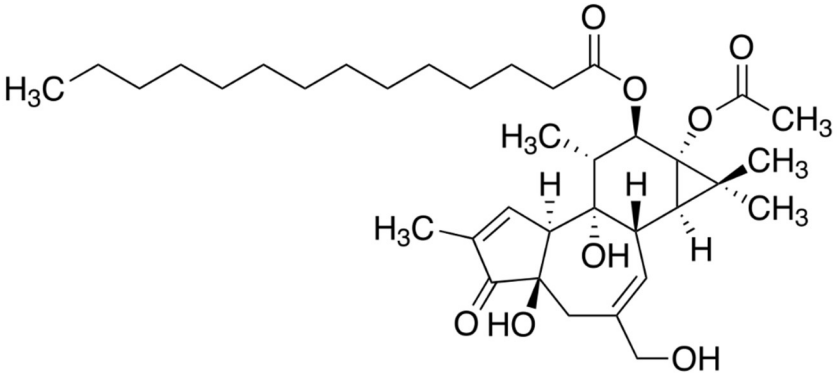


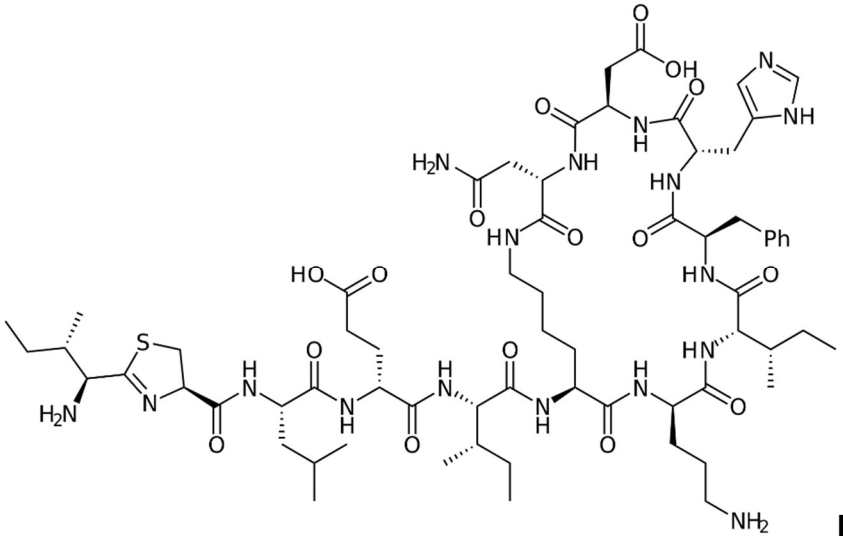
Figure S1. Crystal structure of crosslinked FabA-EcACP (PDB: 4KEH).[1] EcACP in pink and green makes contact with the surface of the FabA dimer (silver and blue).



Amphomycin



Phorbol myristate



Bacitracin A

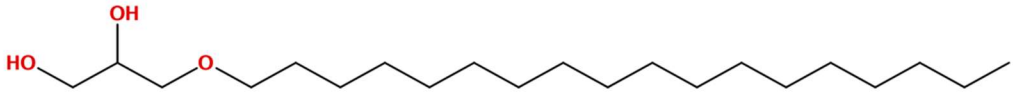


Figure S2. Structures of additional compounds hits produced by MOE virtual screen.

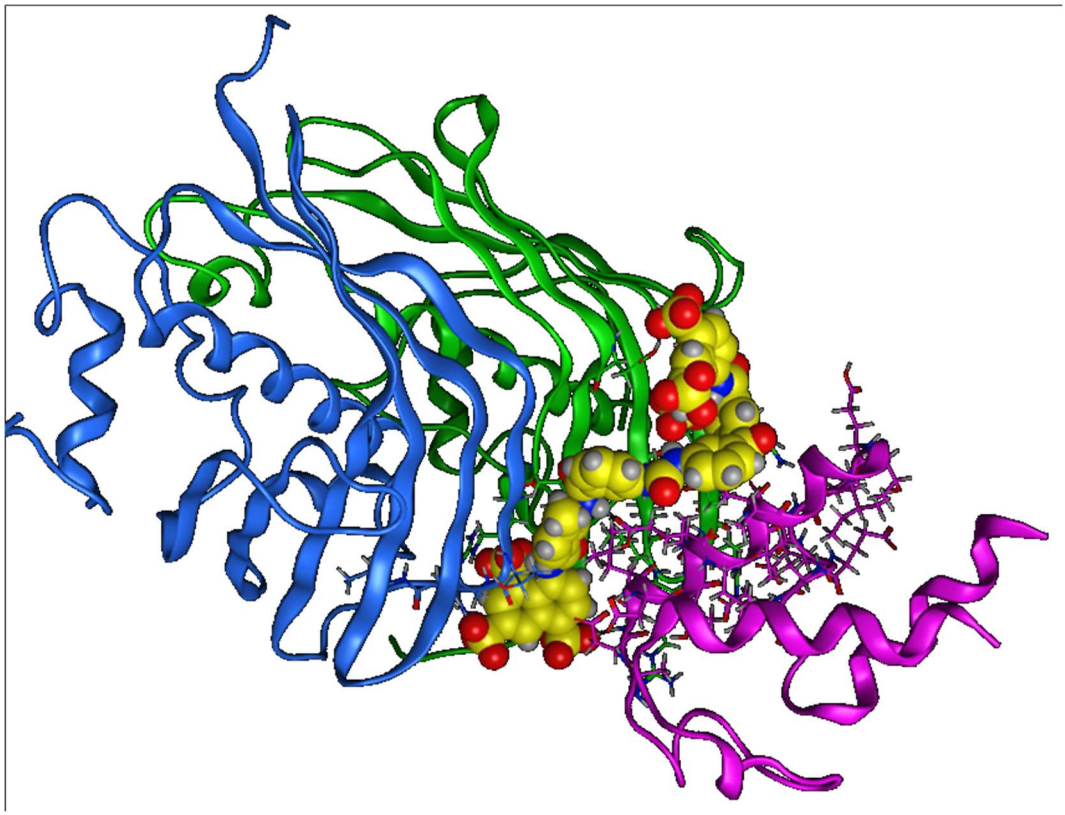


Figure S3. 3D docked structure. Suramin (spheres) on the surface of FabA dimer (blue, green) and EcACP (magenta) with surface representation removed to better visualize suramin mode of docking.

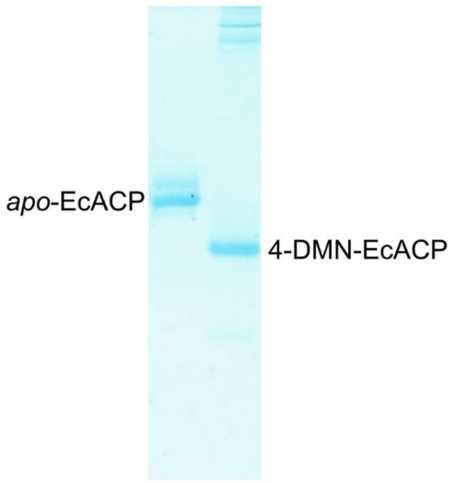


Figure S3.4: 20% acrylamide urea-PAGE. Analysis of *apo*-EcACP (left band) to 4-DMN-EcACP (right band). The gel was run at 150 V for 2 hours and 15 minutes and stained with Coomassie blue. The installation of the 4-DMN probe results in a conformational shift to a more compact protein conformation, causing the 4-DMN-EcACP to run further in the gel. Coa enzymes are visible at the top of the gel in the 4-DMN-EcACP sample.

Figure S3.5: 20% acrylamide urea-PAGE analysis of *apo*-EcACP (left band) to sulfonyl alkyne-EcACP (right band). The gel was run at 150 V for 2 hours and 15 minutes and stained with Coomassie blue. The installation of the sulfonyl-alkyne probe results in a shift to a more compact protein conformation, causing the sulfonyl-alkyne EcACP to run further in the gel.

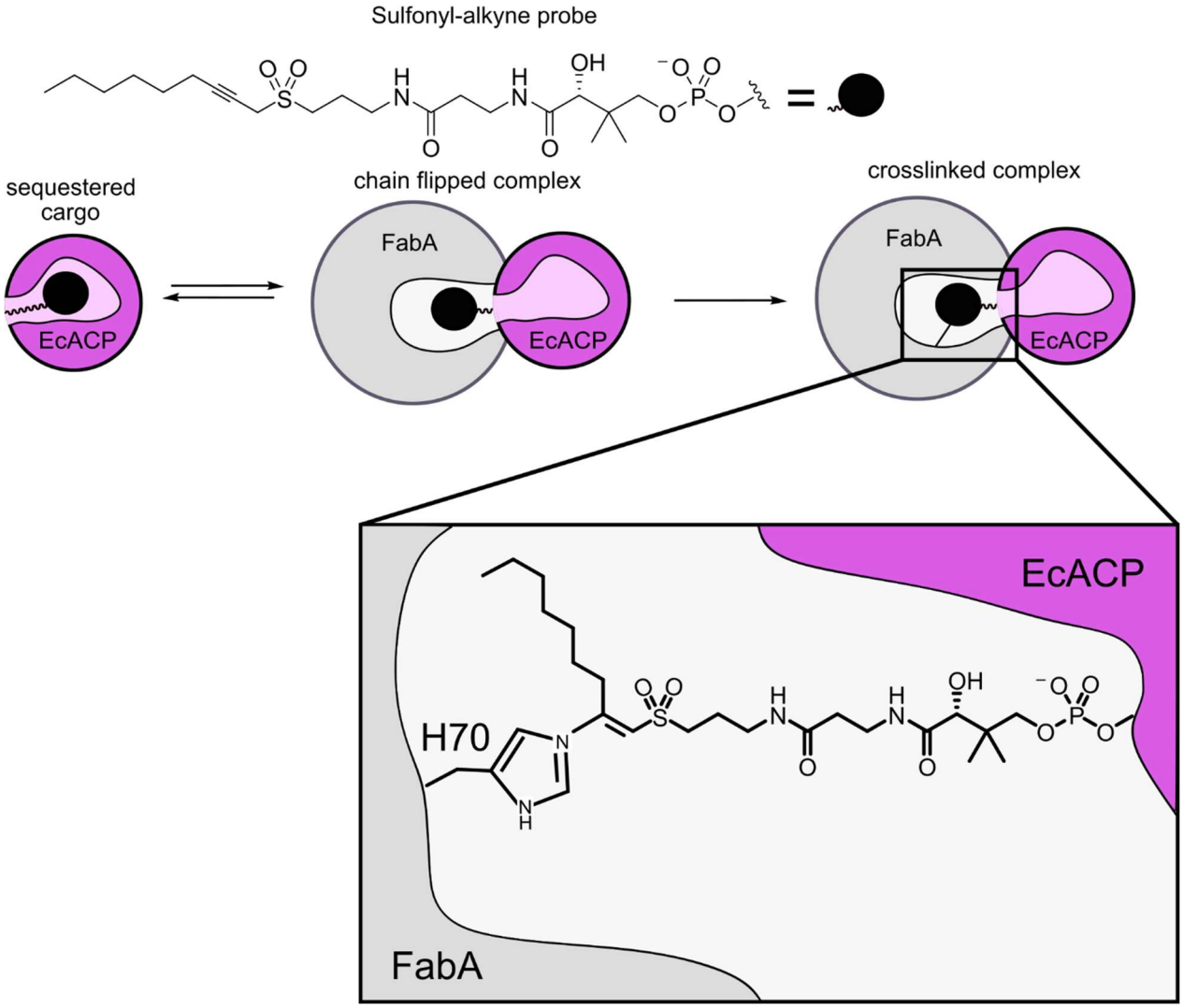


Figure S3.6: Schematic of crosslinking strategy to trap the AcpP-FabA complex. AcpP is chemoenzymatically modified with the sulfonyl-alkyne probe which, upon interaction with FabA, is chain flipped into the active site of FabA. Once in the active site, histidine70 abstracts a proton from the alkyne, thereby forming a ketene which reacts irreversibly with histidine70 forming the covalently trapped complex.

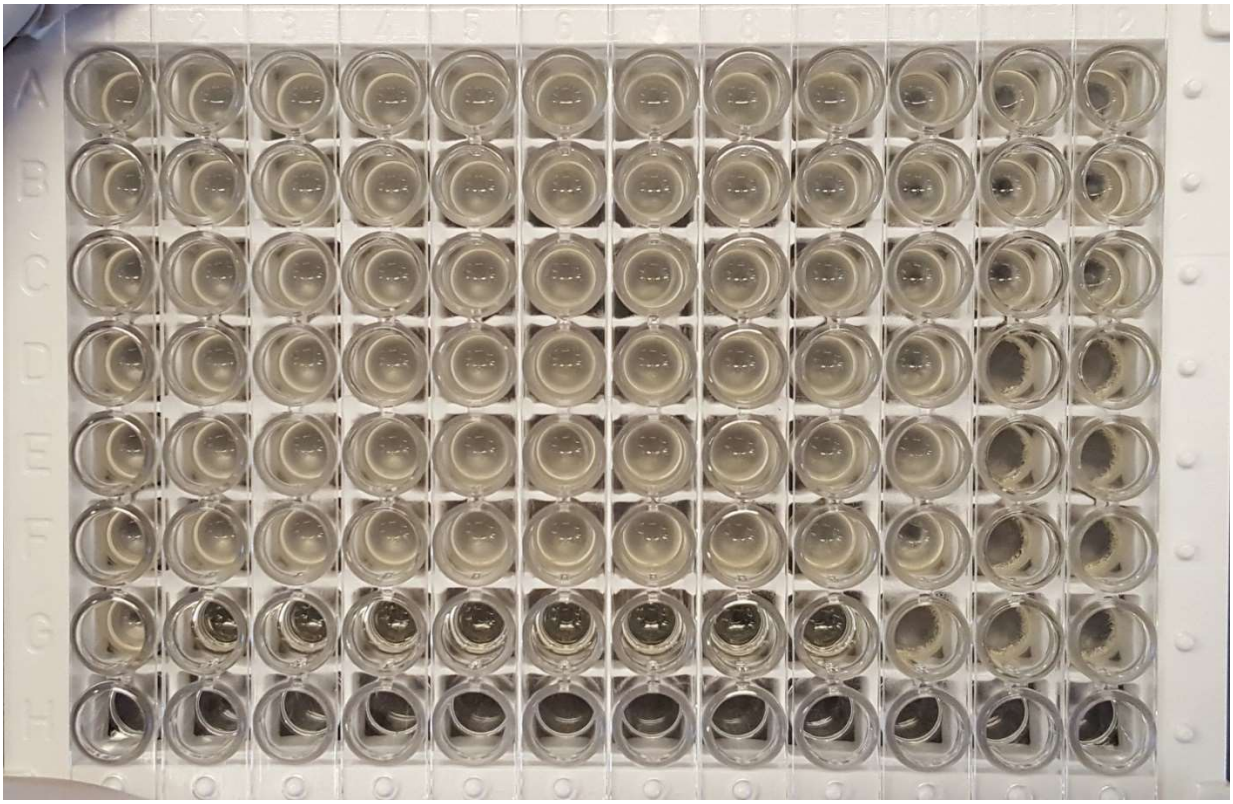


Figure S3.7: Picture of growth plate of *E. coli* K12 dosed with suramin (rows A-F) and ampicillin (row G). The results indicate that suramin does not have an appreciable toxicity to *E. coli* K12 cells, most likely due to the highly charged nature of suramin, making it unable to penetrate the cell membrane.

