Title
Posttranslational modification of natural product biosynthetic enzymes in bacteria and plants

Permalink
https://escholarship.org/uc/item/02s8d2z9

Author
Vickery, Christopher R.

Publication Date
2016

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Posttranslational modification of natural product biosynthetic enzymes in bacteria and plants

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

in

Chemistry

by

Christopher R. Vickery

Committee in charge:

Professor Michael D. Burkart, Chair
Professor Joseph P. Noel, Co-Chair
Professor Elizabeth Komives
Professor Victor Nizet
Professor Faik Tezcan

2016
The Dissertation of Christopher R. Vickery is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

___________________________________________________________________________

___________________________________________________________________________

___________________________________________________________________________

___________________________________________________________________________

Co-Chair

___________________________________________________________________________

Chair

University of California, San Diego

2016
DEDICATION

For always supporting me in my endeavors, I dedicate this thesis to my family. Thank you Mom, Dad, and Michelle for your confidence and encouragement during my years in graduate school. I emerge from this journey as a confident scientist who has not lost the optimism you’ve instilled upon me. Thank you for all you’ve done in shaping who I am and enabling this achievement.
“The only true wisdom is in knowing you know nothing.”

Socrates
TABLE OF CONTENTS

Signature Page ........................................................................................................................................ iii
Dedication ................................................................................................................................................ iv
Epigraph .................................................................................................................................................. v
Table of Contents .................................................................................................................................... vi
List of Abbreviations ............................................................................................................................... viii
List of Figures ........................................................................................................................................ x
List of Tables .......................................................................................................................................... xiv
Acknowledgements ............................................................................................................................... xv
Vita ......................................................................................................................................................... xvii
Abstract of the Dissertation .................................................................................................................... xviii

Chapter 1: Posttranslational modification of natural product biosynthetic enzymes in bacteria and plants................................................................................................................................. 1

References .............................................................................................................................................. 11

Chapter 2: Structure, biochemistry, and inhibition of essential 4’-phosphopantetheinyl transferases from two species of Mycobacteria........................................................................................................... 14

Supplementary Information .................................................................................................................... 26

References .............................................................................................................................................. 45

Chapter 3: Exploring the 4’-phosphopanetheinyl transferase and the cognate acyl carrier proteins of Arabidopsis thaliana ......................................................................................................................... 49

Supplementary Information .................................................................................................................... 66

References .............................................................................................................................................. 78

Chapter 4: Dissecting modular synthases through inhibition: a chemical genetic approach .................................................................................................................................................................................... 82

Supplemental Information ......................................................................................................................... 95

References .............................................................................................................................................. 119
Chapter 5: A combined approach for engineering a heterologous Type III PKS and enhancing polyketide biosynthesis in *S. cerevisiae* .......................................................... 122

Supplemental Information .............................................................................................................................. 146

References ..................................................................................................................................................... 150

Chapter 6: Uncovering enzymatic properties that control aromatic prenyltransferase specificity .......................................................... 152

Supplemental Information .............................................................................................................................. 162

References ..................................................................................................................................................... 165
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty acid synthase</td>
<td>FAS</td>
</tr>
<tr>
<td>Polyketide synthase</td>
<td>PKS</td>
</tr>
<tr>
<td>Non-ribosomal peptide synthetase</td>
<td>NRPS</td>
</tr>
<tr>
<td>Carrier protein</td>
<td>CP</td>
</tr>
<tr>
<td>Acyl carrier protein</td>
<td>ACP</td>
</tr>
<tr>
<td>Phosphopantetheinyl transferase</td>
<td>PPTase</td>
</tr>
<tr>
<td>Coenzyme A</td>
<td>CoA</td>
</tr>
<tr>
<td>Adenosine triphosphate</td>
<td>ATP</td>
</tr>
<tr>
<td>Acyl carrier protein hydrolase</td>
<td>AcpH</td>
</tr>
<tr>
<td>Type III polyketide synthase</td>
<td>T3PKS</td>
</tr>
<tr>
<td>Triacetic acid lactone</td>
<td>TAL</td>
</tr>
<tr>
<td>Aromatic prenyltransferase</td>
<td>aPTase</td>
</tr>
<tr>
<td>Dimethylallyl pyrophosphate</td>
<td>DMAP</td>
</tr>
<tr>
<td>Geranyl pyrophosphate</td>
<td>GPP</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.1: Natural products made by modular and iterative-type synthases ...................... 2

Figure 1.2: Modular biosynthesis centered around the carrier protein ............................. 2

Figure 1.3: Activity catalyzed by the 4’-phosphopantetheinyl transferase ....................... 3

Figure 1.4: Importance of PPTases to species of Mycobacterium ................................. 5

Figure 1.5: Activity catalyzed by 2-PS, a type III polyketide synthase ............................ 7

Figure 1.6: Activity catalyzed by aromatic prenyltransferases ....................................... 9

Figure 2.1: Crystal structure of PptT .................................................................................. 17

Figure 2.2: Surface representation of the CoA binding pocket of a) PptT, b) MuPPT (PDBid 4QJL), c) HsPPT (PDBid 2C43), and d) Sfp (PDBids 4MRT, 1QR0) ....................... 19

Figure 2.3: Activity of PPTases ....................................................................................... 21

Figure S2.1 Phosphopantetheinyl transferases (PPTase) covalently modify conserved Ser residues on target CPs ......................................................................................... 26

Figure S2.2 Mid-point melting temperatures (Tm) of PptT measured using various buffers and pH values ............................................................................................................. 27

Figure S2.3: Tm calculations for PptT using 0.1 M MES (pH 5.8) with additives resulting in a Tm increase ............................................................................................................. 28

Figure S2.4: Purification of C-terminal his-tagged PptT, analyzed by 12% SDS-PAGE gel electrophoresis and stained with coomassie blue ............................................ 29
Figure S2.5: Overall structure of MuPPT ................................................................. 30

Figure S2.6: Alignment of PptT with both Sfp and HsPPT ........................................ 31

Figure S2.7: Phylogenetic tree depicting the evolutionary proximity of Sfp, HsPPT, MuPPT, and PptT based upon sequence conservation ...................................................... 32

Figure S2.8: Comparison of the rotational positioning and coordination of Glu157 relative to the catalytic Mg2+ for PptT (orange), Glu157 of MuPPT (green), and Glu151 of Sfp (blue) .................................................................................................................................... 33

Figure S2.9: Qualitative assessment of mutant activity via the BpsA assay ............... 34

Figure S2.10: Removal of CoA for ITC studies of Sfp and PptT ............................... 35

Figure S2.11: Raw ITC data ....................................................................................... 36

Figure 3.1: Carrier proteins in A. thaliana ................................................................ 50

Figure 3.2: Posttranslational modification of CP substrates with AtPPT .................. 53

Figure 3.3: Labeling of heterologous CP domains by AtPPT ................................. 55

Figure 3.4: Crosslinking between KAS enzymes and CPs ....................................... 57

Figure 3.5: Phylogenetic tree depicting similarities between CPs from a wide variety of organisms .................................................................................................................. 58

Figure S3.1: Alignment of all Arabidopsis ACPs and the E. coli carrier protein AcpP .... 66

Figure S3.2: Comparison of mACP1 and mACP1C98S ............................................ 67
Figure S4.3: Representative growth curves of E. coli BL21 expressing BpsA and Sfp, under supplementation with glutamine, MgCl2 and MnCl2 ................................................... 97

Figure S4.4: Glutamine pantetheinamide as inhibitor of BpsA ................................................... 98

Figure S4.5: Activity of WT BpsA compared to TE mutant ......................................................... 99

Figure S4.6: Chase experiment loading BpsA PCP with cargo .................................................. 100

Figure S4.7: Phylogeny based on TE-domain homology ............................................................. 101

Figure S4.8: Phylogeny based on Ox-domain homology ............................................................. 102

Figure S4.9: Phylogeny based on A-domain homology ............................................................... 103

Figure S4.10: Phylogeny based on PCP-domain homology ....................................................... 104

Figure S4.11: Absorbance curves of all BpsA mutants and mutant combinations indicated in table S5 ........................................................................................................... 105

Figure S4.12: Structures of inhibitors synthesized in this study ................................................ 106

Figure 5.1: Half-lives of 2-PS mutants ......................................................................................... 132

Figure 5.2: Titer values for wild-type and a variety of surface cysteine and active site mutants ................................................................................................................................ 133

Figure 5.3: Titer values of all cysteine mutations ........................................................................ 135

Figure 5.4: Cysteines of 2-PS, and effects of mutating a combination of cysteines ......... 136

Figure 5.5: Detected thiol levels under aerobic and anaerobic growth conditions ............ 138
Figure 5.6: Effect of protease deficient yeast strains on 2-PS production......................... 139

Figure 5.7: Yield of TAL for engineered strain transformed with [C35S,C372S].................. 141

Figure S5.1: TAL production by 2-PS with and without the addition of acetyl-CoA............. 146

Figure S5.2: Purification of 2-PS mutants........................................................................ 147

Figure S5.3: Thermal denaturation of 2-PS mutants ........................................................ 148

Figure 6.1: Prenylation by aPTases.................................................................................. 153

Figure 6.2: Compounds assessed for prenyl donor activity............................................. 155

Figure 6.3: Mutations made in NphB ............................................................................ 159

Figure S6.1: Representative gel of NphB mutant library purification............................... 162

Figure S6.2: Tm of NphB mutants that were assessed for prenylation activity ................. 163
LIST OF TABLES

Table 2.1: Small molecules assayed against PptT for inhibition in a fluorescence polarization assay with both VibB and MAS as the target carrier protein ........................................ 24

Table S2.1: Crystallographic parameters of PptT and MuPPT structures ........................................ 37

Table S2.2: Relative kinetic parameters obtained from the BpsA assay for PptT, MuPPT, and Sfp at 50 nM PPTase ................................................................................ 38

Table S2.3: Primers used for cloning shown in the 5' to 3' direction ........................................ 39

Table S2.4: Structures of inhibitors tested for inhibitory activity measured by IC50 ...... 40

Table S3.1: Sequences of truncated proteins ................................................................................ 77

Table S4.1: Sequence identity matrix of Ox domains using Clustalw ........................................ 107

Table S4.2: Sequence identity matrix of whole synthases using Clustalw ................................. 108

Table S4.3: Sequence identity matrix of TE domains using Clustalw ......................................... 109

Table S4.4: Sequence identity matrix of TE domains using Clustalw ......................................... 110

Table 5.1: List of enzyme mutants used in this study ................................................................. 126

Table 5.2: Kinetic parameters of soluble 2-PS mutants ............................................................. 130

Table S5.1: Melting temperatures (Tm) of mutants purified by affinity chromatography ....... 146

Table 6.1: Results of prenylation reactions ............................................................................... 156

Table S6.1: List of NphB mutants ............................................................................................. 164
ACKNOWLEDGMENTS

I would like to acknowledge Professor Michael D. Burkart and Joseph P. Noel for their support as my advisors and chairs of my committee. Through their guidance and encouragement I have grown substantially as a scientist and been exposed to a breadth of fascinating research experiences.

My lab mate, friend, and fiancée, Lori, who is a constant source of support both in and out of the lab. Her understanding of the struggles of graduate school enabled her to provide essential support and encouragement, and she helped share both setbacks and victories throughout my graduate career.

I also acknowledge all past and current Burkart and Noel Lab members. I’ve learned techniques and practices that will serve me throughout my scientific career. I hope to pass the torch once again for the future of both labs.

Chapter 1, entitled posttranslational modification of natural product enzymes in bacteria and plants, in part, is a reprinting of the material as it appears in Beld, Joris; Sonnenschein, Eva C.; Vickery, Christopher R.; Noel, Joseph P.; Burkart, Michael D. “The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life,” Natural Products Reports, Volume 21, 61-108, 2014. The dissertation author was a shared primary author of the former and secondary author of the latter.

Chapter 2, in full, is the material as it appears in Vickery, Christopher R.; Kosa, Nicolas M.; Casavant, Ellen P.; Duan, Shiteng; Noel, Joseph P.; Burkart, Michael D. "Structure, biochemistry, and inhibition of essential 4'-phosphopantetheinyl transferases from two species of Mycobacteria" ACS Chemical biology, Volume 9, 1939-1944, 2014. 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 Vickery, Christopher R.; Sonnenschein, Eva C.; Beld, Joris; McCulloch, Ian; Noel, Joseph P.; Burkart, Michael D. “Dissecting modular synthases through inhibition: a chemical genetic approach,” The dissertation author was shared primary investigator and author of this material.
Chapter 5, in full, is currently being prepared for submission for publication of the material as Vickery, Christopher R.; Cardenas, Javier; Burkart, Michael D.; Da Silva, Nancy A.; Noel, Joseph P. "A combined approach for engineering a heterologous Type III PKS and enhancing polyketide biosynthesis in *S. cerevisiae*." The dissertation author was a shared primary investigator and author of this material.
VITA

2009 Bachelor of Science in Biochemistry, University of Maryland, College Park
2019-2013 Teaching Assistant, Department of Chemistry and Biochemistry
University of California, San Diego
2011 Master of Science in Biochemistry, University of California, San Diego
2009-2016 Research Assistant, University of California, San Diego
2012-2016 Research Assistant, The Salk Institute for Biological Studies
2016 Doctor of Philosophy in Chemistry, University of California, San Diego

PUBLICATIONS

J Wang, B Stefane, D Jaber, JAI Smith, C Vickery, M Diop, HO Sintim. “Remote C-H
Functionalization: Using the N-O Moiety as an Atom-Economical Tether to Obtain 1,5- and Rare
1,7-C-H Insertions.” Angewandte Chemie International Edition 49 (23), 3964-3968

GH Hur, CR Vickery, MD Burkart. “Explorations of catalytic domains in non-ribosomal peptide
synthetase enzymology.” Natural product reports 29 (10), 1074-1098

C Stewart, CR Vickery, MD Burkart, JP Noel. “Confluence of structural and chemical biology:
plant polyketide synthases as biocatalysts for a bio-based future.” Current opinion in plant biology
16 (3), 365-372

transferases: catalysis of a post-translational modification crucial for life.” Natural product reports
31 (1), 61-108

CR Vickery, NM Kosa, EP Casavant, S Duan, JP Noel, MD Burkart. “Structure, Biochemistry, and
Inhibition of Essential 4’-phosphopantetheinyl Transferases from Two Species of Mycobacteria.”
ACS chemical biology 9 (9), 1939-1944.

FIELDS OF STUDY

Major Field: Chemistry (Biochemistry)
ABSTRACT OF THE DISSERTATION

Posttranslational modification of natural product biosynthetic enzymes in bacteria and plants

by

Christopher R. Vickery

Doctor of Philosophy in Chemistry

University of California, San Diego, 2016

Professor Michael D. Burkart, Chair

Professor Joseph P. Noel, Co-Chair

Carrier proteins (CPs) are essential proteins for many biosynthetic pathways responsible for coordinating the biosynthesis of natural products. However, the CP must be posttranslationally modified by a 4’-phosphopantetheinyl transferase (PPTase) prior to participation in biosynthesis. This essential covalent modification uniquely positions the PPTase as a potential drug target for pathogenic bacteria upon structural understanding of the PPTase. Our lab has solved two PPTase crystal structures from Mycobacterium species, providing important structural information regarding these drug targets for the development of anti-mycobacterial therapeutics. Our lab has also explored the poorly understood eukaryotic PPTases and CPs, using Arabidopsis thaliana as a model organism. Since CP-dependent natural product biosynthesis in plants is organelle specific, the PPTase must access all organelles in order to activate biosynthesis. Since processivity of these interacting proteins still remains unclear, our lab has developed chemical biological tools to elucidate the nature of these interactions.
We have also applied chemical biological and chemical genetics tools to study non-ribosomal peptide synthetases (NRPSs) by focusing on BpsA, an NRPS from *Streptomyces lavendulae*. This model synthase will not only provide a basis for chemical genetic characterization of NRPSs, but will pave the way for elucidating NRPS protein-protein interactions.

Our research has also investigated CP-independent pathways. Previously, the Noel lab has made great strides in studying type III polyketide synthases (T3PKS), which produce a wide variety of plant metabolites from CoA thioester precursors. We investigated the effects of mutations in 2-pyrone synthase, a T3PKS that makes triacetic acid lactone (TAL) from acetyl CoA and malonyl CoA. Mutants were assayed *in vitro* for stability and activity, and for their *in vivo* TAL production when heterologously expressed in yeast. This study has uncovered exploitable enzymatic and organismal properties to increase heterologous production of valuable chemical products.

Additionally, we investigated aromatic prenyltransferases (aPTases), which add prenyl groups onto aromatic molecules, and their activity against a variety of aromatic natural products in order to discover the underlying enzyme properties that control substrate specificity. Knowledge of these properties will enable fine-tuning of aPTase activity to specifically modify molecules with prenyl groups for high-level production of natural prenylated molecules and new, unexplored analogs. This body of work contributes valuable knowledge to the field of natural product biosynthesis, with the duel aims of discovering and characterizing potential therapeutic targets in bacteria and modifying natural product pathways to produce valuable molecules through enzyme engineering.
Chapter 1: Posttranslational modification of natural product enzymes in bacteria and plants

Nature has developed sophisticated metabolic pathways that biosynthesize a plethora of natural products with a wide variety of roles. One common biosynthetic technique is the iterative addition of small molecule building blocks to create larger, macromolecular final products. Three classes of natural products that fall into this category are fatty acids, polyketides, and non-ribosomal peptides (Figure 1.1).\textsuperscript{1,2,3} Fatty acids can be found in all organisms in the Eukaryota and Prokaryota domains, while polyketides and non-ribosomal peptides are often present in bacteria, plants, and fungi. The biosynthetic machinery required for the biosynthesis of each of these product types are fatty acid synthase (FAS), polyketide synthases (PKS), and non-ribosomal peptide synthetases (NRPS) (Figure 1.2).

The central component of FAS, PKS, and NRPS is the carrier protein (CP), a non-catalytic and either a translationally linked domain of a larger polypeptide chain or an independently translated protein.\textsuperscript{4} The CP is responsible for timing and efficiency in shuttling the rapidly changing chemical intermediates due to chain elongation between the structurally diverse multienzyme complexes of these pathways. However, the CP must be post-translationally modified by a 4′-phosphopantetheinytransferase (PPTase) in order to participate in biosynthetic processes.\textsuperscript{5} PPTases mediate the transfer and covalent attachment of a 4′-phosphopantetheine arm from coenzyme A (CoA) to a conserved serine residue of the CP domain through phosphoester bonds (Figure 1.3). These essential post-translational protein modifications convert inactive apo-synthases to active holo-synthases.\textsuperscript{6} PPTases are essential for cell viability across all three domains of life, bacteria, archaea and eukaryota. In general, there exist two major classes of PPTases: AcpS-type, which are most commonly associated with FAS, and Sfp-type PPTases, which modify secondary metabolic pathway CPs, such as PKS and NRPS.\textsuperscript{5} The focus of our research lies in the Sfp-type PPTases of various organisms, the interplay between CPs and their cognate synthases, and the additional enzymes that play large roles in natural product biosynthesis.
Figure 1.1: Natural products made by modular and iterative-type synthases. Fatty acids, type I and II polyketides, type I and II non-ribosomal peptides, and hybrid fatty acid-polyketide molecules all require a CP domain for biosynthesis, whereas type III polyketides and prenylated type III polyketides do not.

Figure 1.2: Modular biosynthesis centered around the carrier protein. FAS, PKS, and NRPS systems all require the CP domain to tether the growing natural product during biosynthesis.
Figure 1.3: Activity catalyzed by the 4'-phosphopantetheinyl transferase. The PPTase modifies a carrier protein domain with the phosphopantetheine moiety of CoA, producing holo ACP from apo ACP.
Due to their key metabolic positions in metabolism, Sfp-type PPTases are considered a prime antibiotic target in medicine as well as agriculture.\(^7\)\(^8\) Previously, we have utilized chemical biology tools to discover small molecule inhibitors for Sfp, the Sfp-type PPTase from \textit{B. Subtilis} and founder of this PPTase class.\(^9\) Using a minimal peptide as a CP surrogate, a large small molecule screen was performed and several potent inhibitors were discovered. Further elaboration on this method led to a nanomolar inhibitor of Sfp.\(^10\) Due to this interest in the Sfp-type as an antibacterial target, we became interested in the PPTases of \textit{Mycobacterium tuberculosis} (Figure 1.4). It was shown that inhibition of the Sfp-type PPTase from \textit{M. Tuberculosis}, named PptT, was essential for organism viability.\(^11\) We sought to understand this protein at a biochemical and structural level, since only two structures of PPTases exist: that of Sfp\(^12\) and the human Sfp-type PPTase.\(^13\) After devising a suitable purification strategy, we measured biochemical parameters and solved the X-ray crystal structure of PptT and MuPPT, the Sfp-type PPTase from the closely related \textit{M. ulcerans}, which is responsible for causing buruli ulcer.\(^14\) These results bring to light the fact that not only do Sfp-type PPTases vary greatly in their primary amino acid sequence between organisms, but their protein structures vary as well, which positions them as enzymes that can be specifically inhibited by small molecules.

While we studied the interaction between bacterial PPTases and CP domains, we were interested in extending our tools and analysis to plants. Apart from a single recent paper detailing some preliminary biochemistry of this protein in \textit{Ricinus communis} and \textit{Spinacia oleracea},\(^15\) PPTases have never been studied in plants. Our goal is to shed light not only on the biochemistry and structure of this enzyme, but on the complex interplay between the single PPTase in the model plant \textit{Arabidopsis thaliana} and CP targets. There are a total of 9 CP substrates in \textit{A. thaliana}: 5 CPs located in the chloroplast, 3 located in the mitochondria, and one cytosol.\(^16\)\(^17\) The chloroplastic and mitochondrial CPs are involved in fatty acid biosynthesis, but the cytosolic CP domain is part of an uncharacterized synthase AtsA that
Figure 1.4: Importance of PPTases to species of Mycobacterium. (A) Pathways that require a PPTase. In red, the PPTases, in purple, the biosynthetic enzymes, and in green, products found in all Mycobacteria strains, and in blue, a product that causes symptoms of M. ulcerans. (B) Crystal structures of the PPTases from M. tuberculosis (PDB id: 4QJK) and M. ulcerans (PDB id: 4QJL) with CoA bound.
shows similarity to lysine metabolic enzymes in other organisms.\textsuperscript{18} Since there exists only a single PPTase in the genome of \textit{A. thaliana},\textsuperscript{19} it is still yet unclear how it accesses 9 total different CP substrates located in three different cellular compartments. Additionally, we hope to characterize the interaction of the 8 \textit{A. Thaliana} CPs and the FAS machinery in both the chloroplast and mitochondria. This will give us a clear picture of FAS in plants, providing a platform for primary and secondary metabolic engineering of CP pathways in plants.

PPTases are also essential to NRPS pathways. While investigating NRPS systems, we came across the single module NRPS BpsA, an enzyme originally discovered in the bacteria \textit{Streptomyces lavendulae}.	extsuperscript{20} Recently described in an assay that measures PPTase activity,\textsuperscript{21} this enzyme makes the blue dye indigoidine by utilizing adenosine triphosphate (ATP) and L-glutamine. Since this synthase is produced from \textit{Escherichia coli} in the inactive apo form, it requires action of an exogenous PPTase and CoA to add phosphopantetheine to its peptidyl carrier protein (PCP) domain in order to become active. This small NRPS serves as an attractive model for studying NRPS domains. We have taken a chemical genetics approach to study each individual domain of this synthase by applying small molecule inhibitors to each step in indigoidine biosynthesis. Additionally, mutation and deletion of several domains has confirmed our inhibitor based activity observations. We hope to build a small “NRPS toolbox” that will be applicable to other NRPS systems whose activity and mechanisms are hitherto unknown.

Until recently, our work has focused mainly on PPTases and CPs that require modification by PPTases. However, we have since then began to investigate the chemical modification of other enzymes in natural product biosynthesis. One such class are the type III polyketide syntases (T3PKS),\textsuperscript{22} utilize CoA bound substrates, as opposed to the fungal and bacterial type I and type II polyketide synthases. Each CoA thioester interacts with an active site cysteine in an interactive fashion, until a certain chain length has been reached, which is encoded into the sequence of the T3PKS.\textsuperscript{22} The starting CoA thioester varies between T3PKS, but the extending CoA thioester is always malonyl CoA.\textsuperscript{23} T3PKS are widely
Figure 1.5: Activity catalyzed by 2-PS, a type III polyketide synthase. (A) One molecule of acetyl CoA and two molecules of malonyl CoA condense and cyclize to form the product triacetic acid lactone. (B) Crystal structure of 2-PS indicating all surface cysteine residues. In purple, the active site, in teal, buried cysteines, and in red, surface cysteines.
distributed among the plant lineage and are responsible for making natural products like resveratrol,\textsuperscript{24} triacetic acid lactone,\textsuperscript{25} and chalcone.\textsuperscript{26} We have recently investigated these enzymes with regard to chemical modification of surface amino acids. Spontaneous post-translational modification can have a large effect on \textit{in vivo} bioproduction. We have recently investigated this phenomenon with regards to 2-pyrone synthase, a CP independent plant PKS responsible for biosynthesizing triacetic acid lactone (TAL) from acetyl CoA and malonyl CoA (Fig 1.5a). In an effort to increase TAL titers in \textit{S. cerevisiae} that contain the 2-PS gene, we introduced point mutations into the gene. While mutations in the active site affected TAL titers, several surface cysteine to serine mutations greatly enhanced TAL production (Figure 1.5b). While initially unclear why these types of modifications affected the \textit{in vivo} production of TAL, a series of biochemical experiments utilizing the mutants provided evidence for the cause of this TAL titer increase.

