Copyright

by

Christopher Dane Fage

2014

The Dissertation Committee for Christopher Dane Fage Certifies that this is the approved version of the following dissertation:

Unveiling the Architectures of Five Bacterial Biomolecular Machines

Committee:

Adrian Keatinge-Clay, Supervisor

David Hoffman

Dean Appling

Christian Whitman

Brent Iverson

Marvin Hackert

Unveiling the Architectures of Five Bacterial Biomolecular Machines

by

Christopher Dane Fage, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2014

For Mom and Dad.

Acknowledgements

First and foremost, I would like to acknowledge those who inspired me to pursue this journey—those excellent educators, such as professors Craig Bieler, Jessica Tischler, Brian Buffin, Charles Dunlop, Robert Bix, Mohamed Daassa, Joseph Rahme, and Jie Song, who did not simply bombard me with facts, but instilled in me a thirst for knowledge and deep sense of wonder. I thank Jianting Zheng, David Hoffman, Borries Demeler, Eta Isiorho-Mansoorabadi, Hung-wen Liu, Yung-nan Liu, Jennifer Brodbelt, Joseph Boll, Joe Cannon, Stephen Trent, Jeremy Henderson, and Dusty Brown for fruitful collaborations. I am grateful to Art Monzingo for his assistance with all aspects of crystallography. I thank Jessica Momb for preserving my sanity during quals and allowing me to float ideas. I thank Empiricist and associated members for serving as a constructive outlet. I am indebted to Adrian Keatinge-Clay for being an incredibly patient and encouraging mentor and role model in my transformation into a research scientist. I am thankful to the dysfunctional family that is the Keatinge-Clay Group for their valuable feedback and comic relief: Darren Gay, Drew Wagner, Constance Bailey, Shawn Piasecki, Amanda Hughes, Eta Isiorho-Mansoorabadi, Andrew Harper, Jianting Zheng, and others. Lastly, I thank Mom, Dad, Lou, Sal, Maggie, Zula, and the rest of my family for all of their love and support.

Unveiling the Architectures of Five Bacterial Biomolecular Machines

Christopher Dane Fage, PhD

The University of Texas at Austin, 2014

Supervisor: Adrian Keatinge-Clay

Natural products represent an incredibly diverse set of chemical structures and activities. Given this fathomless, ever-evolving diversity, a reasonable approach to designing new molecules entails taking a closer look at the biochemistry that Nature has crafted over billions of years on Earth. In particular, much can be learned by unveiling the architectures of proteins, life's molecular machines, through methods like X-ray crystallography. Acquiring the blueprints of an enzyme brings us closer to understanding the mechanism by which the enzyme transforms a simple substrate it into a complex product with biological function, and inspires us to engineer such systems to our own ends.

With a focus on macromolecular structural characterization, this document elaborates on five Gram-negative bacterial biosynthetic enzymes from two categories: Cell-surface modifiers and polyketide synthases. Among the first category are the glycyl carrier protein AlmF and its ligase AlmE of *Vibrio cholerae* and the phosphoethanolamine transferase EptC of *Campylobacter jejuni*. These proteins are responsible for decorating cell-surface molecules (e.g., lipid A) of pathogenic bacteria with small functional groups to promote antibiotic resistance, motility, and host colonization. AlmE and EptC represent potential drug targets and

their structures lay the groundwork for the design of therapeutics against food-borne illnesses. Included in the second category are the [4+2]-cyclase SpnF and two ketoreductase-linked dimerization elements, each from the spinosyn biosynthetic pathway in *Saccharopolyspora spinosa*. The former catalyzes a putative Diels-Alder reaction to form a tricyclic precursor of the insecticide spinosad, while the latter two organize ketoreductase domains within modules of a polyketide synthase. The second category also includes *Ralstonia eutropha* β-ketoacyl thiolase B, a substrate-permissive enzyme that can make or break carbon-carbon bonds with assistance from Coenzyme A or an analogous thiol. Each of these proteins exhibit intriguing structural features or catalyze reactions that show promise for biochemical engineering.

Table of Contents

List of Tables xiii

List of Figures xv

Chapter 1: Introduction 1

1. Bacterial Cell-surface Biosynthetic Enzymes 1

2. Bacterial Polyketide Biosynthetic Enzymes 7

Chapter 2: Antimicrobial Peptide Resistance of *Vibrio cholerae* Results from LPS Modification Pathway Related to Nonribosomal Peptide Synthetases . 17

Abstract 17

Introduction 17

Results and Discussion 21

AlmF is phosphopantetheinylated by Vc2457 21

AlmE adds glycine to the AlmF carrier protein. 24

AlmE has relaxed substrate selectivity for amino acids other than glycine 25

Overall structure of AlmE 26

Residues contacting the glycyl moiety 28

AlmE variant Cys316Ala can efficiently adenylate and transfer L-alanine to AlmF 29

Conclusion 31

Materials and Methods 34

Bacterial growth 34

Recombinant DNA techniques 34

Cloning, expression, and purification of N-terminal His_6-AlmF or co-expression constructs 35

Destabilizing-urea PAGE of purified AlmF samples 36

UVPD-MS analysis of tryptic digested AlmF samples 37

Cloning, expression, and purification of N-terminal His_6-AlmE or C316 & L248 single amino acid variants 38

In vitro AlmE assays 39
Crystallization of AlmE 39
X-ray data collection, processing, phasing, and refinement 40
Polymyxin B and polymyxin E minimal inhibitory concentration assay 40
Matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS of lipid A 41
Acknowledgements 41

Chapter 3: Crystallographic Study of the EptC Phosphoethanolamine Transferase Required for Polymyxin Resistance and Motility in *Campylobacter jejuni* 63
Abstract 63
Introduction 64
Results and Discussion 66
Overall structure of the catalytic domain of EptC 66
Comparison of cEptC with known phospho-form transferase structures 68
Active-site architecture of cEptC 70
Mutational characterization of conserved cEptC active-site residues 71
Conclusion 74
Materials and Methods 75
Strains and plasmid construction 75
Expression, purification, and crystallization of cEptC 75
X-ray diffraction data collection, processing, phasing, and refinement 76
Protein mass spectrometry 78
Motility assays 78
Polymyxin B resistance assays 78
Acknowledgements 79

Chapter 4: The [4+2]-Cyclase SpnF at 1.50-Å Resolution 97
Abstract 97
Introduction 97
Results and Discussion 100
SpnJ and SpnM prepare the macrolactone for [4+2]-cyclization 100
Overall structure of SpnF 101
Substrate and product docking 103
Mutagenesis 104
The cycloaddition reaction 105
Materials and Methods 106
Materials 106
Cloning 107
Protein expression and purification 107
Crystallization, data processing, and structure determination ... 108
SpnF activity assay 109
SpnF mutants 109
Energy minimization 111
Consensus docking 111
Acknowledgements 112
Chapter 5: The Missing Linker: A Dimerization Motif Located within Polyketide Synthase Modules 128
Abstract 128
Introduction 128
Results and Discussion 130
Crystal structure of a DE+KR dimer 130
NMR analysis of a standalone DE 132
Contribution of DE to dimerization 134
Modeling a β-Module 135
Materials and Methods 137
Cloning, expression, and purification 137

Crystallization and structure determination 138
NMR spectroscopy 139
Gel filtration chromatography 140
Analytical ultracentrifugation 141
Modeling a β-module 141
Acknowledgements 142

Chapter 6: The Structure and Activity of a Substrate-permissive β-ketoacyl Thiolase from *Ralstonia eutropha* 154
Abstract 154
Introduction 155
Results and Discussion 157
Intial diffraction analysis 157
Overall structure of *R. eutropha* β-ketoacyl thiolase B 159
Comparison of the binding pockets of *Re*BktB and other thiolases 161
Modelling of substrates and reaction coordinate 163
In vitro thiolytic cleavage of β-ketoacyl-*S*-NACs 166
Conclusion 167
Materials and Methods 167
Cloning, expression, and purification *R. eutropha* β-ketoacyl thiolase B 167
Crystallization, data collection, processing, phasing, and refinement 168
Syntheses of NAC and β-ketoacyl-S-NACs 169
In vitro thiolytic cleavage of β-ketoacyl-S-NACs 169
Acknowledgements 171

Chapter 7: Conclusions and Future Perspectives 190

Appendix A: Introduction to X-ray Crystallography 193
Overview of X-ray Crystallography 193
Crystal Lattices 195

Bragg's Law 197
Fourier Sums and Transforms 199
Calculating Phases 202
References 221

List of Tables

Table 2-1: Crystallographic data and refinement statistics for AlmE 59

Table 2-2: Minimal inhibitory concentration (MIC) determination 60

Table 2-3: Strains and plasmids utilized in this study 61

Table 2-4: Primers utilized in this study 62

Table 3-1: Bacterial strains and plasmids. 94

Table 3-2: Primers used in this study. 95

Table 3-3: Data-collection and refinement statistics for cEptC 96

Table 4-1: Crystallographic data and refinement statistics for SpnF 119

Table 4-2: All AutoDock Vina outputs for SpnF substrate energy-minimized with UFF force field 120

Table 4-3: All DOCK6 outputs for SpnF substrate energy-minimized with UFF force field. 121

Table 4-4: All AutoDock Vina outputs for SpnF substrate energy-minimized with MMFF94s force field 122

Table 4-5: All DOCK6 outputs for SpnF substrate energy-minimized with MMFF94s force field 123

Table 4-6: All AutoDock Vina outputs for SpnF product energy-minimized with UFF force field 124

Table 4-7: All DOCK6 outputs for SpnF product energy-minimized with UFF force field 125

Table 4-8: All AutoDock Vina outputs for SpnF product energy-minimized with MMFF94s force field 126

Table 4-9: All DOCK6 outputs for SpnF product energy-minimized with MMFF94s force field 127

Table 5-1: Crystallization data and refinement statistics for Spn(DE+KR)3 153

Table 6-1: Data-collection and refinement statistics for *Re*BktB..... 187

Table 6-2: Estimation of the twin fraction, α, as calculated by *phenix.xtriage* of *PHENIX*..... 188

Table 6-3: Improvement of residual factors after intensity-based twin refinement of the 2.0-Å dataset in *REFMAC5*..... 189

Table A-1: The 65 possible space groups in macromolecular crystallography 220

List of Figures

Figure 1-1: A dominant glycoform of *Escherichia coli* K-12 lipopolysaccharide (LPS)..... 13

Figure 1-2: Representative polyketides produced by processive Type I PKSs 14

Figure 1-3: Synthesis of the polyketide antibiotic erythromycin A by 6-dEB synthase (DEBS) and associated tailoring enzymes..... 15

Figure 1-4: Polyketides derived from PKS classes other than processive Type I 16

Figure 2-1: Proposed model for the lipid A glycine modification pathway in *V. cholerae*..... 42

Figure 2-2: Structure-sequence analysis of *V. cholerae* AlmF and related carrier protein domains 43

Figure 2-3: Due to low protein yields of recombinant His_6-AlmF, even when over-expressed in *E. coli*, a two-step purification of His_6-AlmF was required: Ni-sepharose (Ni) followed by size-exclusion (SE) 44

Figure 2-4: Co-expression of Vc2457 (VcAcpS) with AlmF (apo-AlmF) in *E. coli* results in the production of phosphopantetheinylated AlmF (holo-AlmF) as assessed by destabilizing urea-PAGE..... 45

Figure 2-5: Incubating holo-AlmF with AlmE does not result in a shift in the destabilizing urea-PAGE migration assay 46

Figure 2-6: Stacked comparison of the base peak chromatograms for reverse-phase HPLC chromatographic profiles of trypsin-digested AlmF samples as analyzed in **Figure 2-7** 47

Figure 2-7: UVPD-MS/MS confirms AlmF phosphopantetheinylation by Vc2457 and subsequent glycylation by AlmE when co-expressed in *E. coli* 48

Figure 2-8: Expanded region of UVPD mass spectrum to confirm phosphopantetheinylation of AlmF 49

Figure 2-9: Mass spectra of AlmF species 50

Figure 2-10: Relative to reactions where Coenzyme A (CoA) or apo-AlmF are provided as aminoacyl acceptors, providing holo-AlmF results in a robust increase in AlmE adenyltransferase activity under our standardized pyrophosphatase-coupled assay conditions 51

Figure 2-11: AlmE can efficiently adenylate glycine and to a lesser extent D-alanine. 52

Figure 2-12: Crystal structure of AlmE with bound glycyl-AMP 53

Figure 2-13: Stereoview of the active sites of AlmE (green, with the glycyl-adenylate in yellow) and *Bacillus cereus* DltA (grey, with the D-alanyl-adenylate also in grey) superposed 54

Figure 2-14: Modeling the AlmE/AlmF complex 55

Figure 2-15: A rationally designed AlmE Cys316Ala variant shows relaxed specificity for L-alanine 56

Figure 2-16: Lipid A was isolated from the C6706 almE mutant (vc1579::Tn5) expressing the AlmE C316A variant 57

Figure 2-17: Sequence alignment of a key specificity-encoding region of glycine-specific A-domains (Rausch *et al.*, 2005), AlmE, and DltAs (*B. subtilis* and *B. cereus*) 58

Figure 3-1: Topology and biological functions of *C. jejuni* phosphoethanolamine transferase EptC 81
Figure 3-2: Structure of the C-terminal, catalytic domain of the pEtN transferase EptC (cEptC) from *C. jejuni* 82
Figure 3-3: Expression and isolation of cEptC proteins 83
Figure 3-4: Topology map of cEptC 84
Figure 3-5: Surface rendition of **(A)** cEptC, compared with **(B)** cLptA (PDB 4KAY) and **(C)** *B. subtilis* cLtaS (PDB 2W8D) 85
Figure 3-6: Phosphothreonine is located on the N-terminal end of the nucleophilic helix (α2) 86
Figure 3-7: Overall structures of pEtN transferases cEptC and cLptA, and *B. subtilis* lipoteichoic acid synthase cLtaS 87
Figure 3-8: Juxtaposed active sites of **(A)** cEptC, **(B)** cLptA (PDB 4KAY), and **(C)** *S. aureus* cLtaS (PDB 2W5S) 88
Figure 3-9: Minimized sequence alignment (Clustal Omega; Sievers & Higgins, 2014) of representative phospho-form transferases in the COG2194 and COG1368 groups, demonstrating conservation of active-site residues 89
Figure 3-10: Metal-binding sites of **(A)** cEptC, **(B)** cLptA (PDB 4KAY), and **(C)** *B. subtilis* cLtaS (PDB 2W8D) are conserved 90
Figure 3-11: Minimal inhibitory concentrations (MICs) of polymyxin B for *C. jejuni* EptC active-site mutants 91
Figure 3-12: EptC active-site mutant motility assays 92
Figure 3-13: Negative F_o-F_c electron density maps reveal radiation damage to the Zn remote-wavelength dataset. 93

Figure 4-1: Production of polyketides containing carbocyles 113

Figure 4-2: Reactions mediated by SpnF and accompanying tailoring enzymes 114

Figure 4-3: SpnF complexed with SAH at 1.50-Å resolution 115

Figure 4-4: Superposition of the two SpnF monomers..... 116

Figure 4-5: Consensus docking of **(A)** substrate and **(B)** product 117

Figure 4-6: Assays of SpnF mutants 118

Figure 5-1: Subunit architecture of the spinosyn PKS 144

Figure 5-2: DE+KR structure..... 145

Figure 5-3: The pseudo-twofold symmetry of the DE dimer..... 146

Figure 5-4: NMR characterization of an isolated DE..... 147

Figure 5-5: Analytical ultracentrifugation 148

Figure 5-6: Size-exclusion chromatography of a DE+KR fragment 149

Figure 5-7: Analytical ultracentrifugation of Amp(DE+KR)2..... 150

Figure 5-8: Model of a PKS module 151

Figure 5-9: Sequence alignment of linkers 152

Figure 6-1: Thiolase reaction mechanisms 172

Figure 6-2: Representative diffraction image for the twinned, 2.0-Å dataset 173

Figure 6-3: *L* Test for acentric data, adapted from *phenix.xtriage* output... 174

Figure 6-4: Superposed **(A)** tetramers and **(B)** monomers from *Re*BktB and *Zr*BktB 175

Figure 6-5: Structure of *Re*BktB..... 176

Figure 6-6: Two views of the βαβαβ motif of an *Re*BktB monomer 177

Figure 6-7: Topology of an *Re*BktB monomer, with α-helices (maize) and β-strands (blue) labeled α1-15 and β1-13, respectively 178
Figure 6-8: *B*-factor heat map of chain A..... 179
Figure 6-9: Conserved CoA binding pocket of *Re*BktB and *Zr*BktB (in stereo) 180
Figure 6-10: *R. eutropha* BktB possesses unique sequences on helices α3 and α5 181
Figure 6-11: Active-site features of *Re*BktB (in stereo) 182
Figure 6-12: Models of acyl-*Re*BktB intermediates, guided by PDBs 1M4T and 1DM3 (in stereo)..... 183
Figure 6-13: Representative reverse-phase HPLC chromatograms (235 nm) for *Re*BktB-catalyzed β-ketoacyl-*S*-NAC thiolysis reactions 184
Figure 6-14: Standard curve of **6** 185
Figure 6-15: *In vitro* thiolytic cleavage of β-ketoacyl-*S*-NAC thioesters by *Re*BktB 186
Figure A-1: (130) planes in a crystal lattice..... 208
Figure A-2: The 7 possible crystal systems, with unit-cell angle and edge definitions..... 209
Figure A-3: The 4 lattice types: Primitive (*P*), body-centered or internal (*I*), side-centered (*C*), and face-centered (*F*)..... 210
Figure A-4: Section of the zero-level plane (*hk*0) from the diffraction data of selenomethionine-derived cEptC, which crystallized in space group $P2_12_12_1$ (see **Chapter 3**) 211
Figure A-5: Illustration of Bragg's law, $2d_{hkl}\,sin\theta=n\lambda$, which describes the requirements for diffraction..... 212

Figure A-6: Bragg's law in reciprocal space 213
Figure A-7: Vector representation of structure factors 214
Figure A-8: Harker diagram for isomorphous replacement 215
Figure A-9: Two-dimensional Patterson maps 216
Figure A-10: K absorption edge of zinc 217
Figure A-11: Vector representation of structure factors for Friedel pairs under anomalous conditions 218
Figure A-12: Harker diagrams for anomalous dispersion of Friedel mates 219

Chapter 1: Introduction

Under selective pressures, life has evolved an impressive network of metabolic pathways for the synthesis and degradation of countless natural products. Primary metabolic pathways, such as those involved in glucose oxidation, are essential and therefore ubiquitous across the tree of life. Secondary metabolic pathways, in contrast, are not strictly required for life but confer specific survival advantages in the organisms in which they emerge. For instance, the soil bacterium *Saccharopolyspora erythraea* utilizes a polyketide synthase pathway to generate the macrolactone erythromycin, an antibiotic that kills competing organisms by binding their ribosomes and bringing translation to a halt (Bulkley *et al.*, 2010). Man has already harnessed the inherent potential of many such natural products; however, if we wish to wage war with rapidly evolving pathogens or assemble new molecules with new activities, we clearly ought to further emulate those reactions so elegantly forged by Nature. After all, Nature has invested nearly 3.5 billion years developing the rich chemistry of life on Earth.

In order to fully appreciate how a biochemical reaction transpires on a physiological timescale, an accurate structural model of the catalyst (usually protein) is required, as the structure of an enzyme defines its function. Additionally, unveiling the architecture of an enzyme can, for example, highlight which atoms make critical contacts with a substrate or facilitate electronic rearrangements, in turn opening avenues for molecular engineering. Although a number of methods are at the disposal of the structural biologist, X-ray crystallography is by far the most widely utilized in enzyme structural analysis. Most auspiciously, this method provides a three-dimensional map of electron density into which atoms and bonds

can be accurately modeled at high resolution, essentially giving the crystallographer a snapshot of a low-energy state for the macromolecule of interest. (See Appendix A for a brief introduction to the principles of X-ray crystallography.)