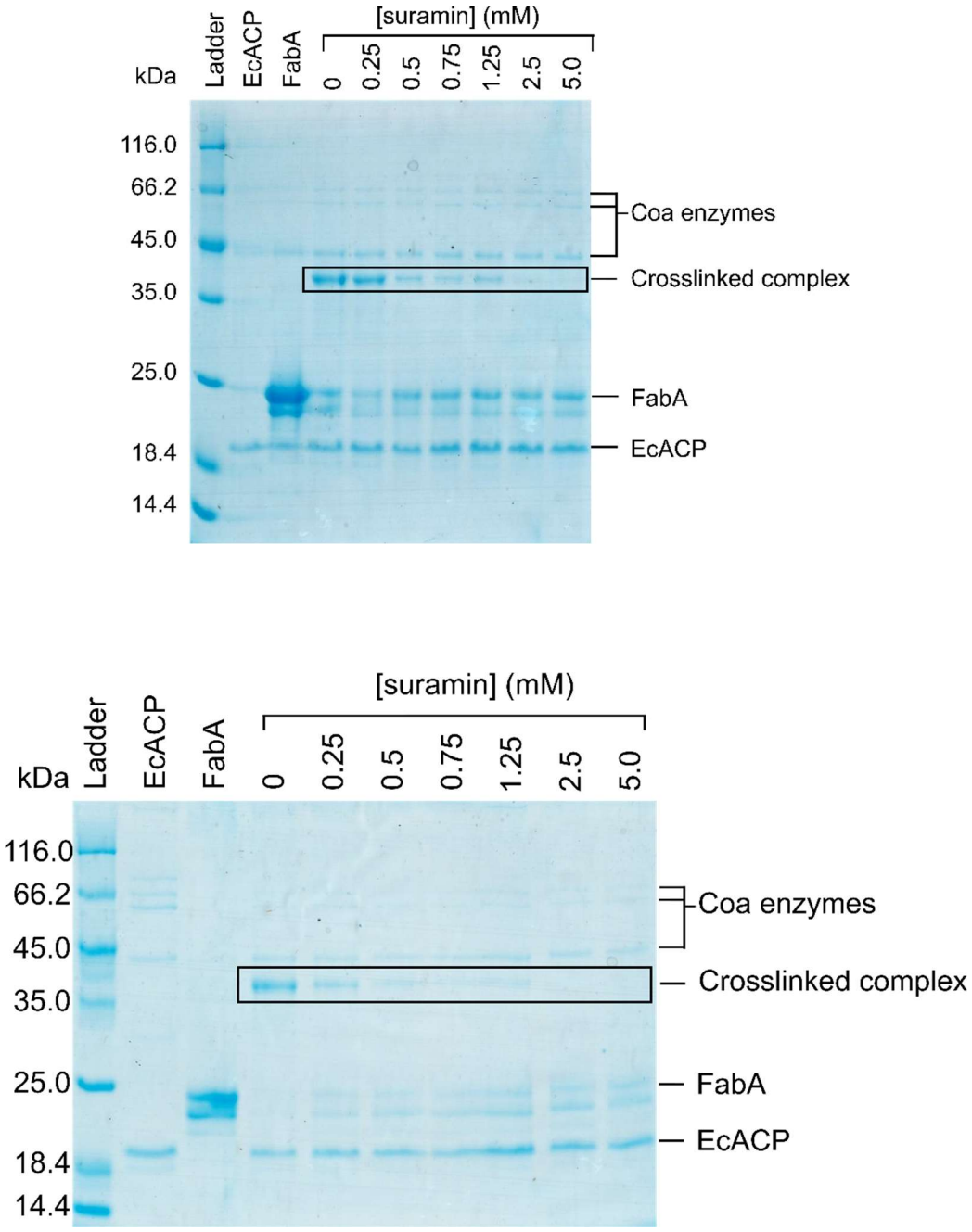


Figure S3.8. 12%SDS-PAGE analysis of repeat of crosslinking experiments.

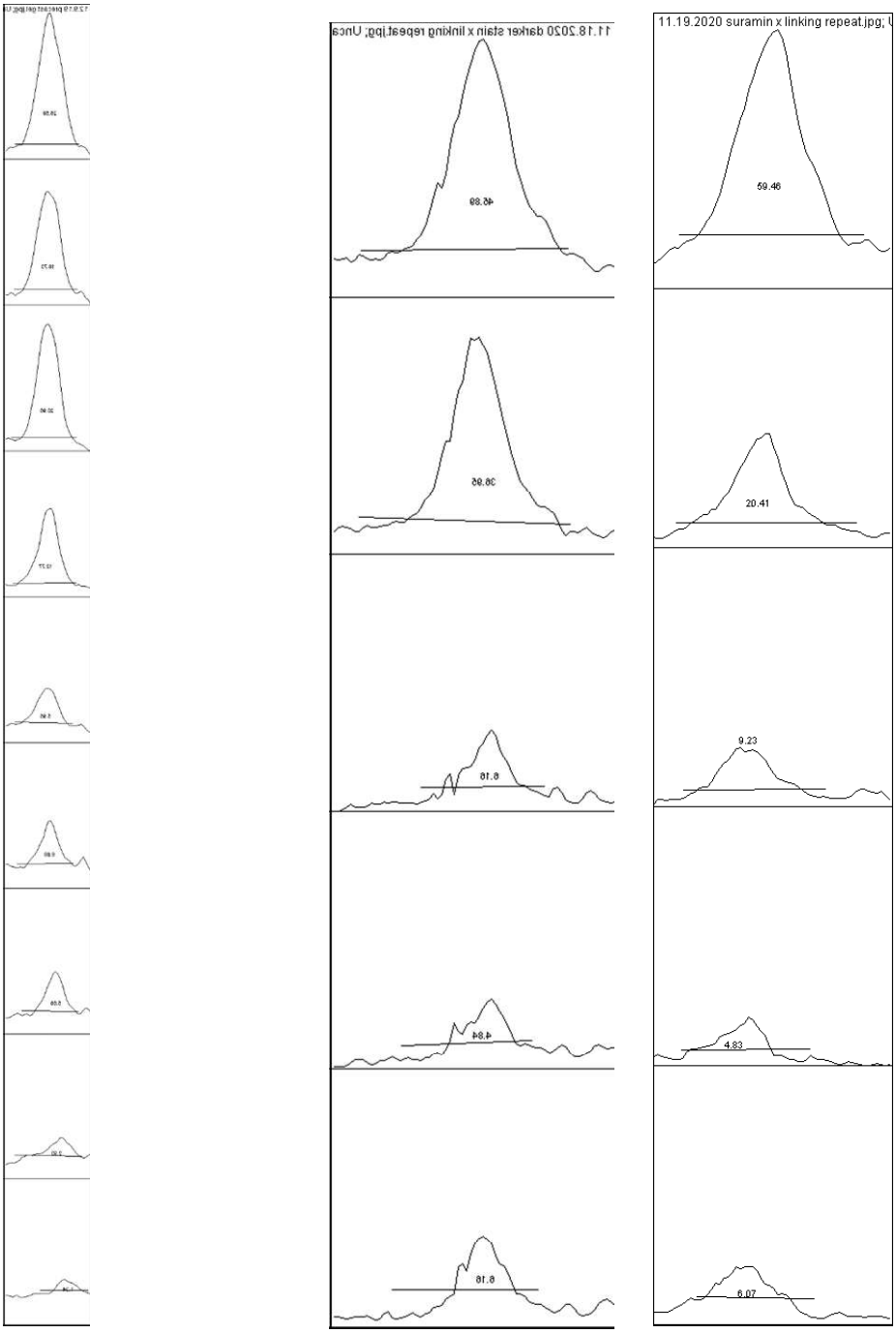


Figure S3.9. ImageJ plots of crosslinking bands. Crosslinking gels were stained with Coomassie blue and scanned using an Epson desktop scanner and analyzed using ImageJ software. To analyze the density of protein in each band, the lane of the gels with 0 mM suramin was used as a 100% crosslinking reference. Remaining crosslinking bands were selected and the area under the curve of each plot calculated by ImageJ. To generate the crosslinking experiment IC_{50} curve, area under the curve percentages were normalized to the 0mM suramin band, with 5 mM suramin yielding a 0% relative area under the curve and the 100% suramin yielding a 1.0 relative area under the curve.

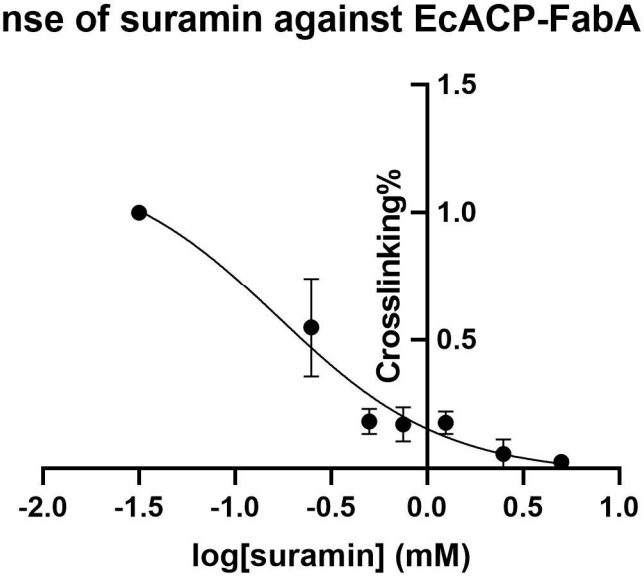


Figure S10. IC_{50} curve from suramin blocking EcACP-FabA crosslinking. Crosslinking experiments as described in the main text were completed in triplicate (additional gels shown in Figure S8). Relative crosslinking percentages were determined using ImageJ software (Figure S9) and normalized to the 0 mM suramin crosslinking band. Crosslinking percentage results were averaged, and error bars represent standard deviation of the average. GraphPad Prism 9 software was used to analyze results. The IC50 was determined to be 0.167 (0.08766 to 0.2842) mM (data reported with 95% confidence intervals).

Acknowledgements:

Chapter 3, in full, has been submitted for publication and is currently under revision after peer review of the material as it may appear: Charov, Katherine; Burkart, Michael D. "*In silico* identification and *in vitro* evaluation of a protein-protein interaction inhibitor of *Escherichia coli* fatty acid biosynthesis." *Chemical Biology & Drug Design.* The dissertation author was the primary investigator and author of this paper.

Chapter 4: Decoupling disruptions to the *Escherichia coli* acyl carrier protein·ketosynthase II equilibrium and chain-flipping mechanism from disruptions to ketosynthase II catalysis

Abstract

Protein-protein interactions (PPIs) can involve large surface areas of proteins but are often mediated by a small subset of residues, often termed hot-spot residues. Engineering chimeric fatty acid, polyketide and non-ribosomal fatty acid biosynthetic pathways has been slowed by insufficient progress in mapping PPIs and the hot-spot residues dictating them, as these pathways are mediated by PPIs formed between a carrier protein and a catalytic partner. Here we present a rapid, fluorescence-based assay for identifying residues on the surface of enzymes that interact with the *Escherichia coli* acyl carrier protein (EcACP). To demonstrate proof of concept, interfacial residues known to mediate PPIs of the fatty acid biosynthesis enzyme FabF, a ketosynthase, with EcACP were evaluated. Mutagenesis followed by screening with the fluorescence-based assay showed that single point mutation of hot-spot residues on the surface of FabF had large effects on the EcACP•FabF interactions. Comparison to previously published kinetic assays of FabF revealed that these mutations usually correlate with a loss of enzymatic activity, but in rare cases identified residues that are solely involved in enzyme catalysis and not in the PPIs themselves, and vice versa. This distinction has implications for the future design of interface inhibitors and protein engineering efforts of FAS enzymes.

Introduction

Protein-protein interactions (PPIs) mediate multiple biological processes, including signaling, regulation and enzymatic activity. Interest in understanding the molecular underpinnings of these interactions has been driven in part by recent advances in structure-aided design of PPI inhibitors. To design effective PPI inhibitors, it is imperative to map the surface resides most essential to interface formation. These can be used to tailor drug scaffolds that target critical residues and inhibit interface formation without blocking the entire PPI surface area, which can reach many hundreds of Å in size.[1,2] PPIs have also been evaluated for protein engineering of biosynthetic pathways in which PPIs play a critical role, such as synthases of fatty acid (FAS), polyketide (PKS) and non-ribosomal (NRPS) pathways.[3,4] Much work has been dedicated to engineering new biosynthetic pathways by swapping enzymatic domains from within these pathways, with a goal of producing new bioactive compounds. However, attempts to coordinate the action of chimeric enzymes may require a fundamental understanding of PPIs and the challenge of readily identifying residues critical for those interactions remains a long-standing bottleneck.

One of the best characterized systems for studying PPIs is *E. coli* fatty acid biosynthesis.[5–8] Fatty acids are biosynthesized in an iterative fashion in both eukaryotes and prokaryotes; each turn of the fatty acid synthase (FAS) cycle extends the growing acyl chain by two carbons. Plants, bacteria and mitochondria use a type II system (FAS-II, Figure 1), in which each enzyme is a discrete entity, while eukaryotic cytosol contains a megasynthase (FAS-I), with each enzymatic domain part of the same polypeptide chain. Central to both FAS-I and FAS-II is the acyl carrier protein (ACP), which serves as a

scaffold and shuttle for the growing fatty acyl chain as it is acted upon by each of the enzymes in the pathway.

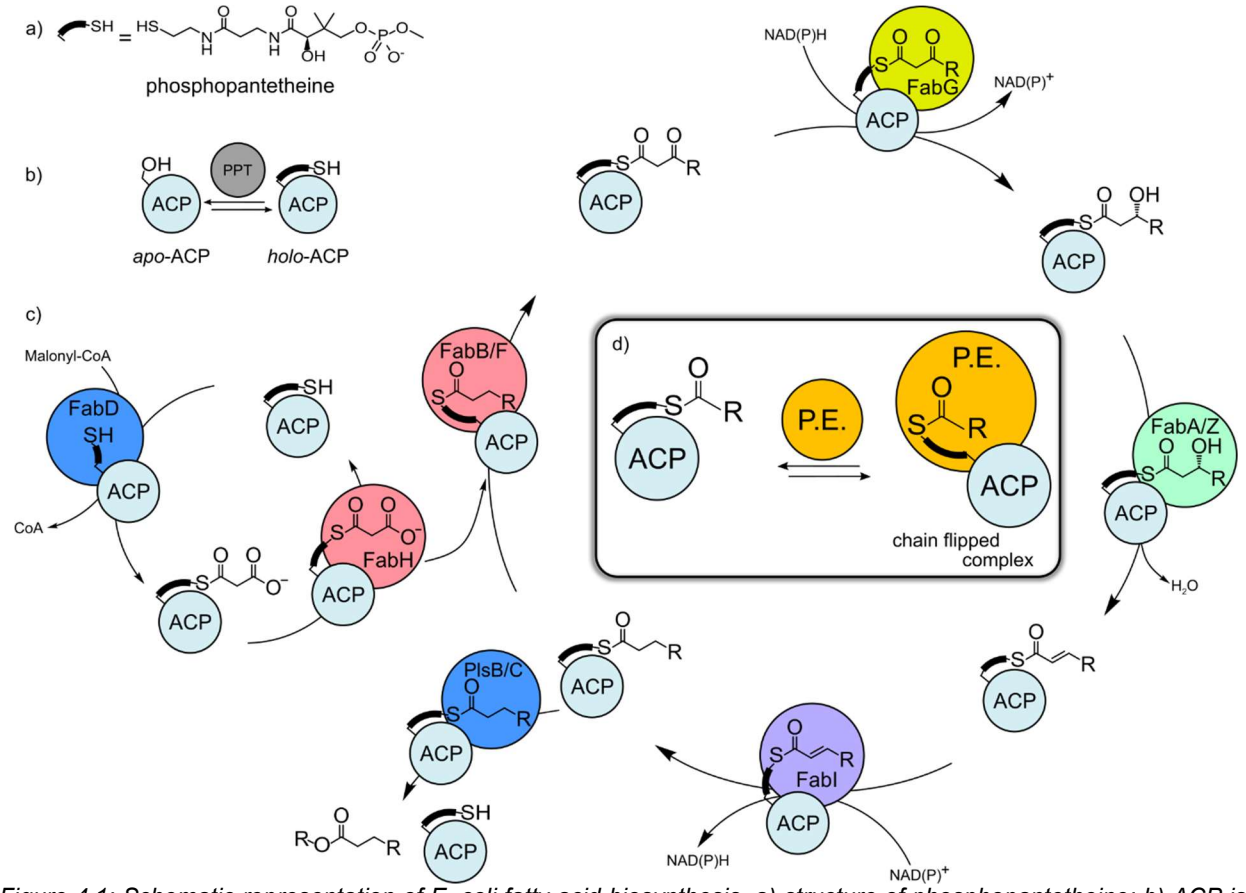


Figure 4.1: Schematic representation of E. coli fatty acid biosynthesis. a) structure of phosphopantetheine; b) ACP is expressed in an inactive apo- state and then post-translationally modified by a phosphopantetheinyl transferase (PPT) to form holo-ACP. The terminal thiol of the pantetheine arm serves to tether the growing acyl substrates to ACP via thioester bond; c) E. coli type II fatty acid biosynthesis starts with the loading of holo-ACP with a unit of malonyl by action of a malonyl-CoA-acyl transferase, FabD. Next, a ketosynthase, FabH, condenses malonyl-ACP with another acyl-ACP to extend the acyl chain by two carbons via a decarboxylative Claisen-like condensation. This beta-keto thioester is then reduced by the ketoreductase, FabG, to form a beta-hydroxy thioester, which in turn is reduced by one of two dehydratases, FabA or FabZ. Next, an enoyl reductase, FabI, fully saturates the acyl chain, which can either rejoin the cycle and condense with another unit of malonyl-ACP or exit the pathway wherein the fully saturated acyl chain is transferred via membrane bound PlsB or PlsC to form a phospholipid; d) Central to all ACP interactions with partner enzymes is the universal chain-flipping mechanism: ACP sequesters its cargo within its alpha helical bundle to protect it from hydrolysis or side reactions but upon interaction with a partner enzyme, places the cargo into the active site of the partner enzyme via chain-flipping mechanism.