We have also investigated “molecular post-translational modification” by way of the aromatic prenyltransferase (aPTase) enzyme family.\textsuperscript{27} These enzymes chemically modify type III polyketides with the addition of a prenyl group,\textsuperscript{28} the molecular equivalent of post-translational modification. While there exist protein PTase enzymes,\textsuperscript{29} our work was focused on small molecule Ptases, in which a prenyl acceptor (small molecule) is modified by the prenyl donor (isporenyl or geranyl pyrophosphate) in an SN1-like mechanism (Figure 1.6).\textsuperscript{30} We assessed the ability of bacterial aPTases to modify plant-derived products with both isoprenyl and geranyl moieties. We first constructed a panel of 4 aPTases and assessed their activity against a large library of small molecules, ranging from type III polyketides to antibiotics, such as erythromycin. Additionally, we investigated in detail the effects of multiple mutations on the aPTase NphB. In collaboration with DNA2.0, a 96 mutant library of NphB was constructed by combining phylogenetic and crystal structure observation in order to modulate its ability to accept different prenyl donor and prenyl acceptor molecules. By assessing a panel of PTases against a variety of substrates, we hope to uncover the poorly understood mechanism of prenyl donor selectivity and prenyl acceptor specificity.
Figure 1.6: Activity catalyzed by aromatic prenyltransferases. (A) An aPTase transfers a prenyl group onto either a carbon or an oxygen of a phenolic compound. (B) Crystal structure of NphB. All residues highlighted in red were modified during our studies.
This body of work details our efforts towards contributing valuable knowledge to the field of natural products. First, characterization of the Mycobacterial PPTases paves the way for antmycobacterial drug discovery. The characterization of Arabidopsis thaliana FAS systems and the NRPS BpsA will provide a strong base of knowledge for engineering and basic metabolic understanding. Engineering heterologous pathways into yeast is an integral industrial process, and discovering ways to increase yields will allow for high level production of valuable chemicals. Finally, understanding aPTases will allow for custom molecular prenylation. This work together aims to have a high impact on the field of natural products by studying the posttranslational modification and activity of natural product biosynthetic enzymes in bacteria and plants.
References


Chapter 1, entitled posttranslational modification of natural product enzymes in bacteria and plants, in part, is a reprinting of the material as it appears in Beld, Joris; Sonnenschein, Eva C.; Vickery, Christopher R.; Noel, Joseph P.; Burkart, Michael D. “The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life,” Natural Products Reports, Volume 21, 61-108, 2014. The dissertation author was a shared primary author of the former and secondary author of the latter.
Chapter 2: Structure, biochemistry, and inhibition of essential 4′-phosphopantetheinyl transferases from two species of Mycobacteria.

Introduction

Widespread sickness throughout the world is caused by the pathogenic bacterial genus Mycobacterium. Although antibiotic regimes for mycobacterial infections have been in place for decades, emerging resistance to current antibiotics strains existing resources and necessitates ongoing development of new treatment strategies. Two of the most prominent pathogenic mycobacterial species are M. tuberculosis, the causative agent of tuberculosis, and M. ulcerans, responsible for Buruli ulcer. Mycobacteria possess a complex and robust cell wall composed of uncommon fatty acids, glycolipids, and polyketide natural products that enable infection and immune evasion. Small molecule virulence factors include mycobactin, a peptidic siderophore found in all mycobacteria that facilitates iron acquisition from hosts required for normal growth, and the small molecule mycolactone, a plasmid-encoded polyketide toxin produced by M. ulcerans.

Biosyntheses of these cell wall components and small molecule virulence factors require large multidomain synthases containing carrier protein (CP) domains. CPs tether the elongating product through a thioester linkage to a post-translationally attached 4′-phosphopantetheinyl group. Post-translational modification is carried out by a magnesium dependent 4′-phosphopantetheinyl transferase (PPTase). PPTase attaches the 4′-phosphopantetheine arm of coenzyme A (CoA) to a conserved serine residue of the CP. The terminal thiol of the CP-bound phosphopantetheine serves as a tether for the nascent product as it visits each catalytic domain for elongation and tailoring. Many bacteria contain two types of PPTases: the acyl carrier protein synthase (AcpS) and a secondary PPTase named for the Bacillus subtilis PPTase associated with surfactin biosynthesis (Sfp). AcpS-type PPTases activate acyl CPs associated with fatty acid synthases (FAS), while Sfp-type PPTases modify CPs involved in the production of secondary metabolites, including polyketide synthases (PKS), hybrid FAS/PKS, and non-ribosomal peptide synthases (NRPS). A single Sfp-type PPTase in Homo sapiens
called aminoadipate semialdehyde dehydrogenase phosphopantetheinyl transferase (HsPPT) activates FAS, lysine metabolism, and tetrahydrofolate biosynthesis. The structure of the AcpS-type PPTase from M. tuberculosis has been solved; however, there are currently no structures of the Sfp-type PPTase from M. tuberculosis, PptT. To date only two Sfp-type PPTases have been structurally characterized: Sfp of B. subtilis and HsPPT of H. sapiens.

Sfp-type PPTase genetic knockouts in Mycobacterium smegmatis and Mycobacterium bovis previously showed that each of these PPTases are essential gene products for cell growth. More importantly, generation of a M. tuberculosis strain with an inducible PptT expression cassette resulted in elimination of cell viability without PptT expression and a reduced bacterial load in mice infected with the mutant M. tuberculosis strain compared to wild type strains. Due to its important role in multiple aspects of mycobacterial secondary metabolism and concomitant bacterial viability, discovery of specific inhibitors of this enzyme will enable new therapeutic leads for the treatment of Mycobacterial infections. We solved the 3D X-ray crystallographic structures of two Sfp-type PPTases from Mycobacteria, PptT from M. tuberculosis and MuPPT from M. ulcerans, providing a valuable comparison of the Sfp-type PPTase family. We also examined in vitro biochemical properties and tested a panel of known PPTase inhibitors to clarify this potential antibiotic target for combating mycobacterial pathogens.

**Results and discussion**

Experimentally, PptT was found to be insoluble upon heterologous expression in E. coli, as observed by Rottier et al. Therefore, we began studying PptT as an N-terminal maltose binding protein (MBP) fusion. In vitro removal of the MBP domain via a thrombin protease cleavage site led to precipitation of PptT. Thermal stability experiments with the MBP-PptT fusion using a dye-binding thermal stability assay led to a modification of purification buffer conditions to promote higher thermal stability. Although PptT has a calculated pI of 6.0 we found that maximum thermal stability was achieved between pH 5.5–5.8 (Figure S 2.2). Sodium chloride, glycerol, and calcium chloride further increased the stability of PptT (Figure S 2.3). These new buffer conditions enabled the expression and purification of a highly soluble C-terminal
hexahistidine tagged PptT construct (Figure S2.4). PptT crystallized in the presence of its two cofactors CoA and Mg\(^{2+}\). Due to low sequence homology between PptT and the Sfp-type PPTases previously characterized, experimental phasing data was obtained using selenomethionine (Se-Met) substituted PptT and single-wavelength anomalous diffraction (SAD) at 1.54 Å resolution. MuPPT shares 80% sequence identity with PptT, and was thus subjected to identical purification strategies. The structure of MuPPT was phased by molecular replacement using the SAD-solved PptT structure. PptT and MuPPT were refined to final resolutions of 1.59 Å and 1.65 Å, respectively (Table S1.1).

Both PptT and MuPPT exhibit a pseudo-dimeric fold characteristic of Sfp-type PPTases (Figure 2.1a, Figure S2.5a). The overall structures of these two PPTases resemble those of Sfp from B. subtilis\(^{27}\) and HsPPT from H. sapiens,\(^{16}\) despite 19% identity and 26% similarity to BsSfp, and 22% identity and 26% similarity to HsPPT (Figure S2.5b–d, Figure S2.6). Since BsSfp and HsPPT are relatively close phylogenetic relatives, while distant from PptT (Figure S2.7), these new structures aid in understanding the structural diversity of Sfp-type PPTase architectures. CoA is bound in the active site located between the two pseudosymmetric halves of PptT and MuPPT. Despite the overall fold similarity between PptT, MuPPT, Sfp, and HsPPT, closer examination reveals several major differences in the active site architectures of the Mycobacterial PPTases.

The most striking difference involves the orientation of the pantetheine arm of CoA. In PptT and MuPPT, pantetheine extends into a deep hydrophobic pocket formed primarily by residues Tyr160, Leu171, Phe173, and Lys156 (Figure 2.1b). Trp170 serves as a “cap” above the entrance of this pocket. Furthermore, a key hydrogen bond between the backbone carbonyl of Leu171 and an amide nitrogen of the pantetheine arm is observed only in the
Figure 2.1: Crystal structure of PptT (PDBid 4QJK). β-sheets are colored orange and α-helices colored blue. Highlighted side chains are depicted as ball and stick models and colored by element. Ligands are depicted as stick models colored by element with polar hydrogens displayed. a) Overall structure of PptT complexed with CoA, exhibiting a pseudo-dimeric structure characteristic of Sfp-type PPTases. b) The interactions of the pantetheine binding tunnel with CoA are highlighted. The hydrogen bond between the backbone carbonyl oxygen of Leu171 and the amide hydrogen of pantetheine is highlighted with a black dotted line. c) Depiction of the amino acids that form a deep hydrophobic pocket at the Adenine binding site. The phosphopantetheine portion of CoA was omitted for clarity. d) Coordination of the 3’-phosphate of CoA by Arg56 and Arg48. Lys75 was overlaid onto the PptT structure and depicted as slightly transparent. Depicted in yellow and green are the residues observed in HsPPT (PDBid 2C43) and Sfp (PDBid 1QR0), respectively, which coordinate the 3’-phosphate.
Mycobacterium PPTases. Sequestration of the pantetheine arm is only partially observed in HsPPT and is absent in Sfp. The adenine base of CoA is buried in a deep, hydrophobic pocket (Figures 2.1c, 2.2a, 2.2b). This pocket is lined by Tyr160, Phe164, and Lys161. In both Sfp and HsPPT, the adenine base sits closer to the surface residing in a shallow pocket lined by fewer hydrophobic residues (Figure 2.2c, 2.2d). Residues that interact with the 3′-phosphate of CoA vary in identity and position along the backbone between the Mycobacterium PPTases, Sfp, and HsPPT (Figure 1d). PptT and MuPPT utilize Arg48 and Arg56, which are found in close proximity on the same α-helix. Lys75 may also serve as an electrostatic link in close proximity to the 3′-phosphate in the MuPPT structure. In comparison, Sfp utilizes residues Lys28 and Lys31, and HsPPT utilizes Arg47 and Arg86. Sequence alignment of PptT, Sfp, and HsPPT reveals no apparent primary sequence conservation among these basic residues. These notable structural differences between the Mycobacterium PPTases and HsPPT should enable specific targeting of pathogenic Mycobacterial PPTases with PPTase inhibitors, a critical factor for therapeutic index of any potential inhibitor candidates.

Although a majority of the active site residues of PptT and MuPPT form divergent interactions with CoA, the residues surrounding the Mg²⁺ ion and diphosphate moiety of CoA are similar to Sfp. Glu116 and Asp114 of PptT and MuPPT are positioned around the Mg²⁺ ion binding site. Lys161 and His93 both play a role in coordinating the α-phosphate of CoA during catalysis. All four of these catalytic residues are generally conserved throughout the Sfp-type PPTase family. Glu157, which putatively coordinates the Mg²⁺ and deprotonates the Ser of the CP, is rotated away from the Mg²⁺ center, and density for Mg²⁺ was not clearly observed in the structure of PptT. Due to the acidity of the crystallization conditions, it is possible that Glu157 is protonated, reducing its ability to effectively coordinate the Mg²⁺ ion. This would allowing rotation of the carboxylate side-chain away from the Mg²⁺ binding site. Similarly, in the structure of MuPPT, Glu157 is found in a different rotameric state (Figure S2.8).
Figure 2.2: Surface representation of the CoA binding pocket of a) PptT, b) MuPPT (PDBid 4QJL), c) HsPPT (PDBid 2C43), and d) Sfp (PDBids 4MRT, 1QR0). CoA is depicted as color-coded bonds on top of a grey protein accessible surface. The CoA conformation observed in the 1QR0 Sfp structure is shown as semitransparent bonds, colored with purple carbons, and overlaid on the Sfp and CoA found in 4MRT.
In activity studies PptT efficiently labeled three recombinant CPs that represent FAS, NRPS, and PKS CP targets. FAS acyl carrier protein AcpP of E. coli, the vibriobactin peptidyl carrier protein VibB from V. cholerae, and mycocerosic acid synthase MAS from M. tuberculosis were labeled by PptT with a synthetic rhodamine-CoA analog (Figure 2.3a).\textsuperscript{28} MAS was used to assess activity against a native target, since it contains a CP domain that requires phosphopantetheinylation by an Sfp-type PPTase.\textsuperscript{29} A previously described assay utilizing the blue pigment producing single module NRPS BpsA\textsuperscript{30,31} was used to measure the relative activities of wild type PptT, MuPPT, and Sfp. The apparent $k_{cat}/K_m$ (min$^{-1}$ µM$^{-1}$) values of PptT and MuPPT were 0.26 ± 0.05 and 0.26 ± 0.06, respectively. Sfp activity was much lower, exhibiting a $k_{cat}/K_m$ of 0.020 ± 0.004 (Figure 2.3b, Table S2.2).

To better understand the catalytic mechanism of PptT, five active site mutants were generated and biochemically characterized (Figure S2.9). Based upon comparisons with the previously published crystal structure of Sfp,\textsuperscript{27} Glu157, Asp114, Glu116, Arg48, and Arg56 of PptT were examined. Removal of the negative charge associated with the putative Mg$^{2+}$ ligand/general base, E157Q, abolished phosphopantetheinylation activity. This activity loss was observed for the corresponding Glu to Gln mutation in both Sfp and HsPPT. The D114N mutation, which eliminates the negative charge of one of the putative Mg$^{2+}$ ligands, also abolished phosphopantetheinylation activity of PptT. These results coincide with the corresponding mutation in both Sfp and HsPPT. A similar isosteric mutation of another Mg$^{2+}$ coordinating residue, E116Q, reduced activity by 500-fold of wild type activity. The E109D mutation was made in Sfp, resulting in no measurable catalytic activity. In AcpS-type PPTases, the positions equivalent to Glu116 in PptT are non-polar residues – specifically Val in M. tuberculosis AcpS. Similarly, this acidic residue is sometimes absent in eukaryotic Sfp-type PPTases. For instance, HsPPT contains a Met at this position, suggesting that an acidic residue at this position is not absolutely essential for phosphopantetheinylation activity, while the acidic residues corresponding to Glu157 and Asp114 of PptT are absolutely conserved.
**Figure 2.3: Activity of PPTases.** a) Gels depicting fluorescent labeling of the three carrier protein targets MAS, VibB, and AcpP with rhodamine CoA (mCoA) with PptT. b) Michaelis-Menten fit of PptT (●), MuPPT (▲), and Sfp (◇) in the BpsA activity assay.
Mutations chosen to disrupt 3´-phosphate binding, R48A and R56A, nonetheless retained activity at 20 fold reduction and 100 fold reduction of wild type activity, respectively. Mutation of the basic residues that coordinate the 3´-phosphate in HsPPT significantly decreased the mutant enzymes' affinity for CoA, with a modest increase in catalytic turnover. While important for CoA binding, Arg48 and Arg56 likely play very small roles in catalysis.

Sfp is known to retain cellular CoA throughout heterologous purification from E. coli. Incubation of recombinant PptT with apo-AcpP was sufficient to generate the holo-AcpP product when monitored via a urea-PAGE gel shift assay, indicating that PptT also co-purifies with tightly bound CoA (Figure S2.10). To weaken the binding of this endogenously bound CoA to Sfp and PptT, calf intestinal alkaline phosphatase was used to remove the 3´-phosphate. Phosphatase treated Sfp exhibited negligible conversion of apo-AcpP to holo-AcpP when compared to untreated enzyme, but PptT retained activity, indicating only partial removal of CoA. Since PptT mutants deficient in coordinating the 3´-phosphate are catalytically active, PptT might retain 3´-dephospho CoA throughout phosphatase treatment and utilize it as a substrate. The overall activity of Sfp and PptT was not affected by this treatment.

Partial removal of endogenous CoA from both Sfp and PptT enabled the measurement of the thermodynamic properties for CoA binding to the PPTases by isothermal titration calorimetry (ITC) (Figure S2.11a). Sfp and PptT have dissociation constants (Kd) of 3.4 ± 0.2 µM and 36 ± 13 nM, respectively. The calculated Kd for Sfp and CoA is comparable to previously reported values. For Sfp, the enthalpy (ΔH, kcal mol$^{-1}$) and entropy (TΔS, kcal mol$^{-1}$) of binding are -12.5 ± 0.15 and -5.07 ± 0.38 respectively, and for PptT, -5.78 ± 0.08 and 4.14 ± 0.16, respectively. Both the enthalpic and entropic values contribute to the Gibbs free energy of binding of CoA to PptT. Combination of the hydrophobic adenine pocket, deep sequestration of the pantetheine arm, and 3´-phosphate coordination observed in the crystal structure might account for the enthalpically and entropically favored interaction between PptT and CoA. PptT and Sfp were also titrated with rhodamine CoA using ITC measurements. Sfp exhibited a Kd of 1.30 ± 0.19 µM, while a Kd between PptT and rhodamine CoA could not be calculated (Figure S2.11b). While Sfp binds
both CoA and rhodamine CoA with comparable affinity, PptT does not. These data suggest that while CoA is buried in the PptT active site, the rhodamine CoA analog is incapable of adopting a similar high affinity interaction. The rhodamine CoA analog may instead adopt a binding conformation similar to that of CoA in the structure of Sfp containing CoA and its CP substrate.\textsuperscript{22} In this latter case, the pantetheine arm extends out of the binding pocket instead of into the pantetheine “tunnel” observed in the PptT crystal structure.

To date, only a small number of Sfp-type PPTase inhibitors have been identified.\textsuperscript{33} We extended the analysis of these compounds to PptT utilizing a recently published fluorescence polarization (FP) assay that measures modification of a CP target with a synthetic fluorescent CoA analog (Table 2.1).\textsuperscript{34} The FP experiments with PptT were conducted using both VibB and MAS as CP targets. The IC\textsubscript{50} values suggest that inhibition of PptT by these molecules is independent of the CP substrate. Inhibitors SCH202676, guanidinyl-naltrindole difluoroacetate, calmidazolium chloride, PD 404,182, and sanguinarine exhibited the greatest inhibition of PptT activity while having no inhibitory activity against HsPPT.\textsuperscript{31} Unlike the structure of PptT described here, HsPPT contains a C-terminal loop-helix motif that abuts the CoA binding pocket at the CP binding interface. These structural differences along with the selective inhibition of PptT bodes well for ongoing efforts of selective inhibitor design against these structurally distinct targets.

In conclusion, we have characterized the Sfp-type PPTases PptT from M. tuberculosis and MuPPT from M. ulcerans both structurally and biochemically. Based on the structural and biochemical similarity between PptT and MuPPT and the structural differences with HsPPT, antimycobacterial drugs that target the Sfp-type PPTase might be generally applicable to other mycobacterial species including M. leperae, M. bovis, and M. avium, further increasing the value of PPTase inhibitors as antimicrobials.
Table 2.1: Small molecules assayed against PptT for inhibition in a fluorescence polarization assay with both VibB and MAS as the target carrier protein. Compounds displaying a reproducible but very small percent inhibition were not calculated (NC). Starred values indicate that large error was encountered while measuring the IC50 value. Structures of each inhibitor can be found in Table S2.4.

<table>
<thead>
<tr>
<th>Compound</th>
<th>MAS</th>
<th>VibB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoA</td>
<td>1.1 ± 0.1</td>
<td>4.7 ± 0.4</td>
</tr>
<tr>
<td>3',5'-PAP</td>
<td>1.6 ± 0.2</td>
<td>0.78 ± 0.20</td>
</tr>
<tr>
<td>2'-deoxy-3',5'-PAP</td>
<td>7.4 ± 0.4</td>
<td>8.5 ± 1.0</td>
</tr>
<tr>
<td>Mitoxantrone 2HCl</td>
<td>Inactive</td>
<td>Inactive</td>
</tr>
<tr>
<td>Benserazide HCl</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>SCH-202676</td>
<td>0.5*</td>
<td>0.8*</td>
</tr>
<tr>
<td>Bay 11-7085</td>
<td>30 ± 4</td>
<td>NC</td>
</tr>
<tr>
<td>6-nitroso-benzopyrone</td>
<td>24 ± 2</td>
<td>17 ± 2</td>
</tr>
<tr>
<td>PD 404,182</td>
<td>19 ± 1</td>
<td>7.1 ± 0.4</td>
</tr>
<tr>
<td>Guanidinyl-naltrindole 2CF₂COOH</td>
<td>12 ± 1</td>
<td>19 ± 1</td>
</tr>
<tr>
<td>Sanguinarine Cl</td>
<td>4.9 ± 0.2</td>
<td>22 ± 2</td>
</tr>
<tr>
<td>Calmidazolium Cl</td>
<td>4.9 ± 0.4</td>
<td>2.0 ± 0.2</td>
</tr>
<tr>
<td>(-)-ephedrine hemisulfate</td>
<td>Inactive</td>
<td>Inactive</td>
</tr>
</tbody>
</table>
Materials and Methods