With the theme of structure begetting function at the forefront, the current document tells the tale of several Gram-negative biosynthetic enzymes whose functions can be broadly divided into two categories: (1) functionalization of cell-surface molecules and (2) production of polyketide metabolites. The first category, which includes i) the *Vibrio cholerae* glycyl carrier protein AlmF and ligase AlmE and ii) the *Campylobacter jejuni* phosphoethanolamine transferase EptC, is relevant in understanding antibiotic resistance and pathogenesis in the organisms responsible for the food-borne illnesses cholera and campylobacteriosis. As described below, AlmE and EptC represent potential drug targets. The second category is comprised of i) the *Saccharopolyspora spinosa* [4+2]-cyclase SpnF and ii) two ketoreductase domains and their associated dimerization motifs, Spn(DE+KR)1 and Spn(DE+KR)3, which, like SpnF, are from the spinosyn biosynthetic pathway. Also included in the second category is iii) the *Ralstonia eutropha* β-ketoacyl thiolase B. The enzymes in this category were investigated for their interesting structural features and biosynthetic capacities.

1. Bacterial Cell-surface Biosynthetic Enzymes

In contrast to Gram-positive bacteria, Gram-negative bacteria envelope the peptidoglycan layer of their cell wall with an outer membrane. The outer leaflet of this membrane is fortified with a complex array of functionalized sugar residues referred to as lipopolysaccharide (LPS) (Raetz *et al.*, 2007; Trent *et al.*, 2006;

Needham & Trent, 2013). LPS consists of three domains (**Figure 1-1**): (1) lipid A, also known as endotoxin; (2) core oligosaccharide, divided into an inner and outer core; and (3) *O*-antigen (some bacteria assemble lipooligosaccharide, LOS, which lacks the *O*-antigen domain). The innermost domain, lipid A, is a disaccharide of glucosamine linked to phosphoryl groups and numerous β-hydroxyacyl chains of 10-16 carbons; this largely hydrophobic domain anchors LPS/LOS to the outer membrane. In addition to the Kdo sugars of the inner-core oligosaccharide, lipid A is necessary for bacterial growth and host immune recognition. While limited activation of the innate immune system via lipid A-TLR4/MD2 interaction is beneficial to the infected host, overactivation can lead to septic shock. Lipid A modifications are important for virulence, as they can affect outer-membrane permeability and limit binding of antimicrobials or host recognition machinery. Core oligosaccharide and *O*-antigen, the remaining LPS domains, endow resistance to host antimicrobials and complement. As one might expect, each domain of LPS/LOS is subject to extensive functionalization or truncation in order to safeguard the bacterium against unfavorable conditions.

Cationic antimicrobial peptides (CAMPs), like host defensins, disrupt the Gram-negative cellular envelope by associating with the anionic lipid A domain of LPS as well as acidic glycerophospholipids (Guilhelmelli *et al.*, 2013). For example, detergent-like aggregates of membrane-bound polymyxins form pores in the associated membrane, disrupting vital bacterial processes (Guilhelmelli *et al.*, 2013). In general, to evade CAMPs, Gram-negative and Gram-positive bacteria have evolved similar strategies to neutralize the net negative charge of major cell-surface molecules. Phosphatidylglycerols aminoacylated with L-lysine, L-alanine, or D-alanine were recently discovered and confer resistance toward CAMPs in a few

species of both Gram-type bacteria (e.g., *Staphylococcus aureus* and *Pseudomonas aeruginosa*) (Roy, 2009). Gram-positive bacteria commonly neutralize their cell walls by transferring D-alanine to surface teichoic acids, the long poly-phosphoribitol or poly-phosphoglycerol chains linked to the glycerophospholipids or *N*-acetylmuramic acid groups of peptidoglycan (Neuhaus & Baddiley, 2003); Gram-negative bacteria do so by transferring phosphoethanolamine or aminoarabinose to phosphoryl groups on the lipid A domain (Needham & Trent, 2013; Raetz *et al.*, 2007). Unexpectedly, aminoacyl esterification of glycine or diglycine to lipid A of the Gram-negative pathogen *Vibrio cholerae* has been observed (Hankins *et al.*, 2012). Lipid A glycylation is a unique Gram-negative strategy necessary for resistance to CAMPs, which resembles the Gram-positive method of modifying teichoic acids with D-alanine.

Chapter 2 of this document explores the roles of glycyl carrier protein AlmF and ligase AlmE in the resistance of O1 El Tor biotype *V. cholerae* to the CAMP polymyxin B. El Tor are responsible for the majority of 3-5 million cases and 100,000-120,000 deaths associated with cholera every year (World Health Organization). In addition, clinical use of polymyxins has emerged as the antibiotic strategy of last resort for infections with multi-drug resistant Gram-negative bacteria. AlmE, AlmF, and AlmG were shown to orchestrate the modification of lipid A with glycine or diglycine (Hankins *et al.*, 2012). Based on sequence homology, AlmE was proposed to activate glycine through adenylation and then transfer it to the 4′-phosphopantetheinyl group of carrier protein AlmF for delivery to glycyl-transferase AlmG. A unique hydroxylauryl chain observed in the lipid A domain from El Tor biotype *V. cholerae* serves as the site of glycine addition (Hankins *et al.*,

2011), making AlmEFG the only known lipid A charge remodeling system that does not involve direct transfer to the 1- or 4′- phosphoryl groups of lipid A.

We present the first detailed molecular mechanism of lipid A glycylation in *V. cholerae*. Mass spectrometry data verify that AlmF is a *bona fide* glycyl carrier protein and, moreover, that AlmE can transfer glycine to its 4′-phosphopantetheinyl group. *In vitro* characterization of AlmE reveals that it also accepts D-alanine, albeit with reduced activity compared to glycine. To further investigate this surprising result, we determined the X-ray structure of AlmE in complex with a glycyl-adenylate intermediate at 2.26-Å resolution. Visualizing the active site of AlmE allowed us to rationally design a single-residue mutant capable of efficiently activating and transferring L-alanine to AlmF. However, *in vivo* evidence suggests that L-alanyl-AlmF is not a suitable substrate for AlmG, indicating that the AlmEFG pathway is strictly specific toward glycine. Selection of glycine as the group AlmEFG adds to lipid A likely balances deleterious effects larger amino acids would have on membrane packing, while providing sufficient electropositive density to effectively resist CAMPs. Our findings provide molecular insight into a complex mechanism of CAMP resistance that incorporates the preexisting chemistry of nonribosomal peptide synthetases into new biosynthetic pathways.

Chapter 3 describes the catalytic domain of the phosphoethanolamine (pEtN) transferase EptC, which promotes polymyxin B resistance, motility, and host colonization in *Campylobacter jejuni*. *C. jejuni* is a Gram-negative enteric pathogen responsible, in large part, for food-borne illness and bacteria-associated diarrhea in the United States and abroad (Scallan *et al.*, 2011; Havelaar *et al.*, 2013). In terms of burden to health and cost, campylobacteriosis ranks highest among food-borne infections (Batz *et al.*, 2012). EptC is unique in that it modifies an array of cell-

surface molecules, including the 1- and 4′-phosphoryl groups of lipid A (Cullen & Trent, 2010), the first heptose residue of the inner-core oligosaccharide of LOS (Cullen *et al.*, 2013), residue T75 of the flagellar rod protein FlgG (Cullen *et al.*, 2012), and *N*-linked glycans of numerous glycoproteins (Scott *et al.*, 2012). Conversely, most mucosal pathogens appear to express several pEtN transferases each dedicated to the modification of one specific target. For example, pathogenic *Neisseria* species produce the enzymes LptA, Lpt3, Lpt6, and PptA that specifically transfer pEtN to only one of the following: lipid A phosphoryl groups (Cox *et al.*, 2003), position 3 of LOS core heptose II (Mackinnon *et al.*, 2002; Wenzel *et al.*, 2010), position 6 of LOS core heptose II (Wright *et al.*, 2004), or the major pilin protein PilE (Naessan *et al.*, 2008), respectively. Likewise, *E. coli* species possess at least three functional pEtN transferases (EptA, EptB, and CptA) that attach pEtN groups to specific positions of LPS (Raetz *et al.*, 2007). Due to its promiscuous nature, EptC is linked to multiple and distinct biological functions in *C. jejuni*. Mutational studies have demonstrated that loss of pEtN modifications cause drastic reduction in flagellar assembly and motility as well as significantly higher sensitivity to a range of cationic antimicrobial peptides (Cullen & Trent, 2010; Cullen *et al.*, 2012). Most importantly, EptC expression is required for proper intestinal colonization of *C. jejuni* in avian and mouse infection models (Cullen *et al.*, 2013).

Sequence analysis places EptC in a family of inner-membrane metalloenzymes containing a 5-helix transmembrane domain and a periplasmic catalytic domain bearing homology to sulfatases (Marchler-Bauer *et al.*, 2013). Placement of the catalytic domain dictates that pEtN modification occurs on the periplasmic side of the inner membrane, where EptC transfers head groups from the phospholipid phosphatidylethanolamine to target molecules. Indeed, no EptC

activity was observed when the enzyme was heterologously expressed in a phosphatidylethanolamine-deficient *E. coli* strain (Cullen & Trent, 2010). To gain insight into the structural basis of enzyme function, we crystallized and determined the atomic coordinates of the catalytic domain of EptC (cEptC) to a resolution of 2.40 Å. Our structural data along with mutational analysis of EptC allowed us to identify zinc-ligand residues, a putative nucleophile, and conserved active-site residues required for *in vivo* activity. Moreover, we juxtaposed the structures of cEptC and the catalytic domains of the related enzymes cLtaS (lipoteichoic acid synthase from *Bacillus subtilis* and *Staphylococcus aureus*) and cLptA (lipid A pEtN transferase from *N. meningitidis*). Observed structural differences around the active sites of cEptC and the lipid A-specific cLptA may provide clues to the promiscuous nature of EptC.

2. Bacterial Polyketide Biosynthetic Enzymes

Modular polyketide synthases (PKSs) are megadalton enzymatic assembly lines that biosynthesize complex polyketides including the antibacterial erythromycin, the antifungal amphotericin, and the insecticide spinosyn, from small coenzyme A derivatives such as malonyl- and methylmalonyl-CoA (**Figure 1-2**) (Khosla *et al.*, 2007; Smith & Tsai, 2007; Meier & Burkart, 2009; Keatinge-Clay, 2012). During the synthesis of a polyketide, each chain extension reaction is catalyzed by a separate module within the PKS (**Figure 1-3**). Thus, the three domains necessary for chain extension are present in all module types: an acyltransferase (AT) that selects an extender unit, a ketosynthase (KS) that catalyzes the condensation of the extender unit with a growing polyketide chain,

and an acyl carrier protein (ACP) that shuttles an extender unit or an extended polyketide chain between active sites of the synthase. Within modules, processing enzymes may also be present, such as a ketoreductase (KR) that reduces the β-keto group of the newly-extended polyketide chain to a β-hydroxy group, a dehydratase (DH) that removes the β-hydroxy group to form a double bond, and an enoylreductase (ER) that reduces the double bond. Finally, a thioesterase (TE) cleaves, and often cyclizes, the polyketide product from the synthase. Tailoring enzymes, typically isolated domains outside of the modular assembly line, may further functionalize the released polyketide with glycosyl, alkyl, or epoxide moieties, for example (Rix *et al.*, 2002).

Notably, unless stated otherwise, when this document discusses polyketide synthases and their products it refers to the modular, processive Type I variety described in the preceding paragraph. Such PKSs have been isolated from bacteria and protozoans. Other varieties of polyketide synthases include: iterative Type I, which employ a single module for many rounds of extension (common in fungi and some bacteria; e.g., lovastatin synthase); Type II, which are complexes of discrete domains (found in bacteria; e.g., doxorubicin synthase); and Type III, which utilize thioesters of CoA instead of acyl carrier proteins (common in plants and some fungi and bacteria; e.g., chalcone synthase) (Hertweck, 2009; Kennedy *et al.*, 1999; Hutchinson, 1997; Austin & Noel, 2003). Representative polyketides from each of these types are illustrated in **Figure 1-4**. PKS modules of all types can also integrate into hybrid assemblies, which can even include modules from the evolutionarily related nonribosomal peptide synthetases and fatty acid synthases.

Chapter 4 investigates the discrete [4+2]-cyclase SpnF from the spinosyn biosynthetic pathway of *S. spinosa*. This pathway is unusual in that it contains a

processive Type I PKS yet generates a carbocyclic product (in general, only polyketides produced by iterative Type I PKSs contain carbocycles) (Kirst, 2010; Crawford & Townsend, 2010; Hertweck *et al.*, 2007; Yu *et al.*, 2012; Keatinge-Clay, 2012). In the construction of the tetracyclic spinosyn core, a desaturated, 22-membered macrolactone is cross-bridged through two carbocyclization reactions (see spinosyn A structure in **Figure 1-2**). While the tailoring enzyme SpnL is believed to catalyze cyclization through a Rauhut-Currier mechanism (Aroyan *et al.*, 2009), the [4+2]-cyclization catalyzed by SpnF has been hypothesized to proceed through a Diels-Alder mechanism (Diels & Alder, 1929), in which an s-*cis*-configured diene is fused with a dienophile (e.g., an olefin) through an exergonic electrocyclization that converts two π-bonds into two σ-bonds (Kim *et al.*, 2011). Although other biosynthetic pathways (e.g., those of lovastatin, equisetin, solanapyrones, and spirotetronates) may contain [4+2]-cyclases, SpnF is the first enzyme to be reported whose sole function is to catalyze such a reaction (Kelly, 2008).

SpnF and SpnL resemble *S*-adenosylmethionine (SAM)-dependent methyltransferases (MTs), yet whether SAM participates in these cyclization reactions is unclear. The carbocyclization catalyzed by SpnF is hypothesized to proceed through a Diels-Alder mechanism, known to be facilitated by the participation of an electron-deficient dienophile. If SpnF indeed catalyzes a Diels-Alder reaction a feature that withdraws electron density from the dienophile may be within its active site. We present the 1.50-Å resolution structure of SpnF bound to *S*-adenosylhomocysteine (SAH). Through consensus docking of the reaction substrate and product, their putative binding modes were interrogated. Mutagenesis studies identified residues important to SpnF-mediated cyclization. The structural results

reported here set the stage for advanced experimental and computational approaches to determine the exact mechanism of cyclization.

Chapter 5 discusses the domain pairs Spn(DE+KR)1 and Spn(DE+KR)3, with focus on the DE (dimerization element) domains. Like the [4+2]-cyclase SpnF, these domains are from the spinosyn biosynthetic pathway of *S. spinosa*; unlike SpnF, however, these structural domains are embedded within the polyketide synthase and contribute to the dimeric architecture of their modules (Sieber *et al.*, 2002; Hur *et al.*, 2012). The only PKS enzyme that requires a dimeric oligomerization state is KS, and the only other enzyme that makes contact across the 2-fold axis of modules is DH (**Figure 1-3**) (Tang *et al.*, 2006; Keatinge-Clay, 2008; Akey *et al.*, 2010). The other domains commonly within modules, AT, KR, ER, and ACP, are monomeric and loop out from the 2-fold axis of the synthase. Accessory motifs have been identified that aid KS and DH in dimerizing the polypeptides of a PKS into its homodimeric subunits: a coiled coil is frequently located at the N-terminal end of a subunit, a four-helix bundle is often present at the C-terminal end, and a dimeric thioesterase (TE) that cleaves the elongated polyketide from the assembly line is usually positioned at the C-terminal end of the final PKS subunit (Broadhurst *et al.*, 2003; Buchholz *et al.*, 2009; Tsai *et al.*, 2001).

The ~55-residue DE is immediately N-terminal to the KR domain in ~50% of modules that contain KR as their sole processing enzyme (β-modules) (Keatinge-Clay, 2012). The 2.6-Å resolution crystal structure of a fragment from the third module of the spinosyn PKS that contains DE and KR, Spn(DE+KR)3, reveals the architecture of the DE dimer as well as a rigid connection between the last helix of DE and the KR domain. NMR analysis of a standalone DE from the first module of the spinosyn PKS is consistent with the crystal structure. Analytical ultracentrifugation

revealed its dimerization constant to be 19.9 ± 1.5 μM. By positioning the C-terminal ends of a structurally characterized KS+AT didomain in proximity with the N-terminal ends of the Spn(DE+KR)3 didomain, a model of a β-module was constructed. Thus, the biophysical characterization of DE has also helped to elucidate the domain organization of a common type of PKS module.

In **Chapter 6**, *Ralstonia eutropha* β-ketoacyl thiolase B (*Re*BktB) is analyzed. While not strictly a polyketide synthase, the β-ketoacyl thiolase is evolutionarily related to the ketosynthase domain that takes part in polyketide chain extension. Both enzymes are members of the condensing enzymes superfamily, whose constituents catalyze the Claisen-like condensation of two thioesters. Thiolase enzymes facilitate the formation or degradation of carbon-carbon bonds in a number of important biological pathways, such as fatty acid synthesis and β-oxidation. Recently, *Re*BktB activity was demonstrated on a range of substrates, including acetyl-, propionyl-, butyryl-, isobutyryl-, and glycolyl-CoA (Martin *et al*., 2013). Martin and colleagues incorporated the thiolase into an engineered biosynthetic route to generate a small panel of chiral, organic building blocks, such as the pharmaceutically important 3-hydroxy-γ-butyrolactone and its precursor 3,4-dihydroxybutyrate, in unprecedented titers. The lactone form of this molecule has served as a starting material in the syntheses of the statin drugs Crestor and Lipitor, the antibiotic Zyvox, and the antihyperlipidemic Zetia (Lee *et al.*, 2009).

In this document, we demonstrate thiolytic cleavage of β-ketoacyl-*S*-*N*-acetylcysteamine substrates by BktB from *Ralstonia eutropha*, with a preference for 3-oxopentanoyl-*S*-NAC when challenged with 3-oxobutanoyl-, 3-oxopentanoyl-, 3-oxohexanoyl-, or 3-oxoheptanoyl-*S*-NAC (acyl- and β-ketoacyl-*S*-NAC thioesters are short, inexpensive mimics of the natural CoA substrates). We also show that the α-

substituted (2*RS*)-2-methyl-3-oxopentanoyl-*S*-NAC is not a suitable substrate for the enzyme. In addition, we present two crystal structures of the thiolase: One, untwinned at 2.8-Å resolution and the other, twinned at 2.0-Å resolution. This glimpse into the active site provides a structural basis for the relatively broad substrate selectively of the thiolase. Our work also lays the foundation for protein engineering efforts to further broaden the substrate selectivity of thiolases, ultimately opening new avenues to valuable chiral molecules from common starting materials.