ACP sequesters its intermediate cargo within its four-alpha helical core to protect the cargo from hydrolysis and cross-reactivity until the correct partner enzyme is encountered, upon which ACP presents its cargo into the active site of the partner enzyme using a universal chain-flipping mechanism.[9] In FAS-II alone, ACP engages with eight

different partner enzymes, each demonstrating unique PPIs. These PPIs are responsible for the correct delivery of cargo by ACP to each enzyme in the pathway and are critical for recognition and catalysis.

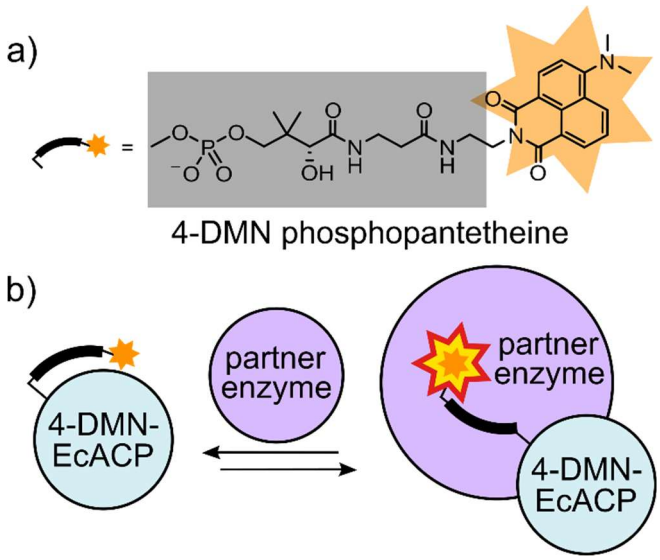


Figure 4.2: Schematic representation of the fluorescence-based assay used to visualize EcACP chain-flipping mechanism. a) structure of 4-DMN-phosphopantetheine, a solvatochromic dye-containing analogue of Coenzyme A that is chemoenzymatically attached to EcACP to form 4-DMN-EcACP. b) When 4-DMN-EcACP interacts with a partner enzyme, it transfers the 4-DMN cargo into the active site of the partner enzyme, which leads to a large increase in fluorescence intensity.

In this work, we developed a universal PPI assay for rapid analysis of hot-spot residues mediating ACP-partner enzyme interactions. This methodology is designed to complement existing methods such as x-ray crystallography and protein NMR, which can be time consuming and challenging to perform. In this assay, EcACP is labeled with a fluorescent reporter, 4-dimethylnaphthalene (4-DMN, Figure 2a), to generate 4-DMN-EcACP. EcACP sequesters the 4-DMN cargo within its alpha helical core, leading to an increase in fluorescence intensity, relative to that of the dye in solution. When 4-DMN-EcACP interacts with a partner enzyme, the 4-DMN cargo undergoes chain-flipping and is delivered to the active site of the partner enzyme (Figure 2b).[10–12] Within the active sites of partner enzymes, the fluorescence intensity of the 4-DMN dye increases further,

relative to 4-DMN-EcACP, due to differences in hydrophobicity within the active site of the partner enzyme and the helical core of EcACP (Figure S1). We recently demonstrated the utility of this fluorescent tool in monitoring multiple EcACP PPIs and in this work sought to expand the scope of the tool in monitoring PPI disruptions.[11]

Results

Analysis of the interaction of E. coli ketosynthase II, FabF, with 4-DMN-EcACP.

Structural analysis of known crosslinked structures revealed three regions at the interface between FabF and EcACP mediate molecular recognition (PDB: 6OKG).[13] Region 1 is comprised of electrostatic interactions formed by Lys65, Arg68, and Lys69 of FabF and Glu13, Gln14, Asp35, and Asp38 of EcACP. Region 1 also includes a flexible loop, with Thr270 which interacts with Asp35 of EcACP. Region 2 of FabF includes Arg127, Lys128 and Arg206 which interact with Glu47, Glu48, Glu53 and Asp56 of EcACP, respectively. Finally, region 3 is a hydrophobic patch in which residues Ile129, Ser130, Pro131, Phe132, Ala205 and Arg206 of FabF interact with Met44, Val40 and Leu37 of EcACP.

We chose to use our solvatochromic assay to rapidly determine which of these residues are critical to EcACP•FabF association by systematically mutating each site to either alanine or to a residue of opposite charge. We reasoned that the effects of changing these residues of the EcACP•FabF interaction could be detected by the solvatochromic PPI assay as a decrease in overall fluorescence intensity, relative to the wild type EcACP•FabF interaction. Altering wildtype PPIs might cause a complete disruption the recognition all together, leading clear lack of a fluorescence intensity change.

Alternatively, mutations at the interface could reduce the stability of EcACP•FabF complexes, shifting the equilibrium between unbound and bound states away from the bound state. This shift would also produce a reduction in fluorescence intensity relative to wild type. Each residue selected for study was subject to either charge swap mutation or mutation to alanine. Double and triple mutants were also generated to study the compounding effects of charge swap or alanine mutations for a few select residues. (Table S1). All FabF variants were expressed in *E. coli* and purified using Nickel chromatography and analyzed using the solvatochromic assay described above. Each FabF variant was tested in triplicate on three separate days for a total of nine replicates each. Results of these assays are summarized in Figure 3.

The negative control C163A variant, with active site cysteine mutated to alanine, displayed no reduction in fluorescence intensity with 4-DMN-EcACP but instead exhibited a larger than wild type fluorescence intensity likely due to the more hydrophobic nature of alanine compared to cysteine and a widening of the active site pocket. Region 1 (R68A, K69N, K65A/K69A, K65A/R68A/K69A) is highly sensitive to mutation, and these mutations generally yielded large reductions in fluorescence intensity, except for T270A, which produced an increase in fluorescence intensity compared to wild type. Mutations to residues in region 2 (K128A, K128N, R206A, R206N, R127A/K128A, R127A/K128A/R206A) were similarly effective in disrupting the interaction or chain-flipping event between 4-DMN-EcACP and FabF, resulting in significant decreases in fluorescence intensity relative to wild type. As anticipated, mutations to alanine were less detrimental to fluorescence intensity than were charge swap mutations, best evidenced by comparing the fluorescence intensity of K128A-FabF (40% wild type fluorescence

intensity) vs. K128N-FabF (10% wild type fluorescence intensity). Finally, mutations to alanine in residues of the hydrophobic region 3 (S130A, F132A, F132L, F132Y) were found to have a more modest effect on the fluorescence intensity compared to wild type, but mutations to F132 to either the bulkier tyrosine or leucine had an intermediate detrimental effect. We also tested the triple cross-region mutant K65A/R127A/K128A which exhibited little fluorescence intensity with 4-DMN-EcACP.

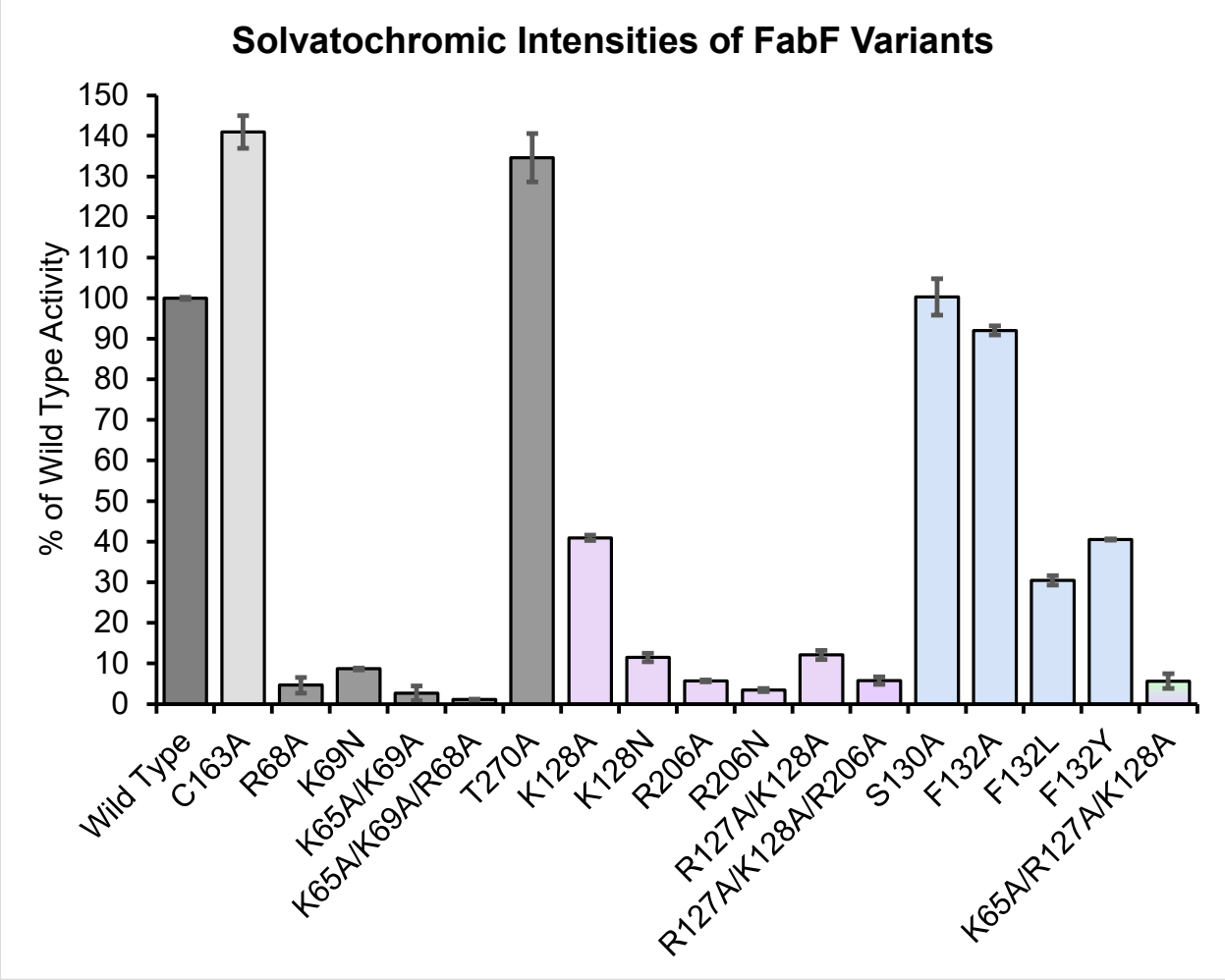


Figure 4.3: Results of the solvatochromic assay of FabF variants with 4-DMN-EcACP. Controls are represented in grey bars. Region 1 mutants are colored green, region 2 in lavender and region 3 in light blue.

Discussion

The chain-flipping mechanism by which ACP transfers cargo from its interior to the active sites of its many partner enzymes is believed to be universal and necessary for the partner enzyme to complete catalysis.[9] A small subset of hot-spot residues on the surface of ACP and its partner enzymes are known to mediate recognition. Mapping these residues is critical to address challenges in drug discovery and protein-protein interaction engineering. The assay demonstrated in this work represents a rapid means to identify these residues that can be readily adapted to study any ACP-partner enzyme interaction, including those found outside of FAS-II.

The step-wise mechanism by which ACP-mediated enzymatic processes occur can be split into five discrete steps: protein-protein recognition, chain-flipping of cargo, chemical reaction, reversal of chain-flipping, and carrier protein dissociation. Enzyme kinetic assays can provide information about the rate of the chemical step, and in the case of ACP-mediated processes offer insights into how interfacial residues affect the ACP-partner enzyme interactions that determine the functions of these proteins. However, it has proven challenging to decouple the effects of mutations at interfacial residues with each of these five steps from one another. On the basis of kinetic data, mutations may appear to diminish the processivity of ACP-partner enzymes and perhaps affect rate acceleration of the chemical step without compromising the ACP•partner enzyme interface. Conversely, the opposite situation is possible, wherein mutations to interfacial residues do not affect the efficiency of an enzyme's ability to carry out a chemical reaction significantly, but do reduce the affinity of ACP for its partner enzyme.

In either case, enzymatic processivity is attenuated and in terms of drug discovery and metabolic engineering it is essential to delineate which of these scenarios is operant.

Comparison of kinetic and solvatochromic assays of FabF

The results presented herein strongly recapitulate the conclusion derived from kinetic analysis of FabF reported by Burkart and Williams.[14,15] A mutation to R206 of FabF lowers the kinetic rate in agreement with our finding that this mutant FabF exhibits in a reduced fluorescence intensity (Figure 3). The reduced intensity indicates that the chain-flipping event or the association of EcACP with FabF has been disrupted, likely the cause for the low kinetic rate. Similarly, kinetic analysis of S130A and F132A FabF showed relatively high kinetic rates and, in this study, also yielded a high fluorescence intensity. These results suggest that S130 and F132 are relatively unimportant for both chain-flipping and FabF catalysis.

More interesting examples involve the study of T270A-FabF. Unpublished kinetics analyses of T270A-FabF from our laboratory show a near complete abolishment of enzymatic activity. Comparison of these data to those generated via solvatochromic assay, however, shows that chain-flipping between 4-DMN-EcACP and T270A-FabF remains intact, suggesting that this residue is of importance to catalysis, even though it is on the surface of the enzyme, but *not* chain-flipping. Examination of the C16-FabF-EcACP crystal structure reveals why this dichotomy is observed: T270 sits on a flexible loop in region 1 of FabF and undergoes a large conformational rearrangement upon acyl-EcACP binding stabilizing the closed, enzymatically active, conformation of FabF. T270 is primarily involved in the catalytic activity of FabF, rather than the recognition and therefore chain-flipping mechanism, or FabF recognition, of EcACP. Conversely, K128A

does not exhibit any reduction in kinetic rate but shows a drop to only 40% of wild type fluorescence intensity. This suggests that K128 is more heavily involved in the chain-flipping action of the EcACP cargo, or the establishment of the EcACP·FabF ⇌ EcACP + FabF equilibrium, rather than the catalytic activity of the FabF enzyme. In the C16-FabF-EcACP crystal structure, K128 on the surface of FabF is shown to primarily interact with E47 of EcACP during binding. Similarly, the double mutant R127A/K128A retains 78% wild type activity in a kinetic assay but exhibits only 12% wild type activity in the solvatochromic assay results presented here. A more subtle example of this is also seen in the case of R68A which has approximately 55% wild type kinetic rate but has very little solvatochromic activity. A comparison of the kinetic activity of FabF variants is summarized in Table 4.1.