Thermofluor guided buffer optimization

MBP-PptT was diluted to 10 uM using 0.1 M solutions of 21 different buffers (Figure S2.2) in a white 96 well plate. Sypro dye (Roche) was added to each well to a 10x final concentration. Each well held a total volume of 20 uL. The temperature was ramped from 25°C to 85°C at a rate of 0.06 degrees/second using a Roche Lightcycler 480 (Roche). The negative first derivative of the raw fluorescence data was used to locate minimum change in slope of the melting curve to determine the mid-point melting temperatures (Tm). The buffers exhibiting the greatest increase in Tm for the PptT portion of the protein fusion were used to expand the screening in two additional dimensions varying the nature and concentration of salts and small molecule additives. Conditions providing the greatest Tm increases were combined and used for purification and storage of both PptT and MuPPT.
Phosphopantetheinyl transferases (PPTase) covalently modify conserved Ser residues on target CPs. The phosphopantetheine moiety is transferred from Coenzyme A (CoA) to the Ser residue of an apo-carrier protein in the presence of Mg²⁺, producing holo-carrier protein and 3',5'-phosphoadenylyl phosphate (3',5'-PAP).
Figure S2.2: Mid-point melting temperatures (Tm) of PptT measured using various buffers and pH values. Melting points were determined as described in the methods section. Buffers used in subsequent stability screens are highlighted in red.
**Figure S2.3:** *Tm calculations for PptT using 0.1 M MES (pH 5.8) with additives resulting in a Tm increase.* The negative derivative of the original Fluorescence versus Temperature plot were used to assign the Tm as the negative peak on the plots above. Blue contains no additives, red contains 1 mM CaCl$_2$, and green contains 10% (v/v) glycerol.
**Figure S2.4:** Purification of C-terminal his-tagged PptT, analyzed by 12% SDS-PAGE gel electrophoresis and stained with coumassie blue.
Figure S2.5: (a) Overall structure of MuPPT (pdbid 4QJL). Overlay of PptT (orange) with (b) MuPPT (red), (c) Sfp (green), and (d) HspPT (blue) as ribbon diagrams. These overlays illustrate and contrast the overall pseudodimeric folds of these enzymes and the orientations of bound CoAs. CoAs are shown in all panels as color-coded bonds.
Figure S2.6: Alignment of PptT with both Sfp and HsPPT. Alignments were calculated using MUSCLE\textsuperscript{35} and visualized with ESPRIPT\textsuperscript{36} (http://espript.ibcp.fr/ESPript/ESPript/).
Figure S2.7: Phylogenetic tree depicting the evolutionary proximity of Sfp, HsPPT, MuPPT, and PptT based upon sequence conservation. Tree was constructed using alignments of known PPTases described in Beld et. al. with Muscle and the tree was constructed with MEGA6. Finally, the resulting phylogenetic tree was visualized using Figtree.
Figure S2.8: Comparison of the rotational positioning and coordination of Glu157 relative to the catalytic Mg$^{2+}$ for PptT (orange), Glu157 of MuPPT (green), and Glu151 of Sfp (blue). CoA and Mg$^{2+}$ from the Sfp structure (4MRT) is shown as a grey sphere and blue bonds, respectively.
Figure S2.9: (A) Qualitative assessment of mutant activity via the BpsA assay. Activity is measured as %WT. (B) Raw data of mutant activity assays. WT is depicted in red, the R48A mutant in green, the R56A mutant blue, and the E116Q mutant purple. Blank and inactive mutants are depicted as black curves. Activities were assessed by measuring the acceleration of production of indigoidine (blue dye) by BpsA and measured at A_{590}.31
Figure S2.10: Removal of CoA for ITC studies of Sfp and PptT. Four samples were run on both SDS-PAGE and Urea-PAGE gels to assess purity, stability, and level of CoA retention. Untreated PPTases are labeled "holo" and CIP-treated PPTases labeled "apo".
Figure S2.11: Raw ITC data. Sfp is depicted on the left, and PptT on the right. Curves were calculated using the single binding site equation in the Origin software package (Microcal/GE Healthcare). Values of n (ratio of ligand to protein), K (Thermodynamic constant), ΔH (enthalpy), and ΔS (entropy) are displayed in tables inset into the titration graphs. (a) Sfp and PptT titrated with CoA after CIP treatment. (b) Sfp and PptT titrated with Rhodamine-CoA. Thermodynamic parameters could not be calculated for PptT with Rhodamine-CoA.
Table S2.1: crystallographic parameters of PptT and MuPPT structures.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PptT (SeMet)</th>
<th>MuPPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range, Å</td>
<td>41.22 - 1.59 (4.87 - 1.59)</td>
<td>37.24 - 1.65 (1.709 - 1.65)</td>
</tr>
<tr>
<td>Space group</td>
<td>C 2 2 2 1</td>
<td>P 2 1 2 1</td>
</tr>
<tr>
<td>a, b, c; Å</td>
<td>99.86, 121.51, 48.78</td>
<td>55.02, 59.75, 74.48</td>
</tr>
<tr>
<td>α, β, γ; (°)</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Total reflections</td>
<td>723703 (49707)</td>
<td>403854 (35853)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>44112 (4201)</td>
<td>25868 (2879)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>9.1 (7.2)</td>
<td>5.6 (2.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.58 (97.6)</td>
<td>85.5 (66.7)</td>
</tr>
<tr>
<td>Mean I/sigma(I)</td>
<td>8.5 (1.1)</td>
<td>9.1 (2.3)</td>
</tr>
<tr>
<td>R-merge</td>
<td>0.171 (2.134)</td>
<td>0.15 (0.388)</td>
</tr>
<tr>
<td>R-meas</td>
<td>0.182 (2.318)</td>
<td>0.163 (0.470)</td>
</tr>
<tr>
<td>R-work</td>
<td>0.1708</td>
<td>0.1599</td>
</tr>
<tr>
<td>R-free</td>
<td>0.2026</td>
<td>0.2085</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>2114</td>
<td>2039</td>
</tr>
<tr>
<td>macromolecules</td>
<td>1788</td>
<td>1760</td>
</tr>
<tr>
<td>ligands</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>water</td>
<td>277</td>
<td>229</td>
</tr>
<tr>
<td>Protein residues</td>
<td>228</td>
<td>228</td>
</tr>
<tr>
<td>RMS(bonds)</td>
<td>0.011</td>
<td>0.031</td>
</tr>
<tr>
<td>RMS(angles)</td>
<td>1.381</td>
<td>1.45</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>99.57</td>
<td>99.12</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Clashscore</td>
<td>2.74</td>
<td>2.50</td>
</tr>
</tbody>
</table>

Statistics for the highest-resolution shell are shown in parentheses.
Table S2.2: Relative kinetic parameters obtained from the BpsA assay for PptT, MuPPT, and Sfp at 50 nM PPTase. $K_m$ is measured with respect to CoA as the variable substrate.

<table>
<thead>
<tr>
<th>PPTase</th>
<th>$K_m$ CoA (µM)</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$k_{cat}/K_m$ (µM$^{-1}$min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PptT</td>
<td>1.600±0.248</td>
<td>0.42 ± 0.012</td>
<td>0.26 ± 0.05</td>
</tr>
<tr>
<td>MuPPT</td>
<td>2.637±0.502</td>
<td>0.66 ± 0.026</td>
<td>0.25 ± 0.06</td>
</tr>
<tr>
<td>Sfp</td>
<td>11.69±2.56</td>
<td>0.18 ± 0.012</td>
<td>0.020 ± 0.004</td>
</tr>
</tbody>
</table>
Table S2.3: primers used for cloning shown in the 5' to 3' direction.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP_F1</td>
<td>AAAAAATCTAGACTCCTTCTTTAAAGTTAATGAAAAATCGAAGAAGGTAAA CTGGTAATC</td>
</tr>
<tr>
<td>MBP_R1</td>
<td>CGTGGAAACCAGTCCGCTGCTCCCCAGGTTTGTG</td>
</tr>
<tr>
<td>MBP_F2</td>
<td>AAAAAATCTAGATTTTTTAAAGGGAGATATAATGAAAAATCGAAGA AG</td>
</tr>
<tr>
<td>MBP_R2</td>
<td>AAAAAACATATGGGATCCTCGTGGAAACCAGTCCGCTGCTCCCCAGG</td>
</tr>
<tr>
<td>E157Q_F</td>
<td>TTGTGCAAAACACGAGCTTTACAAAGCAGATG</td>
</tr>
<tr>
<td>E157Q_R</td>
<td>TAAGTTGCCCTGTATTTGCAAAAAACAGGATA</td>
</tr>
<tr>
<td>E116Q_F</td>
<td>TATCGATGCTCAACCTCATGATGTTCTGCC</td>
</tr>
<tr>
<td>E116Q_R</td>
<td>TCATGAGGTTGAGCATCGATACCTACCGCT</td>
</tr>
<tr>
<td>D114N_F</td>
<td>CGTAGGTATCAAATGCTGAAACCTCATGATG</td>
</tr>
<tr>
<td>D114N_R</td>
<td>GGTTCAGCATTGATACCTACGCATCATGAG</td>
</tr>
<tr>
<td>R48A_F</td>
<td>CGTTGCAAAAGCCCGTAACGATTTATTAC</td>
</tr>
<tr>
<td>R48A_R</td>
<td>AATTCCGTTACCGGGTTTTGCAACGCTACGA</td>
</tr>
<tr>
<td>R56A_F</td>
<td>GAATTCATTACCGGGTTCATTCGCG</td>
</tr>
<tr>
<td>R56A_R</td>
<td>CGCGCAATGAGCCAGGTAATAATTCGTT</td>
</tr>
</tbody>
</table>
Table S2.4: *Structures of inhibitors tested for inhibitory activity measured by IC\textsubscript{50}.*

<table>
<thead>
<tr>
<th>IC\textsubscript{50} (µM) against PptT</th>
<th>Compound</th>
<th>MAS</th>
<th>VibB</th>
<th>Compound</th>
<th>MAS</th>
<th>VibB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image" alt="CoA" /></td>
<td>1.1 ± 0.1</td>
<td>4.7 ± 0.4</td>
<td><img src="image" alt="6-nitroso-benzopyrone" /></td>
<td>24 ± 2</td>
<td>17 ± 2</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="3',5'-phosphoadenylyl phosphate" /></td>
<td>1.6 ± 0.2</td>
<td>0.78 ± 0.2</td>
<td><img src="image" alt="PD 404,182" /></td>
<td>19 ± 1</td>
<td>7.1 ± 0.4</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="2'-deoxy-3',5'-phosphoadenylyl phosphate" /></td>
<td>7.4 ± 0.4</td>
<td>8.5 ± 1.0</td>
<td><img src="image" alt="Guanidinyl-naltindole difluoroacetate" /></td>
<td>12 ± 1</td>
<td>19 ± 1</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Benserazide HCl" /></td>
<td>NC</td>
<td>NC</td>
<td><img src="image" alt="Sanguinarine Cl" /></td>
<td>4.9 ± 0.2</td>
<td>22 ± 2</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Mitoxantrone 2HCl" /></td>
<td>Inactive</td>
<td>Inactive</td>
<td><img src="image" alt="Calmidazolium Cl" /></td>
<td>4.9 ± 0.4</td>
<td>2.0 ± 0.2</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="SCH-202676" /></td>
<td>0.5*</td>
<td>0.8*</td>
<td><img src="image" alt="(-)-ephedrine hemisulfate" /></td>
<td>Inactive</td>
<td>Inactive</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Bay 11-7085" /></td>
<td>30 ± 4</td>
<td>NC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Methods

Cloning and purification MBP-PptT

PptT was codon optimized for *E. coli* and synthesized by DNA 2.0 in pJ201, and subcloned into pET24b (Novagen) using restriction sites Ndel and Xhol. The MBP sequence was duplicated from the pMAL-c2 (New England Biolabs) plasmid with primers MBP_F1 and MBP_R1 to begin addition of the thrombin cleavage site followed by PCR with MBP_F2 and MBP_R2 to finish creation of the thrombin cleavage site, as well as optimize the RBS from the first fragment and add the Ndel site on the 3’ end in preparation for vector insertion. Following double digestion with XbaI/Ndel and treatment with shrimp alkaline phosphatase (NEB), the desired MBP DNA was ligated into the parent pET24b/PptT vector. This MBP construct was transformed into BL-21 DE3 chemically competent cells and grown at 37°C to an O.D. of 0.8 in LB media supplemented with 2g L⁻¹ glucose, induced with 1 mM IPTG, and grown overnight at 18°C. Cells were pelleted and lysed into 50 mM Tris pH 8.0 and 500 mM NaCl. Lysate was passed over amylose resin and washed with lysis buffer. MBP-PptT was eluted from the column with lysis buffer plus 1 mM DTT and 25 mM maltose. The elutant was concentrated and used without further purification.

Cloning, Expression, and Purification of PPTases and carrier protein targets

MuPPT was codon optimized for *E. coli* and synthesized by Genscript. Both PptT and MuPPT were subcloned into pET24b using the restriction sites Ndel and Xhol. BL-21 DE3 chemically competent cells were transformed with the pET24b plasmid containing PptT or MuPPT. Cells were grown to an O.D. of 0.8 in LB media at 37°C, induced with 1 mM IPTG, and grown overnight at 16°C. The cells were lysed into a lysis buffer consisting of 50 mM MES pH 6.2, 500 mM NaCl, 1 mM CaCl₂, 10% (v/v) glycerol, and 20 mM imidazole. PptT was purified over Ni-NTA and eluted with Lysis Buffer containing 250 mM imidazole. The elutant was concentrated and the imidazole was removed using a PD-10 desalting column (GE healthcare) equilibrated with 50 mM MES pH 5.8, 500 mM NaCl, 1 mM CaCl₂, and 10% (v/v) glycerol. MuPPT and all PptT mutants were purified as described for PptT. Se-Met PptT was grown in M9 minimal media supplemented by an amino acid cocktail that included L-Selenomethionine (Sigma). Purification
was carried out in the same manner as described above. Protein destined for crystallization was then further purified on a Sephadex S200 (GE Healthcare) size exclusion column equilibrated with crystallization buffer consisting of 20 mM MES pH 5.8, 100 mM NaCl, and 5 mM MgCl₂, and was concentrated to 8 mg mL⁻¹ measured by Bradford method. 5 mM DTT and 1 mM CoA were added prior to crystallization. Sfp²¹, VibB³⁹, AcpP³⁴, and MAS³⁴ were produced and purified as previously described. Primers for MBP fusion construction and PptT mutants are listed in table S3.

**Crystallization techniques**

Using hanging drop vapor diffusion, medium sized plate-like crystals of PptT formed over a period 6 days in 100 mM Sodium acetate pH 4.5, 200 mM LiSO₄, and 30% (w/v) PEG 8000, which was discovered from Wizard Screen I (Emerald Biosystems). Diffraction quality crystals were obtained when drops consisting of 2 µL protein and 1 µL buffer solution were hung over an empty reservoir. Crystals were frozen in a cryo solution consisting of crystallization buffer plus 15% (v/v) Ethylene Glycol. Large rod-like crystals of MuPPT formed over a period of 1-2 months in drops consisting of 2 µL protein and 1 µL 2M LiCl, 32% (w/v) PEG 8000, pH 5.5. Crystals were frozen in a cryo solution consisting of mother liquor plus 15% (v/v) Ethylene Glycol.

**Data Collection and processing**

Data was collected on beamlines 8.2.1 and 8.2.2 at the Advanced Light Source (Berkeley, CA, USA). Raw data was indexed with mosflm⁴⁰ and scaled with scala using the CCP4 suite.⁴¹ Initial models using phasing data from Se-Met PptT and all refinement was performed with Coot⁴² and Phenix.⁴³ PptT was phased using single wavelength anomalous diffraction data from Se-Met labeled protein. MuPPT was phased using the previously solved PptT structure. All figures depicting crystal structures were prepared using PyMol.⁴⁴

**Gel activity**

Gel based analysis of labeling of MAS, VibB, and *E. coli* ACP were performed as previously described.⁹

**Removal of pre-bound CoA from PPTases**
10 mg of Sfp was treated with 100 U CIP in 5 mL 50 mM TrisCl pH 8, 250 mM NaCl, 10% (v/v) glycerol, 10 mM MgCl₂, along with a sufficient quantity of nickel resin for binding. Incubation of the mixture proceeded with rocking at room temperature for 2 hours. Nickel resin was washed with 50 mM MES pH 6.2, 500 mM NaCl, 10% (v/v) glycerol, 1 mM CaCl₂ and then eluted with 300 mM imidazole in MES buffer. Eluted CIP-treated Sfp was buffer exchanged with a centrifugal filter prior to use. PptT was prepared in the same manner as Sfp, except for a pre-binding of PptT to Ni-NTA resin for 1 hour on ice, prior to buffer exchanging to the Tris CIP reaction buffer for CIP treatment. CIP-treated, desalted PPTase at 20 μM was combined with apo E. coli ACP at 100 μM with 50 mM HEPES pH 7.6 and 10 mM MgCl₂ at 37°C overnight without coenzyme A to qualitatively gauge the removal of pre-bound coenzyme A from PPTases. Additional controls including untreated PPTase demonstrate lower conversion from apo- to holo-CP upon CIP treatment, as well as demonstrate retention of activity with the re-addition of coenzyme A. Samples were run on 20% Urea-PAGE gels to determine relative apo- and holo-ACP amounts.

**ITC experiments**

ITC measurements were performed on a VP-ITC isothermal titration calorimeter (Microcal/GE Healthcare) at 16°C for both Sfp and PptT. CoA analogs and PPTases were diluted to 300 μM and 30 μM, respectively, in 50 mM MES pH 6.0, 250 mM NaCl, 10 mM MgCl₂, and 5% (v/v) glycerol. 30 10 μL injections of 300 μM CoA or rhodamine CoA were added at intervals of 360 seconds while stirring at 300 rev. min⁻¹. An initial injection of 2 μL was performed and was not integrated into the data analysis. Data was fit to a titration curve using the built-in Origin software (Microcal).

**BpsA assay of PPTase kinetics**

The BpsA assay was performed in clear 96-well microplates at 25°C. Reaction conditions consisted of 75 mM phosphate pH 7.8, 5 mM MgCl₂, 8 mM ATP, 8 mM L-Gln, 50 nM PPTase, and CoA varying from 500 nM to 250 μM. To this mixture was added BpsA to a final concentration of 1 μM to initiate the reaction. Total reaction volume was 150 uL. Reactions were monitored for change in absorbance at 590 nm for approximately 30 minutes, with intervals of 13-
14 seconds per data point. Raw data was analyzed in GraphPad Prism as described previously to obtain kinetic parameters." The BpsA assay for qualitative comparison of mutant and W.T. PptT were carried out under the same conditions as above, except that the PPTase concentration was increased to 5 µM and CoA was held constant at 1 mM in order to visualize activity of weakly active mutants. Reactions were monitored at 590 nm for 45 minutes.

**Fluorescence Polarization.**

Fluorescence polarization activity assay proceeded as previously described unless otherwise noted. PptT was implemented at concentrations of 250 nM for inhibitor screening. Substrate concentrations for VibB inhibitor screening was 10 µM carrier protein and 5 µM rhodamine-CoA. Substrate concentrations of 4 µM MAS and 2 µM rhodamine CoA were utilized for inhibitor screening with MAS.
References


Chapter 2, in full, is the material as it appears in Vickery, Christopher R.; Kosa, Nicolas M.; Casavant, Ellen P.; Duan, Shiteng; Noel, Joseph P.; Burkart, Michael D. "Structure, biochemistry, and inhibition of essential 4'-phosphopantetheinyl transferases from two species of Mycobacteria" ACS Chemical biology, Volume 9, 1939-1944, 2014. The dissertation author was the primary investigator and author of this paper.
Chapter 3: Exploring the 4’-phosphopanetheinyl transferase and the cognate acyl carrier proteins of Arabidopsis thaliana

Introduction

Central to the metabolism of almost every organism is fatty acid (FA) biosynthesis. Fatty acids and their downstream products are essential metabolites for cell viability. The biosynthetic machinery responsible for FA biosynthesis is interconnected by way of the acyl carrier protein (ACP). The ACP tethers the growing FA on a 4’-phosphopantetheine arm that is posttranslationally installed by a phosphopantetheinyl transferase (PPTase) via a labile thioester linkage. PPTases covalently transfer the phosphopantetheine portion of Coenzyme A (CoA) onto a conserved serine of the acyl carrier protein, converting it from the apo form to holo form. In many eukaryotes, fatty acid synthesis (FAS) exists as a single mega synthase that contains all required enzymes on a single polypeptide chain, and is termed Type I FAS. Conversely, bacteria contain type II FAS, in which each component of FAS exists as a discrete polypeptide, including the ACP. Carrier proteins are also essential for other biosynthetic processes, including polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), and lysine and folate metabolism in humans, which are also modified by a PPTase.

Plants, like bacteria, have type II FAS systems. Only one functional PPTase is found in the genome of A. thaliana. The only other plant PPTase studies to date involve Spinacia oleracea and Ricinus communis. A complete set of FAS enzymes exists in the chloroplast and mitochondria, including the ACP required for fatty acid biosynthesis. Interestingly, A. thaliana contains multiple carrier proteins, which are nuclear encoded and targeted to the chloroplast and mitochondria. Additionally, the genome of A. thaliana contains a three domain synthase that consists of an adenylation domain, CP domain, and a third domain of unknown function (Figure 1). This synthase shows homology to Lys2 of S. cerevisiae and aminoadipate semialdehyde dehydrogenase (AASDH) of H. sapiens, though the function in A. thaliana is unknown. In this study, we biochemically characterize the Sfp-type PPTase.
Figure 3.1: Carrier proteins in A. thaliana. Models of each of the 8 standalone carrier proteins is shown. *E. coli* AcpP was used as a template for the models. Also depicted is the three domain synthase AtsA, with the CP domain, which lies between two catalytic domains, highlighted by an arrow.
AtPPT (Figure 3.1). We investigate the potential compartmentalization of AtPPT, as well as and assess its ability to modify both native and non-native CPs. We also investigate the interaction between A. thaliana CPs and β-ketoacyl synthases (KAS) enzymes from both A. thaliana and E. coli. This is the most complete biochemical analysis of the PPTase and CPs from A. thaliana to date, and sheds light on the activation and compartmentalization of carrier protein-dependent pathways. Similarly, studying the ACP-partner protein interactions provides insight into both the mechanism and evolution of plant fatty acid biosynthesis. Understanding these complex biosynthetic machineries will enable not only alteration of native metabolic processes, but will pave the way for heterologous production of natural products that contain CP domains in plants.

Results

Cloning of A. thaliana proteins.

The 300 a.a. PPTase AtPPT (AT3G11470) was sub-cloned from the ABRC cDNA clone U21362. Analysis by targetP indicated a putative mitochondrial targeting sequence of 12 amino acids at the N terminus of AtPPT. Therefore, the first 12 amino acids were removed and AtPPT cloned into both a C-terminal and N-terminal Hisx6 vector, and dubbed Δ12AtPPT. A recent paper detailing this enzyme also identified a possible 30 a.a. leader sequence,\(^\text{10}\) which was also constructed and named Δ30AtPPT. Initially, two candidate CPs from A. thaliana, cACP1 (AT3G05020) and mACP1 (AT2G44620), were obtained from the ABRC full length cDNA clones U22975 and U13654, respectively. In order to increase stability \(\text{in vitro}\), the two CPs were truncated to remove the targeting peptide sequences to localize each CP to its target organelle \(\text{in vivo}\). Leader peptide length was estimated using TargetP as well as previous work on Arabidopsis CPs (Figure S3.1).\(^\text{18}\)

While cACP1 expressed as a single band, mACP1 expressed as two distinct bands, which has been observed previously.\(^\text{18}\) A cysteine located near the end of the mACP1 sequence was subsequently mutated to a serine, since cysteine residues in CPs are known to form strong disulfide bonds between monomers.\(^\text{19}\) The A. thaliana carrier proteins cACP2 (AT1G54580), cACP4 (AT4G25050), cACP5 (AT5G27200), mACP2 (AT1G65290), and mACP3 (AT5G47630)
were codon optimized for *E. coli* and expressed in BL-21 cells. CACP3 (AT1G54630) was not cloned, since it differs from cACP2 by a single amino acid at the C-terminus (Figure S3.1). The putative synthetase AtsA (AT5G35930) was cloned from cDNA reverse transcribed from total plant *A. thaliana* Col-3 RNA. The full length construct did not express when transformed into *E. coli* BL-21 (DE3) cells. Therefore, to study AtsA modification by AtPPT, the CP domain was cloned into an *N*-terminal MBP vector containing a thrombin cleavage site between the MBP and AtsA CP.20

**AtPPT Activity against CP targets in *A. thaliana***.

PPTase activity against all cACPs and mACPs was first measured via a conformationally sensitive urea PAGE analysis technique.21 All CPs, except for mACP2, were modified with phosphopantetheine by the PPTase AtPPT and displayed a shift on the gel (Figure 3.2). The mACPs seemed to be partially in the *holo* form based on Urea PAGE results after incubation with CoA and AtPPT. Although the cysteine to serine mutant of mACP exhibited a more clear gel shift in the urea PAGE gel, a band of unknown identity persisted (Figure S3.2).

The cACPs and mACPs were modified with a rhodamine CoA analog (mCoA) in the presence of AtPPT (Figure 3.2).22 A fluorescent band corresponding to labeled CP was observed for nearly all of the CPs, including the mACPs that were predominantly in the *holo* form. The AtsA MBP-CP fusion was also clearly labeled with mCoA by AtPPT.

Both the full length and Δ12AtPPT versions of AtPPT were soluble and active (Figure 3.2). Δ30AtPPT was found to be insoluble with either an *N*-terminal or *C*-terminal His<sub>6</sub> tag. A putative isoform of AtPPT (named AtPPTX1), which is annotated in the TAIR database and is shortened by 69 amino acids, was also insoluble.
Figure 3.2: Posttranslational modification of CP substrates with AtPPT. (A) Urea PAGE depicting conversion of apo to holo based on gel shift. (B) Fluorescent labeling of CPs using a rhodamine CoA analog. (C) Modification of the CP domain from AtsA. (D) Comparison of AtPPT and Δ12AtPPT.
AtPPT successfully labeled the ACP domain from *H. sapiens* FAS and *E. coli* AcpP, which correspond to Type I and Type II FAS systems, respectively (Figure 3.3). AtPPT also labeled the ACP-TE 6 didomain from DEBS and ActACP,23 CPs from type I and type II PKSs, respectively. Although the type II NRPS CP VibB,24 a CP involved in vibriobactin biosynthesis, was successfully modified, the type I NRPS module GrsA,25 which is involved in gramicidin biosynthesis, was not. Additionally, when AtPPT was incubated with the blue pigment producing type I NRPS BpsA,26 no activity was observed. Lastly, the hybrid FAS/PKS protein MAS, involved in mycocerosic acid biosynthesis, was successfully modified by AtPPT.