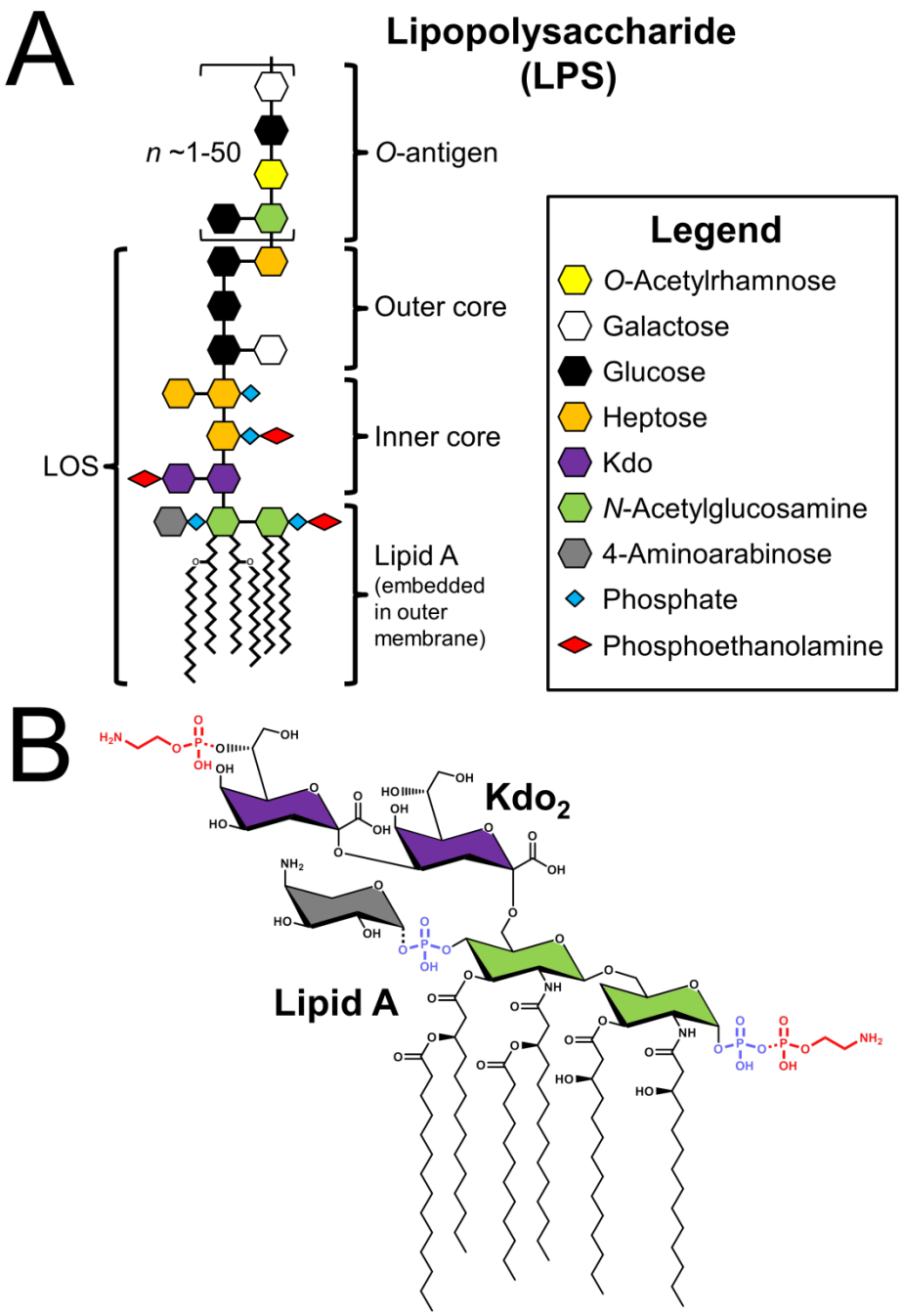


Figure 1-1. A dominant glycoform of *Escherichia coli* K-12 lipopolysaccharide (LPS). **(A)** Schematic of LPS and **(B)** structure of Kdo_2-lipid A portion (Raetz *et al.*, 2007; Müller-Loennies *et al.*, 2003; Stevenson *et al.*, 1994). Optional covalent modifications are indicated by dashed bonds in **B**. K-12 cells normally display LOS (e.g., LPS lacking *O*-antigen) due to a mutation in the *O*-antigen operon. Kdo = 3-deoxy-D-manno-oct-2-ulosonic acid.

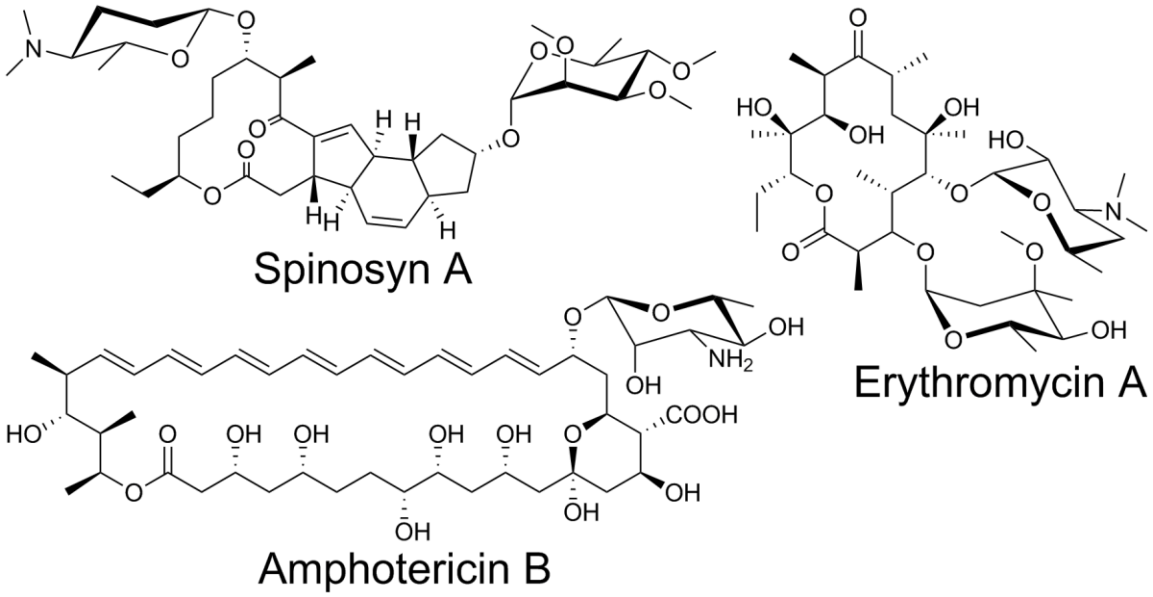


Figure 1-2. Representative polyketides produced by processive Type I PKSs. Polyketides can assume a diverse set of structures and functions. Spinosyn A is an insecticide synthesized by *Saccharopolyspora spinosa*, erythromycin A is an antibiotic synthesized by *Saccharopolyspora erythraea*, and amphotericin is an antifungal synthesized by *Streptomyces nodosus*.

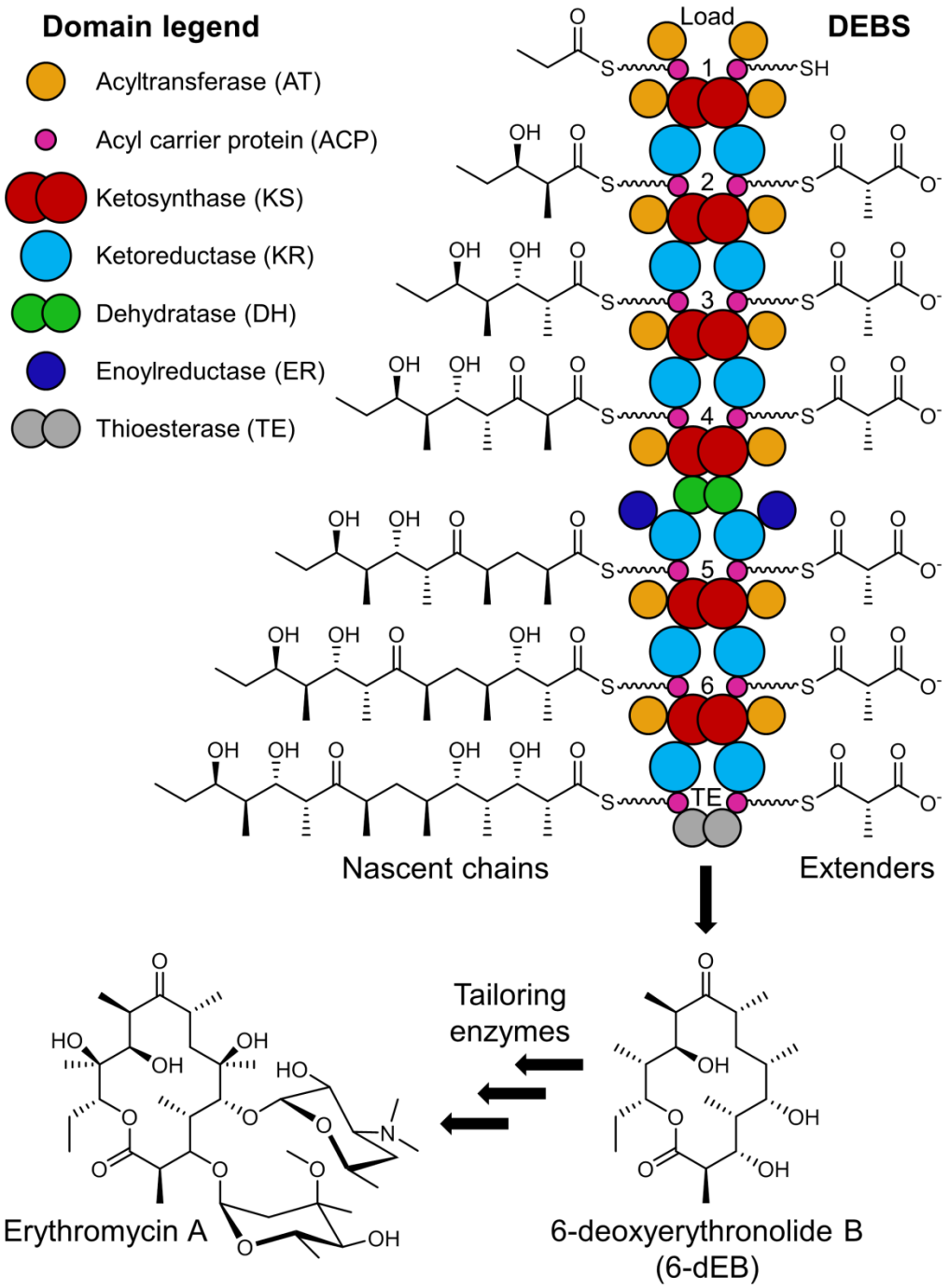


Figure 1-3. Synthesis of the polyketide antibiotic erythromycin A by 6-dEB synthase (DEBS) and associated tailoring enzymes. Note the 2-fold axis across which the KS, DH, and TE domains of separate polypeptides make contact. Modules are labeled Load, 1-6, and TE. DEBS utilizes six methylmalonyl extender units (right) and one propionyl starter unit derived from thioesters of CoA. Squiggly lines represent the pantetheine coenzyme of acyl carrier proteins.

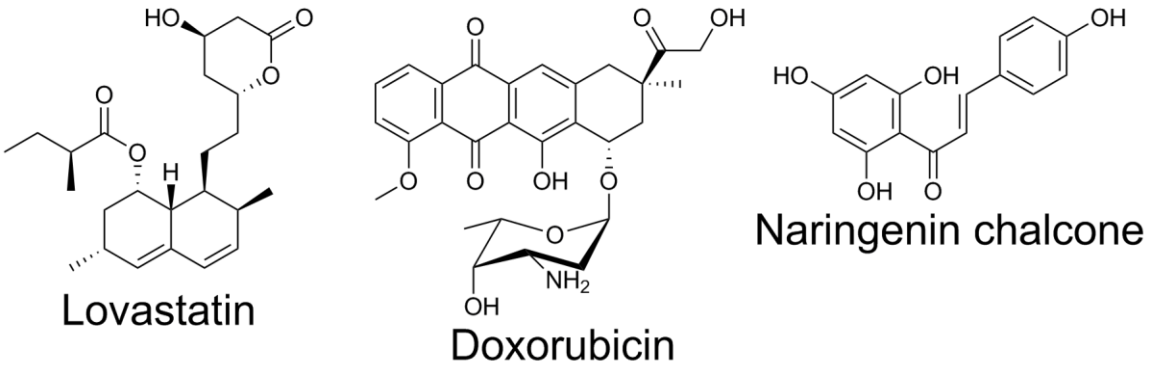


Figure 1-4. Polyketides derived from PKS classes other than processive Type I. Lovastatin is an anticholesterol agent synthesized by an iterative Type I PKS in the fungus *Aspergillus terreus* (Kennedy *et al.*, 1999). Doxorubicin is an anticancer agent synthesized by a Type II PKS in the bacterium *Streptomyces peucetius* (Hutchinson, 1997). Naringenin chalcone is a precursor to flavonoids synthesized widely by a Type III PKS in higher plants (Austin & Noel, 2003).

Chapter 2: Antimicrobial Peptide Resistance of *Vibrio cholerae* Results from LPS Modification Pathway Related to Nonribosomal Peptide Synthetases

Reprinted with permission from *ACS Chemical Biology*.

ABSTRACT

The current pandemic El Tor biotype of O1 *Vibrio cholerae* is resistant to polymyxins, whereas the previous pandemic strain of the classical biotype is polymyxin sensitive. The *almEFG* operon found in El Tor *V. cholerae* confers >100-fold resistance to polymyxins through glycylation of lipopolysaccharide in the outer membrane. Here, we present the mechanistic determination of initial steps in the AlmEFG pathway. We verify that AlmF is an aminoacyl carrier protein and identify AlmE as the enzyme required to activate AlmF as a functional carrier protein. A combination of structural information and activity assays were used to identify a pair of active site residues that are important for mediating the glycine specificity of AlmE. Overall, the structure of AlmE in complex with its glycyl-adenylate intermediate reveals that AlmE is related to Gram-positive D-alanine:D-alanyl carrier protein ligase, while the trio of proteins in the AlmEFG system forms a chemical pathway that resembles the division of labor in nonribosomal peptide synthetases.

INTRODUCTION

Clinical use of polymyxins has emerged as the antibacterial strategy of last resort for infections with multi-drug resistant Gram-negative bacteria such as *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Klebsiella spp.* and *E. coli spp.* (Li *et al.*, 2006; Falagas *et al.*, 2010a; Guilhelmelli *et al.*, 2013). Polymyxins are cationic

antimicrobial peptides (CAMPs) comprised of a cationic/hydrophobic cyclic decapeptide linked to a fatty acyl chain. Like host defensins, CAMPs disrupt the Gram-negative cell envelope by associating with the anionic lipid A membrane anchor of the major cell-surface molecule lipopolysaccharide (LPS) as well as acidic glycerophospholipids (Guilhelmelli *et al.*, 2013). Detergent-like aggregates of membrane-bound polymyxins form pores in the associated membrane, disrupting vital bacterial processes (Guilhelmelli *et al.*, 2013). Accumulating evidence suggests that many bacteria have evolved elegant resistance strategies toward polymyxins and related families of CAMPs (Guilhelmelli *et al.*, 2013; Falagas *et al.*, 2010b). For this reason, a comprehensive understanding of the molecular basis for resistance to CAMPs is paramount in designing new, or modifying current, antibiotics and in developing improved clinical regimens with existing drugs.

In general, to evade CAMP-mediated killing, Gram-negative and Gram-positive bacteria have evolved similar strategies to neutralize the net negative charge of major cell-surface molecules. Phosphatidylglycerols aminoacylated with L-lysine, L-alanine, or D-alanine were recently discovered and confer resistance toward CAMPs in a few species of both Gram-type bacteria (e.g., *Staphylococcus aureus* and *P. aeruginosa*) (Roy, 2009). Gram-positive bacteria commonly neutralize their cell walls by transferring D-alanine to surface teichoic acids, the long poly-phosphoribitol or poly-phosphoglycerol chains linked to the glycerophospholipids or *N*-acetylmuramic acid groups of peptidoglycan (Neuhaus *et al.*, 2003); Gram-negative bacteria do so by transferring phosphoethanolamine or aminoarabinose to phosphate groups on the lipid A domain of LPS (Needham *et al.*, 2013; Raetz *et al.*, 2007). Unexpectedly we observed aminoacyl esterification of glycine or diglycine to lipid A of the Gram-negative pathogen *Vibrio cholerae* (Hankins *et al.*, 2012). Lipid A

glycylation is a unique Gram-negative strategy necessary for resistance to CAMPs and which resembles the Gram-positive method of modifying teichoic acids with D-alanine.

Unlike the O1 classical biotype of *V. cholerae*, which caused the first six cholera pandemics, the O1 El Tor biotype responsible for the most recent pandemic is resistant to polymyxin B (Gangarosa *et al.*, 1967). El Tor strains are responsible for the majority of 3-5 million cases and 100,000-120,000 deaths associated with *V. cholerae* infection every year (World Health Organization). AlmE (Vc1579), AlmF (Vc1578), and AlmG (Vc1577) were identified by our laboratory as three gene products required for polymyxin B resistance in El Tor biotype *V. cholerae* O1 (Hankins *et al.*, 2012). Together, these enzymes orchestrate the modification of *V. cholerae* lipid A with glycine or diglycine (**Figure 2-1**). According to our proposed model, AlmE activates glycine through adenylation and then transfers it to the 4′-phosphopantetheine group of the carrier protein AlmF. AlmG then transfers the glycine substrate from the carrier AlmF to the lipid A anchor. A unique hydroxylauryl chain observed in the lipid A species from El Tor biotype *V. cholerae* serves as the site of glycine addition (Hankins *et al.*, 2011), making AlmEFG the only known lipid A charge remodeling system that does not involve direct transfer to the 1- or 4′- phosphate groups of lipid A. Notably, the classical biotype *V. cholerae*, which does not attach glycine to lipid A and is thus polymyxin-sensitive, encodes a truncated AlmF lacking the predicted phosphopantetheinylation site necessary for its proposed carrier protein functionality (Hankins *et al.*, 2012).

A pair of studies has revealed that the AlmEFG pathway is important during early stages of infection and initial host colonization by *V. cholerae* (LaRocque *et al.*, 2005; Mandlik *et al.*, 2011). In *V. cholerae* isolated from infant mouse small intestine,

the mRNA transcript of the *almEFG* operon is upregulated >10-fold (Mandlik *et al.*, 2011). Moreover, microarray-based comparative gene expression analysis showed a 3.2-fold increase (p-value 1.6×10^{-6}) in AlmE levels between early- and late-stage infections of humans by *V. cholerae* (LaRocque *et al.*, 2005). In that report, AlmE was one of 42 statistically significant, differentially-regulated genes to be identified. As evidence accumulates to demonstrate the importance of the AlmEFG system during infection, detailed mechanistic studies of AlmEFG have the potential to inform further research on therapeutic drugs and vaccines targeting *V. cholerae* infections.

Here, we present the first detailed molecular mechanism of lipid A glycylation in *V. cholerae*. Mass spectrometry data verify that AlmF is a *bona fide* glycyl carrier protein, and moreover that AlmE can transfer glycine to its 4′-phosphopantetheine group. *In vitro* characterization of AlmE reveals that it also accepts D-alanine, albeit with reduced activity compared to glycine. To further investigate this surprising result, we determined the X-ray structure of AlmE in complex with a glycyl-adenylate intermediate at 2.26-Å resolution. Visualizing the active site of AlmE allowed us to rationally design a single-residue mutant capable of efficiently activating and transferring L-alanine to AlmF. However, *in vivo* evidence suggests that L-alanyl-AlmF is not a suitable substrate for AlmG, indicating that the AlmEFG pathway is strictly specific toward glycine. Selection of glycine as the group AlmEFG adds to lipid A likely balances deleterious effects larger amino acids would have on membrane packing, while providing sufficient electropositive density to effectively resist CAMPs. Our findings provide molecular insight into a complex mechanism of CAMP resistance that incorporates preexisting chemistry, commonly found in nonribosomal peptide synthetases, into new biosynthetic pathways.

RESULTS AND DISCUSSION

AlmF is phosphopantetheinylated by Vc2457

Homology-based structural data obtained from the Phyre2 server suggest that AlmF forms the same three-helix architecture observed in carrier proteins such as the Gram-positive D-alanyl carrier protein DltC, and harbors a conserved, nucleophilic serine residue at the N-terminal end of helix 2 (**Figure 2-2**) (Crosby & Crump, 2012). This serine residue is post-translationally modified with 4′-phosphopantetheine (Ppant), a thiol-containing prosthetic group that allows carrier proteins to shuttle substrate molecules as covalently-attached thioesters (Crosby & Crump, 2012). A phosphopantetheinyltransferase appends the Ppant moiety from Coenzyme A directly onto the carrier protein serine residue (Beld *et al.*, 2014). Bacteria encode at least two phosphopantetheinyltransferases: one activates acyl carrier proteins (e.g. AcpS) necessary for primary metabolic reactions such as fatty acid biosynthesis, and another activates carrier proteins of the biosynthetic pathways producing secondary metabolites such as polyketide and nonribosomal peptide antibiotics (Fischbach & Walsh, 2009).