Table 4.1: Comparison of kinetic activities of FabF mutants to solvatochromic intensities of same mutants. While many mutations disrupt both kinetic activity and solvatochromic intensity, some mutations are limited in their effect on either solvatochromic intensity or kinetic activity, suggesting a decoupling of the two steps.

	k_{cat} *(%w.t.)*	*Solvatochromic intensity (%w.t.)*
R68A	55.6	4.6
K65A/K69A	22.2	2.6
K65A/K69A/R68A	4.4	1.1
R206A	11.1	5.7
K128A	100.0	40.9
R127A/K128A	77.8	12.0
S130A	75.0	100.0
F132A	111.1	92.0
T270A	4.4	134.0

Conclusions

The ability to decouple the enzymatic activity of ACP partner enzymes from the chain-flipping/partner enzyme recognition event has important implications for future protein engineering and drug discovery efforts. Using two complementary assays to examine effects on kinetics versus cargo chain-flipping allows identification of residues that are good candidates for mutation to alter substrate specificity without disrupting or destroying the critical PPI between ACP and its partner enzymes. Similarly, identification of critical residues allows for efficient structure-aided design of new drug scaffolds. Here we present a single, fluorescent based assay that can be used to probe the chain-flipping mechanism of EcACP with the ketosynthase II partner enzyme that can be expanded to monitor many more EcACP-partner enzyme interactions. This assay is fast, reproducible and reports purely on disruptions to the chain-flipping mechanism of EcACP or EcACP·FabF ⇌ EcACP + FabF equilibrium, as a function of surface residue mutations. We anticipate the future development of this assay into a high-throughput screen of many different EcACP-partner enzyme interactions to rapidly map the hot-spot interfacial residues and guide future protein engineering and drug discovery efforts.

Materials and methods

4-DMN-EcACP preparation. E. coli acyl carrier protein (EcACP) was expressed and purified from *E. coli*, yielding a mixture of *apo*- and *holo*-EcACP. The mixture of *apo*- and *holo*-EcACP was incubated at 37°C overnight with *Pseudomonas aeruginosa* phosphodiesterase (AcpH) at a final concentration of 0.01mg/mL, to yield 100% apo-EcACP as previously described.[11] The phosphopantetheine arm removal by AcpH was analyzed by partially denaturing 20% urea-PAGE, run at 150 V for 90 minutes. To

generate 4DMN-EcACP, *apo*-EcACP was incubated overnight at 37°C with 0.1 mg/ml each of previously purified CoA-biosynthetic enzymes from *E. coli* (CoaA, CoaD and CoaE), and the phosphopantetheinyl transferase from *Bacillus subtilis* (Sfp), 12 mM MgCl2, 8 mM ATP, 0.1% Triton-X and 2 mM 4-DMN pantetheine dissolved in dimethylsulfoxide, as previously described.[11] 4-DMN-pantetheine was synthesized as previously described.[11] This one pot reaction to generate 4-DMN-EcACP was analyzed by 20% urea-PAGE. 4-DMN-EcACP was further purified using an S75 Superdex size exclusion column with an ÄKTA Pure FPLC in 50mM Tris pH 7.4, 150 mM NaCl, 10% glycerol.

Site-directed mutagenesis of FabF. FabF was mutated as previously described.[14]

Expression and purification of FabF variants. FabF variants were expressed and purified as previously described.[14]

Solvatochromic assay. To analyze the effect on chain flipping of each surface mutation to FabF, we adapted a previously published solvatochromic assay. Here, 4-DMN-EcACP was concentrated to 120 μM in 50mM Tris pH 7.4, 150 mM NaCl, 10% glycerol and 10 μL of the protein was dispensed into individual wells of a black 384-well plate. The fluorescence of 4-DMN-EcACP was read using a ThermoFisher VarioLux plate reader at λ_{ex} = 408 nm and λ_{em} = 512 nm to generate a baseline fluorescence reading for each well. Following the baseline reading, 15 μL of FabF (wild type or variant of interest) was added to each well and the fluorescence of the 4-DMN-EcACP·FabF interaction monitored using a ThermoFisher VarioLux plate reader at λ_{ex} = 408 nm and λ_{em} = 512 nm over the course of 5 minutes to allow the system to come to equilibrium. To provide a blank, 15 μL of FabF (wild type or variant of interest) was added to a well containing 10 μL of just buffer. To

analyze the experiment, first all readings were averaged over the final 60 seconds of the fluorescence reading. Then the blank reading of each FabF variant was subtracted from the fluorescence reading of the 4-DMN-EcACP·FabF interaction reading. To normalize the fluorescence data, this corrected reading was divided by the fluorescence baseline reading of just 4-DMN-EcACP. To compare solvatochromic intensities of the various FabF variants to the wild type fluorescence intensity, each FabF variant reading was divided by the fluorescence intensity of the wild type reading and multiplied by 100 to generate a percentage of wild type fluorescence intensity value. Data were collected in triplicate on three separate days for a total of nine data points for each variant.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

This work was supported by NIH grant RO1 GM095970. K.C. was supported by the National Science Foundation Graduate Assistance in Areas of National Need (NSF GAANN) Fellowship. We thank Dr. Jeffrey Mindrebo for fruitful discussions of FabF-EcACP structures and Dr. Laetitia Misson Mindrebo for site-directed mutagenesis of FabF plasmids and FabF variant expression and purification. We also wish to thank Dr. Ashay Patel for further discussion and aid in manuscript preparation.

References

1. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* **21**, 1102–1114 (2014).

2. Modell, A. E., Blosser, S. L. & Arora, P. S. Systematic Targeting of Protein-Protein Interactions. *Trends Pharmacol. Sci.* **37**, 702–713 (2016).

3. Klaus, M., D'Souza, A. D., Nivina, A., Khosla, C. & Grininger, M. Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chem. Biol.* **14**, 426–433 (2019).

4. Yuzawa, S., Deng, K., Wang, G., Baidoo, E.E.K., Norten, T.R., Adams, P.D., Katz, L., Keasling, J.D. Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol.* **6**, 139–147 (2017).

5. 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* **505**, 427–431 (2014).

6. Tallorin, L., Finzel, K., Nguyen, Q.G., Beld, J., La Clair, J.J., & Burkart, M.D. Trapping of the Enoyl-Acyl Carrier Protein Reductase-Acyl Carrier Protein Interaction. *J. Am. Chem. Soc.* **138**, 3962–3965 (2016).

7. 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.* **116**, 6775–6783 (2019).

8. 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. U. S. A.* **117**, 24224–24233 (2020).

9. Cronan, J. E. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. *Biochem. J.* **460**, 157–163 (2014).

10. Beld, J., Cang, H. & Burkart, M. D. Visualizing the chain-flipping mechanism in fatty-acid biosynthesis. *Angew. Chem. Int. Ed Engl.* **53**, 14456–14461 (2014).

11. Charov, K. & Burkart, M. D. A Single Tool to Monitor Multiple Protein-Protein Interactions of the Escherichia coli Acyl Carrier Protein. *ACS Infect. Dis.* **5**, 1518–1523 (2019).

12. Charov, K. & Burkart, M. D. Quantifying protein-protein interactions of the acyl carrier protein with solvatochromic probes. *Methods Enzymol.* **638**, 321–340 (2020).

13. 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.* **11**, 1727 (2020).

14. 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* **59**, 3626–3638 (2020).

15. Ye, Z. & Williams, G. J. Mapping a ketosynthase:acyl carrier protein binding interface via unnatural amino acid-mediated photo-cross-linking. *Biochemistry* **53**, 7494–7502 (2014).

Acknowledgements:

Chapter 4, in full, is currently being prepared for submission for publication as: Charov, Katherine; Burkart, Michael D. “Decoupling disruptions to the Escherichia coli acyl carrier protein·ketosynthase II equilibrium and chain-flipping mechanism from disruptions to ketosynthase II catalysis.” The dissertation author was the primary investigator and author of this paper.

Chapter 5. Synthase-selective exploration of a tunicate microbiome by activity-guided single-cell genomics

Woojoo E. Kim,[1,†] Katherine Charov,[1,†] Mária Džunková,[2,†] Eric D. Becraft,[3,4,†] Julia Brown,[3,†] Frederik Schulz,[2] Tanja Woyke,[2] James J. La Clair,[1] Ramunas Stepanauskas[3,*] & Michael D. Burkart[1*]

[1] Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0358, United States

[2] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Mail Stop: 91R183, 1 Cyclotron Road, Berkeley, CA 94720, USA.

[3] Bigelow Laboratory for Ocean Sciences, East Boothbay ME 04544, United States

[4] University of North Alabama, Florence AL 35632, United States

* Corresponding authors:

† These authors contributed equally to the work.

Abstract

While thousands of environmental metagenomes have been mined for the presence of novel biosynthetic gene clusters, metagenomics recovers sequences of only the most abundant bacterial species and such computational predictions do not provide

evidence of in vivo biosynthetic functionality. Using a fluorescent in situ enzyme assay targeting carrier proteins common to polyketide (PKS) and non-ribosomal peptide synthetases (NRPS), we applied fluorescence-activated cell sorting to tunicate microbiomes to enrich for microbes with active secondary metabolic capabilities. Single-cell genomics uncovered the genetic basis for a wide biosynthetic diversity in the enzyme-active cells and revealed the bacterial sequences of a marine Oceanospirillales, harboring a novel NRPS gene cluster with high similarity to phylogenetically distant marine and terrestrial bacteria. Interestingly, this synthase belongs to a larger class of siderophore biosynthetic gene clusters commonly associated with pestilence and disease. This demonstrates activity-guided single-cell genomics as a tool to guide novel biosynthetic discovery.

Introduction

Antecedent discoveries of secondary metabolites have been characteristically limited to their direct extraction from macroorganisms or cultured microbes.[1] Nowadays, thousands of genomes of uncultured microbes from a variety of environmental samples such as soil or water are sequenced and mined for the presence of biosynthetic gene clusters, yet challenges still remain in the application of this approach to complex macroorganism-associated microbomes.[2] Once a cluster is identified, the genes within these clusters can be synthesized and their biosynthetic enzymes expressed in vitro in culturable hosts (Supplementary Figure S5.1).[3] While an attractive approach to produce a secondary metabolite without culturing, considerable effort is required before one can reduce these large genomic sets into biosynthetic systems and associated secondary metabolites.[4] One of the largest issues involves the discovery of the appropriate

biosynthetic pathway. While metagenome-based discoveries have revolutionized the way one can access microbes at the genetic level,[5] the presence of a biosynthetic gene cluster does not directly correlate to a functional biosynthetic pathway, and would benefit from methods that enable sample selection based on biomarkers that confirm biosynthetic activity prior to genomic analyses.[6]

Here we apply fluorescent in vivo labeling of biosynthetic proteins as a tool to guide the selection of individual microorganisms expressing secondary metabolic activities of interest directly in their macroorganism host for capture and single-cell genome sequencing. This activity-guided approach identifies targeted pathways in rare microorganisms without a priori knowledge of microorganism identity or cultivation.[7] As shown in Figure 5.1b, this workflow provides a robust complement to current genomic sequencing approaches by applying a fluorescent in situ biochemical readout as a tool for organism selection, a process that can be directly integrated into single-cell genomics workflows.[8]

To demonstrate this approach, we turned to carrier proteins (CPs) and their associated 4'- phosphopantetheinyltransferases (PPTases),[9] an enzyme-substrate pair that plays a key role in the biosynthesis of fatty acids (FAs), polyketides (PKs) and non-ribosomal peptides (NRPs). As shown in Figure 5.1b, this system provides a durable model, as we have previously shown that a variety of synthetic pantetheine analogues (pantetheinamides) can penetrate the cell for functional activity.[10] Once inside the cell, the pantetheinamide hijacks coenzyme A (CoA) biosynthesis for conversion to the corresponding CoA analog and ultimately becomes site-selectively tethered onto the CP by the action of a PPTase (bottom, Figure 5.1b).[9] Importantly, the Sfp-type PPTases,

which are associated with secondary metabolism of PKS and NRPS, are well known to show promiscuity for CoA identity, and therefore are excellent catalysts for labeling CPs in vivo. Here, we demonstrate how this process can be used to fluorescently label CPs of the secondary metabolism by attaching a suitable tag as an amide linkage upon the N-terminus of the pantetheinamide (Figure 5.1b).

Our study began by the selection of a dynamic fluorescent reporter. Developed in 2008,11 4-dimethylnaphthalene (4-DMN) is a solvatochromic tag with a fluorescence response dependent upon the hydrophobicity of its local environment.[12] In our previous studies, we used a 4-DMN-labeled pantetheinamide (DMN-P, Figure 5.1b) to probe protein-protein interactions between Escherichia coli fatty acid synthase (FAS) carrier protein (EcACP) and its multiple partner enzymes.[13] When an EcACP is labeled with DMN-P, the probe is sequestered within the hydrophobic alpha helical core of EcACP,[14] leading to a large increase in fluorescence intensity relative to that unbound in solution. Based on our previous studies,[10] we anticipated that DMN-P would cross the cell membranes and label CPs *in vivo*. The fact that the DMN-P was not fluorescent in solution but fluoresced when protein-bound suggested that this strategy could be used to directly identify cells containing the enzymatic machinery necessary to load a CP with DMN-P, as illustrated in Figure 5.1b. The use of CP-PPTase pairs was ideally suited to develop this method as CP domains are found in the majority PKS and NRPS systems.

We began by testing the labeling strategy using a set of culturable marine microbes. Super-resolution imaging (see Online Methods) of Staphylococcus sp. CNJ924 indicated that specific cells (blue, Figure 5.1c) became fluorescent upon incubation with 500 nM DMN-P. Counterstaining with 1 μM SYTO 9, a green-fluorescent nucleic acid live-

cell stain, indicated that while some cells stained with DMN-P (green and blue, Figure 5.1c), others did not (green only, Figure 5.1c), suggesting that only a subset of cells were biosynthetically active for CP labeling. This observation further suggested that DMN-P staining was not only sufficient to label Staphylococcus sp., but also responded to the metabolic state of individual microbial cells. In other species, such as the rod-shaped Bacillus sp. CNJ803, consistent subcellular localization was observed. As evident in Fig. 1d, Bacillus sp. cells incubated with 500 nM DMN-P displayed blue fluorescence throughout their cytosolic space, in contrast to DNA staining, which appeared within nanometer-sized nucleoid regions. From these examples, we were able to demonstrate cellular and intracellular selective labeling of microbial cells with DMN-P and develop methods to sort them (Figure 5.1e). Most importantly, these studies demonstrated that the DMN-P probe targeted specific cells and did not present uniform labeling expected by targeting a primary metabolic pathway such as FA biosynthesis.

We then applied this labeling on the microbiome associated with the colonial marine tunicate Ciona intestinalis, a marine invertebrate chordate that has provided valuable resource for developmental biology, evolutionary biology and neuroscience,[14,15] and has been suggested as a model for host-microbe interaction studies as marine invertebrates are known to be hotspots for secondary metabolite production.16 It was also one of the first animal to have its genome sequenced.[14] Tunicate Ciona intestinalis (Figure 5.1a) were collected from a dock in the Gulf of Maine. Freshly collected tunicates were incubated in sterile-filtered seawater along with control samples of the proximal water column (WC) with either RedoxSensor Green (RSG), a marker of bacterial cell viability, or the enzyme labeling DMN-P. To ensure sufficient sample of microbiome cells,

a total of four tunicates were pooled for each probe. After homogenization of the tunicates, probe-positive cells were subjected to fluorescence-activated cell sorting (FACS) to deposit fluorescently-labeled cells into a 384-well microplate, one per well (see Supplementary Information).17 Of the ~12,000 µL-1 viable (RSG-positive) microbial cells in its tunicate homogenate, only ~13 µL-1 (0.1%) were labeled with DMN-P in a comparable homogenate, suggesting high probe specificity. This was further supported by the enrichment of sequences coding for PKS and NRPS synthases observed in the DMP-P sorted microbes (Figure 5.2b). The DNA of 146-148 sorted particles per treatment was amplified, sequenced and de novo assembled individually (Table S5.1).17 In total, we obtained the following number of >20 kbp genome assemblies from individual, sorted particles: a) 50 of tunicate microbiome CP-expressing cells (DMN-P probe positives); b) 59 of tunicate microbiome respiring cells (RSG probe positives); and c) 95 of adjacent seawater bacterioplankton (SYTO-9 probe positives).