**Investigating apo vs holo states of mACPs.**

Since the mACPs were expressed predominantly in the holo form, we tested two acyl carrier protein phosphodiesterase (AcpH) enzymes, which convert holo ACPs to apo, for activity with each mACP (Figure S3.3).27 AcpH from *Cyanothece sp.* (CyAcpH) and *Pseudomonas fluorescens* (PfAcpH) were previously shown to act on a wide variety of ACP substrates, and showed activity against the mACPs. Apofication was measured by fluorescent CoA labeling following AcpH treatment. The mACPs were also incubated with CoA prior to incubation with the fluorescent CoA as a positive control for holo carrier protein. MACP1 and mACP3 were minimally affected by AcpH treatment, indicating that the mACP protein preparations contained some amount of apo protein. However, mACP2 was efficiently labeled after treatment with either CyAcpH or PfAcpH, indicating that it was purified from *E. coli* almost completely in the holo form.

**Probing protein-protein interactions between CPs and KAS enzymes.**

Using activity based crosslinking probes, we assessed the ability of both cACPs and mACPs to interact with the *E. coli* β-ketoacyl ACP synthase (KAS) enzyme KASII (FabF), and the *A. thaliana* mitochondrial β-ketoacyl ACP synthase mtKAS (At2g04540). The ACPs were loaded
Figure 3.3: Labeling of heterologous CP domains by AtPPT. AtPPT was incubated with each of the depicted proteins in the presence of a rhodamine CoA analog.
with either a chloroacrylate pantetheine analog or an α-bromo hexanoyl pantetheine analog designed to covalently crosslink ACPs with KS domains when an interaction occurs between KS and CP. As previously observed, the chloroacrylate proved to be a more efficient crosslinking probe for KAS enzymes. The chloroplastic ACPs showed crosslinking with E. coli FabF through the appearance of a band on an SDS-PAGE gel that measures the size of FabF plus the carrier protein (Figure 3.4). However, no crosslinking between the mitochondrial ACPs and FabF was observed. Similar results were obtained with AcpP, all cACPs, and all mACPs were capable of crosslinking to A. thaliana mtKAS (Figure 3.4).

Observing phylogenetic relationships between plant CP domains.

A maximum-likelihood phylogenetic tree was constructed of all detectable chloroplastic and mitochondrial ACPs from the plants A. thaliana, Zea mays, Glycine max, Solanum lycopersicum, Oryza sativa, and Vitis vinifera, the mosses Physcomitrella patens and Selaginella moellendorfii, and the green algae Chlamydomonas reinhardtii, Coccomyxa subellipsoidea, and Chlorella variabilis. (Figure 3.5). ACP sequences from the bacteria Mycobacterium tuberculosis, Bradyrhizobium japonicum, Pseudomonas aeruginosa, Bacillus subtilis, Pseudomonas syringae, and E. coli, the cyanobacteria Nostoc punctiforme, Synechocystis sp. PCC6803, Synechococcus elongatus, and Synechococcus sp. BL107, and the parasite P. falciparum were added for comparison. Additionally, the phylogenetic conservation of AtsA was observed by blasting AtsA against the phytosome database of plants (Figure S3.4).

To compare the Arabidopsis carrier proteins to bacteria, each A. thaliana CP was blasted against bacteria. The top 1000 hits from each CP were collected and merged into a single list of bacteria, and a maximum-likelihood phylogenetic tree was constructed (Figure S3.5). A similar tree was constructed for A. thaliana KAS enzymes, cKASI, cKASII, cKASIII, and mtKAS (Figure S3.6).
Figure 3.4: Crosslinking between KAS enzymes and CPs. (A) Crosslinking of CPs from *E. coli* and *A. thaliana* to *E. coli* KASII. (B) Crosslinking of CPs from *E. coli* and *A. thaliana* to *A. thaliana* mtKAS.
Figure 3.5: *Phylogenetic tree depicting similarities between CPs from a wide variety of organisms.* All CPs were compared based on sequence. In the bottom portion of the tree, the mACPs are divided into “clades” that all contain at least one CP from each plant in the tree.
Docking of CPs to interacting domains.

Molecular models of all cACPs and mACPs were constructed and then docked with the crystal structures of AtmtKAS and EcFabF. Docking of the CPs to mtKAS provided docking poses that model productive interactions between a CP and its interacting domain, with the conserved serine that holds the phosphopantetheine and fatty acyl chain positioned at the entrance to the active site (Figure S3.7). Similarly, all CPs showed favorable docking poses when docked with *E. coli* FabF.

Discussion

In the genome of *A. thaliana*, a BLAST search using *B. subtilis* Sfp produces two hits: AtPPT (At3g11470), a 300 a.a. protein, and AtPPT2 (At2g02770), a 660 a.a. protein. However, due to the lack of EST support for AtPPT2 and no conservation of AtPPT2 among other plants, AtPPT was considered the only PPTase found in the *A. thaliana* genome. Like *H. sapiens*, *A. thaliana* contains a single Sfp-type PPTase, which exists as a pseudodimer.17,2 An AcpS-type PPTase, which exists as a homo-trimer and is associated with type II FAS in bacteria, was not found in any plants when blasted with *B. subtilis* or *E. coli* AcpS. Interestingly, certain plants contain two Sfp-type PPTases (Figure S3.8). It is possible that they may contain other undescribed CP domains that require phosphopantetheinylation.

A genome-wide search of the *Arabidopsis thaliana* for unique carrier protein substrates revealed five chloroplastic ACPs, 3 mitochondrial ACPs, and a single multi-domain synthase, AtsA. All CPs were labeled by AtPPT. While it is unclear why there exists multiple FAS carrier proteins in plants, previous work suggests that they are differentially regulated by light.32 Sequence analysis of the cACPs revealed strong sequence similarity between cACP2 and 3 and cACP1 and 5 (Figure S3.1). This sequence similarity correlates with tissue expression profiles as indicated by the Arabidopsis eFP browser (Figure S3.).33 CACP4 seems is constitutively expressed on all tissues, and knocking down expression leads to a decrease in lipid content.34 The mitochondrial mACP sequences differ significantly from the cACPs. However, they all
possess an identical \( \alpha \)-helix, which is important for protein-protein interactions in type II carrier protein systems.\(^{35}\)

AtPPT modified a wide variety of heterologous CP domains, including those from type I and type II FAS, type I and type II PKS, type II NRPS, and a hybrid FAS/PKS enzyme. Since the Sfp-type PPTase family is known to be tolerant of diverse CP substrates,\(^2\) these results are not particularly surprising. However, it is interesting that AtPPT did not modify type I NRPS systems, especially since the domain arrangement of putative synthase AtsA is reminiscent of type I NRPS-like enzymes, such as \( H. \ sapiens \) AASDH and \( S. \ cerevisiae \) Lys2. Furthermore, AtsA is conserved throughout the green plant lineage, suggesting an important role in plant metabolism (Figure S3.). Strong homology to eukaryotic enzymes like \( H. \ sapiens \) AASDH and \( S. \ cerevisiae \) Lys2 suggests a role in lysine metabolism. Further characterization of AtsA is required to understand both its interaction with AtPPT and its role in plant metabolism.

Since there exists only one PPTase in the genome of Arabidopsis, it was initially unclear how AtPPT could access chloroplast targeted ACPs, mitochondria targeted ACPs, and the synthase AtsA, which is tentatively annotated as a cytosolic protein. However, TargetP and mitoprot analysis of AtPPT indicated a putative mitochondrial targeting sequence measuring 12 amino acids and 30 amino acids, respectively.\(^{36,37}\) Solubility and activity of AtPPT was not compromised when the putative 12 amino acid leader sequence was removed from the N-terminus of AtPPT. Guan et. al. proposed a possible 30 a.a. mitochondrial targeting sequence, which is also predicted by Mitoprot.\(^{10,38}\) Removal of 30 amino acids from the N-terminus rendered the enzyme insoluble. However, if AtPPT is targeted to the mitochondria, it is unclear how it interacts with chloroplastic ACPs and AtsA. Characterization of the cellular localization of PPTase activity from Spinach and castor bean showed PPTase activity is primarily cytosolic.\(^{11}\) Additionally, cACPs that have been phosphopantetheinylated can be transferred into chloroplast.\(^{39}\) This data provides evidence for a cytosolic isoform of AtPPT, which could modify both cACPs headed to the chloroplast and the putatively cytosolic synthase AtsA. An alternative splicing variant of AtPPT is annotated on the TAIR database, and was previously shown to be
inactive against mitochondrial ACPs. This isoform could potentially be the predominant form of AtPPT in the cell, with a portion localized to the mitochondria to specifically modify mACPs. However, upon cloning and expression of this isoform of AtPPT, it was found to be insoluble, and thus activity could not be measured.

Both cACPs and mACPs show a relatively high similarity to E. coli ACP (Figure S3.1). A span of 18 aa that make up the third α-helix the carrier protein are nearly identical between the A. thaliana and E. coli (Figure S3.10). Interestingly, these amino acids are important for interaction of CPs with their partner proteins, suggesting an evolutionarily conserved role of this region for partner protein recognition. To test this hypothesis, we utilized a previously described crosslinking assay to detect a productive protein-protein interaction between CPs and FAS enzymes. Using E. coli AcpP and seven A. thaliana CPs, we assessed the ability of each CP to interact with E. coli FabF and A. thaliana mtKAS. While the chloroplastic ACPs crosslinked with the E. coli KAS enzymes, the mitochondrial ACPs did not crosslink (Figure 3.4). However, all CPs crosslinked with the mtKAS. Is possible that a structural feature prevents a productive protein-protein interaction. A closer look at the crystal structures of E. coli FabF and A. thaliana mtKAS reveals a positively charged helix that lies in the putative ACP binding region of FabF that is not present in the mtKAS structure (Figure S3.11). Based on model structures of the A. thaliana carrier proteins and the structure of E. coli AcpP, this positively charged loop could potentially inhibit the mACPs from binding, since a positive patch on the mACP might clash with this region. However, further analysis and of this protein-protein interaction is required to understand the crosslinking results.

To investigate the relationship between ACPs across a variety of organisms, a maximum-likelihood phylogenetic tree was constructed of all detectable ACPs from 6 plants, 2 mosses, 3 algae, 4 cyanobacteria, 6 bacteria, and one parasite (Figure 3.5). A clear division was observed between eukaryotic chloroplast targeted and mitochondria targeted ACPs. Interestingly, while cACPs tended to group based on organism, mACPs seem to group into specific "clades" (Figure 3.5). This may suggest that while chloroplastic ACPs co-evolve together in each respective plant, the mitochondrial ACPs are subject to fewer changes, and thus are more similar to ancestral
progenitor organisms of each clade of mACP. To observe the phylogenetic relationship between
plant and bacterial carrier proteins, all cACPs and mACPs were blasted against bacteria, and a
phylogenetic tree was constructed (Figure S3.5). The cACPs and mACPs clustered on different
branches of the tree. While mACPs were more similar to cyanobacteria, cACPs were grouped
close to Firmicutes and Fusobacteria. This data supports the hypothesis that the mitochondria
and the chloroplast have different bacterial origins. It was surprising to see the mitochondrial
ACPs clustering closely to the photosynthetic cyanobacteria. It was hypothesized that because of
the photosynthetic nature of cyanobacteria, the carrier proteins in the chloroplast of plants would
more closely resemble cyanobacterial CPs. However, the mitochondrial ACPs clearly group
closer to cyanobacteria than chloroplastic ACPs. Additionally, all KAS enzymes in Arabidopsis,
including KASI, KASII, and KASIII from the chloroplast and the mitochondrial mtKAS, were
blasted against bacteria (Figure S3.6). Interestingly, the pattern observed with the ACPs was not
reflected by the KAS clustering. cKASIII clustered with cyanobacteria, while cKASI and cKASII
were both found to cluster near Bacteroidetes. Finally, mtKAS seemed to cluster closely with the
Proteobacteria.

**Conclusion**

We have characterized, in detail, the 4’-phosphopanetheinyl transferase AtPPT and
carrier proteins from *Arabidopsis thaliana*. While it is still unclear how AtPPT is
compartmentalized in order to post-translationally modify all CP domains in *A. thaliana*, our
biochemical analysis showcases the ability of AtPPT to label all CP domains discovered in *A.
thaliana*. Additionally, AtPPT interacts with CP domains from other organisms as well, further
exhibiting its ability to label multiple types of CP domains.

**Methods**

**Cloning of Arabidopsis proteins**

The full length PPTase AtPPT (AT3G11470) was obtained from the ABRC full length
cDNA clone U21362. Full length, delta12, and delta30 were subcloned from this original template
into pET24b for C-terminal His-tag and pET28b for an N-terminal His-tag. cACP1 (AT3G05020)
and mACP1 (AT2G44620), were obtained from the ABRC full length cDNA clones U22975 and U13654 and subcloned into pET24b vectors. From these pET vector constructs, truncated versions of cACP1 and mACP1 were cloned into pET24b. Initially, AtsA was PCR amplified from the CP domain of AtsA was subcloned into a vector containing an N-terminal MBP fusion containing a C-terminal His<sub>6</sub>-tag. cACP2, cACP4, cACP5, mACP2, mACP3, and a truncated construct of mtKAS<sup>60</sup> were codon optimized for <i>E. coli</i>, synthesized, and cloned into pET24b by Genscript (Piscataway, NJ).

**Expression of proteins**

All proteins were transformed into BL-21 (DE3) (NEB) chemically competent cells. Cultures for protein purification were grown in 1L of LB media or terrific broth at 37°C until reaching an O.D. of 0.8, after which they were induced with 500µM IPTG and grown overnight at 18°C. In general, proteins were lysed into 50 mM HEPES pH 7.5, 250 mM NaCl, 10% glycerol, and 15 mM imidazole and batch bound onto Nickel-NTA resin (Qiagen) for 30 minutes. Protein was eluted in lysis buffer plus 250 mM imidazole. The resulting eluent was either dialyzed or buffer exchanged in a PD-10 column (GE healthcare) into 50 mM HEPES pH 7.5, 250 mM NaCl, and 10% glycerol. This protocol was used for all proteins except for mtKAS, where 10% glycerol was replaced with 25% glycerol in order to increase stability.

**AtPPT CoA modification reactions**

All AtPPT activity assays (both full length and delta12) were carried out in 50 mM HEPES pH 7.5, 10 mM MgCl<sub>2</sub>, 1 µM AtPPT, 1 mM DTT, 1 mM CoA or rhodamine CoA, and between 1 and 10 µM carrier protein. Reactions were initiated by addition of AtPPT, and incubated at 37°C for 1 hour. Reactions were quenched by addition of 2x SDS running buffer. Reactions involving CP modification with CoA were run on 20% urea PAGE gels and stained with coomassie blue. Labeling reactions involving rhodamine CoA were run on 12 or 15% SDS-PAGE gels. After fixing, gels were imaged on a typhoon fluorescence imager.

**AcpH activity assays**
1 µM of each mACP was incubated with approximately 10 µM PfAcpH and CyAcpH in 50 mM Tris pH 8.0, 10 mM MgCl₂, and 1 mM MnCl₂ overnight at 37°C in a total volume of 50 µL. This reaction was centrifuged for 5 minutes at 13,500 rpm in a tabletop centrifuge, and the supernatant was removed. To this supernatant was added 1 uM AtPPT and 1 mM rhodamine CoA, and incubated for 1 hour at 37°C. Reactions were quenched with 2X SDS loading buffer. Additionally, a portion of the supernatant was first incubated with 1 µM AtPPT and 1 mM CoA for 1 hour at 37°C, followed by the addition of 1 mM rhodamine CoA in order to produce an a “holo” sample for the analytical gel.

**Crosslinking reactions**

Approximately 0.5-10 µM ACP was incubated with 15 µM, FabF or mtKAS in 50 mM HEPES pH 7.5, 10 mM MgCl₂, 1 µM AtPPT, 0.2 µM each of CoaA, CoaD, and CoaE, 2.5 µM chloroacrylate pantetheine, 1 mM DTT, and 10 mM ATP in a total volume of 10 µL. This reaction was incubated at 37°C for 1 hour, and then quenched with 2X SDS loading buffer. Reactions were run on 12% SDS-PAGE gels. *

**Phylogenetic tree construction**

All ACPs from *A. thaliana* were BLASTed against all bacteria. The top 1000 hits for each ACP were combined, and duplicates were removed. The Resulting ~5000 sequences were aligned using MUSCLE⁴¹ and then assembled into a maximum-likelihood tree using the default settings in Fasttree.⁴² This tree was then visualized in Dendroscope.⁴³ The same process was repeated, except instead of ACPs, the three chloroplastic KAS enzymes cKASI, cKASII, and cKASIII, and the mitochondrial KAS mtKAS from *A. thaliana* were used.

Phylogenetic analysis of plant PPTases and AtsA-like proteins was performed by utilizing the Phytozome database⁴⁴ to BLAST AtPPT and AtsA. Sequences were hand-parsed for truncations and duplications. Sequences are aligned and trees were constructed using MEGA6.⁴⁵

**ACP docking to KAS enzymes**

Models of all cACPs and mACPs were constructed using swissmodel⁴⁶ with heptanoyl *E. coli* AcpP (PDB id:2FAD) as a template. These CP models were then docked with the crystal
structures of AtmtKAS and EcFabF using the cluspro server. The lowest energy poses for each carrier protein were displayed on the same crystal structure to compare the docking poses.
Figure S3.1: Alignment of all *arabidopsis* ACPs and the *E. coli* carrier protein AcpP. Secondary structure is depicted above the sequence. Helix 2 shows high homology between all CPs.
Figure S3.2: Comparison of mACP1 and mACP1C98S. A similar gel shift is observed with C98S, but the additional band persists.
**Figure S3.3:** Using *P. fluorescens* and *Cyanothece sp.* AcpH enzymes on the mACPs. mACP2 shows the most dramatic difference between untreated and treated samples.
Figure S3.4: Small phylogenetic tree of AtsA-like proteins. For the plants depicted, a tree was constructed to show conservation of AtsA. Below is the portion surrounding the conserved serine (black) to which phosphopantetheine is attached.
Figure S3.5: Phylogenetic tree of A. thaliana ACPs vs. all bacteria. Types of bacteria are highlighted in color. cACPs and mACPs depicted in green and red, respectively.
**Figure S3.6:** Phylogenetic tree of *A. thaliana* KAS enzymes vs. *bacteria*. Types of bacteria are highlighted in color. Chloroplastic KAS enzymes are highlighted in green, and mitochondrial highlighted in red.
Figure S3.7: Docking of ACPs to mtKAS and KASII. (A) docking of all cACPs, mACPs, and AcpP to mtKAS. (B) Docking of all cACPs, mACPs, and AcpP to *E. coli* KASII.
Figure S3.8: *Tree of PPTase conservation*. There are two major clades of PPTases in plants, with most plants with two.
Figure S3.9: Tissue specific expression of ACPs. eFP images of where each ACP of *A. thaliana* is expressed based on transcriptome profiling.
Figure S3.10: Graphical depiction of helix 2 of cACP1, mACP1, and AcpP. Shown in red on the structure is the conserved serine. Below is an alignment of helix 2, depicting the high level of conservation.
Figure S3.11: Comparing mtKAS and KASII interacting faces. Circled in green is the putative positive loop in (A) KASII, which is absent in (B) mtKAS. Circled in yellow is the active site cysteine.
**Table S3.1: Sequences of truncated proteins.** Full length proteins were not used due to insolubility, therefore truncations were made and they were cloned into pET24b.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence of translated protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>cACP1</td>
<td>MAKQETIEKVSAIVKKQLSLTPDKKVVAETKFADLGADSLDTV EIVMGLEEEFNIQMAEEKAQKIATVEQAAELIEELEINEKKEHH HHHH</td>
</tr>
<tr>
<td>cACP2</td>
<td>MAKPETVDKCAYVRKQLSLKEADEITAATKFAALGADSLDT VEIVMGLEEEFGIEMAEKAQSIATVEQAAALIEELLFEEKAKLE HHHHH</td>
</tr>
<tr>
<td>cACP4</td>
<td>MAKAEYQKVSDKIVKEQLALADVPLTAESKFSAALGADSLDT VEIVMALEEKFNSVEESDAQNITIQEAADLIEDLVQKPKAAE TSLEHMMMM</td>
</tr>
<tr>
<td>cACP5</td>
<td>MVKQETEVEKVSEIVKKQLSLTDQKVTAGTFETALGADSLDT VEIVMGLEEEFQTMVNAEERAKEIATVQQAAELIEEELVQEKTAL EHHHH</td>
</tr>
<tr>
<td>mACP1</td>
<td>MSREAAFXRVLDDVKSFPKVDSSKVTPVFQNDLGLDSLD TVEIVMAIEEEFKEIPDKEADKIDCSLAEYVYNHPMSSLEH HHHH</td>
</tr>
<tr>
<td>mACP2</td>
<td>MDKSEVTDRVLSSVKNFQKVDPSKVPKANFQNDLGLDSLD SVEVMALEEFGFEPDEAKISDIALEDFIAAHESPQAKLE HHHH</td>
</tr>
<tr>
<td>mACP3</td>
<td>MGQDQILSRVIELVKKYDKNTNTERADFQKDLSLSDKT EIVMAIEEEFSEIPDEKAADKLCTCCDVTYILSETPTKLEH HHH</td>
</tr>
<tr>
<td>mKAS</td>
<td>MISTSSYHSRRVVVTGLGVTRGLRGGVETTWRLIDGEC GIRGTLDDLKMKSDFEETKLYTFDQLSSKVAAFVPYGSNPG EFDEALWNKAVÅNFIGYVCAADEARLDÆWLPTEEEK ERTGVSIIGGISDICDEAALICEKRLRRSPFFISIPKILVNMA SGHVSMTKYGFOQPNHAATCATGAHSIDARNQIFDGM DAMVAGTRESSIDSAGFSSRALTŠKFNSPQEARSNFU DCDRDGFIVGEVSIVLVEEEYHAKRRAKIYAEILGYM GDHHTQPPEDKGAIVLAMTRALRQSGLPNDYVNAHA TSTQDGADEARAITVFSEHATSGTLAFSSTKGATGHLLGAA GAVEAIFSILAIIHGAPMNLNPKPNPDKRMPMTTTSKKK LVRTAMSNSFGGTNASLLFAILEHHHHHH</td>
</tr>
</tbody>
</table>
References


Modular synthases make up the core machinery of metabolically essential small molecules such as fatty acids, polyketides and non-ribosomal peptides. Besides their involvement in fatty acid anabolism throughout all kingdoms, they are responsible for the production of bioactive compounds crucial for cell survival in bacteria, fungi and plants. These molecules often possess antibiotic, immunosuppressive, cytostatic, and cytotoxic activities and have gained considerable attention in modern medicine and agriculture. Their biosynthesis follows an assembly-line strategy unique among metabolic pathways. Each synthase consists of individual domains possessing distinctive functions, such as the recognition of the substrate (amino acid or acyl-coenzyme A (acyl-CoA)), propagation of the growing chain, tailoring and condensation of intermediates and finally, release of the product. In case of a non-ribosomal peptide synthetase (NRPS), the basic components include adenylation (A), peptidyl carrier protein (PCP), condensation (C) and thioesterase (TE) domains. Due to their complexity, the study of modular synthases remains challenging, from mechanistic and structural studies, to concepts as basic as product identity. With this report, we aim to demonstrate the dissection of a NRPS modular synthase in order to probe the activity of each domain, as well as the activity of the ensemble. Here we provide a case study with the blue pigment synthase A (BpsA), the bacterial source of the blue pigment indigoidine (Figure 4.1).