To determine if AlmF is a *bona fide* carrier protein, AlmF was co-expressed with either predicted phosphopantetheinyltransferase of *V. cholerae*, Vc2457 or Vc0780, and subsequently purified and examined for phosphopantetheinylation. To obtain quantities of AlmF suitable for analysis, proteins were overexpressed in *E. coli* BL21(DE3)pLysS using the pQlink system and purified by a two-step protocol, yielding >95% pure AlmF as evaluated by SDS-PAGE (**Figure 2-3**). To assess phosphopantetheinylation, purified "Ppant-modified" (holo-) and "unmodified" (apo-) carrier proteins, were resolved using a destabilizing-urea (non-denaturing) PAGE gel (Masoudi *et al.*, 2013; Agarwal *et al.*, 2012; Rock *et al.*, 1981). When co-

expressed with Vc2457, AlmF displayed a migration shift consistent with Ppant-addition, compared to AlmF expressed alone or co-expressed with Vc0780 (**Figure 2-4**). *E. coli* strains expressing AlmF alone yielded a band consistent with the predicted migration of holo-AlmF only after supplementing the media with millimolar concentrations of pantothenate (**Figure 2-5**; <50% conversion), consistent with nonspecific activity of *E. coli* AcpS on increased Coenzyme A substrate pools (**Figure 2-5**) (Alberts & Vagelos, 1966). AlmF appears completely phosphopantetheinylated when co-expressed with *V. cholerae* AcpS (Vc2457) without pantothenate supplementation (**Figures 2-4** and **2-5**).

Unequivocal determination of AlmF Ppant-modification was assessed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) of trypsin-digested AlmF. Digests of unmodified (apo-AlmF; expressed alone) or Ppant-modified (holo-AlmF; co-expressed with Vc2457) AlmF were analyzed using the same purified AlmF material assessed by destabilizing urea-PAGE (**Figures 2-4** and **2-5**). Three key peptides, each covering the first 41 amino acids of AlmF, were identified from LC–MS/MS analysis (**Figure 2-6**). The AlmF peptide (FFDEFDSFSFIDIVAK) containing the proposed site of modification, Ser34, eluted around 20 min in the apo-AlmF sample, whereas the same Ser34-containing peptide eluted 2 min later in the holo-AlmF sample, consistent with Ppant-addition (**Figure 2-6**; *m/z* shift of 114 for the $[M+3H^+]^{3+}$ peptide). An increase in overall hydrophobicity is expected upon Ppant-addition, consistent with the delayed elution of the holo-AlmF peptide. The electrospray ionization mass spectrum of the holo-AlmF Ser34-containing peptide showed a mass shift of ~340 Da relative to the apo-AlmF peptide, indicative of Ppant-modification (**Figure 2-7**). The sequence of the unmodified FFDEFDSFSFIDIVAK peptide ($[M+2H^+]^{2+}$) was readily confirmed based on the

presence of the y_3 to y_{15} fragment ions (**Figure 2-7**, top spectrum). MS/MS analysis of the peptide from the holo-AlmF sample resulted in a series of diagnostic *y* sequence ions as well as singly-charged product ions corresponding to pantetheine (Pant, *m/z* 261.1) and Ppant (*m/z* 359.1) in the low-mass region (**Figure 2-7**, middle spectrum). Generation of Pant and Ppant product ions during MS/MS analysis of other Ppant-modified proteins has been previously reported (Kelly *et al.*, 2007). It should be noted that while Ser36 of AlmE could serve as the Ppant-modification site, the observed *y*-type product ions containing Ser36 were matched to their expected, unmodified masses and thus did not exhibit incorporation of Ppant (**Figure 2-7**, middle spectrum). Low-abundance *y*-type product ions (from y_{10} up to y_{15}) that incorporated Ppant were also observed (**Figure 2-8**), implicating Ser34 as the sole Ppant-modified site. Together with results from destabilizing urea-PAGE experiments, these findings convincingly support our structural homology-based prediction that AlmF is a Ppant-containing carrier protein.

Classical biotype O1 *V. cholerae* encodes a truncated version of AlmF that lacks the phosphopantetheinylation site required for carrier protein functionality. Thus, our biochemical determination that AlmF is a glycyl carrier protein provides a rationale for why classical biotype O1 *V. cholerae* does not add glycine to lipid A and exhibits polymyxin sensitivity. Why classical strains appear to have lost carrier protein functionality and thus CAMP resistance is intriguing. Expression data reported from other laboratories show that *almEFG* is upregulated during early stages of *V. cholerae* infection in human volunteers and mouse infection models (LaRocque *et al.*, 2005; Mandlik *et al.*, 2011). Improved CAMP resistance may have led to the emergence of El Tor over classical biotypes as the premier pandemic strain. However, the hypothesis that classical strains lost the ability to add glycine to

the cell surface could mean that a modification of this type is detrimental to *V. cholerae* fitness under certain environmental conditions. Consistent with this hypothesis, our laboratory has recently discovered a two-component system responsible for regulating the AlmEFG system.

AlmE adds glycine to the AlmF carrier protein

To verify the role of AlmF as a glycyl carrier protein, we engineered holo-AlmF-producing *E. coli* strains to express AlmE, the predicted carrier protein ligase. Unexpectedly, destabilizing urea-PAGE revealed no apparent shift in the migration of holo-AlmF when co-expressed with AlmE (**Figure 2-5**). However, deeper analysis of this sample by the same LC-MS/MS protocols described above demonstrated that the Ppant-containing peptide was modified with glycine (*m/z 775.8, z = 3,* **Figure 2-7**, bottom spectrum) and, similarly, the glycyl-Pant and glycyl-Ppant singly-charged fragment ions showed the 57-Da increase expected from glycine addition (**Figure 2-7**, bottom spectrum). The glycylated Ppant-containing peptide eluted 30 s earlier than the Ppant-containing peptide, consistent with the addition of a charged chemical moiety (**Figure 2-6**). By comparing the peak areas from extracted ion chromatograms of the doubly-charged Ppant-containing and glycylated Ppant-containing peptides (*m/z* 1134 vs. *m/z* 1163), we estimate that no more than 5% of the total holo-AlmF pool was glycylated when purified from *E. coli* expressing AlmE (**Figure 2-9**). Overexpression of AlmF alone yields relatively low protein levels (<1 mg/L of *E. coli*), suggesting that expression may be toxic in *E. coli*. Efficient glycylation may be impeded by other factors within *E. coli* or, conversely, another factor may be required for efficient glycylation of AlmF in *V. cholerae*. Because *V. cholerae* biotype El Tor also contains lipid A modified with diglycine, we

investigated whether holo-AlmF shuttles diglycine; however, our highly-sensitive MS/MS protocols were unable to identify peaks associated with more than one glycine attached to AlmF (**Figure 2-9**). As diglycine-modified lipid A is observed when isolated from *V. cholerae* biotype El Tor, it is likely that the terminal enzyme in the pathway, AlmG, can perform two rounds of glycylation. Dependent upon the El Tor strain, diglycine-modified lipid A can range from a minor to the most abundant lipid A species isolated (Hankins *et al.*, 2012). Development of an assay for AlmG is currently underway in our laboratory. These results validate genetic and *in silico* predictions that AlmE activates glycine for transfer to the Ppant-containing carrier protein AlmF.

AlmE has relaxed substrate selectivity for amino acids other than glycine

To further explore the enzymology of AlmE, we developed a rapid *in vitro* assay that couples AlmE-catalyzed glycyl-AMP formation with indirect detection of pyrophosphate, the other AlmE reaction product (see **Figure 2-1**). Pyrophosphatase was employed to convert pyrophosphate to phosphate, which was quantified using a spectrophotometric-based malachite green assay (Geladopoulos *et al.*, 1991). *In vitro* AlmE assay conditions were optimized based on reports of previously characterized aminoacyl adenyltransferase/carrier protein ligases (Osman *et al.*, 2009; Trauger *et al.*, 2000; McQuade *et al.*, 2009). Inclusion of purified holo-AlmF, compared to apo-AlmF, significantly stimulated activity as monitored by our pyrophosphatase-coupled assay (**Figure 2-10**). Among a representative sampling of amino acids, glycine is the preferred substrate for adenylation; however, moderate activity (~25-30%) was observed when D-alanine was provided in AlmE reaction mixtures (**Figure 2-11A**). It should be noted that these assay conditions only

monitor the ability to adenylate an amino acid and provide no information regarding subsequent ligation to holo-AlmF or other candidate acceptor substrates.

To more thoroughly monitor AlmE activity, we developed an alternative to the pyrophosphatase-coupled assay that could directly monitor formation of ^{3}H-aminoacyl-AlmF. Reaction products were TCA-precipitated, resuspended, then separated by SDS-PAGE and transferred to a nitrocellulose membrane for radioisotope detection. Formation of ^{3}H-glycyl-AlmF is indicated by the presence of the ^{3}H-labeled protein band (**Figure 2-11B**), which is not observable when ATP or AlmE are omitted from assay mixtures (**Figure 2-11B**). We also tested other ^{3}H-labeled amino acids, such as ^{3}H-L-alanine and ^{3}H-D-alanine. Consistent with our pyrophosphatase assays, AlmE activated and transferred ^{3}H-L-alanine (<10%; compare **Figure 2-11A** and **2-11B**), but was unable to transfer ^{3}H-D-alanine to holo-AlmF (**Figure 2-11B**). These data suggest that *V. cholerae* preferentially glycylates holo-AlmF, consistent with its mechanistic role in lipid A glycylation. Given the similarity of the active site to the D-alanine-specific enzyme DltA, it is not too surprising that AlmE showed some activity toward D-alanine. However, the activity seems to be restricted to adenylation, where subsequent ligation to carrier AlmF is not observed. As formation of an aminoacyl-adenylate is energetically costly, and AlmE exhibits modest activity towards D-alanine *in vitro*, it remains possible that AlmE can activate and transfer D-alanine to non-AlmF substrates such as other carrier proteins or coenzyme A.

Overall structure of AlmE

To better understand the features that govern substrate selectivity for glycine we pursued the structural characterization of AlmE. No adenylation domains

specific for this smallest amino acid have been deposited in the Protein Data Bank (PDB). AlmE formed narrow rod-shaped crystals with one monomer per asymmetric unit in space group $P3_121$ (**Table 2-1**). Its structure was determined via molecular replacement to a resolution of 2.26 Å. In addition to DltA enzymes, AlmE resembles the adenylation domains (A-domains) of nonribosomal peptide synthetases (NRPSs) (Conti *et al.*, 1997), acyl- and aryl-CoA synthetases (Jogl & Tong, 2004; Gulick *et al.*, 2003), and firefly luciferases (Conti *et al.*, 1996); it is comprised of a large N-terminal body (50.6 kDa, residues 1-446), a smaller C-terminal lid (12.2 kDa, residues 450-556), and a very short hinge region (residues 447-449) linking the two domains (**Figure 2-12A**). Based on studies of several related enzymes, AlmE crystallized in the thiolation conformation in which a channel for the 4′-phosphopanthetheinyl arm of AlmF is created through a ~140° rotation of the C-terminal lid relative to its position in the adenylation conformation (Wu *et al.*, 2009). Formation of the glycyl-adenylate intermediate was achieved by incubating AlmE with glycine, $MgCl_2$, and ATP prior to crystallization (**Figure 2-12B**). No electron density is apparent for the His-tag, residues 1-14, residues 25-29 that connect the N-terminus to the main body of the enzyme, residue 189, residues 533-536, or residues 551-556 of the C-terminus. Density for residues 15-24 is visible adjacent to the C-terminal lid of a symmetry mate, with residues 19-23 forming a short β-strand that runs antiparallel to the last β-strand (residues 523-527) of the neighboring lid.

The active site of AlmE is more similar to those of DltA enzymes (*Bacillus subtilis* DltA, PDB code 3E7W; *Bacillus cereus* DltA, PDB code 3DHV) than those of other adenylation domains whose coordinates have been deposited in the PDB (**Figure 2-12C**) (Yonus *et al.*, 2008; Du *et al.*, 2008). DltA enzymes are D-alanyl

carrier protein ligases that transfer D-alanine to lipoteichoic acids in the Gram-positive cell wall. The *B. subtilis* and *B. cereus* enzymes share 32% and 31% sequence identity with AlmE (Protein BLAST) (Altschul *et al.*, 1997), respectively, and their N-terminal bodies align with AlmE to 1.41 Å r.m.s.d. over 278 C_α atoms and 1.59 Å r.m.s.d. over 289 C_α atoms (**Figure 2-13**) in PyMOL. Like AlmE, *B. subtilis* DltA was observed in the thiolation conformation, while *B. cereus* DltA was visualized in the adenylation conformation. *B. cereus* DltA, like AlmE, was crystallized bound to an aminoacyl-adenylate intermediate. With the AlmE structure, a Phyre2 model of AlmF, the *Streptomyces coelicolor* MatB structure (PDB code 3NYQ) (Hughes *et al.*, 2011b), and the *E. coli* EntB-EntE complex structure (PDB code 3RG2) (Sundlov *et al.*, 2012) as guides, the interaction between holo-AlmF and AlmE could be modeled (**Figure 2-14**).

Residues contacting the glycyl moiety

Several residues make favorable contacts general to all amino acyl groups, while others may play a greater role in enforcing glycine selectivity. The Lys452 amine forms a charged hydrogen bond with the glycyl carbonyl, while the carboxylate of Asp247 as well as the carbonyls of Gly341 and Ile347 make charged interactions with the positively-charged glycyl amine. (Like Val301 of *B. cereus* DltA, Ile347 is in the disallowed region of the Ramachandran plot: $\varphi, \psi = 75°, -81°$.) The residues of AlmE nearest the C_α of the glycyl-adenylate, Leu248 (3.9 Å from $C_{\delta 2}$ to C_α) and Cys316 (3.4 Å from S_γ to C_α), may select against amino acids larger than glycine by sterically clashing with their α-substituents – Leu248 with the side chains of D-amino acids, and Cys316 with those of L-amino acids. The conformation of Leu248 is determined through interactions with the side chains of the neighboring

residues Phe246, Ser242, and Leu287. In *B. cereus* DltA, Ser242 and Leu287 are substituted with Ala192 and Thr240 – their smaller side chains position the C_δ atoms of Leu198 (equivalent to Leu248 of DltA) farther from C_α of the aminoacyl-adenylate than in AlmE, allowing room for the methyl group of D-alanine (**Figure 2-13**). The helix containing Leu248 in AlmE is also straight, in contrast to the same helix in DltA that creates space for the D-alanine methyl group through its bent geometry (Yonus *et al.*, 2008). Structural evidence supports the hypothesis that Leu248 appears to contribute to stereoselectivity against D-amino acids, while Cys316 is positioned to sterically select against L-amino acids with large C_α-substituents.

AlmE variant Cys316Ala can efficiently adenylate and transfer L-alanine to AlmF

By comparing the active site architecture of AlmE with that of D-alanyl carrier protein ligase DltA, we predicted how glycine is selected over other small amino acids such as L- or D-alanine. Based on this analysis, as well as our observation that AlmE adenylates D-alanine and ligates L-alanine onto AlmF *in vitro* (**Figure 2-11**), we decided to investigate the specificity of AlmE through rational mutagenesis.

Single-residue variants of Leu248 and Cys316 were prepared with the hypothesis that less bulky side chains would relax the specificity of AlmE for D- or L-amino acids (**Figure 2-15**). For example, such substitutions are present in the L-phenylalanine-specific PheA (PDB code 1AMU), L-leucine-specific SrfA-C (PDB code 2VQS), and L-valine-specific CytC1 (PDB code 3VNS) NRPS A-domains. Using site-directed mutagenesis, we prepared variants Cys316Ser, Cys316Ala, Leu248Val, and

Leu248Ala and tested their specificity for glycine, D-alanine, L-alanine, L-serine, L-cysteine, and L-valine.

As expected Leu248Val and Leu248Ala mutants did not improve AlmE specificity toward L-amino acids, since Leu248 is positioned to sterically exclude side chains of D-amino acids (**Figure 2-15B**). However, these mutants also showed no activity toward D-alanine or glycine. This could indicate that Leu248 substitution significantly altered the geometry of the amino acid binding site, disrupting a number of the favorable binding interactions described above (**Figure 2-13**). Our prediction that adequate space could be provided for an L-amino acid by mutating Cys316 appears valid. Replacement of the Cys316 sulfur atom with the relatively smaller oxygen atom, as in Cys316Ser, or removal of the thiol group, as in Cys316Ala, resulted in L-alanine adenylation. The Cys316Ser variant displayed reduced but modest activity for glycine and L-alanine relative to wild-type AlmE, while the Cys316Ala variant was equally able to adenylate glycine or L-alanine at levels below wild-type AlmE with glycine (**Figure 2-15B**). Furthermore, the AlmE Cys316Ala variant was equally able to form ^{3}H-glycyl-AlmF and ^{3}H-L-alanyl-AlmF, as determined by the destabilizing PAGE/radioisotopic assay described above (**Figure 2-15C**). As both Cys316Ser and Cys316Ala variants were less active for glycine adenylation than wild-type AlmE, it appears that substituting smaller side chains removes optimal packing interactions maintained between native AlmE and the C_{α} of the glycyl group (**Figure 2-15A**). These results further indicate that the Cys316 thiol of wild-type AlmE sterically occludes all L-amino acids from the active site.

As the Cys316Ala mutant of AlmE was capable of adenylating and subsequently transferring glycine and L-alanine with nearly equal specificity, one could predict that L-alanyl-AlmF could serve as a substrate for AlmG for transfer

onto lipid A. However, our MS analysis of *V. cholerae* lipid A expressing AlmE Cys316Ala did not show mass shifts consistent with L-alanine modification (**Figure 2-16**). Nevertheless, this variant appears to be active, as we observed glycine-modified lipid A and restored polymyxin resistance in the polymyxin-sensitive *almE* mutant (**Table 2-2**). Based on the absence of alanine-modified lipid A in the Cys316A mutant, coupled with the fact that low levels of L-alanyl-AlmF are generated by this variant in our *in vitro* assays, we propose that AlmG displays strict specificity for glycyl-AlmF.

CONCLUSION

The discovery of amino acid-modified *V. cholerae* lipid A revealed an interesting link between Gram-negative and Gram-positive cell surface modification systems, not only because the molecular pathways are remarkably similar, but because they confer a shared phenotypic outcome leading to polymyxin resistance. Here, we report the biochemical characterization of the initial steps of the AlmEFG lipid A glycylation pathway required for polymyxin resistance in El Tor biotype *V. cholerae* O1, the causative agent of the ongoing cholera pandemic. By destabilizing urea PAGE and mass spectrometry, we verified that AlmF is 4′-phosphopantetheinylated at Ser34 by Vc2457, a predicted 4′-phosphopantetheinyltransferase of the AcpS superfamily. With the same techniques, we were able to show that co-expression of AlmE in this strain results in the formation of glycyl-AlmF, consistent with our hypothesis that AlmF serves as a glycyl carrier protein (**Figure 2-1**). Furthermore, we carried out two different *in vitro* assays to corroborate how AlmE efficiently catalyzes the transfer of glycine to

holo-AlmF. Taking into account that AlmG ultimately transfers glycine from glycyl-AlmF, we find that the overall pathway displays remarkable parallelism to the division of labor employed by nonribosomal peptide synthetases (**Figure 2-1**).

Our *in vitro* assays suggest that the substrate selectivity of AlmE is somewhat relaxed, adenylating D-alanine at a reduced level compared to glycine. In light of this finding, as well as the absence of structural data for an adenylate-forming enzyme with glycine specificity, we pursued the X-ray structure of AlmE. The enzyme crystallized in the thiolation conformation in which it is prepared to transfer a glycyl moiety to the 4′-phosphopantheteine of holo-AlmF. Electron density for the glycyl-adenylate intermediate was observed in the N-terminal body of AlmE. Given the similarity of the active site to the D-alanine-specific enzyme DltA, it was not too surprising that AlmE showed some activity toward D-alanine in the pyrophosphate release assay (**Figures 2-11A**, **2-13**, and **2-17**). However, we did not observe subsequent transfer of D-alanine onto AlmF (**Figure 2-11B**), and various attempts to substitute D-alanine for glycine in our crystallization conditions failed to produce AlmE crystals.