The phylogenetic composition of the analyzed microbial cells from the adjacent water column (WC) was typical of the Gulf of Maine prokaryoplankton,[18] with a predominance of pelagic lineages SAR11 (Alphaproteobacteria) and Flavobacteriales (Bacteroidetes) (Figure 5.2a; Table S5.1). The viable microorganisms in the tunicate microbiome were dominated by Rhizobiales, Kordiimonadales (Alphaproteobacteria), Flavobacteriales (Bacteroidetes) and Campylobacteriales (Epsilonproteobacteria), which have been previously shown to be associated with C. intestinalis.[19] In contrast, cells from the tunicate microbiomes labeled with DMN-P were dominated by Campylobacteriales; Flavobacteriales; Alteromonadales, Cellvibrionales, Vibrionales and Oceanospirillales (Gammaproteobacteria). Genomic sequencing indicated that while particles labeled with

SYTO-9 and RSG were individual microbial cells, many particles labeled with DMN-P contained multiple, phylogenetically divergent co-sorted cells. This potentially suggests that active CP labeling by PPTase may be greatest when the producer cell is in direct contact with a different microorganism.

As shown in Figure 5.2b, antiSMASH 5.020 identified 0.77 metabolite pathways per Mbp in the RSG probe-positive single amplified genomes (SAGs) and 1.17 metabolite pathways per Mbp in DMN-P positive SAGs, a level that was considerably higher than the 0.47 metabolite pathways per Mbp that were observed in the adjacent water column (p-value 0.024). This observation confirmed that activity-based probes, such as DMN-P, can select microbes with biosynthetic potential.

To provide an experimental evidence that the detected CP had been truly targeted by the DMN-P probe and to confirm the CP specificity, we explored two CPs from two different SAGs. The first CP1, and its associated PPTase, PPT1, were obtained from a putative NRPS/Type I hybrid gene cluster (this gene cluster was observed only in this SAG) in an Oceanospirillales (Gammaproteobacteria) genome AH-491-C20 (red, Figure 5.2c), which had the highest 16S rRNA gene similarity (97%, 80% length overlap) to Amphritea spongicola MEBiC0546.[21] CP1 was chosen as it provided an excellent example of a hybrid NRPS-PKS, had a proximal PPTase, and it contained a unique domain architecture (Figure 5.2c) suggesting it was a novel NRPS. Analysis of tetramer frequencies and contig binning showed that the contig containing the NRPS genes was similar to other contigs with Oceanospirillales marker genes, which excludes the possibility that the NRPS gene cluster belonged to another co-sorted microbe or contaminating DNA (Supplementary Figure S5.3). A second, unrelated CP2, was

obtained from a predicted NPRS/NRPS-like/Type I PKS cluster in a Cellvibrionales (Gammaproteobacteria) genome AH-491-D14, with Oceanicoccus sagamiensis NBRC107125 as the most closely related cultured isolate in Genbank (96% 16S rRNA gene identity, 81% overlap).[22] CP2 was chosen as it also contained and NRPS-PKS and would provide an excellent comparison to CP1.

Escherichia coli codon optimized genes were synthesized for CP1 (Supplementary Figure S5.4), CP2 (Supplementary Figure S5.5), and PPT1 (Supplementary Figure S5.6), inserted into pET28a vectors, and their associated proteins were prepared by recombinant expression in *E. coli* followed by His6-tagged purification (Supplemental Figure S5.7a). Applying the method in Figure 5.1b *in vitro*, samples of CP1 and CP2 were screened for their ability to be fluorescently labeled. Recombinant Sfp,9 a member of the surfactin-type PPTase known to have a broad scope in CP labeling, was able to load DMN-P onto CP1 under conditions established to label the EcACP, a positive control (Figure 5.2d, Supplementary Figure S5.7a). Under the initial experimental conditions, CP2 was not labeled, nor was CP1 or CP2 labeled with the PPT1. Concerned that unfolding the proteins under SDS-PAGE would destroy the environmentally sensitive fluorescence from the DMN-P probe, we repeated the labeling process (Figure 5.1b) with a non-solvatochromic dye labeled CoA, TAMRA-CoA.[23] Here we observed labeling of CP1 with Sfp (Figure 5.2e, Supplementary Figure S5.7b-d). Interestingly, PPT1 was only able to label CP1, which comes from the same bacterial species. Remarkably, while Sfp could label EcACP and CP1, the fact that PPT1 only labeled its native substrate CP1, and not CP2 or EcACP, demonstrates the unique selectivity found within PPTases.[9]

To explore the diversity and presence of biosynthetic gene clusters related to the Oceanospirillales bacterium AH-491-C20 NRPS across Bacteria and Archaea in publicly available genomic datasets, we used IMG/M database24 (March 2020, containing 77,808 genomes from isolates, SAGs and metagenome assembled genomes from diverse environments) and 12,715 SAGs from the GORG-Tropics data set from marine prokaryoplankton.[25] Surprisingly, the AH-491-C20 NRPS gene cluster was not present in any known members of the Oceanospirillales order (Figure 5.2f), with the exception of Zymobacter palmae, an ethanol-fermenting species isolated from palm sap,[26] which had moderate amino acid sequence similarity (40-60%) to this gene cluster, but with a different domain organization (Figure 5.2c). An amino acid sequence similarity below 60% and different domain organization was also found in other members of Proteobacteria and Firmicutes phyla (Figure 5.2f). The most similar domain organizations were found in pathogenic bacteria, such as Vibrio cholerae and Vibrio mimicus[27] and in marine bacteria, such as Shewanella psychrophila and Photobacterium profundum[28] (Figure 5.2c) belonging to different orders in the Proteobacteria. This yersiniabactin- and vibriobactin-like cluster was not found in any of the SAGs generated using the RSG probe in this study, which indicates the cluster's low abundance in the total active tunicate microbiome and highlights the utility of the activity-guided single-cell genomics for bioactive molecules.

Overall, this study demonstrates how activity-based fluorescent labeling can be coupled with single-cell genomics to enrich for organisms expressing specific biosynthetic activity. The usage of environmentally sensitive fluorescent probes designed to label CPs in situ enabled the identification of previously undiscovered CP substrate and compatible

PPTase enzyme partners in the same, uncultured microbial cell. The use of this solvatochromic DMN-P probe provided an enhanced response (increased fluorescence when attached to a CP9) that enabled selective detection of bacteria presenting CP-containing synthases. Using model bacteria, we were able to show that the DMN-P probe selectively labeled microbial cells and we were able to use this signal to gate cell sorting for cells that would most likely contain an active NRPS or PKS CP-PPTase pair. While one cannot rule out that FAS pathways may have also been labeled, our data suggest that the application of sort gates with strict fluorescence intensity thresholds selects for microbial cells that contain added CP-PPTase activity associated with NRPS/PKS biosynthesis. We illustrated the enrichment for bacteria with biosynthetic activity with an example of a novel NRPS gene cluster found in an Oceanospirillales species from tunicate microbiome, which was below detection in the non-labeled control and also absent in large public repositories of microbial genomes. While well recognized in their human health context (see homology to Yersinia pestis and Vibrio cholerae in Figure 5.2c), the discovered yersiniabactin and vibriobactin biosynthesis pathways have not been documented in Oceanospirillales nor tunicate microbiome previously. This indicates that such novel biochemistry would remain undiscovered without the use of our activity-guided single-cell genomics approach.

Acknowledgments

This study was supported by NIH R21 AI 134037 to M. D. B. and R. S. R. S. was also supported by NSF 826734, 1441717 and 133581. We thank Prof. William Fenical (Scripps Institution of Oceanography, UC San Diego) for samples of Staphylococcus sp. CNJ924 and Bacillus sp. CNJ803 used to develop the labeling conditions and Prof. Jörn

Piel (ETH) for the Bacillus subtilis strains used as positive and negative controls. Confocal imaging was possible through the assistance of Jennifer Santini (UC San Diego) and supported by funding from the NIH P30 NS047101.

Author contributions

K. C. prepared the samples of DMN-P; W. E. K., K. C. and J. J. L. conducted the super-resolution imaging; E. D. B., J. B. and R. S. collected the tunicates and associated water column; E. D. B. and J. B. prepared the microbial extracts from the tunicates; E. D. B., J. B. and R. S. conducted the FACS analyses; E. D. B., J. B. and R. S. conducted the single-cell genomic sequencing; E. D. B. and R. S. conducted the genome assemblies and conducted biosynthetic pathway searches; E. D. B., K. C., W. E. K., M. D. B., J. J. L. and R. S. mined and selected the CP and PPTase genes; K. C. and W. E. K. cloned and expressed the CP and PPTases; K. C. and W. E. K. performed the CP labeling studies; M. D. and F. S. performed sequence analysis of the NRPS gene cluster; and T. W., J. J. L, M. D. B. and R. S. organized the study. All authors contributed to the writing.

Additional information

Experimental methods for the DMN-P FACS analyses, single-cell DNA amplification and sequencing, SAG classification and assembly, and biosynthetic pathway analyses have been provided. The supplementary information also contains enlarged microscopic images, gene and protein sequences of CP1, CP2 and PPT1, and raw images of the SDS-PAGE gels. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

References

1. Kenshole, E., Herisse, M., Michael, M. & Pidot S. J. Natural product discovery through microbial genome mining. *Curr. Opin. Chem. Biol*. 60, 47–54 (2021).

2. Palaniappan, K., Chen, I.M.A., Chu, K., Ratner, A., Seshadri, R., Kyrpides, N.C., Ivanova, N.N., & Mouncey, N.J. IMG-ABC v.5.0: an update to the IMG/atlas of biosynthetic gene clusters knowledgebase. *Nucleic Acids Res*. 48, D422–D430 (2020).

3. Wang G, Zhao Z, Ke J, Engel Y, Shi YM, Robinson D, Bingol K, Zhang Z, Bowen B, Louie K, Wang B, Evans R, Miyamoto Y, Cheng K, Kosina S, De Raad M, Silva L, Luhrs A, Lubbe A, Hoyt DW, Francavilla C, Otani H, Deutsch S, Washton NM, Rubin EM, Mouncey NJ, Visel A, Northen T, Cheng JF, Bode HB, Yoshikuni Y. CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat Microbiol*. 2019 Dec;4(12):2498-2510.

4. Kalkreuter E, Pan G, Cepeda AJ, Shen B. Targeting Bacterial Genomes for Natural Product Discovery. *Trends Pharmacol Sci*. 2020 Jan;41(1):13-26

5. Miller IJ, Chevrette MG, Kwan JC. Interpreting Microbial Biosynthesis in the Genomic Age: Biological and Practical Considerations. *Mar Drugs*. 2017;15(6):165. 2017 Jun 6.

6. Li L, Liu X, Jiang W, Lu Y. Recent Advances in Synthetic Biology Approaches to Optimize Production of Bioactive Natural Products in Actinobacteria. *Front Microbiol.* 2019;10:2467. 2019 Nov 5.

7. Martinez-Garcia, M., Brazel, D. M., Swan, B. K., Arnosti, C., Chain, P. S., Reitenga, K. G., Xie, G., Poulton, N. J., Lluesma Gomez, M., Masland, D. E., Thompson, B., Bellows, W. K., Ziervogel, K., Lo, C. C., Ahmed, S., Gleasner, C. D., Detter, C. J., & Stepanauskas, R. Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia*. PloS one*, 7(4), (2012).

8. Doud DFR, Woyke T. Novel approaches in function-driven single-cell genomics. *FEMS Microbiol Rev*. 2017 Jul 1;41(4):538-548.

9. Beld J, Sonnenschein EC, Vickery CR, Noel JP, Burkart MD. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life*. Nat Prod Rep*. 2014;31(1):61-108.

10. Reyes, C. P., La Clair, J. J. & Burkart, M. D. Metabolic probes for imaging endosymbiotic bacteria within toxic dinoflagellates. *Chem. Comm*. 46, 8151–8153 (2010).

11. Loving, G & Imperiali, B. A Versatile Amino Acid Analogue of the Solvatochromic Fluorophore 4-N,N-Dimethylamino-1,8-naphthalimide: A Powerful Tool for the Study of Dynamic Protein Interactions. *J. Am. Chem. Soc*. 130, 13630–13638 (2008).

12. Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Acc. Chem. Res.* 50, 366–375 (2017).

13. Charov, K. & Burkart, M. D. A Single Tool to Monitor Multiple Protein-Protein Interactions of the *Escherichia coli* Acyl Carrier Protein. *ACS Infect. Dis*. 5, 1518–1523 (2019).

14. Satou, Y., Nakamura, R., Yu, D., Yoshida, R., Hamada, M., Fujie, M., Hisata, K., Takeda, H., & Satoh, N. A Nearly Complete Genome of Ciona intestinalis Type A (C. robusta) Reveals the Contribution of Inversion to Chromosomal Evolution in the Genus Ciona. *Genome Biol. Evol.* 11, 3144–3157 (2019).

15. Cao, C., Lemaire LA, Wang W, Yoon PH, Choi YA, Parsons LR, Matese JC, Wang W, Levine M, Chen K. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. Nature. *Nature* 571, 349–354 (2019).

16. Leigh, B. A., Liberti, A. & Dishaw, L. J. Generation of Germ-Free Ciona intestinalis for Studies of Gut-Microbe Interactions. *Front. Microbiol.* 7, 2092 (2016).

17. Stepanauskas, R., Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun.* 8, 84 (2017).

18. Giovannoni, S. J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans *Ann. Rev. Mar. Sci.* 9, 231–255 (2017).

19. Blasiak, L. C., Zinder, S. H., Buckley, D. H. & Hill, R. T. Bacterial diversity associated with the tunic of the model chordate Ciona intestinalis. *ISME* J. 8, 309–320 (2014).

20. Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S., Suarez, H.G., de los Santos, E.L.C., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., weber, T., & Medema, M.H. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res*. 45(W1), W36–W41 (2017).

21. Jang, H., Yang SH, Seo HS, Lee JH, Kim SJ, Kwon KK. Amphritea spongicola sp. nov., isolated from a marine sponge, and emended description of the genus Amphritea. *Int. J. Syst. Evol. Microbiol*. 65, 1866–1870 (2015).

22. Park, S., Yoshizawa, S., Kogure, K. & Yokota, A. Oceanicoccus sagamiensis gen. nov., sp. nov., a gammaproteobacterium isolated from sea water of Sagami Bay in Japan *J. Microbiol.* 49, 233–227 (2011).

23. Tallorin, L. Wang J, Kim WE, Sahu S, Kosa NM, Yang P, Thompson M, Gilson MK, Frazier PI, Burkart MD, Gianneschi NC. Discovering de novo peptide substrates for enzymes using machine learning. *Nat. Commun*. 9, 5253 (2018).

24. Chen, I. A. Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. 47, D666–D677 (2019).

25. Pachiadaki, M. G., Brown JM, Brown J, Bezuidt O, Berube PM, Biller SJ, Poulton NJ, Burkart MD, La Clair JJ, Chisholm SW, Stepanauskas R. Charting the Complexity of the Marine Microbiome through Single-Cell Genomics. *Cell*. 179, P1623–1635.E11 (2019).