Although the blue pigment resulting from indigoidine was first observed in 1890, the indigoidine synthase was only discovered in 2002 as IndC in the plant pathogen *Dickeya dadantii* (formerly known as *Erwinia chrysanthemi*). The wildtype strain does not make the blue pigment (the synthase gene is cryptic), but a pectate master regulator (PecS) mutant does. In the latter background, an insertional library identified IndA, -B and -C, from which IndC is the synthase. IndB shows similarities to phosphatases, whereas IndA has unknown function. The synthase IndC is conserved in many organisms, and is annotated as IndC,
Figure 4.1: Pipeline for the biochemical characterization of a modular pigment-producing synthase. After heterologous expression of the synthase gene and purification of the protein, the activity of each individual domain is inhibited by specific small molecules and analyzed in vitro and in vivo by pigment formation. Additionally, mutant proteins of each individual domain are generated and analyzed subsequently. Phylogenetic analysis provides further information on the diversity and versatility of the synthase.
BpsA, IgiD or as putative synthase (Figure S4.1, S4.2), but the accompanying genes are less conserved.\textsuperscript{6} A phylogenetic analysis of IndC homologs shows a wide distribution of these synthases, with conservation putatively even reaching to some eukaryotes (Fig. S4.1).\textsuperscript{4} IndC is a single protein multi-domain NRPS module, with an A domain predicted to utilize L-glutamine as building block.\textsuperscript{4} Bacterial A domains show uniquely predictable active sites that are conserved within groups determined by the amino acid they load onto the PCP, enabling computational prediction of the selectivity.\textsuperscript{7} In 2007, a homologous synthase was identified and characterized in Streptomyces lavendulae, called blue pigment synthase A (BpsA). In vitro experiments showed that BpsA can only accept L-Gln over all proteinogenic amino acids, and that the synthase oxidizes and dimerizes the amino acid.\textsuperscript{8} As predicted by bioinformatics, BpsA is a single module NRPS, with an oxidation (Ox) domain integrated into the A domain (Figure 4.2). The Ox domain harbors a flavin cofactor, as shown by UV absorbance measurements,\textsuperscript{9} and purified protein has the characteristic yellow coloration of flavoproteins. Like all modular synthases, BpsA requires post-translational modification by a 4'-phosphopantetheinyl transferase (PPTase).\textsuperscript{10} The PPTase transfers the 4'-phosphopantetheine moiety of the cofactor CoA onto a conserved serine residue of the PCP, enabling the PCP to tether the natural product via a labile thioester linkage. This enzymatic transformation was recently utilized by Ackerley and others for the development of novel reporter systems.\textsuperscript{11}

We expressed and purified the flavin-bound apo-BpsA and post-translationally modified the protein using CoA and the promiscuous PPTase Sfp. holo-BpsA showed robust activity in vitro,\textsuperscript{8} and \textit{E. coli} BL21 cells co-expressing Sfp and BpsA produced indigoidine when grown with liquid or solid media (Figure 4.4).\textsuperscript{1} Interestingly, supplying the double transformant with Gln in the growth media boosts the production of the blue pigment, and Mn\textsuperscript{2+} supplementation is superior to Mg\textsuperscript{2+}, the latter presumably involved in PPTase activity (Figure S4.3).\textsuperscript{12}
Figure 4.2: (A) Genomic organization of BpsA (gene cluster: AB250063, protein: BAE93896) in Streptomyces lavendulae subsp. lavendulae including domain-specific inhibitors. orfA is a putative ribose-phosphate pyrophosphokinase, bpsA codes for indigoidine synthase, bpsB for a regulator, and bpsC for a putative S-adenosylmethionine synthase. bpsA = 3,849 bp = 1,282 aa. (B) 1 L-Gln-sulfamoyladenosine inhibitor of A-domain, 2 thiazole inhibitor of flavin-dependent Ox-domain, 3 Gin-pantetheinamide, a substrate for in situ transformation into a Coenzyme A analog and subsequent loading onto a carrier protein using a PPTase, 4 and 5 are sulfonylfluoride-containing inhibitors of proteases and thioesterases and 6 is 6-NOBP, a general PPTase inhibitor. * denotes that a domain mutant is available.
To confirm the bioinformatic prediction of domains, thereby clarifying the biosynthetic properties of this synthase, we set out to design and synthesize individual inhibitors for each domain within the synthase. The A-domain is an adenylate-forming enzyme that loads L-Gln onto the PCP. Interestingly, BpsA contains an interrupted A domain, in which the Ox domain separates a large N-terminal from a small C-terminal portion of the enzyme (Figure 4.2B). Adenylate-forming enzymes, including aminoacyl-tRNA synthetases, are inhibited by sulfamoyl-containing analogs of their respective aminoacyl-AMP intermediates. We synthesized and analyzed 5'-O-(N-(L-glutaminyl)-sulfamoyl)adenosine 1 as an inhibitor of BpsA. Only upon addition of millimolar amounts of this inhibitor resulted in a substantial decrease in the formation of blue pigment in vitro, with an IC₅₀ = 3.8 mM (Figure 4.1A). The high concentrations required suggest a competitive, non-covalent mechanism of inhibition, as previously observed in inhibition of the biosynthesis of mycobactin with salicyl-amino-sulfonamide-adenosine. However, why 1 does not inhibit BpsA to the same extent as other sulfonamides inhibit A-domains, luciferase, tRNA synthetases or acyl-CoA ligases, is an open question. We hypothesize that the molecule is unstable in aqueous environment, and due to its polarity has low affinity for the A-domain of BpsA. Alternatively, interrupted A-domains might bind this class of inhibitor differently compared to classical A-domains.

The second domain of BpsA is a flavin-dependent oxidation domain. The Ox domain is most likely responsible for oxidizing either directly glutamine tethered to the PCP or a cyclized glutamine intermediate (Figure 4.5). Takahashi et al. demonstrated by point-mutational analysis of the Ox domain that binding of the cofactor flavin mononucleotide FMN is essential for indigoidine production. Other flavin-dependent oxidation domains in NRPSs have been studied by the Walsh lab, including EpoB, the oxidation domain involved in epothilone biosynthesis. EpoB oxidizes the dihydro heterocyclic thiazolinyl ring to a heteroaromatic oxidation state, forming methylthiazolylcarboxy-S-EpoB. The Ox domain of BpsA shares 29% identity with EpoB (AAF62884) (Figure S4.3). Although some NRPS
Figure 4.3: Inhibition of individual BpsA domains. (A) Inhibition of the PPTase Sfp by 5'-O-[N-(L-phenylalanyl)sulfamoyl] adenosine 1. (B) Inhibition of A domain by 6-NOBP 6. (C) Inhibition of TE domain by PMSF 5 and (D) AEBSF 4. (E) Activity of BpsA by varying the ratio of CoA to the non-hydrolysable L-Gln-Pantethenamide analog 3 during preincubation of BpsA with Sfp.
Ox-domains have been characterized, few inhibitors of these oxidases are known. Inspired by the work of Walsh and co-workers, we synthesized 4H-benzo[d][1,3]oxathiin-2-one (BOTO) 2 as Ox-domain inhibitor.\(^\text{17}\) Micromolar amounts are sufficient to fully inhibit the enzyme.\(^\text{17}\)

The PCP, the third domain of BpsA, is not an enzyme but a carrier protein, making traditional inhibition impossible. However, our lab has developed a methodology to attach non-hydrolyzable analogs of thioester bound substrates to carrier proteins through the preparation of pantetheine analogs elaborated by CoA biosynthetic enzymes and the promiscuous PPTase Sfp.\(^\text{18}\) Facile synthesis of L-Gln-pantetheinamide 3 allowed for the labeling of the BpsA PCP with a non-hydrolyzable substrate mimic. This crypto-BpsA showed no activity \textit{in vitro} (Figure S4.4), and in direct competition assay between CoA and Gln-pantethenamide-CoA, decreased BpsA activity was clearly observed when as the ratio of CoA to Gln-pantethenamide decreased (Fig. 1E).

In order tether substrates to NRPSs for catalysis, all PCPs require post-translational modification by a PPTase. PPTases have only recently been recognized and investigated as potential anti-microbial targets.\(^\text{19}\) In our work on the discovery of inhibitors for the \textit{B. subtilis} PPTase, Sfp, we found the general PPTase inhibitor 6-NOBP 6 shows modest inhibition against a variety of PPTases.\(^\text{19a}\) Based on these findings Owen \textit{et al.}\(^\text{11a}\) used 6 as PPTase inhibitor as demonstrated by indigoidine production with BpsA. Indeed, 6 inhibited BpsA with an IC\(_{50}\) of 25 \(\mu\)M using 50 nM Sfp, 1 \(\mu\)M BpsA, and 30 \(\mu\)M CoA. (Figure 4.1B).

The fourth domain of BpsA is the TE, which presumably releases the product from the PCP. Speculation of this activity differs, but the most recent review proposes release of 5-aminopyridinedione, followed by spontaneous dimerization of two products, forming either colorless leuco-indigoidine or blue indigoidine (Figure 4.5). Numerous thioesterase inhibitors have been developed over the past decades that rely on the nucleophilicity of the active site serine or cysteine.\(^\text{20}\) To inhibit the TE of BpsA, we used the general protease inhibitor phenylmethylsulfonyl fluoride (PMSF) 5 and a derivative of the reactive phenylsulfonylfluoride moiety, 4-(2-
aminoethyl)benzenesulfonylfluoride (AEBSF) 4 (Figure 4.3C, 4.3D). Blue pigment formation was abrogated by micromolar concentrations of either inhibitor, suggesting that although its function is speculative, the TE domain is critical for catalysis. A BpsA mutant lacking the TE domain activity exhibits no turnover when compared to wild type (Figure S4.5A). In an effort to discern the importance of the TE domain, we incubated the TE mutant of BpsA with a previously described type II thioesterase, TycF. TycF is known to remove substrates from the phophopantetheine arm of PCPs. However, the TE mutant was unable to produce blue pigment with or without TycF (Figure S4.5B), suggesting that the TE domain of BpsA directly participates in the formation of indigoidine, not simply hydrolyzing the substrate off of the PCP.

Interestingly, 5 showed ten-fold greater inhibition of blue pigment production than 4, indicating that steric or polarity of the substrates that enter the TE is important. We also used 4 to synthesize a pantetheinamide analog with a reactive sulfonylfluoride warhead. This pantetheinamide probe was similar to the L-Gln-pantetheinamide probe 3 loaded onto BpsA. The loading of this probe eliminates formation of blue pigment and prevents loading of a fluorescent pantetheinamide analog in a chasing experiment (Figure S4.6). This probe could offer new opportunities for mechanistic crosslinking of domains in multi-domain synthases.

A BLAST search of the BpsA TE domain results in hits (24% homology, Table S4.3) in atromentin synthase and TdiA, in which TEs are involved in cyclization/dimerization and dimerization of two activated indolepyruvic acid monomers, respectively. Fascinatingly, the human FAS (FASN) thioesterase also appears as potential hit, with 21% homology, whereas the entire synthase has low identity (Table S4.4). All four thioesterases have the conserved His residue (2481FASN), and the level of sequence similarity is further demonstrated in the phylogenetic tree (Fig. S7). Interestingly, constructing phylogenetic trees based on A, PCP, Ox and TE domains (Figure S4.8-S4.10 and Table S4.1-S4.4), reveals that many putative synthases have similar domains, and that homologs of the individual BpsA domains can be found in very different synthases.
To verify our \textit{in vitro} results, we characterized a set of BpsA mutants designed with selectively inactivated domains. We expressed and purified a BpsA mutant with a K598E mutation in the Ox-domain, preventing the binding of flavin, leading to a colorless-protein; a BpsA mutant with an excized TE domain; a stand-alone Ox-domain; and BpsA mutants with the TdiA\textsuperscript{24} PCP swapped into the synthase (see SI). We verified that each of these mutants are inactive. To assess whether two BpsA proteins can act \textit{in trans} to produce indigoidine, a matrix of inactive mutants of BpsA was made (Table S4.5, Figure S4.11). Interestingly, no significant activity was observed for any of the mutant combinations, indicating that the L-Gln substrates loaded onto the PCP of BpsA are incapable of interacting with catalytic domains on a different monomer of BpsA to produce indigoidine.

After demonstrating systematic inhibition of each domain of BpsA, we set out to evaluate the behavior of these inhibitors \textit{in vivo}. The inhibitor activities were evaluated using an \textit{E. coli} strain co-expressing BpsA and Sfp, which translates to blue pigmentation of bacterial colonies on agar plates. We spotted the inhibitors on sterile filter paper discs that were placed upon a lawn of cells overexpressing both BpsA and Sfp. Whereas the A-domain inhibitor 1 shows activity \textit{in vitro}, \textit{in vivo} this molecule does not show any effect. It is known that some acyl-sulfonamides are hampered in their uptake\textsuperscript{14f} and combined with the instability of this polar molecule, 1 is a poor inhibitor \textit{in vivo}. On the other hand, the Ox-domain inhibitor 2 shows a marked effect \textit{in vivo} (Figure 4.4). With decreasing concentration of 2, we observe a shift from growth inhibition (close to the center of the disk) to growth without formation of pigment, to growth with blue pigment production (edge of plate). Thus, compound 2 shows \textit{in vivo} toxicity at high dosage, inhibition at medium dosage and no inhibition at low dosage. Pantetheinamide 3 has no effect on the formation of blue pigment when supplemented to \textit{E. coli} overexpressing BpsA/Sfp and the PPTase inhibitor 6 shows toxicity but no clear absence of blue pigmentation. How Sfp (or the native PPTase EntD) escapes inhibition by 6 remains unclear. The thioesterase inhibitors 4 and 5 show severe toxicity to the bacteria at the high concentrations we employed (data not shown). However, at lower
Figure 4.4: Inhibition of pigment production of E. coli strains expressing BpsA and Sfp. Clockwise from top left: A) 2 inhibits the Ox-domain and shows a halo of alive, but white colonies (1 µl of 1 M, 0.1 M and 0.01 M); B) 6 inhibits PPTases and 10 µl spots of 10 mM, 1 mM and 0.1 mM show toxicity, but no apparent white halo; both C) 4 (10 µl of 10 mM, 1 mM and 0.1 mM) and D) 5 (10 µl of 1 mM and 0.1 mM) show toxicity, but also the presence of white colonies.
concentrations, both general inhibitors show slight toxicity, but also the presence of white colonies. Although only two of the inhibitors show promising results in vivo, this approach can be a useful addition to our toolbox for the interrogation of synthases.

Taken together, this in vitro and in vivo data show that using this small toolbox of inhibitors, we can quickly deduce the enzymatic activities of (and in) an unknown synthase. From these straightforward experiments, we learned that the interrupted A-domain is functional and active, the Ox-domain is flavin-dependent, the PCP-domain can be loaded with various unnatural cargo in vitro, and the TE-domain is essential for blue pigment production and is inhibited by cysteine-targeting inhibitors. We, and others, can now speculate on the hypothetical enzymatic or non-enzymatic route to the final product (Figure 4.5). We are confident that our systematic approach opens up many new paths to surgically interrogate the individual steps of a multi-domain synthase.

Recent studies further demonstrated the applicability of blue pigment synthases as model system. E.g. they were utilized to develop blue-white screens in Streptomyces species. Furthermore, within an International Genetically Engineered Machine (iGEM) synthetic biology project, NRPS domain swapping recently lead to production of novel indigoidine-containing peptides. As part of this project, the PCP domain was replaced with various other PCPs, resulting in large variability in the amount of product. However, Owen et al. had previously co-expressed a PCP replacement of BpsA with various carrier proteins showing no (or severely reduced) indigoidine formation. In contrast to this split system, Beer et al. found that many of the PCP replacement mutants do produce indigoidine, as opposed to the Walsh TdiA construct (Figure S4.11). It appears that the linker regions between domains (here between Ox-A-PCP and PCP-TE) are crucial for productive catalysis and stable protein expression. However the biological role of these synthases remains speculative, even so many years after their discovery. These synthases appear to be present in many different organisms, ranging from bacteria to eukaryotes, and their genes are often cryptic, such as in the example in the bacterial insect pathogen Photobabds luminsecens.
Here we have shown a systematic chemical genomic approach for interrogating a synthase and its domains. Although current technology facilitates bioinformatical discovery of genes and proteins involved in biosynthesis of secondary metabolites, there is a clear and urgent need to verify predictions with experimental data. Selective inhibition of synthase domains is one of the key implements in the toolbox of the chemical biologist, and the functional elucidation and mechanism of assembly-line synthases can be addressed by these methods.

Experimental

BpsA activity assays

All assays were performed in Costar flat bottom, half area clear plastic 96 well plates. Reactions were monitored at 590 nm at times ranging from 10 mins to 4 hrs on a Perkin-Elmer HTS 7000 microplate reader. Activity assays were adapted from Owen et. al.11а. 1 µM BpsA was incubated in a 100 µL mixture containing 75 mM potassium phosphate pH 7.8, 1 mM CoA, 10 mM MgCl₂, and 2 µM Sfp. This mixture was incubated at 37°C for 10 mins. The reaction was initiated by the addition of a 50 µL mixture containing 75 mM potassium phosphate pH 7.8, 5 mM L-Gln, and 10 mM ATP. The reaction was monitored at 590 nm for ~10 mins. Reaction velocities were calculated by taking the slope of the linear portion of the initial activity. All assays involving mutants of BpsA were conducted in the same manner as described above.

Inhibition assays

A serial dilution of each inhibitors was prepared in a concentration range as shown in Figure 3. Each serial dilution measured 50 µL in volume. Next, a 50 µL mixture containing 2 µM BpsA, 75 mM potassium phosphate pH 7.8, 10 mM MgCl₂, 1 mM CoA, and 2 µM Sfp was added to each serial dilution and incubated at 37°C for 10 minutes. The reaction was initiated by the addition of a 50 µL mixture containing 75 mM potassium phosphate pH 7.8, 5 mM L-Gln, and 250 µM ATP. The reaction was monitored at 590 nm for ~30 mins. Reaction velocities were calculated by taking the slope of the linear portion of the initial activity. For further details, see supplementary information.
Figure 4.5: Proposed biosynthetic mechanism of indigoidine production.
Supplementary Information

Figure S4.1: Phylogenetic tree of BpsA. Neighbour joining tree based on 500 protein sequences obtained by DELTA-BLAST of BAE93896 using MUSCLE (Edgar 2004 Nuc Acids Res) and MEGA 5.2.2 (Tamura 2011 Mol Biol Evol). Marked in blue: branch of BpsA-related proteins used in Figure S4.2.
Figure S4.2: Phylogenetic tree of BpsA clade. Neighbour joining tree of BpsA-related protein sequences generated in Dendroscope. Sequences were marked according to their taxonomic position: yellow: actinobacteria, pink: alpha-proteobacteria.
Figure S4.3: Growth Curves. Representative growth curves of *E. coli* BL21 expressing BpsA and Sfp, under supplementation with glutamine, MgCl₂ and MnCl₂. A) In orange a culture not supplemented with Gln and in blue a culture supplemented with 3.3 mg/ml L-Gln. Inset shows a picture of the two cultures. B) In blue, red and green and aqua cultures supplemented with a dilution series of MgCl₂ ranging from 3.3, 1.7, 0.8, 0.4 mg/ml. C) In blue, green and purple cultures supplemented with a dilution series of 0.2, 0.1 and 0.05 mg/ml MnCl₂, whereas 0.4 mg/ml appeared to be toxic to *E. coli* (flat growth curve and clear first picture).
**Figure S4.4:** Glutamine pantetheinamide as inhibitor of BpsA. In vitro activity assay of BpsA activity by monitoring the formation of blue pigment. In blue, BpsA incubated with Sfp and CoA, followed by addition of L-Gln and ATP; in red, BpsA incubated with glutamine pantetheinamide, Coa-A,-D,-E and Sfp, followed by addition of L-Gln and ATP.
Figure S4.5: Activity of WT BpsA compared to TE mutant. (A) No activity was observed for the TE mutant. (B) Assay of the TE mutant with and without TycF, a type II thioesterase. No activity was observed when BpsA was incubated with TycF.
Figure S4.6: Chase experiment loading BpsA PCP with cargo. Top: fluorescent image of 8% SDS PAGE gel. Bottom: image of coomassie stained gel. BpsA was loaded with either glutamine-pantetheinamide probe or 4DMN-pantetheinamide probe (described in 30).
Figure S4.7: Phylogeny based on TE-domain homology. A phylogenetic tree was constructed based on a sequence alignment of the combined psi-blast results of the thioesterase domains of atromentin, tdiA and bpsa synthases (in total 2000 sequences). Within the atromentin/TdiA clade one other hit, besides many other putative hits, is annotated as PksJ from *B. subtilis* involved in bacillaene synthesis. Within the bacterial clade in orange, besides pyoverdine synthase, HctF, DidJ, surfactin and ZmaQ synthases can be found. Sequences were aligned using Muscle and tree constructed using Fasttree. Tree was visualized using Figtree and Dendroscope.
Figure S4.8: Phylogeny based on Ox-domain homology. A phylogenetic tree was constructed based on a sequence alignment of the combined psi-blast results of the oxidase domains of Mtac, Mtad, Blmiii, EpoB and BpsA synthases (in total 1500 sequences). Sequences were aligned using Muscle and tree constructed using Fasttree. Tree was visualized using Figtree and Dendroscope.
Figure S4.9: Phylogeny based on A-domain homology. A phylogenetic tree was constructed based on a sequence alignment of the psi-blast results of the A domain BpsA (in total 5000 sequences). Sequences were aligned using Muscle and tree constructed using Fasttree. Tree was visualized using Figtree and Dendroscope.
Figure S4.10: Phylogeny based on PCP-domain homology. A phylogenetic tree was constructed based on a sequence alignment of the psi-blast results of the PCP domain BpsA (in total 500 sequences). Sequences were aligned using Muscle and tree constructed using Fasttree. Tree was visualized using Figtree and Dendroscope.
Figure S4.11: Absorbance curves of all BpsA mutants and mutant combinations indicated in table S5. No significant measurable activity was recorded for any mutants when compared to wild type. Variations in absorbance may be due to protein aggregation or cofactor degradation.
Figure S4.12: Structures of inhibitors synthesized in this study.
Table S4.1: Sequence identity matrix of Ox domains using Clustalw

<table>
<thead>
<tr>
<th></th>
<th>EpoB_ox</th>
<th>BlmIII_ox</th>
<th>MtaD_ox</th>
<th>MtaC_ox</th>
<th>BpsA_ox</th>
</tr>
</thead>
<tbody>
<tr>
<td>EpoB_ox</td>
<td>100</td>
<td>37.11</td>
<td>43.92</td>
<td>36.08</td>
<td>23.08</td>
</tr>
<tr>
<td>BlmIII_ox</td>
<td>37.11</td>
<td>100</td>
<td>42.86</td>
<td>36.08</td>
<td>23.89</td>
</tr>
<tr>
<td>MtaD_ox</td>
<td>43.92</td>
<td>42.86</td>
<td>100</td>
<td>44.16</td>
<td>27.66</td>
</tr>
<tr>
<td>MtaC_ox</td>
<td>36.08</td>
<td>36.08</td>
<td>44.16</td>
<td>100</td>
<td>25.79</td>
</tr>
<tr>
<td>BpsA_ox</td>
<td>23.08</td>
<td>23.89</td>
<td>27.66</td>
<td>25.79</td>
<td>100</td>
</tr>
</tbody>
</table>
Table S4.2: Sequence identity matrix of whole synthases using Clustalw

<table>
<thead>
<tr>
<th></th>
<th>MtaC</th>
<th>MtaD</th>
<th>EpoB</th>
<th>BlmIII</th>
<th>BpsA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MtaC</td>
<td>100</td>
<td>45.37</td>
<td>38.84</td>
<td>26.02</td>
<td>10.17</td>
</tr>
<tr>
<td>MtaD</td>
<td>45.37</td>
<td>100</td>
<td>36.36</td>
<td>25.6</td>
<td>12.07</td>
</tr>
<tr>
<td>EpoB</td>
<td>38.84</td>
<td>36.36</td>
<td>100</td>
<td>20.95</td>
<td>9.69</td>
</tr>
<tr>
<td>BlmIII</td>
<td>26.02</td>
<td>25.6</td>
<td>20.95</td>
<td>100</td>
<td>9.88</td>
</tr>
<tr>
<td>BpsA</td>
<td>10.17</td>
<td>12.07</td>
<td>9.69</td>
<td>9.88</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>atromentin_te</td>
<td>TdiA_te</td>
<td>BpsA_te</td>
<td>3TJM_hs_FAS_TE</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>---------</td>
<td>---------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>atromentin_te</td>
<td>100</td>
<td>45.56</td>
<td>24.5</td>
<td>15.32</td>
<td></td>
</tr>
<tr>
<td>TdiA_te</td>
<td>45.56</td>
<td>100</td>
<td>21.05</td>
<td>16.33</td>
<td></td>
</tr>
<tr>
<td>BpsA_te</td>
<td>24.5</td>
<td>21.05</td>
<td>100</td>
<td>20.46</td>
<td></td>
</tr>
<tr>
<td>3TJM_hs_FAS_TE</td>
<td>15.32</td>
<td>16.33</td>
<td>20.46</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Table S4.4: Sequence identity matrix of whole synthases using Clustalw

<table>
<thead>
<tr>
<th></th>
<th>atromentin</th>
<th>TdiA</th>
<th>BpsA</th>
<th>FASN</th>
</tr>
</thead>
<tbody>
<tr>
<td>atromentin</td>
<td>100</td>
<td>34.78</td>
<td>19.78</td>
<td>13.09</td>
</tr>
<tr>
<td>TdiA</td>
<td>34.78</td>
<td>100</td>
<td>18.39</td>
<td>13.58</td>
</tr>
<tr>
<td>BpsA</td>
<td>19.78</td>
<td>18.39</td>
<td>100</td>
<td>12.85</td>
</tr>
<tr>
<td>FASN</td>
<td>13.09</td>
<td>13.58</td>
<td>12.85</td>
<td>100</td>
</tr>
</tbody>
</table>
Table S4.5: *Matrix depicting combination of BpsA mutants in in vitro activity assays*

<table>
<thead>
<tr>
<th></th>
<th>BT1</th>
<th>BT2</th>
<th>BT3</th>
<th>TEMUT</th>
<th>OXMUT</th>
<th>OX</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BT2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BT3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMUT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OXMUT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Methods

Protein expression and purification

The BpsA gene was subcloned from pCDFDUET into pET28a. BpsA was expressed and purified using standard techniques. Briefly, BL21(DE3) cells carrying pET28a-BpsA were grown at room temperature to an OD600 of 0.6 and expression was induced with 0.5 mM IPTG at room temperature for 16 hours. The inactive apo-protein was purified by Ni²⁺-NTA (Novagen) affinity chromatography and the bright yellow (flavin containing) eluent dialyzed against 50 mM phosphate pH 7.8 containing 20% glycerol. Aliquots were stored at -80 °C. Mutant BpsA clones were expressed similarly. CoaA, CoaD, CoaE and Sfp were expressed and purified as previously described.