How D-alanine is transferred from carrier proteins to teichoic and lipoteichoic acid polymers in Gram-positive cell walls remains unknown but involves proteins DltB and DltD (Reichmann *et al.*, 2013). Similarly, characterization of AlmG (~30 kDa, membrane-associated, homologous to lysophospholipid acyltransferases and lipid A lauroyl/myristoyl acyltransferases) will be necessary to clarify its role in the AlmEFG pathway (Hankins *et al.*, 2012). The entire *almEFG* operon is conserved in the emerging marine pathogen *Vibrio nigripulchritudo,* although further study is required to verify whether the system can modify lipid A in this organism (Goudenège *et al.*, 2013). Glycine residues have been observed on

core oligosaccharide of LPS in *Campylobacter jejuni*, *Haemophilus influenzae*, and *Shigella flexneri* (Dzieciatkowska *et al.*, 2007; Engskog *et al.*, 2011; Molinaro *et al.*, 2007). It remains possible that these and other bacteria employ machinery similar to AlmEFG to decorate the carbohydrate and antigenic portion of LPS, not just the endotoxin portion of LPS as observed in *V. cholerae*.

Cationic antimicrobial peptides (CAMPs) are a therapy of last resort in treatment of many bacterial infections. This report characterizes initial steps of a unique LPS modification strategy, employed by and necessary for CAMP resistance in *Vibrio cholerae*. As the AlmEFG system resembles Gram-positive cell-surface remodeling mechanisms, this study reveals how distantly related bacteria have evolved similar modes of antibiotic resistance through chemical modification of conserved surface features. Recent reports suggest AlmEFG is important for host colonization by current pandemic *V. cholerae* El Tor. Due to the fitness advantage contributed by AlmEFG, as well as the nonfunctional operon encoded by pandemic classical *V. cholerae,* AlmEFG likely contributed to the emergence of El Tor as the current pandemic strain. As the role of glycine-modified LPS in host-pathogen dynamics and cholera disease progression becomes clearer, our structural evaluation of AlmE may provide the basis for rationally designed inhibitors of AlmE as a cholera therapeutic. Altogether, the structural and functional characterization of *V. cholerae* AlmE and AlmF presented here brings us closer to a comprehensive understanding of bacterial cell wall modification systems.

MATERIALS AND METHODS

Bacterial growth

Bacterial strains, plasmids, and oligonucleotides used in this study are listed in **Tables 2-3** and **2-4**. *V. cholerae* and *E. coli* were routinely grown at 37 °C on LB or LB agar unless otherwise indicated. 50 μg/mL kanamycin, 100 μg/mL ampicillin and 10 μg/mL streptomycin were used as appropriate.

Recombinant DNA techniques

Plasmids (**Table 2-3**) were isolated using the QIAprep Spin Miniprep Kit (Qiagen). Custom primers were obtained from Integrated DNA Technologies (**Table 2-4**). PCR reagents were purchased from Takara, New England Biolabs, or Stratagene, and PCR products were isolated using a QIAquick PCR Purification Kit (Qiagen). DNA fragments were isolated from 0.7% agarose gels using the QIAquick Gel Extraction Kit (Qiagen). All other DNA-modifying enzymes were purchased from New England Biolabs and used according to the manufacturer's protocols. When necessary, purity and yield of extracted DNA were monitored using a NanoDrop instrument (GE Healthcare). To create pQlink plasmids used for co-expression of multiple genes, reagents purchased from Invitrogen were used for ligation-independent cloning (LIC) and as previously described (Scheich *et al.*, 2007). All plasmids constructed in this study were transformed into a chemically competent *Escherichia coli* XL-1 Blue storage strain before transformation into the strain used in experiments (**Table 2-3**). Plasmid promoter and gene-insert sequences were verified by DNA sequencing at the ICMB Core Facility at the University of Texas at Austin.

Cloning, expression, and purification of N-terminal His_6-AlmF or co-expression constructs

Gene *almF* was PCR-amplified from El Tor biotype *V. cholerae* O1 strain N16961 genomic DNA (gDNA). *Bam*HI or *Hind*III restriction sites were engineered to flank the resultant PCR product using primers FpQlinkHVc1578 and RpQlinkHVc1578 (**Table 2-4**). Both plasmid pQlinkH (Scheich *et al.*, 2007) and PCR-amplified *almF* insert were restriction-digested, and plasmid DNA was treated with Antarctic phosphatase before ligation reactions were used to produce pQlinkHVc1578. The same cloning strategy was used to obtain pQlinkNVc0780, pQlinkNVc2457, and pQlinkNVc1579, where each gene was amplified using the corresponding primers by PCR with *V. cholerae* N16961 gDNA template (**Table 2-4**). Co-expression constructs pQVc1578-Vc0780, pQVc1578-Vc2457, and pQVc1578-Vc2457-Vc1579 (**Table 2-3**) were created using the standardized protocol as previously described (Scheich *et al.*, 2007).

E. coli BL21(DE3)pLysS cells containing the appropriate expression plasmid were grown in LB (1.5 L, 37 °C, 180 rpm shaking). Upon reaching OD_{600} of ~0.6, cultures were briefly chilled on ice to ~18 °C and expression was induced by addition of a sterile IPTG solution (100 mM) to a final concentration of 1 mM. After 20 h (18 °C, 180 rpm) cells were pelleted, resuspended in ice-cold buffer A (500 mM NaCl, 50 mM HEPES pH 7.5, 10% (v/v) glycerol, 40 mM imidazole, 1 mM DTT) supplemented with 10 μM PMSF, and lysed via French press (20k psi). Lysate was clarified by centrifugation (4 °C, 20k × *g*, 30 min) and the soluble portion was syringe-filtered (0.4 μm) before injection over an AKTA-compatible Ni-sepharose HiTrap FF column (GE Healthcare) pre-equilibrated with buffer A. The column was washed with buffer A (10 column volumes) before gradient elution over 30 column

volumes to 100% buffer B (500 mM NaCl, 50 mM HEPES pH 7.5, 10% (v/v) glycerol, 400 mM imidazole, 1 mM DTT). 1-mL fractions were collected where the peak Abs_{260} fraction typically eluted between ~130-210 mM imidazole. Corresponding fractions with >80% AlmF, as determined by SDS-PAGE analysis (AlmF MW 8.9 kDa), were pooled and concentrated via filter centrifugation (Amicon Ultra, 3 kDa cutoff). To further purify AlmF, size-exclusion chromatography was performed using a Superdex S75 column (GE Healthcare) equilibrated with 1:5 dilution of buffer A (no imidazole). Concentrated Ni-sepharose-purified AlmF was diluted 1:5 in water before addition to S75 column. Peak fractions were analyzed by SDS-PAGE, and fractions containing >95% AlmF were pooled and concentrated via filter centrifugation (Amicon Ultra, 3 kDa cutoff).

Destabilizing-urea PAGE of purified AlmF samples

Urea-PAGE gels were manually cast using a mini-PROTEAN casting stand and frame (Bio-Rad). Separating gel consisted of 13% polyacrylamide (19:1; Sigma-Aldrich), 2.5 M urea, 375 mM Tris pH 9.5, 0.05% (w/v) APS, and 0.05% (v/v) TEMED. Stacking gel consisted of 4% polyacrylamide (19:1; Sigma-Aldrich), 2.5 M urea, 125 mM Tris pH 6.8, 0.05% (w/v) APS, and 0.1% (v/v) TEMED. 20 μg of total protein was loaded per lane. Gels were run with prechilled running buffer (25 mM Tris base, 192 mM glycine, 15% (w/v) urea, dissolved in ddH_2O) in a 4 °C cold room for ~3-4 h at a constant 100 V. Visualization of protein bands was performed by Coomassie-staining for 24 h (45% (v/v) methanol, 10% (v/v) glacial acetic acid, 45% (v/v) ddH_2O, and 3 g/L Coomassie R-250) and 24-48 h, due to the high % polyacrylamide, with destain solution (40% (v/v) ethanol, 10% (v/v) glacial acetic

acid, and 50% (v/v) ddH_2O). Gels were stored in ddH_2O until photographically imaged.

UVPD-MS analysis of tryptic digested AlmF samples

Apo-AlmF, holo-AlmF (AlmF + Vc2457), and glycyl-AlmF (AlmF + Vc2457 + AlmE) samples were buffer-exchanged three times into 100 mM ammonium bicarbonate using 3 kDa molecular weight cutoff filters. Samples were then digested with trypsin (1:50 enzyme-to-substrate ratio) overnight and desalted. Peptides were evaporated to dryness and reconstituted in LC solvent A (98%/2%/0.1%, water/acetonitrile/formic acid). For LC profiling, peptides were separated on a Dionex RSLCnano (ThermoDionex) using a gradient of 0-15% solvent B (100%/0.1%, acetonitrile/formic acid) over the course of 30 min, and then 15-70% solvent B over 15 min. Flow rate was set to 750 nL/min through a 150 × 0.075 mm C18 analytical column packed in-house. The three most abundant multiply-charged ions were selected for fragmentation using collision-induced dissociation (NCE 35) and subsequent detection in a Velos Pro Dual-Pressure Linear Ion Trap Mass Spectrometer (ThermoFisher). Ultraviolet photodissociation spectra of the same peptides were acquired on an Orbitrap Elite mass spectrometer (ThermoFisher) after elution using the same gradient described previously on an Eksigent NanoLC Ultra (Eksigent) at a flow rate of 300 nL/min. UVPD was performed using 3 pulses of a 193 nm excimer laser (Coherent Excister XS) (at 3 mJ per pulse). The Orbitrap Elite mass spectrometer was modified and adapted for UVPD as described previously (Shaw *et al.*, 2013). All spectra were manually interpreted.

Cloning, expression, and purification of N-terminal His_6-AlmE or C316 & L248 single amino acid variants

The *almE* gene was PCR-amplified from pVc1579 (Hankins *et al.*, 2012) with primers F- and R-AlmE (**Tables 2-3** and **2-4**), cleaved with *Nhe*I and *Eco*RI, and ligated into pET28b (Novagen) to yield pETAlmE. C316S and L248V mutants were prepared by carrying out site-directed mutagenesis on pETAlmE (**Table 2-3**) with the associated primers (**Table 2-4**). C316S and L248V were further mutated to generate C316A and L248A with the associated primers (**Tables 2-3** and **2-4**).

E. coli BL21(DE3) cells containing the appropriate expression plasmid were grown in LB (1 L) with 50 mg/L kanamycin at 37 °C (180 rpm shaking). Upon reaching OD_{600} of ~0.5, cultures were briefly chilled on ice to ~18 °C and expression was induced by addition of a sterile IPTG solution (100 mM) to a final concentration of 1 mM. After 18 h (18 °C, 180 rpm) cells were pelleted, resuspended in ice-cold buffer A (500 mM NaCl, 50 mM HEPES pH 7.5, 10% (v/v) glycerol, 40 mM imidazole) supplemented with 10 μM PMSF, and lysed via French press (20k psi). Lysate was clarified by centrifugation (4 °C, 20k × *g*, 30 min) and the soluble portion was syringe-filtered (0.4 μm) before injection over an AKTA-compatible Ni-sepharose column (GE Healthcare) pre-equilibrated with buffer A (no PMSF). The column was washed with buffer A (10 column volumes), before gradient elution over 30 column volumes to 100% buffer B (500 mM NaCl, 50 mM HEPES pH 7.5, 10% (v/v) glycerol, 400 mM imidazole). 1-mL fractions were collected where the peak Abs_{260} fraction typically eluted between ~150-200 mM imidazole. Corresponding fractions with >95% AlmE, as determined by SDS-PAGE analysis (AlmE MW 63 kDa), were pooled and concentrated via filter centrifugation (Amicon Ultra, 3 kDa cutoff). For crystallization studies, Ni-sepharose-purified AlmE was

further purified on a Superdex 200 gel filtration column (GE Healthcare) equilibrated with 150 mM NaCl, 10 mM HEPES pH 7.5, 10% (v/v) glycerol.

In vitro AlmE Assays

AlmE/pyrophosphatase-coupled assay conditions are as follows: 50 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 1 mM ATP, 1 mM amino acid (glycine, D-alanine, L-alanine, serine, cysteine, or valine), 0.5 U yeast inorganic pyrophosphatase (Sigma-Aldrich), 8 μM purified holo-AlmF, 100 μM DTT, and 800 nM purified AlmE incubated at 25 °C for 2.5 or 3 h as indicated in the Results and Discussion section and/or figure legends. Percent activity corresponds to the percent conversion of [ATP] during AlmE-catalyzed aminoacyl-AMP formation, where detected [phosphate]/2 equals [ATP] consumed. AlmF ^{3}H-aminoacylation assays with AlmE were performed under identical reaction conditions as described above (with omission of DTT and pyrophosphatase, inclusion of 0.3 μCi ^{3}H-amino acid) and were incubated for 18 h at 18 °C. AlmF in assay mixtures was concentrated by TCA precipitation and resolved by SDS-PAGE. Resolved protein was transferred to a nitrocellulose membrane, which was exposed to a ^{3}H-sensitive screen for 5 days.

Crystallization of AlmE

Protein was concentrated to 10 mg/mL in 10% (v/v) glycerol, 25 mM NaCl, 10 mM HEPES pH 7.5 and incubated with 10 mM glycine, 10 mM ATP, and 1 mM $MgCl_2$ for 10 min before preparing crystal trays. Narrow rod-shaped crystals grew in 50 mM $MgCl_2$, 100 mM HEPES pH 7.0 by sitting-drop vapor diffusion at 22 °C. Prior to flash freezing in liquid nitrogen, crystals were soaked for 10 min in cryosolution consisting of 20% (v/v) glycerol, 50 mM $MgCl_2$, 100 mM HEPES pH 7.0. While

crystals also grew in the absence of glycine and ATP, their quality and size were significantly diminished.

X-ray data collection, processing, phasing, and refinement

Data were collected at 100 K at Advanced Photon Source beamline 23-ID-D and processed in HKL2000 (**Table 2-1**) (Otwinowski & Minor 1997). One crystal diffracted to 2.26 Å. Molecular replacement (MR) computations were performed in Phaser (McCoy *et al.*, 2005) of the CCP4 software suite (Winn et al., 2011), with the N-terminal body of one monomer of PheA (PDB code 1AMU, residues 17-428) as a search model (Conti *et al.*, 1997). One AlmE monomer was found per asymmetric unit in space group $P3_121$. This solution was submitted to the ARP/wARP server for automated refinement (Langer *et al.*, 2008) before iterative refinement was carried out in phenix.refine of the PHENIX software suite (Adams *et al.*, 2010), REFMAC5 (Murshudov *et al.*, 1997) of CCP4, and Coot (Emsley & Cowtan, 2004). The structure factor amplitudes and atomic coordinates of AlmE were deposited in the RCSB Protein Data Bank under PDB code 4OXI.

Polymyxin B and polymyxin E minimal inhibitory concentration assay

V. cholerae strains were grown overnight in LB media and then diluted 1:100 in fresh LB (1mM IPTG, 100 μg/mL). Cells were grown to OD_{600} of ~0.6, diluted 1:100 and then applied to LB agar with a sterile cotton tip applicator. Quantitative minimal inhibitory concentration (MIC) values were determined using Etest gradient polymyxin strips (AB Biodisk) after 24 h, using the manufacturer's guidelines.

Matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS of lipid A

Lipid A was extracted, purified, and analyzed by MS as previously described (Hankins *et al.*, 2012; Hankins *et al.*, 2011).

ACKNOWLEDGEMENTS

Funding from NIH (grants AI064184 & AI76322 to M.S.T., GM103655 to J.S.B., & GM106112 to A.T.K.), the Welch Foundation (F-1155 to J.S.B.) & the Army Research Office (grant W911NF-12-1-0390 to M.S.T.) is gratefully acknowledged. Instrumentation & assistance provided by Macromolecular Crystallography Facility, financially supported by the College of Natural Sciences, the Office of the Executive Vice President and Provost, & the Institute for Cellular and Molecular Biology at UT Austin. Use of the Advanced Photon Source was supported by the U.S. DOE (Contract No. DE-AC02-06CH11357). We thank Thermo Fisher Scientific for help on the Orbitrap Elite mass spectrometer to allow UVPD.

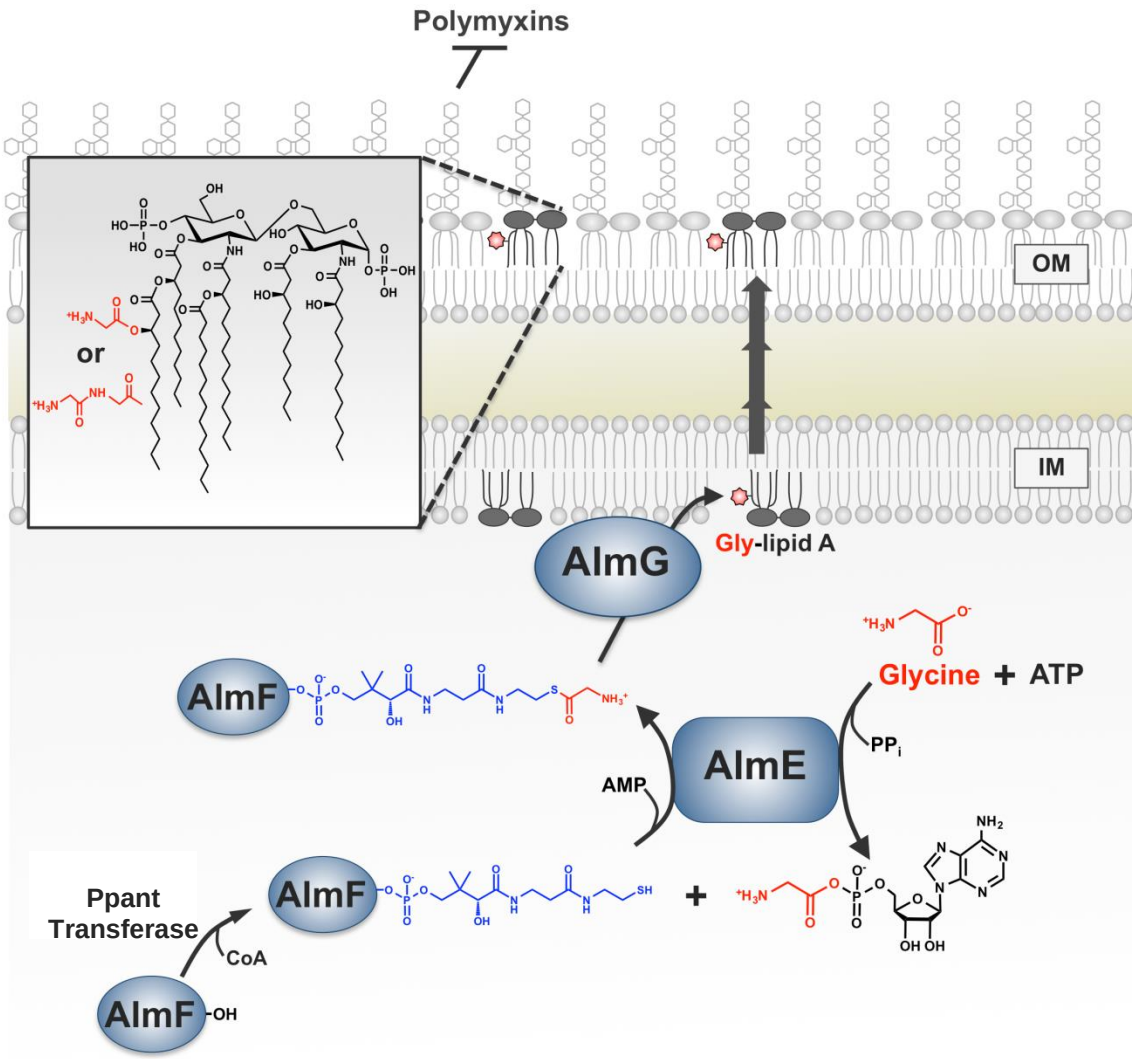


Figure 2-1. Proposed model for the lipid A glycylation pathway in *V. cholerae.* AlmE (Vc1579) generates glycyl-AMP and pyrophosphate from glycine and ATP. Glycine is then ligated onto carrier protein holo-AlmF (Vc1578), with concomitant release of AMP. Holo-AlmF is generated after the 4′-phosphopantetheinyl (Ppant) moiety of coenzyme A is transferred onto apo-AlmF by the phosphopantetheinyltransferase. At the inner membrane, glycyl-AlmF serves as the aminoacyl donor to AlmG (Vc1577), which esterifies glycine onto the secondary hydroxylauryl acyl chain of *V. cholerae* hexa-acylated lipid A. Glycine-modified lipid A is then transported to the bacterial surface to provide resistance against antimicrobial peptides such as polymyxin. OM, outer membrane; IM, inner membrane.