26. Okamoto, T., Taguchi, H., Nakamura, K., Ikenaga, H., Kuraishi, H., and Yamasato, K. Zymobacter palmae gen. nov., sp. nov., a new ethanol-fermenting peritrichous bacterium isolated from palm sap. *Arch. Microbiol.* 160, 333–337 (1993).

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

28. Oku, N., Kawabata, K., Adachi, K. et al. Unnarmicins A and C, New Antibacterial Depsipeptides Produced by Marine Bacterium Photobacterium sp. MBIC06485. *J. Antibiot.* 61, 11–17 (2008).

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

30. Khater, S., Gupta M, Agrawal P, Sain N, Prava J, Gupta P, Grover M, Kumar N, Mohanty D. SBSPKSv2: structure-based sequence analysis of polyketide synthases and non-ribosomal peptide synthetases. *Nucleic Acids Res.* 45(W1), W72–W79 (2017).

31. Bachmann, B. O. & Ravel J. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data *Methods Enzymol.* 458, 181–217 (2009)

32. Miller, B. R., Sundlov JA, Drake EJ, Makin TA, Gulick AM. Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* 82, 2691–2702 (2014).

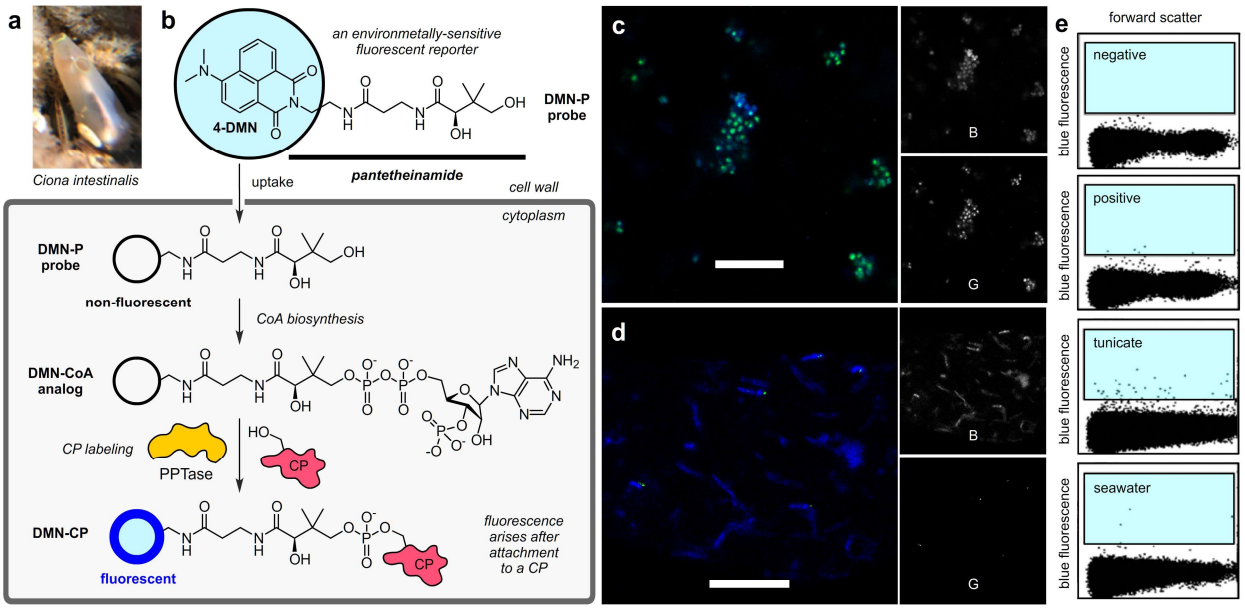


Figure 5.1: Activity-guided microbial single-cell genomics. a) An image of the tunicate Ciona intestinalis specimen explored in this study. b) A schematic representation of the fluorescent CP labeling method used in this study. A fluorescent pantetheinamide DMN-P, designed to mimic pantetheine, is taken up by a microbial cell, where it is converted to a fluorescent-CoA analog. An environmentally sensitive fluorescent tag is used to improve the detection of protein-conjugation in vivo, as the 4-DMN dye used in the DMP-P probe has been shown to undergo an increase and shift in fluorescence once appended to a CP by a PPTase. c)-d) Super-resolution images of c) Staphylococcus sp. CNJ924 or d) Bacillus sp. CNJ803 incubated for 3 h with 500 nM DMN-P (blue) and then stained with 1 μM nuclear SYTO-9 (green) prior to fixation and imaging. Bar denotes 10 μm. Enlarged images are provided in Supplementary Fig. S2. e) Cells (black dots) sorted and selected (blue regions) by flow cytometry with DMN-P staining for tunicate and seawater samples. Negative (Bacillus subtilis 168, a mutant deficient in Sfp) and positive (Bacillus subtilis 3610 DSM10, wild type) controls indicate DMN-P stained cells with active CP - PPTase pairs. (2 columns)

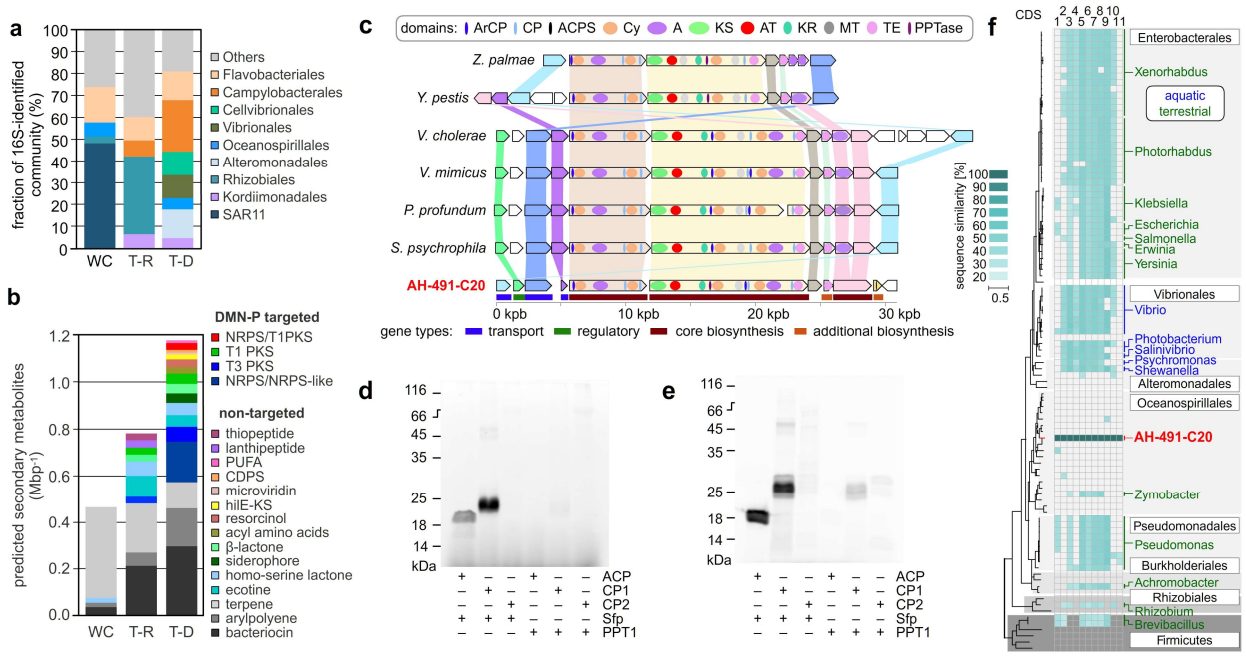


Figure 5.2: Application of activity-guided single-cell genomics to mine selective CP and PPTase interactions from a tunicate microbiome. a) Bacterial composition and b) predicted secondary metabolite clusters per megabase (Mbp-1) of the total microbial community in the water column (WC) compared to the viable (T-R) and DMN-P-responsive (T-D) cells in the C. intestinalis microbiome. In vitro CP-labeling analyses on the cloned and expressed CP1, CP2 and PPT1 mined from the genomic data. c) Visualization of the NRPS gene cluster from SAG AH-491-C20 and examples of alignments to other bacterial species which had >30% amino acid similarity on >75% of the query sequence length. d)-e) SDS-PAGE gel depicting the fluorescence in CPs after labeling by Sfp or PPT1 with d) DMN-P or e) TAMRA-CoA. E. coli FAS ACP (EcACP) was used as a positive control. f) Presence of the AH-491-C20 NRPS gene cluster across different phylogenetic groups. The phylogenetic tree includes genomes, which contained at least 7 out of 11 genes of this cluster and their closest relatives. Heatmap illustrates the level of sequence similarity for each gene. (2 columns)

ONLINE METHODS

General materials. Unless described otherwise, supplies and materials were obtained from VWR or Fischer Scientific and used as is. The DMN-P[13] and TAMRA-CoA[23] were prepared by chemical synthesis. *E. coli* FAS ACP (EcACP) was prepared using established methods.[13]

Microbial isolate labeling studies. A single colony of each strain (Staphylococcus sp. CNJ924 or Bacillus sp. CNJ803) was suspended in 200 µL of A1 media. Cells were then treated at 23 °C for 3 h with 500 nM DMN-P (Figure 5.1b) from a 100 µM stock solution of DMN-P dissolved in DMSO. Following incubation with DMN-P, nuclear stain, SYTO-9

(Thermo Fisher Scientific), was added to the cultures for 15 min at a final concentration of 1 μM. Cells were spun down at 2,000 rpm for 1 min and the supernatant was removed. Cell pellets were mixed with 200 μL of 3:1 EtOH/AcOH (fixing solution) and incubated for 10 min. Cells were centrifuged at 2,000 rpm for 1 min again and the supernatant was removed. The cells were re-hydrolyzed with 100 μL of water. A 20 μL aliquot of each final sample was loaded onto glass slides for super-resolution microscopy.

Super-resolution microscopy. Imaging was conducted on a Zeiss LSM 880 confocal laser scanning microscope with FAST Airyscan equipped with a Plan-Apochromat 63x/1.4 Oil DIC M27 objective. Blue fluorescence from the DMN-P was obtained using a 405 nm laser with beamsplitters (MBS 488/561/633, MBS_InVis MBS -405, DBS1 mirror, FW1 rear) and an additional 410-477 nm filter. Green fluorescence from the SYTO-9 control was obtained using a 488 nm laser with beamsplitters (MBS 488/561/633, MBS_InVis MBS-405, DBS1 mirror, FW1 rear). Pinhole sizes were kept between 40-90 μm. Images were collected in Zen (Zeiss) with recommended gains of 700-1100, digital gain of 1, depth of focus between 0.65-0.71 μm, and pixel time of 2-5 μs and image sizes at 2048 × 2048 pixels. Images were processed offline and rendered from Zen Blue (Zeiss). Copies of original CZI files can be provided upon request.

Specimen collection. Live Ciona intestinalis were collected from 20.4 °C seawater at a depth of ~0.5 m on the side of a floating dock in Boothbay Harbor Maine (Latitude 48.8, Longitude -69.6) between 8:00-9:00 AM on August 29, 2018. To ensure sufficient sample size, a total of four tunicates (these tunicates were selected same location and were neighboring and attached to the same substrate) were pooled for each study. Immediately after collection, the specimens were placed in 50 mL centrifuge tubes containing ambient

seawater. Additionally, ambient seawater samples were collected adjacent to tunicates in 50 mL centrifuge tubes. The samples were transported to Bigelow Laboratory for Ocean Sciences at in situ temperature in the dark.

Activity-guided single-cell genomics. To target cells with CP activity, four tunicates were incubated in ambient seawater amended with 10 μM DMN-P at in situ temperature in the dark for 2 h. To target all viable microbial cells, four other tunicates were incubated in 80 mL of ambient seawater amended with 1 μM RedoxSensor Green (RSG, Thermo Fisher Scientific) at in situ temperature in the dark for 20 min. The tunicates of each treatment were rinsed in sterile Sargasso Sea water, pooled and homogenized together with 50 mL sterile Sargasso Sea using a Ninja Professional 900 W blender until the majority of the biomass was visibly disintegrated. The homogenate was spun down at 2,000 rpm for 1 min and the supernatant was passed through a 100 μm mesh filter. To assess the composition of the sortable microorganisms in the seawater around tunicates, ambient seawater samples were labeled with a 5 μM SYTO-9 live nucleic acid stain (Thermo Fisher Scientific) for 20-40 min. Immediately before cell sorting, samples were diluted 10× in sterile Sargasso Sea water. Sort gates for DMN-P probe-positive cells were defined along blue fluorescence and forward scatter axes and adjusted for background noise using negative (Bacillus subtilis 168, a mutant deficient in Sfp) and positive (Bacillus subtilis 3610 DSM10, wild type) controls (methods for microbial growth was described in the above section entitled Microbial isolate labeling studies). Probe-positive cells were sorted into 384-well plates, lysed, their DNA amplified with WGA-X, and genomes sequenced and quality-controlled.17 Genome assemblies originating from multiple co-sorted cells were identified and parsed using a combination of nucleotide tetramer

principal component analysis and homology searches in the NCBI nr database.29 Biosynthetic pathways were identified with antiSMASH20 using KnownClusterBlast, ActiveSiteFinder and SubClusterBlast options. If antiSMASH predicted multiple metabolite pathways in the same coding region, then all possible products were reported.

Heterologous expression of select genes/domains. Selected domains and di-domains were chosen based upon evaluation of sequencing data with antiSMASH (versions 4.2 and 5.0 were used).20 In particular, we selected a PPTase (PPT1) and two CP (CP1 and CP2) containing sequences for recombinant evaluation. The predicted protein sequences of the full selected genes were further evaluated with BLAST alignment to identify homologues and better understand domain organization. In the case of CPs, disconnection locations for excised domains were determined by online domain organization tools30,31 and by comparison with recent crystal structures.32,33 The genes for the resulting proteins were synthesized (Twist Bioscience) and cloned into a pET28a vector. With the exception of the PPTase, all genes were cloned with a 3'- stop codon to provide an N-terminal His6 tag for immobilized metal affinity chromatography purification. The PPTase gene was cloned without a stop codon to provide a C-terminal His6 tag based upon our prior experience with loss of activity with N-terminal fusions.11 The identified protein sequences, gene sequences synthesized and protein sequences are provided in Supplementary Figures S5.4-S5.6.

Expression of CP1, CP2, and PPT1. The His6-tagged proteins were expressed in *E. coli* (BL21) and grown in Terrific Broth. Cells were grown in the presence of 50 mg/L kanamycin, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (ITPG) at OD600 = 0.8, and incubated at 16 °C for 16 h. The cell culture was spun down by centrifugation

at 2,000 rpm for 30 min and the collected pellets were lysed by sonication, followed by another centrifugation at 10,000 rpm for 1 h to clear the lysate. The proteins were purified using Ni-NTA resin (ThermoFisher). Purified proteins were collected and concentrated to 2-4 mg/mL using 3 kDa (CP1 and CP2) or 10 kDa (PPT1) Amicon Ultra centrifuge filters (Millipore).

Carrier protein labeling studies. The CP labeling studies were conducted in a 30 µL reaction containing final concentration of 100 µM of the respective CP1 or CP2, 1 mM of TAMRA-CoA, and 1 µM of Sfp or PPT1 in 50 mM HEPES, 10 mM MgCl2 pH 7.2. The mixture was gently shaken at 23 °C overnight (12 h). The resulting reactions were analyzed by 12% SDS-PAGE gel and imaged on Typhoon TRIO Variable Mode Imager (GE Healthcare BioSciences). Imaging was conducted using Cy2 (Excitation 473 nm, Emission 520 nm) for DMN-P and Cy3 (Excitation 532 nm, Emission 580 nm) for TAMRA-CoA.