Inhibition of A domain

A 50 µL serial dilution of 1 was made in concentrations ranging from 10 mM to 20 µM. A 50 µL mixture containing 1 µM BpsA, 75 mM potassium phosphate pH 7.8, 10 mM MgCl₂, 1 mM CoA, and 2 µM Sfp was added to each serial dilution and incubated at 37°C for 10 mins. The reaction was initiated by the addition of a 50 µL mixture containing 75 mM potassium phosphate pH 7.8, 5 mM L-Gln, and 250 µM ATP. The reaction was monitored at 590 nm for 30 mins.

Inhibition of TE domain

A serial dilution of either PMSF or AEBSF was made in 50 µL potassium phosphate pH 7.8 with concentrations ranging from 200 µM to 400 nM for PMSF and 2.5 mM to 10 µM for AEBSF. First, 1 µM BpsA was incubated in a 50 µL mixture containing 75 mM potassium phosphate pH 7.8, 1 mM CoA, 10 mM MgCl₂, and 2 µM Sfp. This mixture was incubated at 37 °C for 10 minutes. To these reactions were added the serial dilution of either PMSF or AEBSF and allowed to incubate at 37 °C for 5 minutes. The reaction was initiated by the addition of a 50 µL mixture containing 75 mM potassium phosphate pH 7.8, 5 mM L-Gln, and 10 mM ATP. Reactions were allowed to proceed for approximately 10 minutes.

Inhibition of PPTase by 6-NOBP
A serial dilution of 6-NOBP was made in concentrations ranging from 500 µM to 1 µM. 50 µL of a mixture containing 75 mM potassium phosphate pH 7.8, 10 mM ATP, 5 mM L-Gln, 10 mM MgCl₂, 30 µM CoA, and 50 nM Sfp was added to each serial dilution. This mixture was allowed to stand at 25 °C for 10 minutes. The reaction was initiated by the addition of 1 µM BpsA and the reaction was monitored at 590 nm for approximately 30 minutes.

**In vivo inhibition assay**

BL21(DE3) cells carrying pET28a-BpsA and pUC8-Sfp were grown overnight at 37°C and 200 µl was plated on an LB agar plate containing 50 mg/L kanamycin and 100mg/L ampicillin. 1 cm filter disk were placed upon surface and 10 µl of compound was added. The plates were incubated o/n at 37°C or 24 hrs at room temperature.

**Phylogenetic analysis of BpsA**

The BpsA protein sequence from *Streptomyces lavendulae* (BAE93896) was subjected to DELTA-BLAST with default parameters increasing the aligned sequence number to 500. Sequences were aligned in MUSCLE. Using MEGA 5.2.2, a neighbour joining tree was generated using default settings. The subtree of highly conserved BpsA-like proteins was identified using Dendoscoope and processed into a separate neighbour joining tree.

**Phylogenetic analysis of BpsA**

Sequences were collected from the NCBI database using Psi-Blast. Domain borders were identified using Interpro and based on literature. Large sequence collections were sorted for duplicates using Duplicatefinder and names shortened using Speciesidentifier. Muscle was used to align sequences and Fasttree to assemble large phylogenetic trees. Dendropscope and Figtree were used to visualize trees.

**Sequence identity matrices**

Known domains were aligned using ClustalW and the sequence identity matrix shown.

**Synthesis of Molecular inhibitors**

**General**
Chemicals were obtained from various sources (Fluka, Sigma-Aldrich, Fisher, TCI and Acros). All reactions were carried out under an argon atmosphere in dry solvents and constant magnetic stirring. TLC analysis was performed using silica gel 60 F254 plates (EM Scientific) and visualized using an appropriate stain: 2,4-dinitrophenylhydrazin for PMB protecting group containing compounds, cerium molybdate for azides, ninhydrin for amines and potassium permanganate for pantetheine analogs. Flash chromatography was carried out with Silicycle 60 230-400 mesh. UPLC analysis was performed on a Waters Acquity system using an Acquity binary solvent manager, an Acquity column manager, a 2777c robotic autosampler, an Acquity TUV detector and a SQ detector. Compounds were separated on a BEHC18 1.7μm 2.1 x 50mm UPLC column at 0.8 ml/min using a linear gradient of 95% H2O containing 0.1% formic acid to 85% acetonitrile containing 0.1% formic acid in 2 minutes. The column was maintained at 55 °C. High-res ESI mass spectra were obtained at the UCSD mass spectrometry facility using a Micromass Quattro Ultima Triple Quadrupole MS. Small molecule NMR spectra were obtained on a 400 Mhz Varian Mercury Plus spectrometer, a 500 Mhz JEOL ECA 500spectrometer and a 500 Mhz Varian VX500 equipped with a XSens 2-channel cold probe. NMR spectra are shown at the end of this document.

**L-Gln-sulfonamide-adenosine:**

L-Gln sulfonamide-adenosine was synthesized in four linear steps from adenosine. Adenosine was protected, reacted with synthesized sulfamoyl chloride, coupled with NHS-activated and Boc protected L-Gln, followed by global deprotection.

\((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methanol\)\(^{39}\) (S1)

Adenosine (1 g, 3.7 mmol) was suspended in 50 ml acetone under argon. Dropwise addition of 800 µl of HClO\(_4\) gave a clear solution. After 2h the solution was treated with ammonia until the pH was 7. The solvent was evaporated and the product purified by flash chromatography using CH\(_2\)Cl\(_2\)/MeOH (95%). 1H NMR (400 MHz, ddmso) \(\delta = 1.30\) (3H, s), 1.53 (3H, s), 4.24 (2H,
Sulfamoyl chloride\textsuperscript{40} (S2)

Chlorosulfamoyl isocyanate (1 g, 7 mmol) was placed in a dry three-neck flask, equipped with a CaCl\textsubscript{2} drying tube. Formic acid (270 µl) were added dropwise by syringe, while cooling the flask with ice. A vigorous reaction occurred and the mixture turned turbid. The reaction was stirred for 30 minutes at room temperature and the reaction was diluted with benzene and evaporated to yield the desired product which stained with ninhydrin (70%), and was used without further purification.

\((3aR,4R,6R,6aR)-6-(6\text{-amino-9H-purin-9-yl})-2,2\text{-dimethyltetrahydrofuro}[3,4-d][1,3]dioxol-4-yl)methyl sulfamate (S3)

To a magnetically stirred solution of NaH (84 mg of a 60% suspension in mineral oil, 2.5 mmol) in dioxane, under argon at 0 °C, was added 2',3'-O-isopropylideneadenosine S1 (500 mg, 1.6 mmol) dissolved in dioxane. After 1h, sulfamoyl chloride (3.2 mmol, 370 mg, S2) was added and the mixture was allowed to warm to room temperature and stirred for 24h. Methanol was added and the solvent evaporated. The crude material was purified by flash chromatography using CH\textsubscript{2}Cl\textsubscript{2}/MeOH. (75%) 1H NMR (400 MHz, ddmso) δ = 1.29 (3H, s), 1.52 (3H, s), 4.22 (1H, m), 4.95 (1H, m), 5.31 (2H, m), 6.12 (1H, s), 7.36 (1H, bs), 7.43 (1H, bs), 8.15 (1H, s), 8.33 (1H, s); ESI-MS m/z 387.1

\((3aR,4R,6R,6aR)-6-(6\text{-amino-9H-purin-9-yl})-2,2\text{-dimethyltetrahydrofuro}[3,4-d][1,3]dioxol-4-yl)methyl ((tert-butoxycarbonyl)-D-glutaminyl)sulfamate\textsuperscript{41} (S4)

Boc-L-Gln-OH (1 g, 4 mmol) was dissolved in CH\textsubscript{2}Cl\textsubscript{2} at 0° C and N-hydroxysuccinimide (500 mg, 4.3 mmol) and EDC (800 mg, 4.2 mmol) were added. The reaction mixture was maintained at 4 °C for 16h and the solvent evaporated. The crude material was dissolved in ice water, carefully acidified, extracted with ethyl acetate, and the aqueous layer basified, extracted with ethylacetate and combined extract dried with Na\textsubscript{2}SO\textsubscript{4} and evaporated. Compound S3 (200 mg, 0.5 mmol) was dissolved in dry DMF and DBU (80 µl, 1eq.) added. The NHS-activated ester
(151 mg, 1 eq.) was dissolved in DMF and dropwise added to the solution, followed by stirring for 16h at room temperature. The mixture was diluted with water and extracted 3x with ethylacetate. The combined organic extracts were washed with water and brine and the product purified by flash chromatography using CH$_2$Cl$_2$/MeOH (70%) 1H NMR (400 MHz, CDCl$_3$) $\delta$ = 1.21 (s), 1.29 (s), 1.48 (s), 2.17 (m), 2.55 (m), 4.12 (m), 4.34 (s), 5.82 (m), 5.95 (brs), 7.26 (s), 7.79 (s), 7.91 (s), 8.08 (s), 10.06 (s); ESI-MS m/z 615.2

\((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl (L-glutaminyl)sulfamate$^{41}$ (1)

Compound S4 (40 mg, 0.07 mmol) was dissolved in CH$_2$Cl$_2$ and chilled to 0°C. An equal amount of 20% TFA in CH$_2$Cl$_2$ was added and the reaction stirred for 2h. The reaction was evaporated and used as is. (90%) 1H NMR (500 MHz, ddmso) $\delta$ = 10.70, 8.31, 8.09, 7.33, 5.84, 4.57, 4.12, 3.93, 3.12, 2.55, 2.03, 1.93, 1.68; 13C NMR (100 MHz, ddmso) $\delta$ = 27, 31, 49, 51, 62, 71, 74, 88, 116, 118, 140, 149, 152, 156, 174, 176

Oxidation domain inhibitor

\(4H\)-benzo[d][1,3]oxathiin-2-one$^{42}$ (2)

A solution of 2-mercaptobenzoic acid (0.5 g, 3.2 mmol) in THF (10 mL) was added dropwise to a mixture of LiAlH$_4$ (0.3 g, 7.9 mmol) in THF (20 mL) at 0°C. The mixture was stirred at room temperature under argon. After 3 hours, dH$_2$O (10 mL) and 6M HCl (5 mL) was added dropwise at 0°C. The resulting mixture was poured into additional H$_2$O (50 mL) and extracted with EtOAc (3 x 100 mL). The organic layer was dried over anhydrous Na$_2$SO$_4$, filtered, and concentrated to give the alcohol as a brown oil (0.44 g, 98%).

To avoid oxidation of the thiophenol, the crude oil was carried on without purification and was dissolved in CH$_2$Cl$_2$ (50 mL). Carbonyl diimidazole (0.56 g, 3.44 mmol) was added to the mixture portionwise over 1 min. The reaction mixture stirred under argon at room temperature. After 14 hours, dH$_2$O (50 mL) was added and the aqueous layer was extracted with CH$_2$Cl$_2$ (2 x 50 mL). The organic layers were collected, dried over anhydrous Na$_2$SO$_4$, filtered, and concentrated to give a crude yellow oil. Flash chromatography with silica gel and CH$_2$Cl$_2$ gave
4H-benzo[d][1,3]oxathiin-2-one as a white crystalline solid (0.39 g, 74% yield, 2 steps). 1H NMR (400 MHz, CDCl₃) δ 5.29 (s, 2H), 7.30–7.41 (m, 4H). 13C NMR (101 MHz, CDCl₃) δ 166.6, 131.2, 129.9, 128.9, 127.8, 126.8, 126.5, 72.0. FTIR: 1687, 1581, 1471, 1447, 1388, 1248, 1248, 1218, 1161, 1137, 1108, 1062, 1000, 971, 939, 870, 837, 751. mp 58-60°C. HRMS (APCI-TOF) m/z [M + H]^+ calcd for C₈H₇O₂S, 167.0167; found, 167.0156.

**Glutamine-pantotheinamide**

The synthesis of Pmb protected pantotheine amine S5 has been described elsewhere. Here, we coupled Boc and trityl protected L-Gln to S5 and deprotected the final product.

**tert-butyl (1-((4R)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxan-4-yl)-1,5,10,14-tetraoxo-16,16,16-triphenyl-2,6,9,15-tetraazahexadecan-11-yl)carbamate (S6)**

S5 (100 mg, 0.2 mmol) was dissolved in 2 ml dry DMF, the solution chilled in ice, and 80 mg (0.2 mmol) Boc-Gln-OH, 63 mg (2 eq.) HOBt and 80 mg (2 eq.) EDC were added. DIPEA (70 µl, 2 eq.) was added and the pH of the solution controlled by further small additions of DIPEA. The reaction mixture was allowed to warm to room temperature and stirred overnight. Water was added and the suspension extracted 3x with ethyl acetate. The combined organic layers were extracted with water and brine, dried with Na₂SO₄ and evaporated. The crude material was purified by flash chromatography using CH₂Cl₂/MeOH. (90 %)

**2-amino-N1-(2-(3-((R)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)pentanediamide (3)**

Compound S6 (50 mg) was dissolved in 1 ml CH₂Cl₂ and chilled. 2 ml of 20% TFA in CH₂Cl₂ was added and the reaction stirred for 2h at room temperature. The reaction was evaporated and 3 used as is. (99%)

**AEBSF-pantotheinamide**

The commercially available AEBSF (Pefabloc) with its amine-terminated extension, was used to synthesize a reactive pantetheinamide analog. For spacing, we extended pantothenic acid with aminobutyric acid.
2,5-dioxopyrrolidin-1-yl 3-((4R)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamido)propanoate (S7)

PMB-protected pantothenic acid (300 mg, 0.9 mmol) and NHS (100 mg, 0.9 mmol) were dissolved in THF (2 ml) and chilled. EDC (0.9 mmol) and DIPEA (0.9 mmol) were added and the reaction mixture stirred o/n. The solvent was evaporated and S7 purified with flash chromatography using CH2Cl2/MeOH (90%).

4-(3-((4R)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamido)propanamido)butanoic acid (S8)

S7 (200 mg, 0.5 mmol) was dissolved in CH2Cl2, 50 mg (0.5 mmol) aminobutyric acid added, and the reaction mixture stirred o/n at room temperature. The product was purified by flash chromatography using CH2Cl2/MeOH.

2,5-dioxopyrrolidin-1-yl 5-((3-(4R)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamido)propanamido)pentanoate (S9)

S8 (200 mg, 0.5 mmol) and NHS (57 mg, 0.5 mmol) were dissolved in THF (2 ml) and chilled. EDC (0.9 mmol) and DIPEA (0.9 mmol) were added and the reaction mixture stirred o/n. The solvent was evaporated and S9 purified with flash chromatography using CH2Cl2/MeOH (90%).

4-(2-(4-(3-((4R)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamido)propanamido)butanamido)ethyl)benzenesulfonyl fluoride (S10)

S9 (40 mg, 0.08 mmol) was dissolved in chilled CH2Cl2, followed by addition of 15 mg of AEBSF and stirred o/n while being allowed to warm to room temperature. S10 was purified using flash chromatography using CH2Cl2/MeOH (90%).

(R)-4-(2-(4-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)butanamido)ethyl)benzenesulfonyl fluoride (5)

Compound S10 (40 mg, 0.08 mmol) was dissolved in 80% aqueous acetic acid and stirred o/n at room temperature (90%).
References


(17) I. McCulloch, M. D. Burkart, *manuscript under revision* 2015.


(37) J. D. Thompson, T. Gibson, D. G. Higgins, *Current protocols in bioinformatics* 2002, 2.3. 1-2.3. 22.


(42) I. McCulloch, M. D. Burkart, *manuscript under revision* 2015.


Chapter 4, in full, is currently being prepared for submission for publication as Vickery, Christopher R.; Sonnenschein, Eva C.; Beld, Joris; McCulloch, Ian; Noel, Joseph P.; Burkart, Michael D. “Dissecting modular synthases through inhibition: a chemical genetic approach,” The dissertation author was shared primary investigator and author of this material.
Chapter 5: A combined approach for engineering a heterologous Type III PKS and enhancing polyketide biosynthesis in \textit{S. cerevisiae}


troduction

A variety of metabolic systems in existence today rely heavily on the use of heterologous enzymes (Nevoigt, 2008). This is often due to the need to leverage well-known, valuable industrial strains for use in producing molecules that are not native to the host organism. With a growing interest in biobased products, the demand for novel semisynthetic and biosynthetic routes toward unique compounds is ever increasing. Polyketides have gained acceptance as excellent target compounds, as this class of molecules naturally form a diverse group with highly differentiated properties. Type III polyketide synthases (PKSs) are a class of enzyme classically associated with plant metabolism. These homo-dimeric enzymes utilize a variety acyl-CoA substrates to produce a plethora of natural products with varying structures and properties (Hertweck, 2009). Recently, there has been great interest in utilizing these synthases for the production of chemically diverse and industrially relevant molecules (Stewart 2013). The relatively small size (~43 kDa) and ability to act without any other cofactors or protein partners makes this enzyme class ideal for heterologous production of polyketide products (Pfeifer and Khosla, 2001). The enzyme 2-PS is responsible for making the lactone core of the \textit{G. hybrida} natural products gerberin and parasorboside (Eckermann 1998). Synthesis of this precursor, triacetic acid lactone, has been observed in vitro using one molecule of acetyl-CoA and two molecules of malonyl-CoA. Previous work has demonstrated the usefulness of triacetic acid lactone (TAL) as a potential biologically-derived intermediate for conversion to various relevant chemicals that rely on non-renewable fossil fuel-based processes (Cardenas and Da Silva, 2014; Chia et al., 2012; Tang et al., 2013; Xie et al., 2006).

Recent reports suggest multiple type III PKSs retain the ability to produce the molecule TAL separate from the \textit{G. hybrida} 2-PS. We have shown that a similar Type III PKS can be readily modified to behave like 2-PS, specifically the naringenin-producing chalcone synthase (Jez et al., 2000a). Others have shown how modified Type I PKSs which require a separate
activation machinery can also be used for the production of TAL (Xie et al., 2006). Specifically, crystallographic studies on Medicago sativa chalcone synthase (CHS2) and Gerbera hybrida 2-pyrone synthase (2-PS) suggest that a very simple organization of few residues can dictate the entire specificity and selectivity of the enzyme (Jez et al., 2002).

We have previously demonstrated that these enzymes are equally capable of producing TAL relative to the wildtype enzyme in S. cerevisiae when using a high-copy overexpression vector in vivo (Cardenas and Da Silva, 2014). Therefore, rational enzyme engineering was performed using the native G. hybrida 2-PS in order to attempt to increase the performance of this enzyme in vivo. Structure based site-directed mutagenesis led to the testing of 41 variants in vivo. These results were coupled with in vitro characterization of several variants in order to uncover the enzymatic properties that conferred increased production of TAL. This novel approach of coupling in vitro with in vivo analysis facilitated the creation of a useful and stable enzyme as applied in a biological fermentation process. These data provide invaluable information for designing heterologous synthase systems for improved catalytic efficiency and stability in microbial systems, as demonstrated in Saccharomyces cerevisiae for the purposes of this work. Feasibility of such a strategy was further validated by performing fed-batch fermentation with a variant 2-PS and engineered yeast strain and demonstrating increased performance through improved titers and yields. Our engineering technique should be a valuable tool for optimization of industrial bioprocessing of economically relevant chemicals.

Materials and Methods

In vitro enzyme assays

In vitro assessment of 2-PS activity were carried out as previously described (Jez 2000) in a reaction containing 100 mM Sodium phosphate buffer, 60 µM malonyl CoA, and varying concentrations of acetyl CoA. However, after quenching with acetic acid, reactions were directly analyzed on a Shimadzu LCMS-2020 UPLC-MS using a phenomenex Kinetex C18 reverse phase column. Samples were monitored on the negative ion channel at 125.11 m/z, and the area under the mass peak was used to quantify TAL.
Linear activity for each mutant was assessed as previously described (Jez 2000). For each mutant, a suitable reaction time was determined (6 to 15 minutes), and 0.5—1 µg enzyme was incubated with 60 µM CoA and concentrations ranging from 250nM to 32 µM acetyl-CoA for the determined length of time and quenched with 5% acetic acid. Integration of the peaks corresponding to TAL were compared against a standard curve of TAL consisting of concentrations ranging from 50 nM to 1 µM. Reactions were performed in triplicate and analyzed using the Michaelis-Menten equation in GraphPad Prism 5 to obtain kinetic parameters.

For the stability experiments performed in E. coli, each mutant assayed was incubated at room temperature in 100mM Sodium phosphate buffer pH 7.0 for intervals of time from 0 minutes to 180 minutes. At each time point, an activity assay was performed. Assays were performed as described above, except that acetyl-CoA was omitted from the reactions, and all reactions were incubated for 6 minutes. The resulting activity was plotted and analyzed using a single decay equation to obtain the half-life of each mutant.

Additional stability studies were performed following fermentation of Saccharomyces cerevisiae using a standard DTNB assay. The reaction consisted of 150 mM tris buffer (pH 8.0), 4mM DTNB, and 2 nmol of purified protein extract. Extracts were performed at 24h, 36h, and 48h during both aerobic and anaerobic yeast cultivation. Using a modified purification scheme as done previously (Jez et al., 2000b), additional protection of purified protein was achieved via nitrogen-sparging prior to the assay. The free thiol levels were reported using DTT standards from 0-20 µM and detecting absorbance at 312 nm (Shimadzu UV-2450 spectrophotometer; Columbia, MD).

**Strains and plasmids**

Escherichia coli strain XL1-Blue (Stratagene, Santa Clara, CA) was used for amplification of yeast shuttle vectors, and strain BL21 (DE3) (EMD Millipore, Billerica, MA) for expression of 2-PS and its variants for E. coli in vitro studies following transformation by the Quickchange protocol (Qiagen, Valencia, CA). Protease-deficient S. cerevisiae strain BJ5464 (Jones, 1991) was used as the base strain for TAL production. The engineered strain BJΔpyc2Δnte1 (Cardenas
and Da Silva, 2014) was used for fed-batch fermentation studies following the initial characterization efforts.

The g2ps1 gene encoding the 2-PS and all variants from Gerbera hybrida were PCR amplified from the pHIS8 cassette (Jez et al., 2000a). As previously done, the pXP vector system (Shen et al., 2012) was used to carry the 2-PS 2-PS enzyme variants. Specifically, the high copy 2 \( \mu \)-based pXP842 vector was used, harboring the glucose-repressed ADH2 promoter and CYC1 terminator with loxP-flanked URA3 selection marker. Yeast cells were transformed as previously described (Gietz et al., 1992; Hill et al., 1991) using selective SDC-A plates. Cell colonies were allowed to grow for 3-5 days before generating inoculum cultures or glycerol stocks for long-term storage.

Following SpeI and XhoI digestion, the gene was inserted into pXP842 using the Rapid DNA Ligation Kit (Thermo Scientific, Waltham, MA). A list of all plasmid constructs used in this study can be found in Table 5.1. Plasmid recovery was performed using the GeneJetTM Plasmid Miniprep Kit (Thermo Scientific, Waltham, MA) and DNA sequence analysis confirmed the correct sequence of all PCR-amplified inserts (GeneWiz, South Plainfield, NJ; Eton Biosciences, San Diego, CA).

The KOD Hot-start polymerase (EMD Chemicals, San Diego, CA) was used in PCR reactions for plasmid constructs. Restriction enzymes, T4 DNA ligase, Taq DNA polymerase, and deoxynucleotides were purchased from New England Biolabs. Oligonucleotide primers were purchased from IDT DNA (San Diego, CA). A list of plasmids used in this study can be found in Table 5.1.
Table 5.1: List of enzyme mutants used in this study.