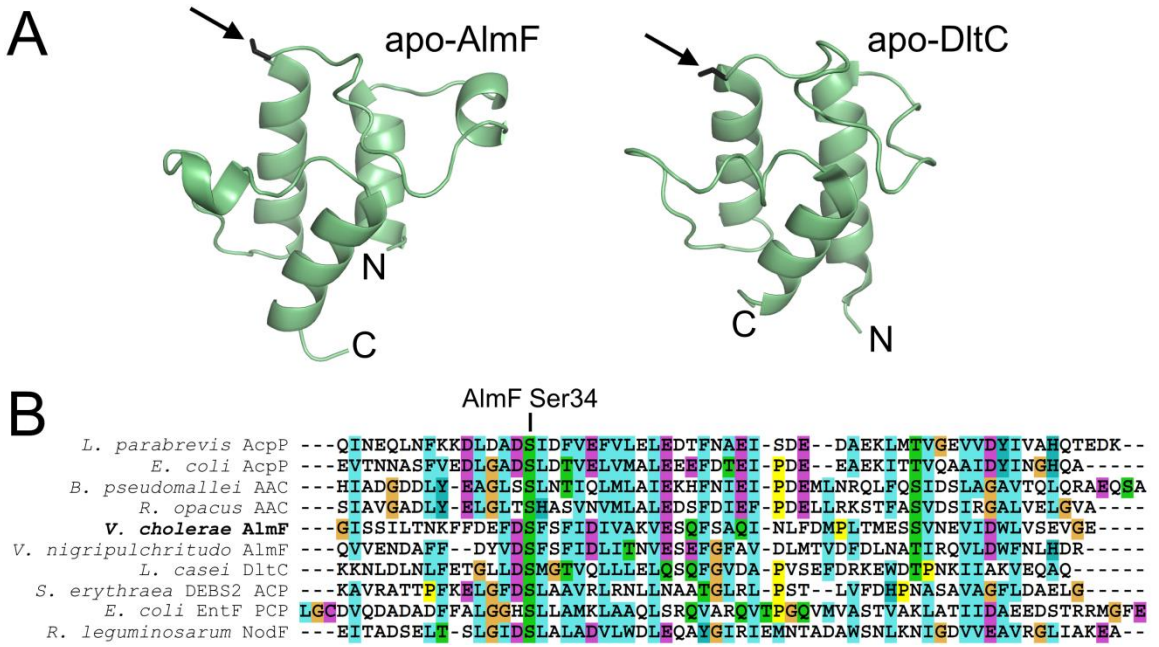


Figure 2-2. Structure-sequence analysis of *V. cholerae* AlmF and related carrier protein domains. **(A)** Soluble carrier proteins adopt three-helix bundle architecture. Apo-AlmF from *V. cholerae* and apo-D-alanyl carrier protein (DltC) from *Lactobacillus casei* (PDB code 1HQB) are depicted as cartoons. The structural model for AlmF was created using coordinates generated by Phyre2 (Kelley & Sternberg, 2009). Both structures lack the 4′-phosphopantetheine prosthetic group. A conserved serine residue, represented as black sticks in the model, serves as the point of attachment for 4′-phosphopantetheine, indicated by an arrow. **(B)** The amino acid sequence alignment shows that a conserved serine (Ser34 of AlmF) is readily identifiable even across a broad array of carrier protein domains, which can be single domains of multidomain nonribosomal peptide synthetases or individual soluble proteins. Protein sequences were aligned using the Clustal Omega algorithm (Sievers & Higgins, 2014) and include the following: *Lactobacillus parabrevis* AcpP (WP_020088992), *Escherichia coli* AcpP (YP_489362), *Burkholderia pseudomallei* aminoacyl carrier protein (AHK63940), *Rhodococcus opacus* aminoacyl carrier protein (AHK29595), *Vibrio cholerae* Vc1578/AlmF (NP_231218), *Vibrio nigripulchritudo* A2555/AlmF homolog (YP_008621728), *Lactobacillus casei* DltC (AAB17659), *Saccharopolyspora erythraea* module 2 acyl carrier protein of 6-deoxyerythronolide B synthase (DEBS; PDB code 2JU2), truncated sequence of *E. coli* EntF peptidyl carrier protein domain (ZP_06938669), *Rhizobium leguminosarum* NodF (AFQ99236).

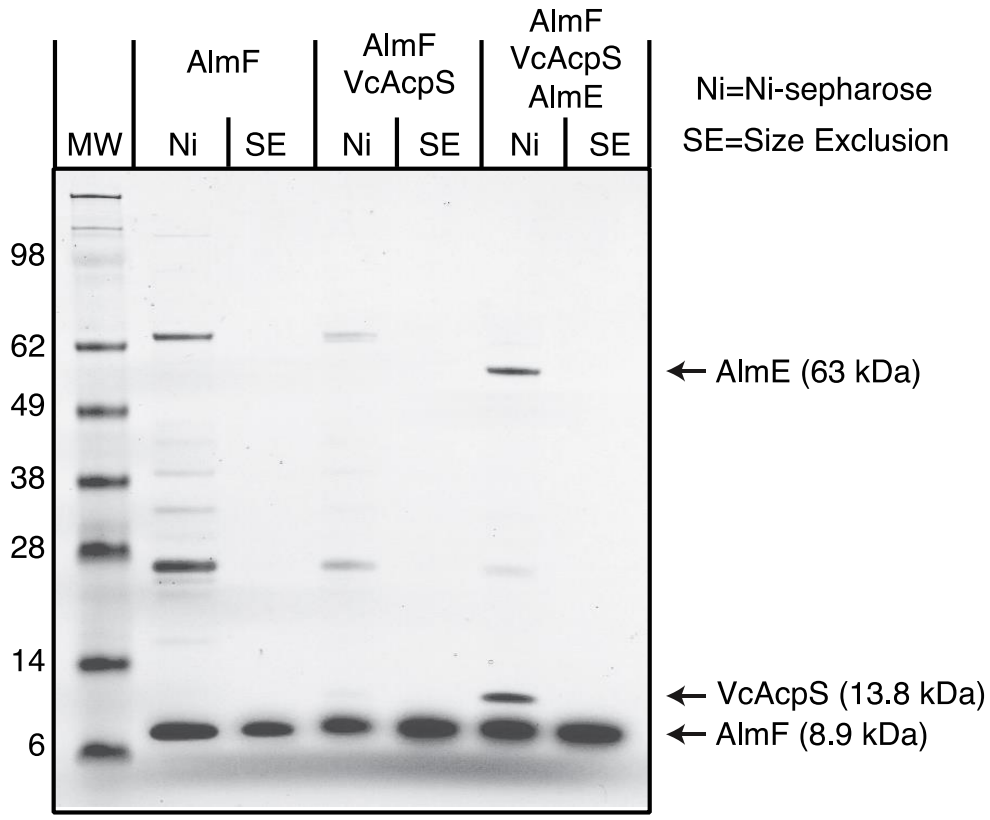


Figure 2-3. Due to low protein yields of recombinant His_6-AlmF, even when over-expressed in *E. coli*, a two-step purification of His_6-AlmF was required: Ni-sepharose (Ni) followed by size-exclusion (SE). An SDS-PAGE gel shows concentrated purification fractions from N-terminal His_6-AlmF overexpressed individually (AlmF), co-expressed with Vc2457, or co-expressed with both Vc2457 and AlmE. 10 µg total protein was loaded per lane. Purified samples were characterized by UVPD-MS/MS to assess phosphopantetheine- (Ppant) and $Ppant_{Gly}$-modification status (**Figure 2-11**).

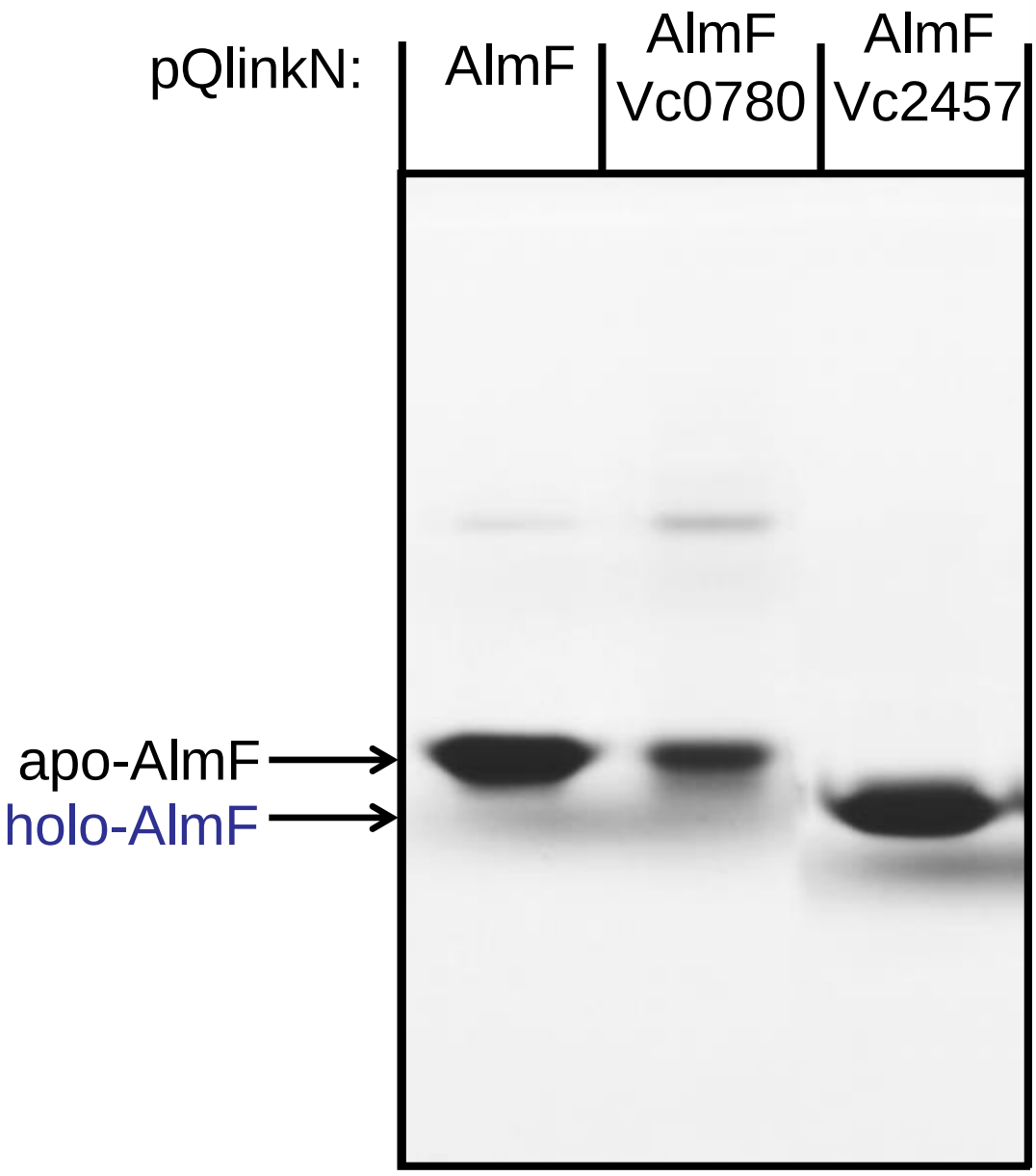


Figure 2-4. Co-expression of Vc2457 (VcAcpS) with AlmF (apo-AlmF) in *E. coli* results in the production of phosphopantetheinylated AlmF (holo-AlmF) as assessed by destabilizing urea-PAGE. Consistent with previous reports of other carrier proteins, holo-AlmF migrates farther in the gel than apo-AlmF as a result of increased packing upon Ppant modification. Vc0780 does not exhibit phosphopantetheinyltransferase activity towards AlmF.

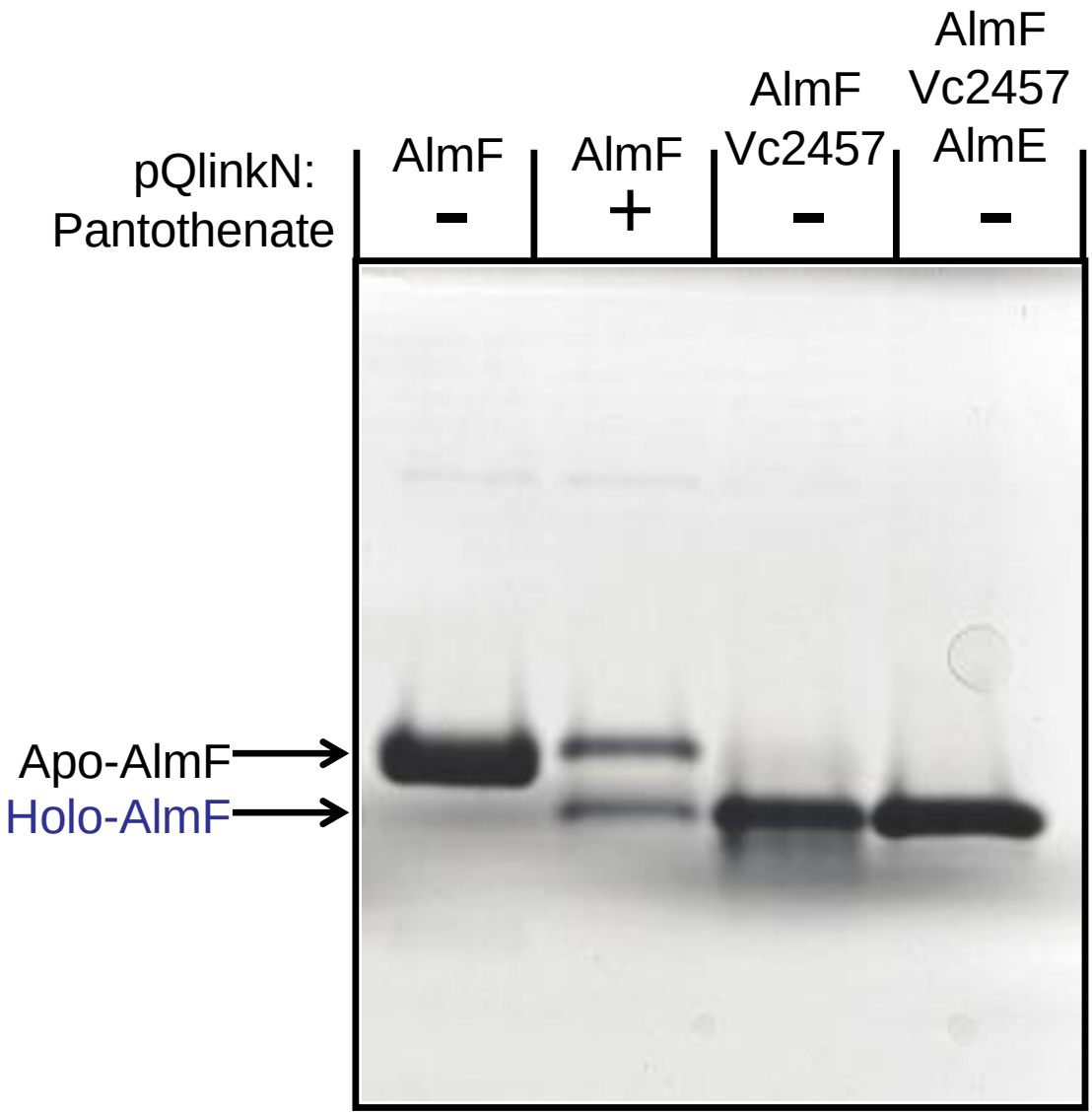


Figure 2-5. Incubating holo-AlmF with AlmE does not result in a shift in the destabilizing urea-PAGE migration assay. Note the single band in the far right lane. However, deeper analysis with UVPD-MS/MS shows that AlmE is capable of transferring glycine to holo-AlmF (**Figure 2-11**). As Vc2457 is a homolog of *E. coli* AcpS, supplementing the *E. coli* AlmF overexpression strain with Ca^{+2}-pantothenate (1 mM) results in modest conversion of apo-AlmF to holo-AlmF. Thus, although the *E. coli* AcpS is not nearly as efficient as Vc2457 at holo-AlmF production, supplementing with pantothenate increases the substrate Coenzyme A pools to push the non-specific reaction forward (**Figure 2-1**).

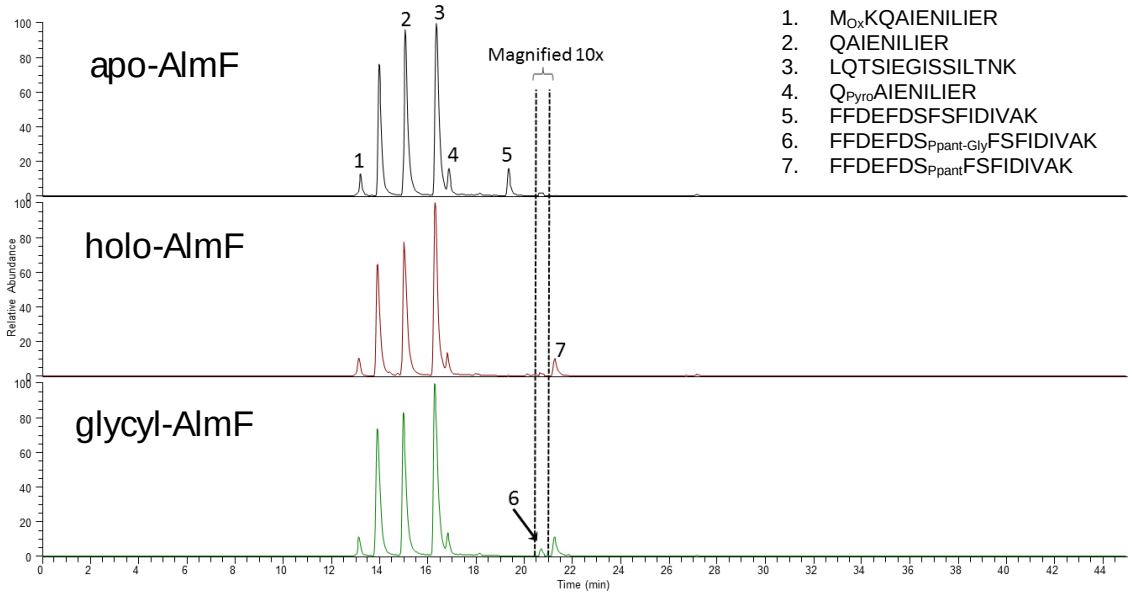


Figure 2-6. Stacked comparison of the base-peak chromatograms for reverse-phase HPLC chromatographic profiles of trypsin-digested AlmF samples as analyzed in **Figure 2-7**. Samples are unmodified (apo-AlmF), Ppant-modified (holo-AlmF), or $Ppant_{Gly}$-modified (glycyl-AlmF). The peptide assignments are shown in the legend in the upper right.