Supplementary Methods

DMN-P Guided Fluorescent Activated Cell Sorting (FACS). Bigelow Laboratory Single Cell Genomics Center's (scgc.bigelow.org) standard workflow was employed to generate and sequence single amplified genomes.[17] Prior to FACS, samples were diluted to below 105 cells mL-1 with filter-sterilized Sargasso Sea water and pre-screened through a 40 µm mesh filter (Becton Dickinson). FACS was performed using a BD InFlux Mariner flow cytometer equipped with 365 and 488 nm lasers for excitation and either a 70 µm nozzle orifice (Becton Dickinson, formerly Cytopeia).[17] The cytometer was triggered on side scatter, and the "single-1 drop" mode was used for maximal sort purity. Gates for the sorting were defined based on particle blue (DMN-P) or green (SYTO-9 and RSG probes) fluorescence, light side scatter, and the ratio of green versus red fluorescence (for improved discrimination of cells from detrital particles).

Single-Cell DNA Amplification and Genome Sequencing. Cells were lysed and their DNA was denatured by 2 freeze-thaw cycles, the addition of 700 nL of a lysis buffer consisting of 0.4 M KOH, 10 mM EDTA and 100 mM dithiothreitol, and a subsequent 10 min incubation at 20 °C. The lysis was terminated by the addition of 700 nL of 1 M Tris•HCl, pH 4. The 10 µLWGA-X reactions for genomic DNA amplification contained 0.2 U µL-1 Equiphi29 polymerase (Thermo Fisher Scientific), 1× Equiphi29 reaction buffer (Thermo Fisher Scientific), 0.4 µM each dNTP (New England BioLabs), 10 µM dithiothreitol (Thermo Fisher Scientific), 40 µM random heptamers with two 3'-terminal phosphorothioated nucleotide bonds (Integrated DNA Technologies), and 1 µM SYTO-9 (Thermo Fisher Scientific). These reactions were performed at 45 °C for 12-16 h, then inactivated by a 15 min incubation at 75 °C. In order to prevent WGA-X reactions from

contamination with non-target DNA, all cell lysis and gDNA amplification reagents were treated with UV in a Stratalinker (Stratagene). An empirical optimization of the UV exposure was performed in order to determine the length of UV exposure that is necessary to cross-link all detectable contaminants without inactivating the reaction. Cell sorting, lysis and gDNA amplification setup were performed in a HEPA-filtered environment conforming to Class 1000 cleanroom specifications. Prior to cell sorting, the instrument, the reagents and the workspace were decontaminated for DNA using UV irradiation and sodium hypochlorite solution. To further reduce the risk of DNA contamination, and to improve accuracy and throughput, Bravo (Agilent Technologies) and Freedom Evo (Tecan) robotic handlers were used for all liquid handling in 384-well plates.

Libraries for SAG genomic sequencing were created with Nextera XT (Illumina) reagents following the manufacturers instructions, except for purification steps, which were done with column cleanup kits (Qiagen), and library size selection, which was done with BluePippin (Sage Science, Beverly, MA), with a target size of 500±50 bp. Libraries were sequenced with NextSeq 500 (Illumina) in 2×150 bp mode using v.2 reagents. For each experimental treatment, 146-148 SAGs were subject to Low Coverage Sequencing (LoCoS),17 resulting in a variable number of 2x150bp reads per SAG, with an average of ~300k paired-end reads. Subsequently, 52 SAGs generated from the DMN-P-positive particles and 51 SAGs generated from RSG-positive particles were selected based on their fastest WGA-X kinetics on a SAG plate and subject to deeper, post-LoCoS sequencing, producing at least two million 2x150bp reads per SAG.[17] The obtained sequence reads were quality-trimmed with Trimmomatic v0.3234 using the following

settings: -phred33 LEADING:0 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:36. Human DNA (≥95% identity to the H. sapiens reference assembly GRCh38) and low complexity reads (containing <5% of any nucleotide) were removed. For de novo assemblies, quality-filtered reads were digitally normalized with kmernorm 1.05 (http://sourceforge.net/projects/kmernorm) using settings -k 21 -t 30 -c 3 and then assembled with SPAdes v.3.0.0 using the following settings: --careful --sc --phred-offset.35 Each end of the obtained contigs was trimmed by 100 bp, and then only contigs longer than 2,000 bp were retained. Based on best blast hits to Ciona intestinalis seven SAGs were found to belong to the eukaryote host and were excluded from further analyses. Key technical data related to SAG sequencing is provided in Table S1 (provided as a separate XLS file). This workflow was evaluated for assembly errors using three bacterial benchmark cultures with diverse genome complexity and %GC, indicating no non-target and undefined bases in the assemblies and average frequencies of mis-assemblies, indels and mismatches per 100 kbp: 1.5, 3.0 and 5.1.17

Standard functional annotation was first performed using Prokka36 with default Swiss-Prot databases supplied by the software. The output of Prokka and the secondary annotation were joined into a single, tab-delimited table with headers identifying the origin of the assignment. The genome completion estimates were performed on conserved protein-coding genes by CheckM.37 The 16S rRNA gene regions longer than 500 bp were identified using local alignments provided by BLAST against CREST's38 curated SILVA reference database SILVAMod v128; and taxonomic assignments were based on a reimplementation of CREST's last common ancestor algorithm.

Phylogenomics. Protein sequences of the NRPS gene cluster detected by antiSMASH in SAG AH-491-C20 were aligned to all publicly available microbial genomes in IMG/M24 (genomes accessed in January 2020) and GORG Tropics data set25 (accessed in March 2020) by Diamond v0.9.26 blastp.39 Only hits with a sequence similarity of at least 1e-5 and 30% alignment fraction were considered for further analysis. A set of 56 universal single copy marker proteins40 was used to build a phylogenetic tree of the genomes that had at least 7 different hits against the 11 genes in the AH-491-C20 gene cluster and of their closely related genomes regardless of the number of blastp hits, in particular all close relatives of the Oceanospirillales order (de-replicated based on 95% average nucleotide identity using FastANI 1.341). Marker proteins were identified with hmmsearch v3.1b2 (hmmer.org) using a specific HMM for each of the markers. For every marker protein, alignments were built with MAFFTS10 v7.294b42 and subsequently trimmed with BMGE using BLOSUM30.43 Single protein alignments were then concatenated and maximum likelihood phylogenies were inferred with IQ-TREE 1.6.1244 using the LG4X+F model.

The assembly AH-491-C20 harboring the NRPS gene cluster was then analyzed in details to check that the contig containing the NRPS belongs to the Oceanospirillales host. All contigs were binned by MetaBAT 2 v2.12.1.45 As the result, the contigs containing Oceanospirillales marker genes (from the previous step) and the contig containing NRPS gene cluster were found in the same genome bin. The similar genomic content of these contigs was confirmed also by the analysis of tetranucleotide frequencies, calculated for each contig on 1600 bp windows (with 200bp steps) as part of the Bigelow Laboratory Single Cell Genomics Center’s analysis pipeline.

Supplementary References:

34. Bolger, A.M., Lohse, M. & Usadel, B. Bioinformatics 30, 2114–2120 (2014).

35. Bankevich, A., et al. SPAdes: J. Comp. Biol. 19, 455–477 (2012)

36. Seemann, T. Bioinformatics 30, 2068–2069 (2014).

37. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. PeerJ PrePrints 2, e1346 (2015).

38. Lanzén, A. et al. PLoS One, 7, e49334 (2012)

39. Buchfink, B., Xie. C. & Huson, D. H. Nat. Methods. 12, 59–60 (2015).

40. Yu, F. B. et al. eLife 6, e26580 (2017).

41. Jain, C., Rodriguez-R, L. M, Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. Nat. Commun. 9, 5114 (2018).

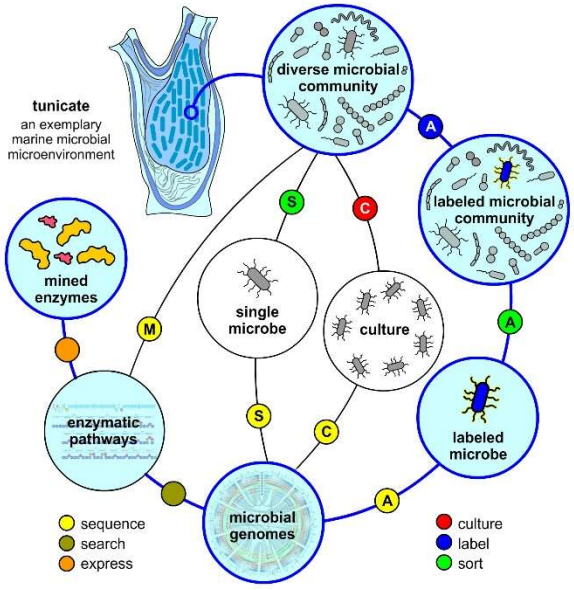
42. Katoh K & Standley D. M. Bioinformatics. 32, 1933–1942 (2016).

43. Criscuolo, A. & Gribaldo, S. BMC Evol. Biol. 10, 210 (2010).

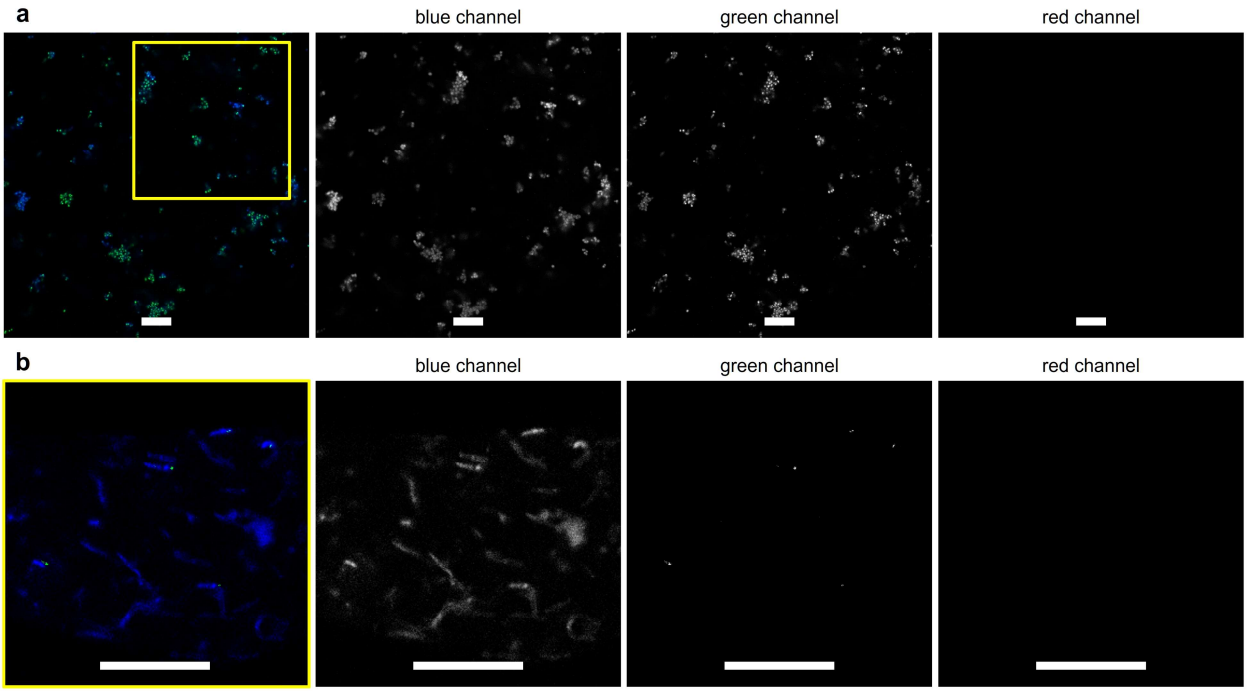
44. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. Mol. Biol. Evol. 32, 268–274 (2015).

45. Kang, D. et al. PeerJ Preprints. 7, e27522v1 (2019).

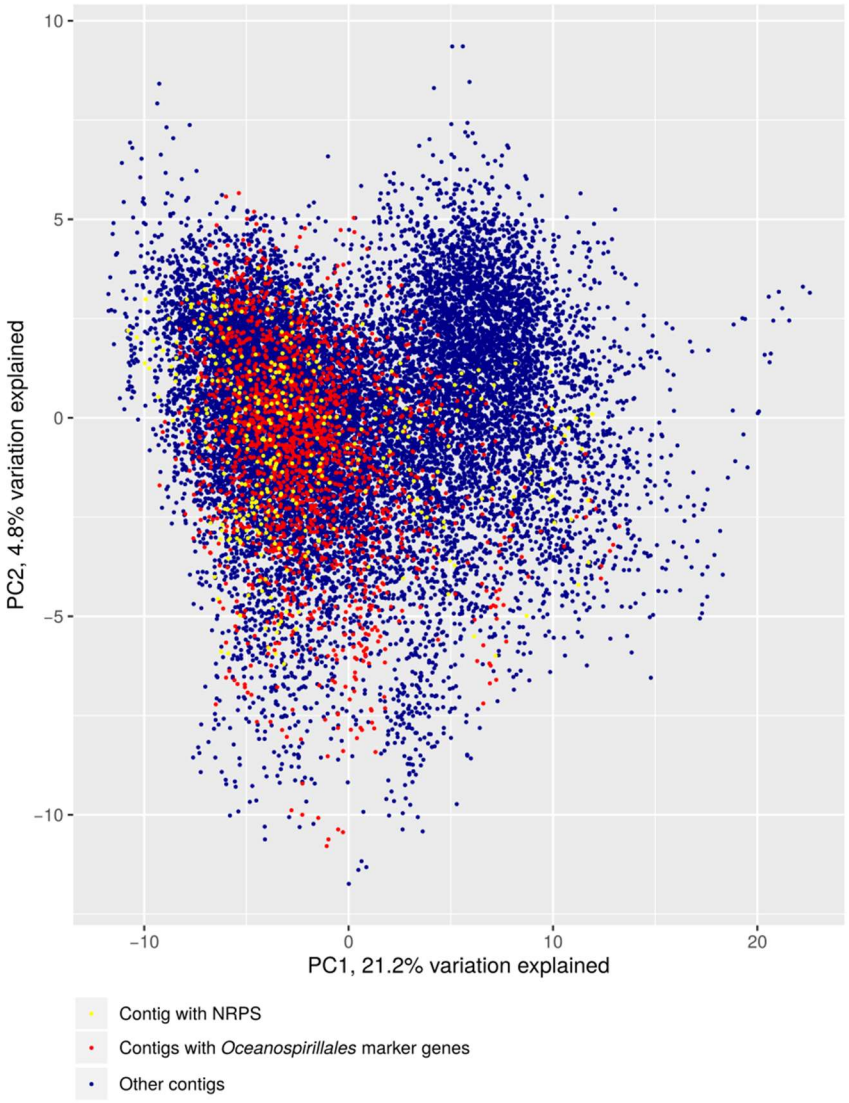
Supporting Information



Supplementary Figure S5.1: Multiple approaches exist to mine enzymes from microbial communities, including: sequencing of cultured microbial strains (C); metagenomic analyses (M); and, single-cell genomics (S). Shown as in a cycle (blue lines and light blue shading), activity-guided single-cell genomics (A) labels a targeted enzyme within the microbial community to guide the sorting of enzyme active microbes. While each pathway provides genomic data that can be searched for enzymatic pathways, the activity-guided approach enriches for the targeted enzyme prior to genome sequencing.



Supplementary Figure S5.2: Super-resolution AiryScan images. a) Image of Staphylococcus sp. CNJ924 incubated for 3 h with 500 nM DMN-P (blue) and then stained with 1 μM nuclear SYTO-9 (green) prior to fixation and imaging. b) Super-resolution images of Bacillus sp. CNJ803 incubated for 3 h with 500 nM DMN-P (blue) and then stained with 1 μM nuclear SYTO-9 (green) prior to fixation and imaging. Bars denotes 10 μm. AiryScan imaging were collected under optical settings at which the blue fluorescence in unstained Staphylococcus sp. CNJ924 and Bacillus sp. CNJ803 was indistinguishable from background. The red channel is provided as a control. Images were cropped to generate that shown in Fig. 1c-d. Raw images are shown rendered with histograms set for optimal regions for each color ZEN 2 blue edition (Zeiss). Raw files are available upon request.