<table>
<thead>
<tr>
<th>Construct Name</th>
<th>Mutations present</th>
</tr>
</thead>
<tbody>
<tr>
<td>me1</td>
<td>C35A</td>
</tr>
<tr>
<td>me2</td>
<td>C35S</td>
</tr>
<tr>
<td>me3</td>
<td>L268F</td>
</tr>
<tr>
<td>me4</td>
<td>C35S L202M</td>
</tr>
<tr>
<td>me5</td>
<td>C35S L202F</td>
</tr>
<tr>
<td>me6</td>
<td>C35S L268M</td>
</tr>
<tr>
<td>me7</td>
<td>C35S L268F</td>
</tr>
<tr>
<td>me8</td>
<td>C35S I343M</td>
</tr>
<tr>
<td>me9</td>
<td>C35S I343F</td>
</tr>
<tr>
<td>me10</td>
<td>C35S M259F</td>
</tr>
<tr>
<td>me11</td>
<td>C35S T137L</td>
</tr>
<tr>
<td>me12</td>
<td>C35S T137M</td>
</tr>
<tr>
<td>me13</td>
<td>C35S T137F</td>
</tr>
<tr>
<td>me14</td>
<td>C35S I201L</td>
</tr>
<tr>
<td>me15</td>
<td>C35S I201M</td>
</tr>
<tr>
<td>me16</td>
<td>C35S I201F</td>
</tr>
<tr>
<td>me17</td>
<td>C35S L202F L268M</td>
</tr>
<tr>
<td>me18</td>
<td>C35S L202F L268F</td>
</tr>
<tr>
<td>me19</td>
<td>C35S L268M I343M</td>
</tr>
<tr>
<td>me20</td>
<td>C35S C65A</td>
</tr>
<tr>
<td>me21</td>
<td>C35S C65S</td>
</tr>
<tr>
<td>me22</td>
<td>C35S C89A</td>
</tr>
<tr>
<td>me23</td>
<td>C35S C89S</td>
</tr>
<tr>
<td>me24</td>
<td>C35S C135A</td>
</tr>
<tr>
<td>me25</td>
<td>C35S C135S</td>
</tr>
<tr>
<td>me26</td>
<td>C35S C195A</td>
</tr>
<tr>
<td>me27</td>
<td>C35S C195S</td>
</tr>
<tr>
<td>me28</td>
<td>C35S C346A</td>
</tr>
<tr>
<td>me29</td>
<td>C35S C346S</td>
</tr>
<tr>
<td>me30</td>
<td>C35S C372A</td>
</tr>
<tr>
<td>me31</td>
<td>C35S C372S</td>
</tr>
<tr>
<td>me32</td>
<td>C89S C372S</td>
</tr>
<tr>
<td>me33</td>
<td>C35S C89S C372S</td>
</tr>
<tr>
<td>me34</td>
<td>C89S C346A C372S</td>
</tr>
<tr>
<td>me35</td>
<td>C35S C89S C346A C372S</td>
</tr>
<tr>
<td>me36</td>
<td>C65S C89S C372S</td>
</tr>
<tr>
<td>me37</td>
<td>C35S C65S C89S C372S</td>
</tr>
<tr>
<td>me38</td>
<td>C89S C195S C346A C372S</td>
</tr>
<tr>
<td>me39</td>
<td>C35S C89S C195S C346A C372S</td>
</tr>
<tr>
<td>me40</td>
<td>C65S C89S C346A C372S</td>
</tr>
<tr>
<td>me41</td>
<td>C35S C65S C89S C195S C346A C372S</td>
</tr>
</tbody>
</table>
**Media and cultivation**

Luria-Bertani (LB) media was used for proliferation of XL1-Blue cells with 150 mg/L ampicillin for selection of plasmid-containing strains (Sambrook and Russell, 2001). Complex YPD media (1% dextrose; 1% Bacto yeast extract; 2% Bacto peptone) was used for 48h cultivations, and selective SDC(A) media (1% dextrose, 0.67% yeast nitrogen base, 0.5% Bacto casamino acids, 0.5% ammonium sulfate and 100 mg/L adenine) was used for generating overnight inoculum cultures. For the inoculum cultures, S. cerevisiae strains were grown for 16 h overnight in 5 mL selective SDC(A) media in an air shaker (New Brunswick Scientific) at 250 rpm and 30°C, and used to inoculate cultures to an initial cell density (OD600) of 0.3 (Shimadzu UV-2450 UV-VIS Spectrophotometer, Columbia, MD). During the cultivation, a 48h sample was taken where cell densities were determined and the samples were centrifuged at 3,000 rpm (2,600 g) for 5 min at 4°C (Beckman GS-6R Centrifuge, Brea, CA). The supernatants were stored at 4°C for HPLC analysis of TAL levels in the culture broth.

**HPLC Assay**

The concentration of triacetic acid lactone was measured by HPLC using a Shimadzu HPLC system: LC-10AT pumps (Shimadzu), UV-VIS detector (SPD-10A VP, Shimadzu), Zorbax SB-C18 reversed-phase column (2.1x150 mm, Agilent Technologies). Acetonitrile buffered in 1% acetic acid was used as the mobile phase, while HPLC grade water buffered in 1% acetic acid was used as the aqueous phase. A gradient program using a 95-85% Pump B gradient (H2O with 1% acetic acid) provided an elution time of approximately 12 minutes (flow rate 0.25 mL/min, column temperature 25°C).

**Fed-batch fermentations**

A New Brunswick BioFlo III system equipped with a 2.5 L vessel was employed to maintain cultivation parameters during fed-batch operation. Initial 5mL overnights of our engineered strain were grown to inoculate a 50mL shake flask culture that would be used as the seed inoculum. Following inoculation to an OD600 of 0.3 with 1% YPD, the fermentor pH was
controlled at pH 6 by automatic supply of either 6M sodium hydroxide or hydrochloric acid. Throughout the fermentation Antifoam SE-15 (Sigma) was pumped for maintenance of foaming, controlled by the antifoam probe. Dissolved oxygen was initially maintained by using a 400 rpm agitation speed, and supplying sparged air at a rate of 0.2 vvm in order to maintain dissolved oxygen levels to 20%. While agitation was kept constant, the aeration rate was allowed to increase in order to ensure DO levels could be maintained, and no additional control was employed for this batch mode operation. The cultivation proceeded under these conditions for 12 h for glucose-fed and 24 h for ethanol-fed operation before the selected carbon source feed was initiated at 1.5 mL/h (glucose feed, 3.6M) and 2.5 mL/h (ethanol feed, 16% v/v). Nitrogen-containing amino acids were supplied with the glucose feed in order to support a 35:1 (g/g) carbon-to-nitrogen ratio, ideal for high fermentative capacity (Albers et al., 1996; Manikandan et al., 2010) as follows: 869 mg/L L-arginine, 960 mg/L L-aspartic acid, 12.4 mg/L L-glutamine, 360 mg/L L-glutamic acid, 84.5 mg/L L-lysine-HCL, 100 mg/L adenine sulfate. The glucose feed rate varied to maintain DO levels at 20% with constant agitation (400 rpm) and aeration (0.8 vvm). The ethanol feed rate was kept constant throughout the fermentation; therefore, DO levels were maintained using an agitation-coupled DO control using a maximum agitation rate of 425 rpm and constant aeration rate of 0.2 vvm.

**Yield calculations**

Yield was calculated as the total TAL formed divided by the total carbon supplied at time of sampling. Maximum theoretical yield was calculated assuming all carbon was converted to TAL, previously determined to be 0.47 g/g on glucose, and 0.91 g/g on ethanol. Therefore, the percent theoretical yield of TAL on glucose or ethanol was calculated as: \[
\text{Percent Theoretical Yield} = \left( \frac{\text{TAL (g/L)}}{\text{Carbon Source Fed (g/L)}} \right) \times \left( \frac{\text{X (g/g Carbon)}}{100} \right).
\]

**Results**

Following expression and purification from E. coli, recombinant 2-PS produces TAL when incubated in the presence of the precursors acetyl-CoA and malonyl-CoA (Fig 1). Additionally, when 2-PS was incubated only with malonyl-CoA, TAL formation was also observed due to the
ability of 2-PS to decarboxylate malonyl-CoA into acetyl-CoA (Eckermann et al., 1998). Kinetic analysis was performed with varying concentrations of acetyl-CoA and a fixed concentration of malonyl-CoA as previously described (Jez, 2000). Although some background activity was present in all reactions due to the decarboxylation event, TAL formation was greater in the presence of acetyl-CoA (Figure S5.1) (Eckermann 1998). 2-PS was measured to have a Km of 3.1 ± 2.5 µM, a kcat of 0.089 ± .001 s⁻¹, and kcat/Km of 471 M⁻¹s⁻¹, similar to previously reported values (Jez, 2000).

Using our previous studies which identified steric hindrance as a method of defining specificity and selectivity in Type III PKSs, in vivo studies were designed for evaluation of 2-PS active site mutations, including a C35S intended to reduce oxidation sensitivity after purification. Based on residues inside the active site location, amino acid changes were selected in order to create size restriction, and therefore larger, bulkier residues were chosen for the study. For initial characterization of kcat and Km, the enzymes were expressed in E. coli and subsequently purified (Table 5.2). Additional characterization was performed by determining TAL biosynthesis levels, where 2-PS enzymes were expressed via high-copy plasmid (Table 5.1) in S. cerevisiae strain BJ5464 using glucose as the carbon source via 1% YPD, and TAL titer detected by HPLC.

In the mutants chosen for preliminary evaluation, only eight (including wildtype) produced significant and comparable quantities of soluble protein for kinetic analysis (Figure S5.2). The most active mutant following in vitro analysis was [C35S,L202F], with kcat/Km of 1.38x104 M⁻¹s⁻¹ (Table 5.2). Mutants with similar catalytic efficiencies to WT included [C35SL268F], [L268F], and [C35S,I201M]. The [C35S,I343F], [C35A], and [C35S] exhibited 1.7, 6.1, and 6.6-fold higher catalytic efficiencies than WT, respectively. Additionally, 2-PS mutants exhibited varying degrees of stability after affinity purification from E. coli (Figure S5.3)
### Table 5.2: Kinetic parameters of soluble 2-PS mutants

Mutants were assayed for activity and kinetic parameters were calculated using the Michaelis-Menten steady state equation.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>$K_m$ (µM)</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$k_{cat}/K_m$ (s$^{-1}$M$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>3.1 ± 2.5</td>
<td>0.089 ± 0.001</td>
<td>471</td>
</tr>
<tr>
<td>C35SL343F</td>
<td>2.2 ± 0.4</td>
<td>0.11 ± 0.001</td>
<td>819</td>
</tr>
<tr>
<td>C35SC372S</td>
<td>0.3 ± 0.07</td>
<td>0.416 ± 0.001</td>
<td>19908</td>
</tr>
<tr>
<td>C35SL202F</td>
<td>0.2 ± 0.04</td>
<td>0.164 ± 0.001</td>
<td>13790</td>
</tr>
<tr>
<td>C35SL268F</td>
<td>3.0 ± 0.9</td>
<td>0.056 ± 0.001</td>
<td>311</td>
</tr>
<tr>
<td>L268F</td>
<td>8.5 ± 4.8</td>
<td>0.123 ± 0.001</td>
<td>240</td>
</tr>
<tr>
<td>C35A</td>
<td>0.3 ± 0.1</td>
<td>0.058 ± 0.001</td>
<td>2891</td>
</tr>
<tr>
<td>C35S</td>
<td>0.6 ± 0.1</td>
<td>0.105 ± 0.001</td>
<td>3108</td>
</tr>
<tr>
<td>C35SI201M</td>
<td>0.5 ± 0.1</td>
<td>0.0126 ± 0.001</td>
<td>429</td>
</tr>
</tbody>
</table>
In order to assess the stability of these mutants, they were subjected to both thermofluor analysis (Niesen et al., 2007) and time-course activity assays. Five representative mutant enzymes were selected to be assayed for stability over time. They were incubated at room temperature in reaction buffer and assayed at several time points, beginning at the initial time of incubation. The [C35S] variant displayed increasing half-life values, while [C35A] showed a decrease in half-life (Figure 5.1). Surprisingly, [L268F] exhibited a greater half-life than WT and [C35S]. The [C35S,L202F] variant possessed the greatest half-life of nearly 100 minutes.

Interestingly, the biologically-produced TAL levels generated from employing these enzymes in S. cerevisiae did not always correlate with the kinetics (Figure 5.2). Relative to wildtype levels (0.48g/L), over 2-fold improvement in TAL titer was observed for 2-PS variants. Specifically, [C35S,L268F] and [C35S,T137F] both increased titers 2.2-fold to ~1.1 g/L TAL. The [C35S,T137F] variant did not purify well for kinetics studies, and therefore does not have additional parameters available for comparison. The [C35S] and [C35S,L202F] showed comparable increases in TAL titer to ~0.85 g/L TAL. This result is interesting when considering the robustness of [L202F] relative to [C35S] during kinetics and stability studies. The importance of [C35S] is further highlighted when comparing levels for enzymes [C35A] or [L268F]. Although kinetics were improved for the [C35A] variant, the enzyme performed poorly in yeast. Additionally, the L268F substitution in the active site is only relevant in combination with the C35S modification, as [L269F] alone also performs poorly. Introduction of a bulky residue via I343F decreased activity by 80%, which also contrasts the kinetics data showing improved efficiency. More closely correlated with the kinetics, [C35S,I201M] was a poor performer in S. cerevisiae, with titers just detectable at ~50 mg/L. In order to additionally ensure expression levels in yeast were not an issue or contributing to these observations, representative variants were screened by Western Blot. These results illustrate the complexity of translating kinetics of purified protein to producing chemicals in a biological environment.
Figure 5.1: Half-lives of 2-PS mutants. Each mutant was incubated at R.T. and measured for activity every 15 to 20 minutes. Curves were analyzed by linear regression, and a half-life value was obtained.
Figure 5.2: Titer values for wild-type and a variety of surface cysteine and active site mutants. Multiple mutations were made at some positions to test the flexibility of each site.
For cysteines present in the 2-PS enzyme, excluding the conserved C169 residue in the catalytic triad, residues were modified to either alanine or serine (Figure 5.3). Based on the 2-PS crystal structure, the cysteines most exposed on the enzyme’s surface were C65, C372, C135, and C89, all of which had the highest observed TAL increases for a serine substitution outside the active site (Figure 5.4). The best performer produced 1.23 g/L TAL using [C35S, C372S]. Kinetics and stability were selectively performed on this enzyme, revealing a $k_{cat}/K_m$ 42-fold (1.99 x104 M-1s-1) and half-life 2.1-fold (46 min) higher than that of wildtype. It was surprising to see significant improvement in performance by incorporating two amino acid substitutions not inside the active site pocket. Substitutions at other cysteine positions also had an effect on TAL levels. Position 135 has the potential to interact closely with C195, and exhibited less of an increase in TAL production over the other three cysteine residues. Substitution by alanine never provided a large increase, while the serine substitution demonstrated improvements over WT. For cysteines at buried, less exposed positions, such as C195 and C346, an alanine substitution was preferred over the serine mutation, although neither substitution yielded large increases over WT and [C35S]. These results would suggest that cysteine placement on the 2-PS enzyme can be manipulated to improve TAL-synthesizing potential. Due to the fact that various cysteine-modified enzymes were difficult to purify, additional studies would need to be performed in order to determine which kinetic parameters, if any, are implicated in the observed TAL increases.

Additional mutations were sequentially incorporated onto external cysteines using either the wildtype 2-PS enzyme or the [C35S] variant to generate a collection of mixed cysteine mutants, and TAL production in S. cerevisiae was evaluated (Figure 5.4). Selection of the [C35S] enzyme in addition to the 2-PS was done based on our previously described data and its importance in improving 2-PS performance. Cysteine variants built off of the 2-PS WT, and not harboring the C35S substitution, exhibited a marked decrease in TAL production as additional serines were introduced, until only 50% of wildtype levels were achieved on a per cell basis (g/g...
Figure 5.3: Titer values of all cysteine mutations. The greatest effects on titer values was obtained in the presence of C35, while additional cysteine mutations exhibited varying effects on titer values.
Figure 5.4: Cysteines of 2-PS, and effects of mutating a combination of cysteines. (A) Depiction of the structure of 2-PS and all cysteines. Cysteines in purple are located within the folds of the enzyme, red are located at or very near the surface, and teal is the active site cysteine. CoA is depicted in stick form. (B) Titer effects of cysteine mutant combinations. The greatest effect is observed when C35S and C372S mutations are combined.
However, the same cysteine mutations performed on the [C35S] enzyme further increased TAL yields, consistently leading to increased 2-fold TAL production relative to the wildtype. It is important to note that although TAL levels were sustained in this enzyme subset, no further TAL increases in yeast beyond that of [C35S,C72S] were observed. Overall these results would correspond well with previous kinetics and stability studies in suggesting that reduced cysteine levels on this enzyme can improve catalytic efficiency and/or stability.

To further understand possible effects of surface modifications on the biological and cellular processes, an additional rapid assay for free thiol determination was used. Following native purification of in vivo produced 2-PS wildtype in yeast, a DTNB assay was performed to quantify the free thiol levels. This assay was performed on protein that was purified under nitrogen-sparged conditions following 24 h, 36 h, and 48 h cultivation using either aerobic or anaerobic cultivation conditions (Figure 4.5). For both aerobic and anaerobic cultivations, a distinct loss in free thiols was observed after 36 h. In addition, the level of free thiols is significantly reduced at all times during anaerobic cultivation. Previous work using this yeast expression system (Cardenas and Da Silva, 2014) shows TAL being produced up to 48 h in batch cultivation. The free thiol data would therefore suggest that 2-PS is underperforming, and it appears to be coupled to a process independent of the presence of oxygen. It has been reported that oxygen-stress leads to overall poor health of the culture (Koc et al., 2004; Steels et al., 1994), potentially explaining the observation that anaerobic conditions yielded overall lower free thiols (per mg protein) than the aerobic system. Additionally, yeast reactive oxygen species (ROS) systems which are triggered in later stages of growth may also contribute to protein thiol oxidation (Grek et al., 2013).

These results demonstrate that surface cysteines can, and should be, altered for increased performance. The [C35S] variant alone improved catalytic efficiency, and was necessary for maintaining 2-PS activity with reduced surface cysteines (Figure 5.4). Initial stability assay results (Figure 5.1) also suggest that reducing thiols supports enzyme stability (via
Figure 5.5: Detected thiol levels under aerobic and anaerobic growth conditions. A spike in free thiol levels is detected after 36 hours of culture, which may contribute to heterologous 2-PS cysteine modification, and thus decreased enzyme activity and lower titer.
Figure 5.6: Effect of protease deficient yeast strains on 2-PS production. While protease deficient strains increase titers for WT and C35S mutants, there is a minimal effect on the high producing C35S C372S mutant.
increased half-life). The temporal changes in free thiol levels when 2-PS is expressed in yeast (Figure 5.5) additionally validate cysteine reduction for 2-PS could be advantageous. Proteolytic activity is one means for repairing or recycling damaged protein, including those which become oxidized (Costa et al., 2007). Previous studies using wildtype 2-PS showed that heterologously expressed 2-PS was absolutely targeted for proteolytic activities by vacuole (Cardenas and Da Silva, 2014). Therefore, 2-PS variants with altered cysteine residues were tested to determine whether cysteine might be involved in proteolytic targeting of 2-PS.

To test the direct impact proteases have on 2-PS variants, [C35S] and [C35S,C372S] variants were selected for transformation into four strains with disruption in two major yeast vacuolar proteases, Pep4 and Prb1. Following cultivation in 1% YPD after 48 h, samples were collected and TAL titer (g/L) determined (Figure 5.6). For strains expressing the WT 2-PS, we found that removal of these vacuolar proteases improved TAL levels (Cardenas and Da Silva, 2014). When the [C35S] variant was expressed, removal of proteases still increased TAL levels, although the increase was much lower. Therefore in this case, protease activity is still a limitation on TAL production. For the [C35S,C372S] variant, titers significantly improved in all strains (~1.3 g/L), including over 5-fold for strain BYt with both proteases intact. Therefore, under these conditions the [C35S,C372S] supported increased levels of TAL, consistent with screening of the 2-PS variants done previously. The results further suggest that the proteolysis taking place in yeast is directly affecting 2-PS, and that cysteine modification of the enzyme surface can reduce the impact. Cysteine residue changes are potentially an avenue for creating stable and more active polyketide synthases in S. cerevisiae.

In a recent study, metabolic engineering of S. cerevisiae for the production of triacetic acid lactone provided various improvements using genetic disruptions in significant pathways including central carbon metabolism (Cardenas and Da Silva, 2014). In addition, the use of glucose-fed batch fermentation was employed to elevate TAL titers. As previously shown, the BJ5464 background strain with the modified Δpyc2Δnte1 genotype produced significantly higher titers over the control strain (Cardenas and Da Silva, 2014). This strain generated yields of
Figure 5.7: Yield of TAL for engineered strain transformed with [C35S,C372S].
0.21 g/g DCW, a 2.8-fold increase over the base strain. Therefore in this work, this engineered strain was chosen for coupling to expression of an engineered 2-PS variant using the modified 2-PS [C35SC372S].

Our previous work had demonstrated the use of glucose for carbon supplementation during the fermentation (Cardenas and Da Silva, 2014). With the use of a glucose-repressed promoter like ADH2p, we chose to additionally make use of ethanol as a carbon source. Therefore, two carbon feeds were supplied to the cell in order to evaluate the extended fed-batch cultivations. First, an ethanol feed was modeled around the maximum ethanol consumption rate in microaerobic fermentation in our system, 0.16 g/L/h. An equally supplying feed rate was implemented at the time when half of endogenously produced ethanol was consumed, or approximately 24 h after batch inoculation. Second, the glucose feed was modified to support an appropriate carbon:nitrogen ratio as described previously. The optimal nitrogen source was selected as a mixture of nitrogen-containing amino acids based on literature comparing to ammonium sulfate or glutamic acid (Albers et al., 1996). To initiate the fermentations, an overnight shake-flask culture was inoculated in selective SDC(A) from 5mL cultures. This seed flask culture was used to inoculate the fermentor to 1.5L starting volume and a cell density (OD600) of 0.3. Depending on the feed strategy implemented, the volume was 1.7-2.0 L at the end of cultivation. Following fed-batch operation for up to 144 h, maximum theoretical yield, productivity, and TAL titer were determined. Incorporating [C35S,C372S] in our strain led to a 1.6-fold increased TAL titer to 4.2 g/L at 30% theoretical yield when using the ethanol feed. An additional fed-batch fermentation using a supplemented glucose / nitrogen feed as mentioned previously was also used to ensure the supplied carbon controlled the C/N ratio to 35:1. Using previous studies as a reference, this C/N ratio was implemented using an amino acid-based nitrogen source (see Materials and Methods) found to support high ethanol yield on glucose (Albers et al., 1996; Larsson et al., 1993; Manikandan et al., 2010). This strain achieved 28 g/L DCW biomass and 10.4 g/L TAL titer (38% theoretical yield), a nearly 5-fold increase relative to wildtype 2-PS expressed in the same strain by glucose fed-batch fermentation (Cardenas and Da
Silva, 2014) (Figure 5.7). This change was primarily due to the extended time during which glucose was consumed, and subsequent boost in biomass formation. From these results, further optimization could be designed to suppress biomass formation in favor of product formation.

**Discussion**

Our approach to engineering TAL production in S. cerevisiae coupled both in vivo titer analysis and in vitro enzyme kinetics that led to the discovery of 2-PS mutations that were not predicted to change TAL production. While alteration of the active site of 2-PS did increase TAL production as well as alter the kinetic properties, the mutation of surface cysteine residues distal from the active site influenced the in vivo and in vitro properties of 2-PS. Further modification of the yeast host boosted titers to the highest TAL titers observed thus far. Furthermore, utilization of these newly identified 2-PS mutants were used in tandem with yeast metabolic modifications, leading to the highest TAL production levels to date.

In vitro analysis of 2-PS and its mutants was complicated by two factors. First, decarboxylation of malonyl CoA to acetyl CoA made kinetic analysis difficult, and it is unclear whether 2-PS mutants decarboxylate malonyl CoA at the same rate as wildtype. Second, oxidation of the active site cysteine, which was observed in the crystallization of 2-PS (Jez 2002), may occur at different rates between the mutant enzymes, which would alter the kinetic profile. Nevertheless, apparent kcat and Km values were calculated to compare 2-PS mutants. Interestingly, [C35SC372S] proved to be the most effective mutant. This could be due to the potential tolerance of this enzyme to oxidative inactivation taking place at either the active site or through protein – protein disulfide bond formation.

Following in vivo yeast cultivations and evaluating TAL production, there was a clear distinction between those variants that had a positive effect on 2-PS and those that decreased enzyme activity. The observation that cysteine residues would be critical in the stability of 2-PS in yeast was not expected. This study demonstrated the importance of surface cysteine residues and the preferential substitution to a serine, a conservative substitution that retains similar steric and chemical properties. Observation of the crystal structure of 2-PS reveals that both C35 and
C372, the two cysteines associated with greatest TAL production, sit at highly solvent exposed positions, whereas all other cysteines are relatively more buried within the enzyme.