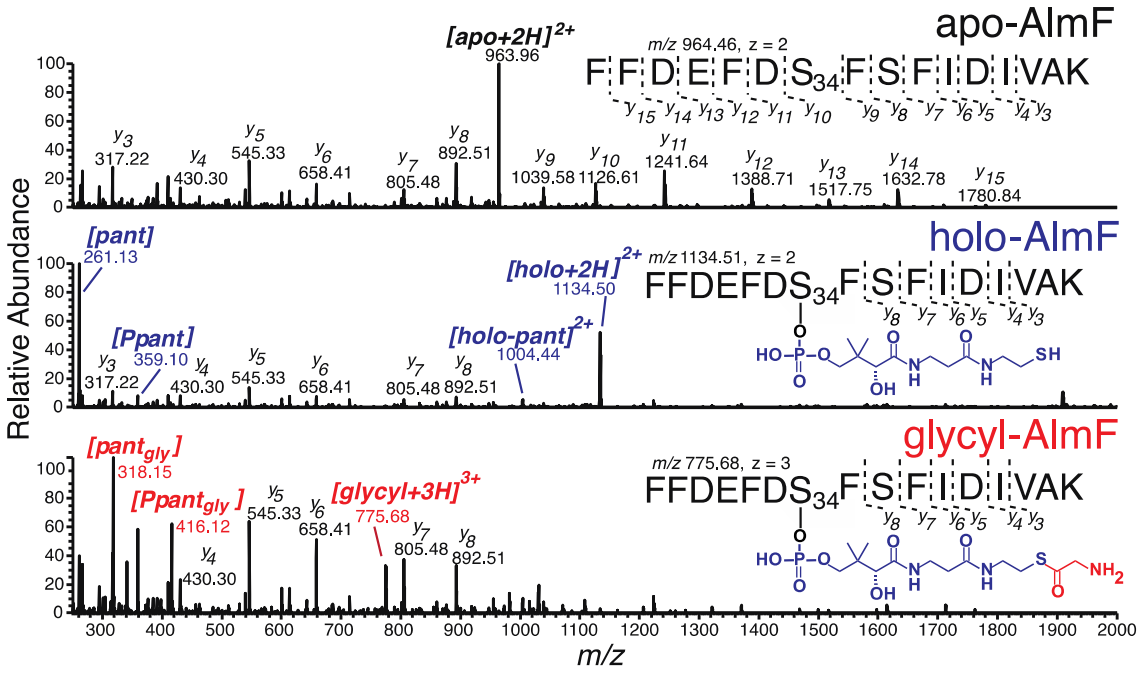


Figure 2-7. UVPD-MS/MS confirms AlmF phosphopantetheinylation by Vc2457 and subsequent glycylation by AlmE when co-expressed in *E. coli*. Shown are MS/MS spectra resulting from 193 nm ultraviolet photodissociation of precursor ions corresponding to the trypsinized AlmF peptide FFDEFDS$_{34}$FSFIDIVAK: **(top)** doubly charged, unmodified (apo-AlmF; precursor *m/z* 964), **(middle)** doubly charged, Ppant-modified (holo-AlmF; precursor *m/z* 1134), and **(bottom)** triply charged, Ppant$_{Gly}$-modified (glycyl-AlmF; precursor *m/z* 776). These samples were obtained from purified AlmF, AlmF co-expressed with Vc2457, or AlmF co-expressed with Vc2457 and AlmE, respectively. Ser34 is the Ppant modification site as the *y*-type ion series through Ser36 *(y_3-y_8)* is conserved for both **(middle)** Ppant-modified AlmF and **(bottom)** Ppant$_{Gly}$-modified AlmF. Cleavage on either side of the phospho group results in diagnostic Ppant modification ions in the low mass region: **(middle)** Pant (*m/z* 261) and Ppant (*m/z* 359) or **(bottom)** Pant$_{Gly}$ (*m/z* 318) and Ppant$_{Gly}$ (*m/z* 416). The most abundant fragment ions are labeled.

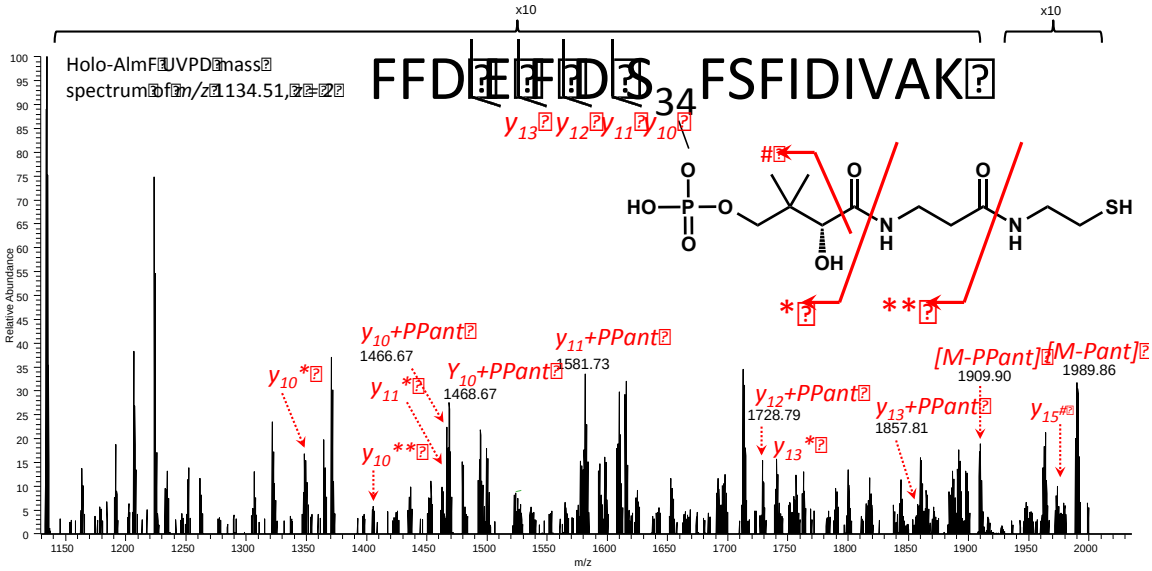


Figure 2-8. Expanded region of UVPD mass spectrum to confirm phosphopantetheinylation of AlmF. This spectrum shows the higher-mass portion of the MS/MS spectrum from **Figure 2-7 (middle)** resulting from 193 nm ultraviolet photodissociation of the doubly-charged phosphopantetheinylated precursor (holo-AlmF; precursor *m/z* 1134). The ions containing the phosphopantetheine modification have been highlighted. 10x magnification was used for all ions with *m/z* values greater than the precursor ion with the exception of the ion of *m/z* 1909.90, which corresponds to the peptide after loss of the Ppant moiety. The asterisks and hash symbol indicate partial losses of the Ppant modification accompanying the formation of certain *y* ions.

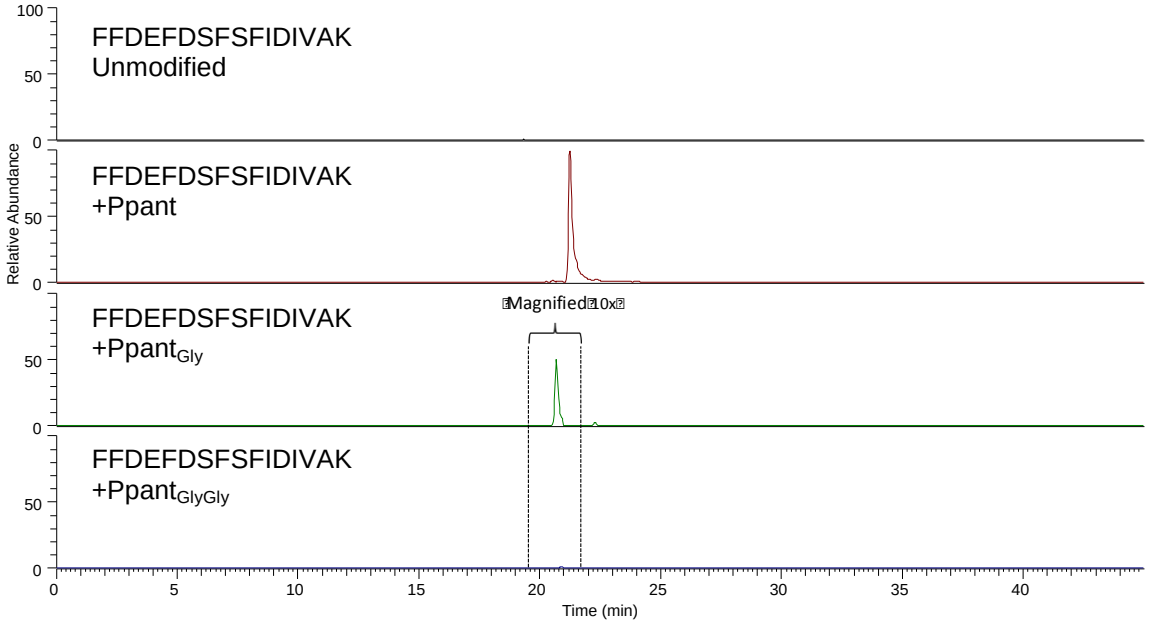


Figure 2-9. Mass spectra of AlmF species. Shown from top to bottom are extracted ion chromatograms for the doubly-charged molecular ions corresponding to unmodified (*m/z* 964), Ppant-modified (*m/z* 1134), Ppant$_{Gly}$-modified (*m/z* 1163), and Ppant$_{GlyGly}$-modified (*m/z* 1191) peptide, FFDEFDS$_{34}$FSFIDIVAK, from the tryptic digest of AlmF co-expressed with Vc2457 and AlmE. The extracted ion chromatogram for the Ppant$_{Gly}$-modified species has been magnified to show glycylation of the Ppant-modified FFDEFDS$_{34}$FSFIDIVAK peptide.

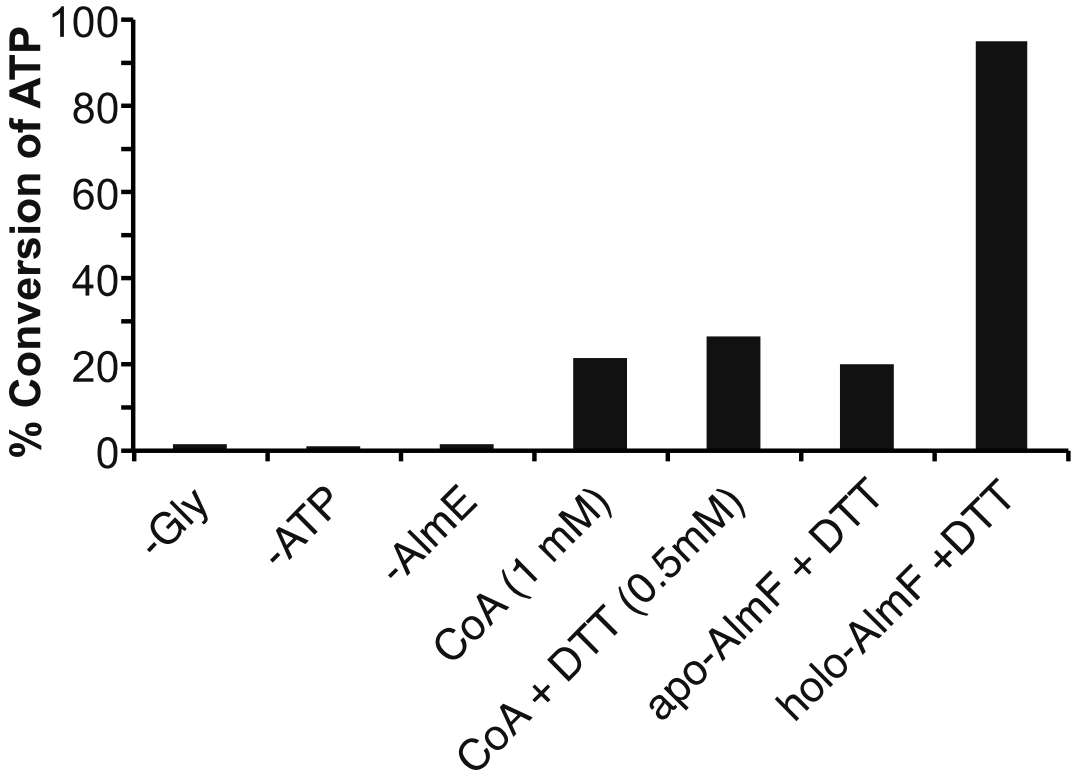


Figure 2-10. Relative to reactions where Coenzyme A (CoA) or apo-AlmF are provided as aminoacyl acceptors, providing holo-AlmF results in a robust increase in AlmE adenyltransferase activity under our standardized pyrophosphatase-coupled assay conditions. AlmE activity requires substrates glycine and ATP. Activity measurements represent the percent conversion of ATP (1 mM) supplied in the assay, which was allowed to proceed to completion (4 h).

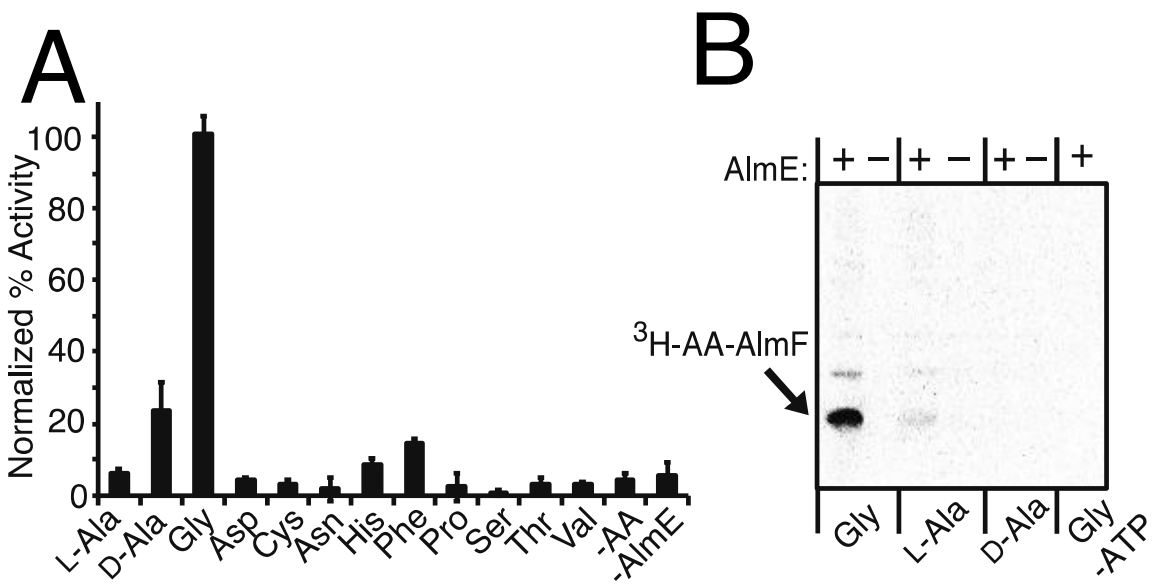


Figure 2-11. AlmE can efficiently adenylate glycine and to a lesser extent D-alanine. However, only glycine is efficiently transferred onto holo-AlmF under the described *in vitro* conditions. **(A)** Various amino acids (1 mM) were provided as substrates in pyrophosphatase-coupled AlmE (800 nM) assays. Control reactions are included where no amino acid was provided (–AA) or no enzyme was provided in the reaction (–AlmE). Percent activity was determined within the observed linear range of activity (2 h), and normalized to glycine-containing reactions (~35% activity overall). Activity measurements represent the percent conversion of ATP (1 mM) supplied in the assay. Error bars denote the standard deviation of reactions as performed in triplicate. **(B)** Destabilizing urea-PAGE analysis of holo-AlmF, incubated with AlmE and the indicated ^{3}H-amino acid *in vitro*, shows that AlmE can efficiently transfer activated glycine onto holo-AlmF, with trace activity observed when ^{3}H-L-alanine is provided and no transferase activity when ^{3}H-D-alanine is provided.

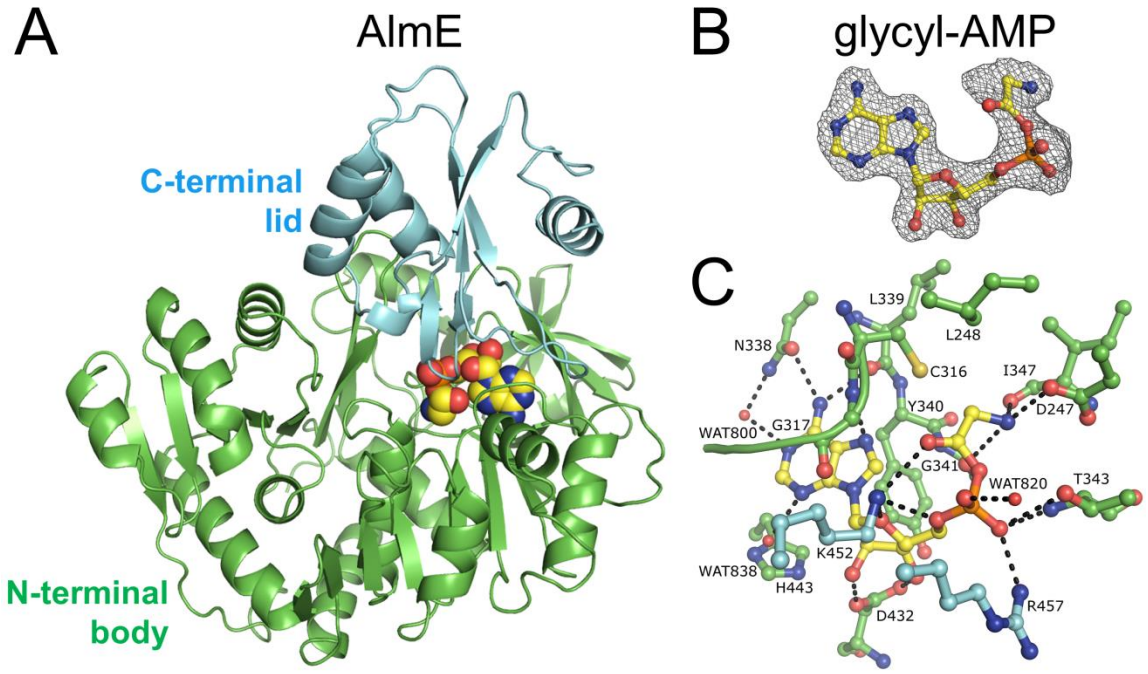


Figure 2-12. Crystal structure of AlmE with bound glycyl-AMP. **(A)** Cartoon representation of AlmE with glycyl-AMP (yellow spheres) bound between the N-terminal body (green) and C-terminal lid (cyan). AlmE is in the thioester-forming conformation in which the C-terminal lid has rotated ~140° after forming the aminoacyl-adenylate intermediate. In this conformation, AlmE is prepared to transfer glycine from glycyl-AMP to the Ppant arm of carrier protein AlmF. **(B)** F_o-F_c electron density omit map of glycyl-AMP, contoured at 3 r.m.s.d. **(C)** Binding mode of glycyl-AMP (yellow). The adenine ring rests between His443, Ile365 (not pictured), and conserved Tyr340 on one side, and the Gly317-containing loop on the other. Residues contributed by the N-terminal body are green and those by the C-terminal lid are cyan. Hydrogen bonds are shown as dashed black lines. Some main chain atoms are hidden for clarity.