Supplementary Figure S5.3: Tetranucleotide frequency analysis of the AH-491-C20 assembly. Principal component analysis (PCA) of tetranucleotide frequencies calculated on 1600 bp sliding windows (200 bp steps) of each AH-491-C20 contig. Each dot represent a single sliding window. The first principal component (x-axis) shows that contigs containing Oceanospirillales marker genes (red) and the contig containing NRPS genes (yellow) have consistent tetranucleotide composition.

Carrier Protein 1 (CP1), a putative peptidyl-CP from well AH-491-C20, Cluster 23, Ctg 2642_6

Sequences identified antiSMASH search on Ctg 2642_6:

TFWERLLEQPVTPDSDFFRLGGDSLLATRMVVELSKHLGQPVSLQHVFEQPVLQAFC
AALE

TLTGIWEDILEQPVNAQTDFFRAGGDSLSATRMVVAAQQAGFSQISLQQVFEHPRFSD
FCQVLGT

Synthesized CP1 gene:

GTCTCCGCTGATCTGCTGTTACAAGTTCAAACCTTCTGGGAACGCCTGCTGGAGCA
GCCTGTTACGCCGGATAGCGATTTTTTTCGTTTAGGTGGTGATTCGTTGTTGGCCA
CCCGTATGGTAGTCGAACTGAGCAAACATCTCGGCCAGCCGGTGTCTCTGCAACA
CGTGTTTGAACAGCCAGTGTTGCAAGCATTTTGCGCCGCATTGGAGGCCCCCGTC
ACCGAGGCCGCAGGCGACGATGTACTCTTATCAGAAACCCTCACAGGAATTTGGG
AGGATATCCTCGAACAACCAGTTAACGCGCAGACCGACTTCTTCCGCGCAGGCGG
CGACTCGCTGTCCGCCACCCGCATGGTGGTCGCGGCGCAACAGGCTGGGTTTTC
GCAAATTAGCTTACAACAAGTATTCGAACATCCGCGGTTTTCTGATTTCTGCCAGGT
ATTGGGTACAGCGGAAATTTCCGAGGTTGCAGAACCGGAGGGGGAAGTCACTGC
GACAGAAACCTAA

Sequence of Expressed CP1 protein:

MGSSHHHHHHSSGLVPRGSHMVSADLLLQVQTFWERLLEQPVTPDSDFFRLGGDSL
LATRMVVELSKHLGQPVSLQHVFEQPVLQAFCAALEAPVTEAAGDDVLLSETLTGIWE

DILEQPVNAQTDFFRAGGDSLSATRMVVAAQQAGFSQISLQQVFEHPRFSDFCQVLG

TAEISEVAEPEGEVTATET

Supplementary Figure S5.4: Carrier Protein 1 (CP1). (top) Sequence identified from the Ctg 2642_6 were used to (middle) synthesize a gene and (bottom) prepare the CP1 protein. Protein sequences are colored with: (black) amino acid sequence added from the pET28a vector; (green) amino acid sequence of the gene not reported by antiSMASH but added from the sequenced gene to provide a complete protein; (blue) amino acid sequence of the first of the two carrier proteins identified by antiSMASH, and; (red) amino acid sequence of the second of the two carrier proteins identified by antiSMASH.

Carrier Protein 2 (CP2), a putative acyl-CP from c03335_AH-491-D14, Cluster 28, Ctg 3327_22.

Sequences identified from Ctg 3327_22:

TYEQLHKSSSKLANFLSVHGVKPSDVIGICLDRSVELMICLIGILKMGAAYVPIDPNYPKE

RTNFILNNSKLRLLLSESKYSGNLSEFDGETIYLDTACNYECASDDFSPGVVSLNALMY

MIFTSGSTGQPKGVRVSHRSAANFILAMKNRPGMSSRDRLLAVTTLSFDIAVLELFLPIV

VGASIVLATQEEAGDGRALSRLISAKSVTMMQATPHVWRLLLDTKWVSPSKFVALCGG

EAFPPDLVEPLVGCCESVWNLYGPTEATVWSTCYQVEAGDRVVPIGSPIDNVSCYVV

NEGMQLQPVGVCGELLVGGEGVSDGYHQQEELTEEKFVANSFDSLVPNSRLYKTGD

LVRWRENGLLEYVERIDTQVKIRGYRVELGEIESALVQHRAIQRAVC

EVSKMWKQILKIDTVSLEDNFFDLGGHSLLAMKFIDAIMSDYGVQLNANSLV

Synthesized CP2 gene:

TTGCAAAAGGTGTCGCATACGCCGTCGAAGGTAGCCGTAGAAGCCGGAGACCAAC

GTATTACCTATGAACAATTACATAAGAGTTCATCAAAGCTTGCGAACTTCCTCTCAG

TACACGGTGTTAAGCCGTCTGATGTGATTGGCATTTGTCTGGACCGCTCAGTTGAA

TTAATGATTTGTCTTATTGGCATTCTGAAGATGGGCGCAGCATACGTTCCAATTGAT

CCAAACTATCCAAAGGAGCGTACCAACTTCATCTTAAATAACAGCAAGTTACGCCTT

CTCCTTTCTGAGTCCAAGTATAGCGGCAATCTTTCTGAGTTCGACGGCGAGACTAT

CTACTTAGACACTGCCTGTAACTATGAGTGCGCTTCTGACGATTTCAGTCCGGGCG

TTGTTAGCCTCAATGCGCTGATGTACATGATCTTTACGTCCGGTTCTACGGGACAA

CCCAAGGGCGTACGCGTGAGCCATCGCAGCGCAGCAAATTTCATTTTGGCGATGA

AGAACCGCCCGGGCATGAGTTCCCGCGACCGTCTTTTAGCCGTGACAACTTTAAG

TTTCGACATTGCAGTTTTAGAGTTGTTTCTGCCAATCGTTGTCGGCGCGAGCATTG

TATTAGCAACTCAAGAGGAAGCGGGTGATGGGCGCGCCCTGAGCCGTTTGATCAG

TGCAAAGAGCGTCACTATGATGCAGGCAACTCCACATGTGTGGCGTCTTTTATTGG

ACACCAAGTGGGTTAGTCCTAGCAAATTCGTCGCATTGTGCGGTGGTGAGGCCTTT

CCGCCGGATTTAGTGGAGCCCCTTGTTGGGTGCTGTGAAAGCGTATGGAATCTCT

ATGGTCCAACTGAGGCAACCGTCTGGTCTACGTGTTACCAGGTTGAAGCGGGCGA

CCGCGTGGTGCCAATCGGCTCGCCGATTGATAATGTTTCTTGCTACGTCGTTAACG

AGGGCATGCAATTACAACCCGTCGGTGTTTGCGGCGAGCTGTTGGTGGGCGGGG

AAGGGGTTTCAGACGGCTATCACCAGCAAGAAGAGCTCACAGAGGAGAAGTTTGT

GGCGAACTCGTTTGACAGCTTGGTGCCGAATTCCCGCTTATATAAGACCGGCGAC

CTGGTTCGCTGGCGCGAAAATGGCCTGTTAGAGTATGTGGAACGTATCGACACAC

AAGTCAAGATCCGTGGCTACCGCGTGGAGCTGGGAGAGATCGAGTCAGCACTTGT

CCAACACCGTGCAATCCAGCGTGCGGTTTGCTGCGTGATTGCAATCGATAAGAAC

GACGCACGTCTTGTGGCCTATACGAAGCTGGCACGCGGTGAGAATATTACCGGTA

CTGAGATTCGTAAGTACCTGCGTTTAAAGATCCCGGAGTACATGATCCCCAACTCG

TTCGTCGAGATTGATGACTTCCCATTGACCGCGAATGGTAAGATCGACCGCTCGAA

GTTGCCGAGTCCTTATGCAATTGAGACCAAGTTCGTTACTGCGCCGATCGGTGGA

GTACACGAGGAAGTAAGTAAAATGTGGAAGCAAATCTTAAAGATTGACACAGTGAG

CTTAGAAGACAATTTCTTTGACCTGGGCGGCCATAGTCTTTTAGCGATGAAGTTTAT

TGATGCCATTATGAGCGACTATGGCGTTCAATTAAATGCCAATTCATTAGTTATTGA

TAGTCTGGGAAAGATCGTTGCGACGATCGAATCATCCAAGGGATCCCGTGCCAAA

ATCCCGTACGCGTCCATTCTCAGCCTGGTTAAGCGCATCCTCGTAGGCATGCGCA

ACTGTCGCGAGATGTTCCGCTTTCATCAGCAGTAA

CP2 protein:

MGSSHHHHHHSSGLVPRGSHMLQKVSHTPSKVAVEAGDQRITYEQLHKSSSKLANFL

SVHGVKPSDVIGICLDRSVELMICLIGILKMGAAYVPIDPNYPKERTNFILNNSKLRLLLS

ESKYSGNLSEFDGETIYLDTACNYECASDDFSPGVVSLNALMYMIFTSGSTGQPKGVR

VSHRSAANFILAMKNRPGMSSRDRLLAVTTLSFDIAVLELFLPIVVGASIVLATQEEAGD

GRALSRLISAKSVTMMQATPHVWRLLLDTKWVSPSKFVALCGGEAFPPDLVEPLVGC

CESVWNLYGPTEATVWSTCYQVEAGDRVVPIGSPIDNVSCYVVNEGMQLQPVGVCG

ELLVGGEGVSDGYHQQEELTEEKFVANSFDSLVPNSRLYKTGDLVRWRENGLLEYVE

RIDTQVKIRGYRVELGEIESALVQHRAIQRAVCCVIAIDKNDARLVAYTKLARGENITGTE

IRKYLRLKIPEYMIPNSFVEIDDFPLTANGKIDRSKLPSPYAIETKFVTAPIGGVHEEVSK

MWKQILKIDTVSLEDNFFDLGGHSLLAMKFIDAIMSDYGVQLNANSLVIDSLGKIVATIES

SKGSRAKIPYASILSLVKRILVGMRNCREMFRFHQQ

Supplementary Figure S5.5: Carrier Protein 2 (CP2). Protein sequences are colored with: (black) amino acid sequence added from the pET28a vector; (green) amino acid sequence of the gene not identified by antiSMASH but added from the sequenced gene to provide a complete protein; (blue) amino acid sequence of the adjacent adenylation domain identified by antiSMASH, and; (red) amino acid sequence of the carrier proteins identified by antiSMASH.

4’-Phosphopantetheinyltransferase 2 (PPT1), a putative PPTASE from c02649_AH-491-C20, Cluster 23, Ctg 2642_10.

Sequences identified from Ctg 2642_10:

EWLLLQSYHCRRELLTVAFSVKEAFYKAIYP

Synthesized PPT1 gene:

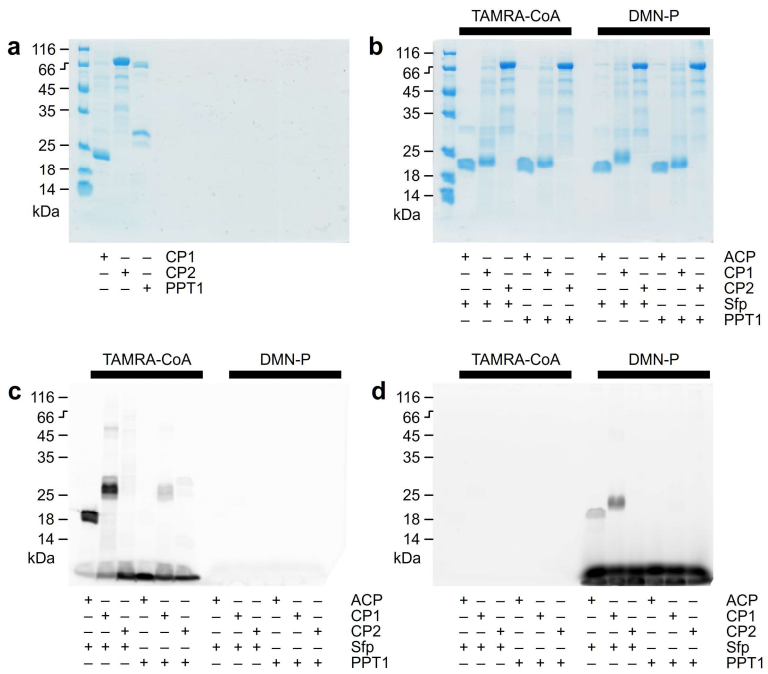
CAGTCCCGGCGGATCTCCAAGGCGCGGTTCTTAAACGTCAGTGCGAATATTTAGC
GGGGCGCTACTGCGCGGTCCAAGCGATGGGTATGGCCGGAATTTTAAATCCTGAG
CCACCATTGCGCGGCTCAGGTGGTGCCCCAATCTGGCCCGCGGGAATCCGCGGC
TCTATCTCACACAGCTGCCAACAGGCCGTCGCTATTCTTTGCCGTACGGACACTGG
CTGTGCGGGACTGGGTGTGGACATCGAGGGCTTTATTGCAGACGATAGTCTTGAC
ACCGTCATGCCGCTCGTATTGCAGACGGACGAGTGGCTGCTGCTGCAGTCCTACC
ATTGCCGTCGTGAGTTATTAACTGTGGCCTTCAGCGTTAAAGAGGCCTTCTACAAA
GCAATCTATCCCTATATCTGCCGCTTTGCCGATTTCCATGAGGTTGTCATCACCGG
TTTGGGCAAGGGACGCGCAGCGTTGACGCCTCGTGCCGGACTCGCAGCAGAGTT
ACCAAAGAATATTCTCATTCAGGCCAGTTACCGTATGGGTAGTTGCGGTGTCGAAA

GCCGTGTCGTTATGATGCGTCGTGCGGAAAAGGGAGAGGTGCCAGAAACCTTTTG
TGAGGGCTCCCGCCTTATCCAGACCCAGTGGGCGGGT

CP2 protein:

MAVPADLQGAVLKRQCEYLAGRYCAVQAMGMAGILNPEPPLRGSGGAPIWPAGIRGS
ISHSCQQAVAILCRTDTGCAGLGVDIEGFIADDSLDTVMPLVLQTDEWLLLQSYHCRRE
LLTVAFSVKEAFYKAIYPYICRFADFHEVVITGLGKGRAALTPRAGLAAELPKNILIQASY
RMGSCGVESRVVMMRRAEKGEVPETFCEGSRLIQTQWAGLGHHHHHH

Supplementary Figure S5.6: 4'-Phosphopantetheinyltransferase 2 (PPT1) a putative PPTASE from c02649_AH-491-C20, Cluster 23, Ctg 2642_10. Protein sequences are colored with: (black) amino acid sequence added from the pET28a vector; (green) amino acid sequence of the gene not identified by antiSMASH but added from the sequenced gene to provide a complete protein; (blue) amino acid sequence of the PPTase domain identified by antiSMASH.



Supplementary Figure S5.7: Labeling experiments. a) Gel depicting the expressed and purified CP1, CP2 and PPT1. PPT1 was observed as monomeric and dimeric form by SDS-PAGE. b) Total protein content for the labeling of CP1 and CP2 by Sfp or PPT1 with TAMRA-CoA (left) and DMN-P (right). c) TAMRA fluorescence observed in the gel in b) for the labeling of CP1 by Sfp or PPT1. This gel was also reported in Fig. 2d. d) DMN-P fluorescence observed in the gel in b) for the labeling of CP1 by Sfp or PPT1. This gel was also reported in Figure 5.2e.

Acknowledgements:

Chapter 5, in full, was resubmitted after revision for publication as it appears: Woojoo E. Kim, Katherine Charov, Mária Džunková, Eric D. Becraft, Julia Brown, Frederik Schulz, Tanja Woyke, James J. La Clair, Ramunas Stepanauskas, Michael D. Burkart "Synthase-selective exploration of a tunicate microbiome by activity-guided single-cell

genomics.” The dissertation author is a primary co-author of this manuscript along with Dr. Woojoo Eunice Kim, Dr. Mária Džunková, Dr. Eric D. Becraft, and Dr. Julia Brown.