With respect to TAL production in protease-deficient strains, any correlation between cysteine changes and yeast protease activity was evaluated. The data suggests that improved stability in 2-PS variants was consistent with increased TAL production in strains harboring either partially or fully active protease systems. Owing to the oxygen-independent inactivation and oxidation of free thiols on the 2-PS surface and the loss in TAL levels when utilizing cysteine-rich variants is highly suggestive of targeted degradation, potentially from the native glutathione machinery in yeast. As suggested previously (Ohmori et al., 1999), S. cerevisiae strains, regardless of oxygen level, exhibit similar activity of enzymes pertaining to glutathione consumption, one of two systems proposed to protect against oxidative stress (Meister and Anderson, 1983; Spector et al., 1988). With the presence of these systems, the opportunity to support enzyme stability could prove useful for many other enzymes which harbor similar cysteine residues, including the related naringenin-producing chalcone synthase (7 cysteines per monomeric subunit) and Type I PKS 6-methylsalicylic acid synthase (17 cysteines).

We hypothesize that activity both in vitro and in vivo is altered by two factors. First, based on time-course stability assays, shrinking the active site with bulky residues may modulate active site oxidation, supported by the fact that [L268F] showed a greater half-life than wildtype. Second, overall enzyme stability and activity is likely modulated by the oxidation state of surface cysteine residues. This is illustrated by the contrasting results of the [C35A] and [C35S] mutants, which showed increased stability over WT, with [C35S] being capable of significant TAL titer increases in yeast but not [C35A]. By increasing TAL production using exposed cysteine residues, as well as observing a preference for nonpolar alanine substitutions at buried positions, these results are highly suggestive for an in vivo mechanism impacting 2-PS performance.

Conclusion

The increasing availability and almost routine use of sequencing is rapidly expanding the repertoire of novel enzymes for the manufacture of enzymes and attractive metabolites. In this
study, an enzyme and metabolic engineering approach was used to increased productivity for the model target triacetic acid lactone in S. cerevisiae. Through in vitro characterization, novel 2-PS variants were constructed using steric modulation to reduce cavity size in the active site in order to prevent long intermediate extensions and to limit promiscuity in the loading moieties. While there was overlap between in vitro and in vivo results, significant changes in TAL production were observed when in vitro data did not predict this behavior. Further investigation led to the understanding that surface-exposed cysteine residues are highly reactive and targeted residues in the microbial host. In addition, the C35 position was identified as not only improving $k_{cat}/K_m$ of the 2-PS, but also was critical in the stability of both unmodified as well as cysteine-deficient 2-PS variants in vivo. Combing a strong TAL-producing variant [C35S,C372S] with the engineered BJ5464Δpyc2Δnte1 strain led to the high level synthesis of triacetic acid lactone. Following fed-batch glucose fermentation, over 10 g/L TAL was produced at 38% yield, the highest demonstration published to date. The modifications evaluated in this work using 2-PS and the biosynthesis of TAL should stand as a model for other polyketide synthase systems.
**Figure S5.1:** TAL production by 2-PS with and without the addition of acetyl-CoA. While malonyl CoA can spontaneously decarboxylate to form acetyl CoA to give background TAL formation activity, the addition of acetyl CoA (1 µM and 60 µM here) prior to reaction initiation shows a measurable increase in TAL production.
Figure S5.2: Purification of 2-PS mutants. A band at approximately 42 kDa was present for all mutants, but some mutants are more soluble than others. Mutants 8-11 were not sufficiently soluble to obtain activity data.
Figure S5.3: Thermal denaturation of 2-PS mutants. Raw data of enzymes melted at 10 μM protein and 10X SYPRO dye in 50 mM potassium phosphate, pH 7.0.
Table S5.1: Melting temperatures ($T_m$) of mutants purified by affinity chromatography. Melting temperatures are the average of two duplicates.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35S372S</td>
<td>58.84</td>
</tr>
<tr>
<td>35S202F</td>
<td>54.69</td>
</tr>
<tr>
<td>35S201M</td>
<td>61.79</td>
</tr>
<tr>
<td>35A</td>
<td>59.48</td>
</tr>
<tr>
<td>35S343F</td>
<td>55.06</td>
</tr>
<tr>
<td>35S</td>
<td>59.53</td>
</tr>
<tr>
<td>WT</td>
<td>60.68</td>
</tr>
<tr>
<td>35S202F268F</td>
<td>42.08</td>
</tr>
<tr>
<td>6xCys</td>
<td>45.18</td>
</tr>
<tr>
<td>35S137F</td>
<td>46.18</td>
</tr>
<tr>
<td>5xCys</td>
<td>45.23</td>
</tr>
<tr>
<td>268F</td>
<td>46.18</td>
</tr>
<tr>
<td>35S268F</td>
<td>54.48</td>
</tr>
</tbody>
</table>
References


Chapter 5, in full, is currently being prepared for submission for publication of the material as Vickery, Christopher R.; Cardenas, Javier; Burkart, Michael D.; Da Silva, Nancy A.; Noel, Joseph P. “A combined approach for engineering a heterologous Type III PKS and enhancing polyketide biosynthesis in S. cerevisiae.” The dissertation author was a shared primary investigator and author of this material.
Chapter 6: Uncovering enzymatic properties that control aromatic prenyltransferase specificity

Introduction

Prenylated small molecules are an essential aspect of both primary and specialized metabolism in a variety of organisms, including bacteria, fungi, and plants. Aromatic prenyltransferases (aPTases) transfer a mevalonate pathway-derived prenyl diphosphate molecule onto a receiving phenolic aromatic compound. Many prenylated compounds derived from plants are commercially and industrially valuable, and are currently used as therapeutics and intoxicants (Figure 6.1). Plant prenyltransferases are often membrane-bound enzymes, and are generally difficult to characterize in vitro. Conversely, bacterial and fungal aPTases involved in specialized metabolism are soluble, monomeric enzymes. Specifically, these aPTases adopt an αββα fold that has been characterized extensively. Previous work on fungal aPTases has demonstrated their promiscuous nature with respect to the prenyl acceptor molecule. However, aPTases are often specific with respect to the size of the prenyl molecule transferred onto the prenyl acceptor. Currently, there exists a single example of an aPTase that is flexible with respect to the prenyl donor size. While the structures of several aPTases exists, the active site features that control prenyl donor chain length and prenyl acceptor specificity are not fully understood.

We have set out to uncover the features of aPTases that control all aspects of the prenyltransferase reaction. We first assessed the ability of a wide variety of aPTases, many of which that have been characterized structurally, to prenylate a more diverse set of molecules. From this, we can draw conclusions about the active site residues that control prenyl acceptor specificity across all aPTases. We then constructed a 96 member library of triple mutants of one aPTase in order to modulate its prenyl donor and prenyl acceptor specificity. Previous single-mutant analysis did not clarify the prenyl donor or prenyl acceptor specificity, and thus a combinatorial mutant library approach was explored. These studies will allow for modulation of
**Figure 6.1: Prenylation by aPTases.** (A) Prenylation reaction catalyzed by aPTases. Depicted is the prenylation of naringenin with dimethylallyl pyrophosphate DMAPP to produce 8-prenyl naringenin and the byproduct pyrophosphate.
aPTase activity via active site mutation, and pave the way for designer aPTases that can specifically prenylate a given molecule with differently sized prenyl groups.

**Results**

A panel of 28 phenolic compounds was tested for prenylation by the 4 aPTases NphB, CloQ, SCO7190 (HypSc), and ppzP (Figure 6.2). Activity was assessed by observing UV spectra and SIM mass spectra via LC-MS (table 6.1). The compound library included phenylpropanoid molecules of plant origin and other phenolic compounds that to date have not been assayed with aPTase enzymes. All activity with NphB was assessed with geranyl pyrophosphate (GPP) and prenyl acceptor, while CloQ, HypSc, and ppzP were assayed with dimethylallyl pyrophosphate (DMAPP) and prenyl acceptor. In general, each PTase tended to prenylated molecules that roughly resembled their natural substrate. Initial experiments were qualitative only, and were not designed to quantify the activity of each aPTase against the tested substrates.

The byproduct of the prenylation reaction, pyrophosphate (PPi), was tested for inhibition of NphB during prenylation of compound 28. A concentration range from 0 to 1 mM PPi was added to the prenylation reaction. Interestingly, no decrease in activity was observed, even at high concentrations of PPi. Therefore, while PPi most likely binds to NphB in some capacity, is not an inhibitor of prenylation activity.

NphB was analyzed via visual inspection of the crystal structure to determine residues putatively involved in binding of the GPP prenyl donor and the 1,6 DHN prenyl acceptor. Simultaneously, a phylogenetic comparison study was performed by DNA2.0 (Menlo Park, CA) to identify residues that remained conserved across aPTase sequences from different species. A library of 96 triple mutants was constructed that combined mutations of the residues identified by each analysis method (Table 6.1).
Figure 6.2. Compounds assessed for prenyl donor activity.
Table 6.1: *Results of prenylation reactions*. N.D. indicates the substrate was not tested with the indicated aPTase. INC. Denotes that the result was inconclusive, and further analysis is required to determine prenylation activity.

<table>
<thead>
<tr>
<th>COMPOUND</th>
<th>ENZYME</th>
<th>NphB</th>
<th>CloQ</th>
<th>HypSc</th>
<th>ppzP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>N.D.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>N.D.</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>N.D.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>No</td>
<td>INC.</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Yes</td>
<td>Inc.</td>
<td>Yes</td>
<td>INC.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>INC.</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>INC.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>INC.</td>
<td>INC.</td>
<td>INC.</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>No</td>
<td>INC.</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>No</td>
<td>No</td>
<td>INC.</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>N.D.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>N.D.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>N.D.</td>
<td>INC.</td>
<td>INC.</td>
<td>INC.</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
Mutants was expressed at a 40 mL scale and affinity purified with Ni-NTA resin. Each mutant was analyzed by SDS-PAGE to assess expression and solubility (Figure S6.1). 41 of the 96 mutants have been successfully purified. However, 24 of the mutants either expressed insoluble or no detectable protein. The remaining mutants have yet to be expressed.

25 of the 96 mutants were assayed for thermal stability (Figure S6.2) prenylation activity. Initial activity studies aimed to assess whether the mutants retained the WT activity of prenylating 11 with GPP (Figure 6.3). From these results, only six mutants retained WT activity: C2, D4, E1, E7, F2, and H7.

Discussion

While previous studies on aPTases has focused on prenylation of native substrates and some plant phenylpropanoids, we expanded the scope of in vitro prenylation to include a wider variety of prenyl acceptors, including the antibiotic tetracycline and the glycosylated molecule aloin, a component of Aloe vera extract. While some molecules tested were not prenylated by any enzymes, it was observed that each of the 4 prenyltransferases assayed often prenylated molecules that resemble their native substrates. Both NphB and HypSc has similar prenylation profiles, with activity detected for most planar, aromatic substrates. CloQ, which normaly prenylates hydroxyphenylpyruvic acid, seemed to prefer charged, non-planar molecules. PpzP was the only aPTase to successfully prenylate aloin, an athroquinone glycoside, which resembles the native substrate of ppzP, dihydrophenazine carboxylate. However, in our current study, the position of prenylation was not determined. However, as observed previously, some reactions contained more than one peak corresponding to the correct mass of the prenylated product. This was previously attributed to a mixture of C- and O-prenylation.

In order to better understand the enzyme properties that control prenylation, we constructed a 96 member mutant library of NphB. Based on our own observation of the previously solved crystal structure and phylogenetic analysis by DNA2.0, a series of triple mutants was constructed. Mutations were made in positions that are putatively involved in binding the prenyl donor, binding the prenyl acceptor, and stabilizing the protein interally or at the surface (Figure
6.3). While single-mutant analysis has been performed previously (unpublished data), these were not sufficient to provide information about the prenylation mechanism. These mutants were subsequently transformed and expressed in *E. coli*. Some mutants proved to be insoluble (Figure S6.1), which also provides information about residues that confer stability. After purification, the proteins were subjected to thermal denaturation to measure the melting temperature \( T_m \). The mutations exerted different effects on the \( T_m \) of NphB, with some mutants exhibiting profiles that contained multiple “melting events.” Additionally, many inactive mutants exhibited an increase in \( T_m \) when compared to WT. There was no clear correlation between enzyme solubility and \( T_m \).

Activity of 25 mutants was assessed with 11 as the target substrate and GPP as the prenyl donor. This substrate was chosen because of the high prenylation activity with NphB. Many mutants proved to be inactive towards 11, but six exhibited WT enzyme activity. These active mutants contained mutations around the active site, and could potentially modulate substrate specificity. Further analysis of these mutants using different prenyl donors and prenyl acceptors will provide information regarding the mechanistic “switches” imposed by the mutations of each member of the library. Exploring DMAPP as a prenyl donor and examining other prenyl acceptors will provide a more complete picture of the effect of the triple mutants on activity.

**Conclusion**

We have expanded the knowledge of aPTase substrate tolerance and mechanism through a combination of substrate screening and mutational analysis. We have shown that substrates for aPTases are more diverse than previously shown, with a variety of phenolic scaffolds serving as suitable prenyl acceptors for the panel of aPTases tested. Additionally, biophysical and biochemical analysis of NphB mutants provided valuable information about the importance of residues located in, around, and distant from the active site. Further analysis of the mutant library and structural characterization of mutants possessing interesting
Figure 6.3. *Mutations made in NphB.* Highlighted are portions of the protein that contribute to prenyl acceptor binding (right) and prenyl donor binding (left).
biochemical properties will further contribute to our knowledge of aPTase structure-function relationships.

**Methods**

**Protein cloning and expression**

All aPTases were grown and expressed as previously described.\(^4\)\(^,\)\(^10\)\(^,\)\(^8\)\(^,\)\(^11\)

**Prenylation reaction**

Prenylation reactions were conducted at a 25 uL scale, and contained 25 mM HEPES pH 7.5, 10 mM prenyl acceptor, 1mM prenyl donor, 10 mM MgCl\(_2\), and 20 µg enzyme. Reactions were allowed to proceed overnight at room temperature. Reactions were quenched by addition of 25 µL acetonitrile. Precipitated protein and insoluble material was removed via centrifugation, and the resulting soluble supernatant was loaded onto a Shimadzu LCMS-2020 equipped with a 50 mm phenomenex kinetex C18 analytical column. Reactions were analyzed between the wavelengths of 210 nm and 500 nm. Additionally, SIM was employed to detect prenylated products by monitoring the mass of the parent compound and the mass of the prenylated reaction product. Reactions were compared to negative controls containing no prenyl donor.

**Growth and purification of mutants**

Triple-mutants of NphB were grown in 40 mL LB media, and induced at R.T. overnight with 1 mM IPTG when the O.D. at 600 nm reached 0.6. Cells were pelleted and lysed by addition of 5 mL B-PER solution (Thermo Fischer Scientific) supplemented with 50 mM Tris pH 7.5, 250 mM NaCl, and 10% glycerol. Lysates were batch-bound with 100 µL Nickel-NTA resin, washed with lysis buffer + 20 mM imidazole, and eluted with 50 mM Tris 7.5, 250 mM NaCl, 10% glycerol, and 250 mM imidazole. Imidazole was not removed from the purified protein prior to activity tests to avoid loss. The addition of 250 imidazole to the prenylation reaction did not have any detrimental effect (data not shown). Prenylation reactions were performed as described above.

**Thermal denaturation experiments**

A Roche Lightcycler 486 thermocycler was used for thermal denaturation experiments. 10X SYPRO orange dye was mixed with 1 µM protein in 50 mM Tris 7.5 and 250 mM NaCl in a
final volume of 20 µL. Temperature was increased at 0.06°C/sec from 25 to 85°C. Melting
temperature was determined as previously described.\textsuperscript{13}
Figure S6.1: Representative gel of NphB mutant library purification. Lanes are as follows: (1) protein standard (2) NphB standard (3-5) B5 total lysate, insoluble fraction, soluble, (6-8) B7 total lysate, insoluble fraction, soluble, (9-11) B8 total lysate, insoluble fraction, soluble, (12-14) B9 total lysate, insoluble fraction, soluble, (15) NphB standard.
**Figure S6.2:** $T_m$ of NphB mutants that were assessed for prenylation activity. Values are an average of two thermofluor $T_m$ measurements.
Table S6.1. List of NphB mutants.

<table>
<thead>
<tr>
<th>Variant # and Mutations</th>
<th>Code</th>
<th>Variant # and Mutations</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>V10_V49A_S177Y_C209G</td>
<td>a1</td>
<td>V54_Q161A_V271E_Q295W</td>
<td>e1</td>
</tr>
<tr>
<td>V11_K118N_C119A_V271E</td>
<td>a2</td>
<td>V55_V49S_Y216A_V294N</td>
<td>e2</td>
</tr>
<tr>
<td>V12_L174V_C230N_L298Q</td>
<td>a3</td>
<td>V56_F123A_M162F_S214G</td>
<td>e3</td>
</tr>
<tr>
<td>V13_K119D_S177E_L219F</td>
<td>a4</td>
<td>V57_C25V_A232S_V271E</td>
<td>e4</td>
</tr>
<tr>
<td>V14_A53E_E112G_R228E</td>
<td>a5</td>
<td>V58_K118Q_L174V_R228Q</td>
<td>e5</td>
</tr>
<tr>
<td>V15_A53E_F213M_R228Q</td>
<td>a6</td>
<td>V59_V49S_S214G_V294A</td>
<td>e6</td>
</tr>
<tr>
<td>V16_A53Q_S177W_L219F</td>
<td>a7</td>
<td>V5_Q161A_S214G_V294F</td>
<td>e7</td>
</tr>
<tr>
<td>V17_V49L_F123A_Y283L</td>
<td>a8</td>
<td>V60_E112D_K119A_N173D</td>
<td>e8</td>
</tr>
<tr>
<td>V18_C25V_Y121W_M162F</td>
<td>a9</td>
<td>V61_F123W_Q161W_D227E</td>
<td>e9</td>
</tr>
<tr>
<td>V19_V49L_S214R_V271E</td>
<td>a10</td>
<td>V62_A53T_N173D_S214R</td>
<td>e10</td>
</tr>
<tr>
<td>V1_A108G_Q295N_L298Q</td>
<td>a11</td>
<td>V63_F123W_M162F_C209G</td>
<td>e11</td>
</tr>
<tr>
<td>V20_D227E_C230N_Q295W</td>
<td>a12</td>
<td>V64_M106E_M162A_Y216A</td>
<td>e12</td>
</tr>
<tr>
<td>V21_G205L_Y216A_C230S</td>
<td>b1</td>
<td>V65_V49A_Q161S_V294A</td>
<td>f1</td>
</tr>
<tr>
<td>V22_Q161S_V294F_L298W</td>
<td>b2</td>
<td>V66_C25V_F213M_Y216A</td>
<td>f2</td>
</tr>
<tr>
<td>V23_L219F_Y283L_L298W</td>
<td>b3</td>
<td>V67_A108G_K119D_L298A</td>
<td>f3</td>
</tr>
<tr>
<td>V24_A17T_F213M_S214R</td>
<td>b4</td>
<td>V68_K118N_C209G_R228Q</td>
<td>f4</td>
</tr>
<tr>
<td>V25_L219F_V294N_Q295A</td>
<td>b5</td>
<td>V69_A53T_M106E_Q161S</td>
<td>f5</td>
</tr>
<tr>
<td>V26_A53E_A108G_K118N</td>
<td>b6</td>
<td>V6_F123W_S177W_Q286E</td>
<td>f6</td>
</tr>
<tr>
<td>V27_Q38G_E112G_M162F</td>
<td>b7</td>
<td>V70_Q38G_D166E_Q295A</td>
<td>f7</td>
</tr>
<tr>
<td>V28_A53T_D166E_Q295W</td>
<td>b8</td>
<td>V71_M106E_G205L_C209G</td>
<td>f8</td>
</tr>
<tr>
<td>V29_L174V_C230S_V294A</td>
<td>b9</td>
<td>V72_E112G_G205M_L298W</td>
<td>f9</td>
</tr>
<tr>
<td>K119A_Q161S_S177E</td>
<td>b10</td>
<td>V73_V49S_K118Q_S177E</td>
<td>f10</td>
</tr>
<tr>
<td>M106E_G286E_L298W</td>
<td>b11</td>
<td>V74_M106E_Y121W_D166E</td>
<td>f11</td>
</tr>
<tr>
<td>D227E_R228E_L298Q</td>
<td>b12</td>
<td>V75_A53Q_L274V_Q295A</td>
<td>f12</td>
</tr>
<tr>
<td>M162A_C209G_Y288H</td>
<td>c1</td>
<td>V76_V49A_F123A_Y288H</td>
<td>g1</td>
</tr>
<tr>
<td>A17T_C25V_E112G</td>
<td>c2</td>
<td>V77_C25V_F123H_Q295N</td>
<td>g2</td>
</tr>
<tr>
<td>A53Q_Y121W_A232S</td>
<td>c3</td>
<td>V78_K119D_Q161W_L298Q</td>
<td>g3</td>
</tr>
<tr>
<td>A53Q_S177Y_Y288H</td>
<td>c4</td>
<td>V79_V49A_Y121W_C230S</td>
<td>g4</td>
</tr>
<tr>
<td>F123H_L274V_L298A</td>
<td>c5</td>
<td>V7_K118N_N173D_Q295N</td>
<td>g5</td>
</tr>
<tr>
<td>S177W_S214F_Y283L</td>
<td>c6</td>
<td>V80_M106A_N173D_S214F</td>
<td>g6</td>
</tr>
<tr>
<td>A53E_G205M_L274V</td>
<td>c7</td>
<td>V81_V49L_D166E_L274V</td>
<td>g7</td>
</tr>
<tr>
<td>A53T_K118N_S214F</td>
<td>c8</td>
<td>V82_V49S_K118D_F213M</td>
<td>g8</td>
</tr>
<tr>
<td>V49S_M162A_Y283L</td>
<td>c9</td>
<td>V83_E112D_L219F_V294F</td>
<td>g9</td>
</tr>
<tr>
<td>S177E_S214R_R228E</td>
<td>c10</td>
<td>V84_F123H_L174V_S177E</td>
<td>g10</td>
</tr>
<tr>
<td>K118Q_S177Y_D227E</td>
<td>c11</td>
<td>V85_E112G_F123H_S214G</td>
<td>g11</td>
</tr>
<tr>
<td>D166E_S177Y_S214F</td>
<td>c12</td>
<td>V86_F123A_S177W_Q205L</td>
<td>g12</td>
</tr>
<tr>
<td>Q161A_M162F_Q295A</td>
<td>d1</td>
<td>V87_S177W_Y288H_V294N</td>
<td>h1</td>
</tr>
<tr>
<td>A53E_Q161A_V294N</td>
<td>d2</td>
<td>V88_A108G_Q161S_G205M</td>
<td>h2</td>
</tr>
<tr>
<td>A108G_F123W_Q295A</td>
<td>d3</td>
<td>V89_Y121W_S177Y_Q286E</td>
<td>h3</td>
</tr>
<tr>
<td>A53Q_G205L_Q295N</td>
<td>d4</td>
<td>V90_K119A_Q161A_R228Q</td>
<td>h4</td>
</tr>
<tr>
<td>K118Q_F123A_R228E</td>
<td>d5</td>
<td>V90_A17T_F123W_L298A</td>
<td>h5</td>
</tr>
<tr>
<td>V49L_E112D_G295E</td>
<td>d6</td>
<td>V91_N173D_F213M_V294F</td>
<td>h6</td>
</tr>
<tr>
<td>G205L_R228E_C230N</td>
<td>d7</td>
<td>V92_A53T_E112D_G205M</td>
<td>h7</td>
</tr>
<tr>
<td>K118Q_Q161W_S214F</td>
<td>d8</td>
<td>V93_Q38G_L174V_S214R</td>
<td>h8</td>
</tr>
<tr>
<td>M162A_D227E_C230S</td>
<td>d9</td>
<td>V94_A17T_V49A_C230N</td>
<td>h9</td>
</tr>
<tr>
<td>V49L_K119D_G205M</td>
<td>d10</td>
<td>V95_A17T_Q161W_A232S</td>
<td>h10</td>
</tr>
<tr>
<td>K119A_S214G_L298A</td>
<td>d11</td>
<td>V96_Q161W_R228Q_Q295W</td>
<td>h11</td>
</tr>
<tr>
<td>Q38G_Y216A_A232S</td>
<td>d12</td>
<td>V9_Q38G_E112D_F123H</td>
<td>h12</td>
</tr>
</tbody>
</table>
References