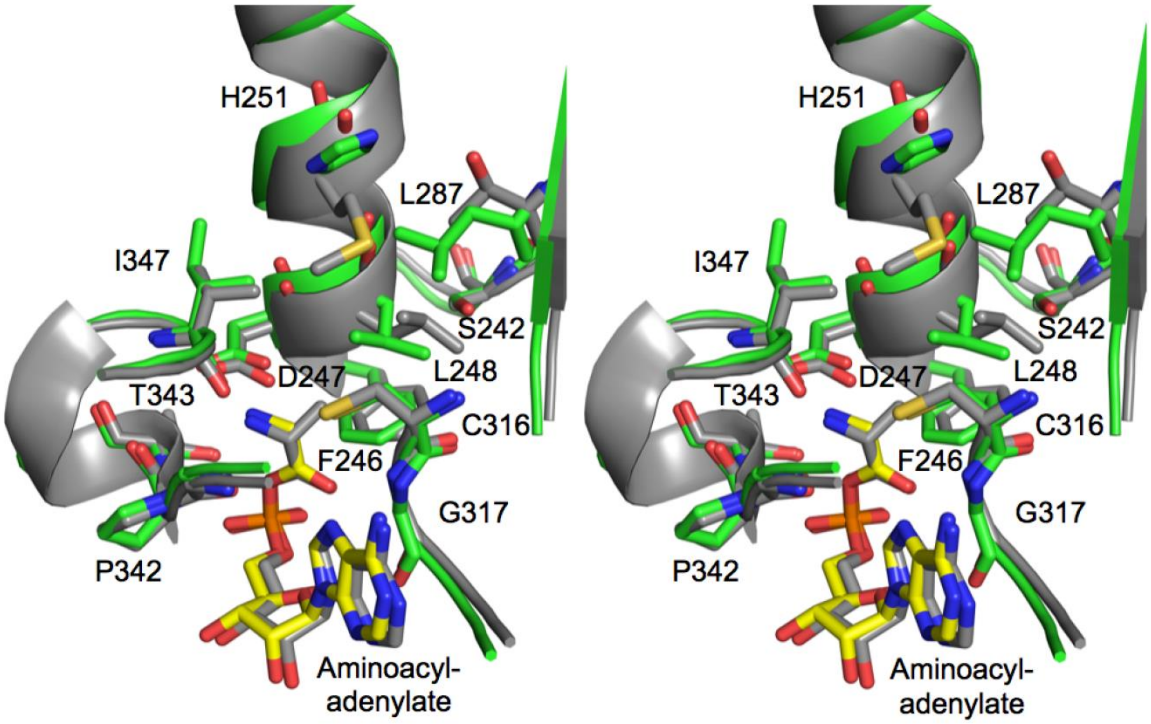


Figure 2-13. Stereoview of the active sites of AlmE (green, with the glycyl-adenylate in yellow) and *Bacillus cereus* DltA (grey, with the D-alanyl-adenylate also in grey) superposed. Small differences in the positions and identities of residues near the aminoacyl groups presumably result in the different specificities of the two enzymes. For example, S242, F246, and L287 affect the position of L248, which sterically clashes with side chains of D-aminoacyl groups in AlmE. S242 and L287 are substituted with A192 and T240 in DltA; additionally, H251 is replaced with M201. For clarity, only residues and secondary structural features near the aminoacyl groups are shown. AlmE residues are labeled; corresponding DltA residues include A192 (S242), F196 (F246), D197 (D247), L198 (L248), M201 (H251), T240 (L287), C269 (C316), G270 (G317), P296 (P342), T297 (T343), and V301 (I347). The N-terminal domains of these enzymes overlap with 1.59 Å r.m.s.d. over 289 C_α atoms.

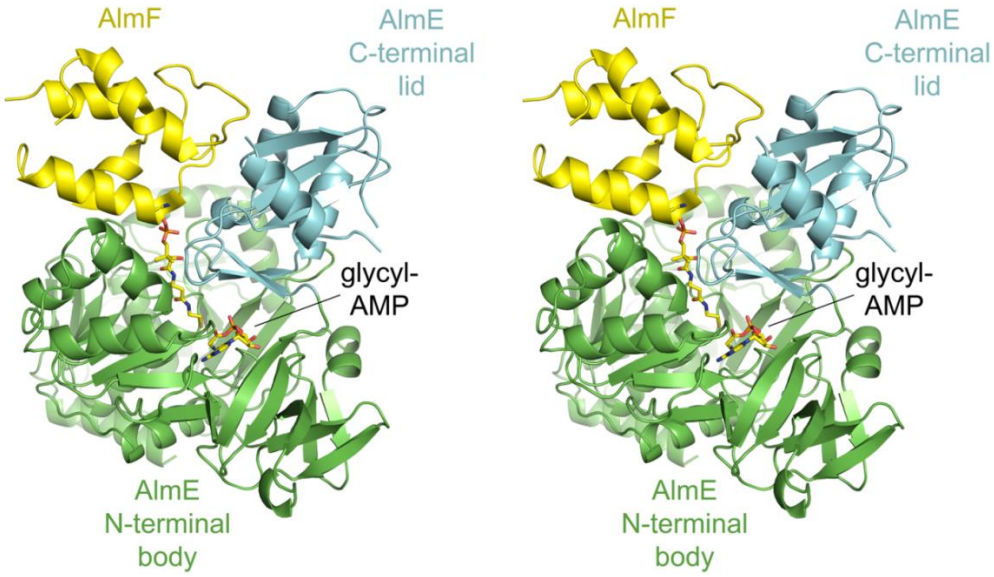


Figure 2-14. Modeling the AlmE/AlmF complex. X-ray structures of NRPS adenylation domains in complex with their cognate carrier proteins have been obtained utilizing an irreversible inhibitor and fusion proteins (Sundlov *et al.*, 2012; Mitchell *et al.*, 2012). These structures revealed the major sites of interaction between the adenylation domain/carrier protein pairs. AlmE and AlmF were modeled according to the EntB-EntE pair (PDB code 3RG2), as they are expected to associate in a similar fashion. The AlmE structure was superposed with the adenylation domain EntE (2.43 Å r.m.s.d. over 366 C_α atoms) and the Phyre2 model of AlmF was superposed with the carrier protein EntB (2.88 Å r.m.s.d. over 57 C_α atoms). Aside from the 4′-phosphopantetheinyl arm, the bulk of the interactions are expected to involve the nucleophilic helix (e.g., residues Phe35, Ile38, Asp39, Lys43, and Glu45) and portions of the loop N-terminal to it (e.g., residues Gln14-Ser16, Glu18, Phe29-Asp33) of AlmF. The 4′-phosphopantetheine arm of AlmF is expected to bind AlmE in a mode analogous to the pantetheinyl moiety observed bound to the *Streptomyces coelicolor* malonyl-CoA synthetase MatB (PDB code 3NYQ, MatB also crystallized in the thiolation conformation). After superposing AlmE and MatB, the positions of backbone atoms that form hydrogen bonds with the 4′-phosphopantetheinyl arm are very similar: Gly455 CO (binds NP1) and Ala454 CO (binds NP2 or OP3). The side chain of Ser242 may rotate to form a hydrogen bond with OP1 (occupied by water in MatB); those of Asp243 and Gln456 likely rotate as well for polar interactions. Lys271 is positioned to form a hydrogen bond with OP2. Tyr241 and Val289 may also contribute steric bulk. In order for the 4′-phosphopantetheinyl arm to reach the aminoacyl-adenylate, the phenyl ring of Phe246 must rotate to match the position of the imidazole ring of His187 of MatB, as in its observed position it obstructs the path of the incoming thiol. Phe246 is in such a conformation in *B. subtilis* DltA (PDB code 3E7W). Lys522 may replace Arg461 of MatB in binding the 4′-phosphoryl group of the coenzyme.

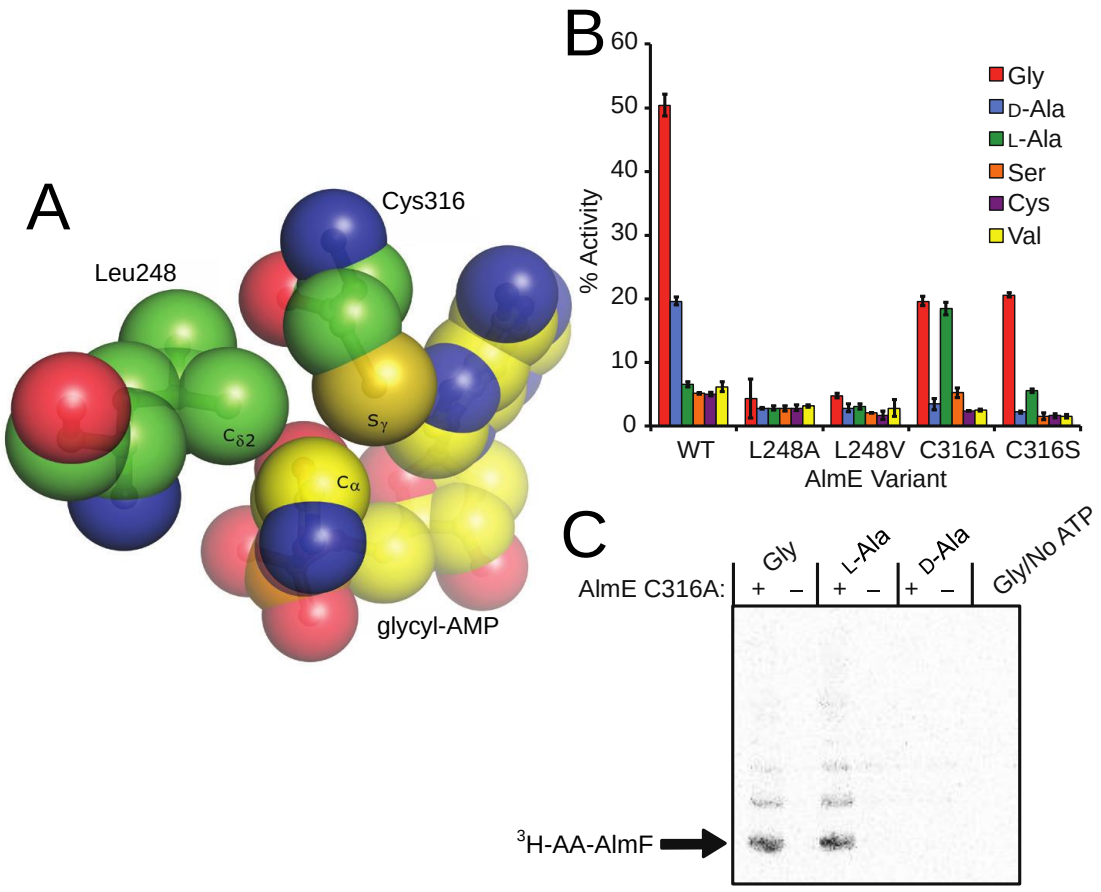


Figure 2-15. A rationally designed AlmE Cys316Ala variant shows relaxed specificity for L-alanine. **(A)** The side chains of Cys316 and Leu248, positioned above C_α of the glycyl-adenylate in the image, are hypothesized to exclude large amino acid chains via steric hindrance. Mutation of Cys316 and Leu248 to smaller amino acids might allow for relaxed specificity toward L- or D- amino acids, respectively. **(B)** All single-site AlmE variants display modest reduction in activity towards glycine; however, AlmE Cys316Ala shows relaxed specificity for adenylation of L-alanine. AlmE variants (800 nM) were incubated with small amino acid substrates (1 mM) under standard pyrophosphatase-coupled assay conditions (see Materials and Methods). Activities were measured in the linear range of wild-type AlmE for glycine. Activity measurements indicate the percent conversion of ATP (1 mM) supplied in the reaction, taken at 2.5 h. Error bars represent the standard deviation of reactions, measured in triplicate. **(C)** Destabilizing urea-PAGE analysis of holo-AlmF incubated with AlmE Cys316Ala and the indicated ^{3}H-amino acid in vitro shows that the Cys316Ala variant can efficiently transfer activated glycine or L-alanine to holo-AlmF.

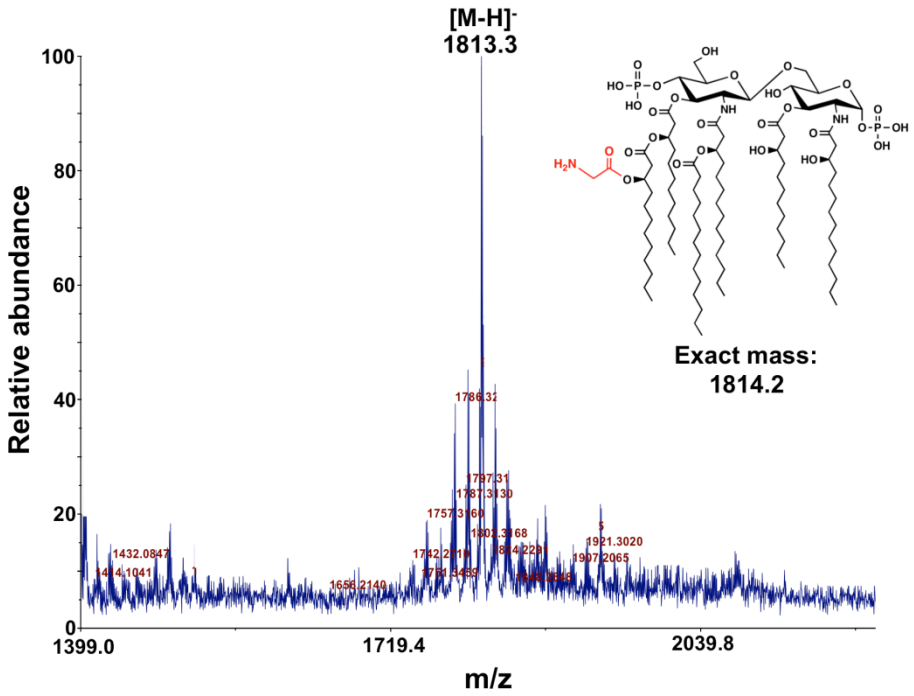


Figure 2-16. Lipid A was isolated from the C6706 *almE* mutant (*vc1579*::Tn5) expressing the AlmE C316A variant. The lipid A was analyzed by MALDI-TOF mass spectrometry and showed glycine addition to lipid A as reported for wild type *V. cholerae* El Tor strains (Hankins *et al.*, 2012).

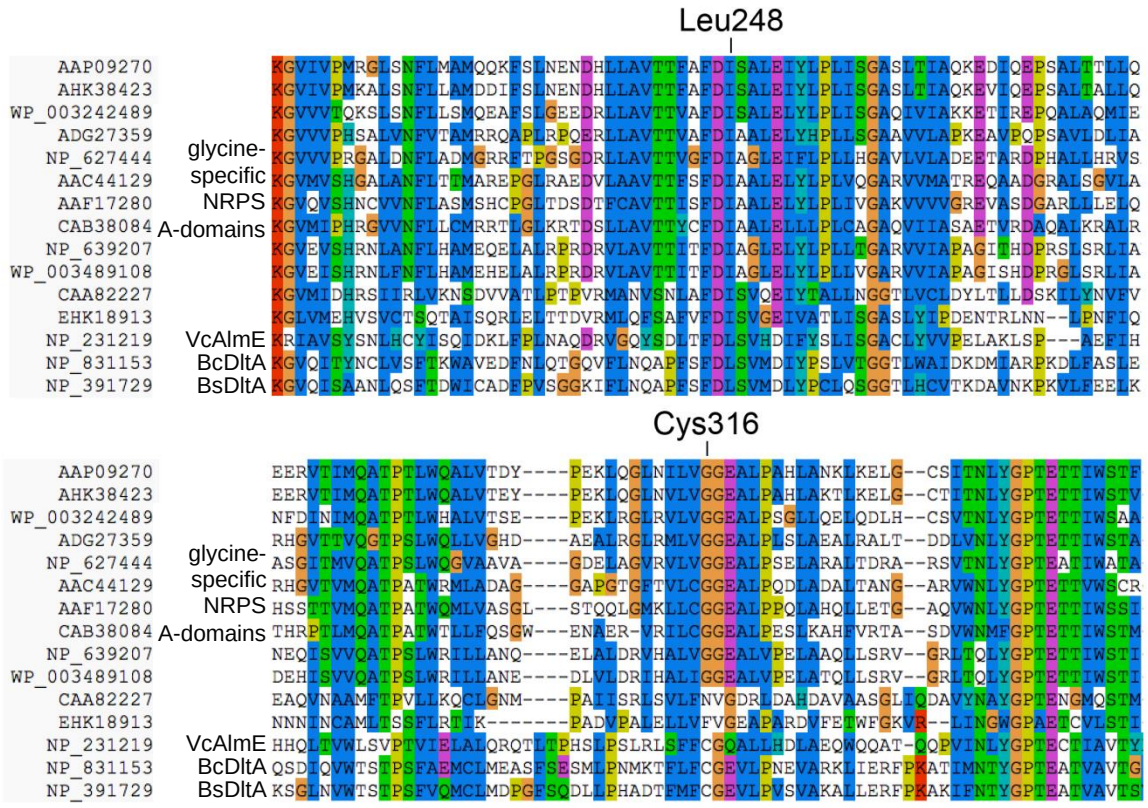


Figure 2-17. Sequence alignment of a key specificity-encoding region of glycine-specific A-domains (Rausch *et al.*, 2005), AlmE, and DltAs (*B. subtilis* and *B. cereus*). Although the chemical steps employed by the AlmEFG trio resemble NRPSs, AlmE is more closely related to DltA enzymes than the A-domains of NRPSs. The two residues closest to the C_α of the aminoacyl-adenylate in the glycine-specific AlmE, Leu248 and Cys316, are conserved in the D-alanine-specific DltAs but are replaced by isoleucine and glycine in glycine-specific A-domains from NRPSs.

Table 2-1. Crystallographic data and refinement statistics for AlmE.

Data collection	
Space group	$P3_121$
Cell dimensions	
a, *b*, *c* (Å)	116.2, 116.2, 99.6
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	50-2.26 (2.30-2.26)
R_{merge}	0.107 (0.916)
$I/\sigma(I)$	16.6 (3.6)
No. of reflections	36632 (1800)
Completeness (%)	99.9 (99.5)
Redundancy	10.7 (10.0)
Refinement	
Resolution (Å)	50-2.26 (2.32-2.26)
R_{work}/R_{free}	0.166/0.193 (0.207/0.238)
No. of atoms	
Protein	4214
Ligand	27
Water	157
Average *B*-factors (Å^2)	
Protein	33.9
Ligand	28.6
Water	33.5
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.085
PDB code	4OXI

Values in parentheses refer to the highest-resolution shell.

Table 2-2. Minimal inhibitory concentration (MIC) determination.

Mueller-Hinton Plates + 1mM IPTG	**Polymyxin B MIC (μg/mL)**	**Polymyxin E MIC (μg/mL)**
C6706	128	128
C6706 Vc1579::Tn5 pQlinkN	1	1
C6706 Vc1579::Tn5 pQlinkN AlmE	128	96
C6706 Vc1579::Tn5 pQlinkN AlmE C316A	128	128

Table 2-3. Strains and plasmids utilized in this study.

Strain or plasmid	Genotype or description	Source or reference
Strains		
Vibrio cholerae		
E7946	O1 El Tor biotype	Hankins *et al*., 2012
N16961	O1 El Tor biotype	S. Payne, University of Texas, Austin, TX
Escherichia coli		
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q}*ZΔ*M15::Tn*10* (Tet^{r})]	Stratagene
BL21(DE3)	Protein expression strain, F^{-} *omp*T *hsd*S$_{B}$ (r$_{B}^{-}$ m$_{B}^{-}$) *dcm*, *gal* λ(DE3)	Novagen
BL21(DE3)pLysS	Protein expression strain, F^{-} *omp*T *hsd*S$_{B}$ (r$_{B}^{-}$ m$_{B}^{-}$) *dcm gal* λ(DE3) pLysS Cam^{r}	Novagen
Plasmids		
pET28b	IPTG-inducible expression vector, Kan^{r}	Novagen
pQlinkN	IPTG-inducible expression vector, Amp^{r}	Scheich *et al.,* 2007
pQlinkH	IPTG-inducible expression vector, N-terminal His$_{6}$ tag, Amp^{r}	Scheich *et al.,* 2007
pQlinkHVc1578	pQlinkH containing *almF*	This study
pQlinkNVc0780	pQlinkN containing *vc0780*	This study
pQlinkNVc2457	pQlinkN containing *vc2457*	This study
pQlinkNVc1579	pQlinkN containing *almE*	This study
pQVc1578-Vc0780	pQlinkH for co-expression of N-terminal-His$_{6}$-AlmF & non-tagged Vc0780	This study
pQVc1578-Vc2457	pQlinkH for co-expression of N-terminal-His$_{6}$-AlmF & non-tagged Vc2457	This study
pQVc1578-Vc2457-Vc1579	pQlinkH for co-expression of N-terminal-His$_{6}$-AlmF & non-tagged Vc2457 and Vc1579 (AlmE)	This study
pETAlmE	pET28b containing *almE*	This study
pETAlmE$_{C316S}$	C316S mutant derivative of pETAlmE	This study
pETAlmE$_{C316A}$	C316A mutant derivative of pETAlmE	This study
pETAlmE$_{L248V}$	L248V mutant derivative of pETAlmE	This study
pETAlmE$_{L248A}$	L248A mutant derivative of pETAlmE	This study

Table 2-4. Primers utilized in this study.

Restriction sites are underlined and nucleotide substitution mutations are lower case.

Name	**Sequence**
FpQlinkHVc1578	5'-GCGGCCGGATCCATGAAACAAGCCATTGAAAATATCC-3'
RpQlinkHVc1578	5'-GCCGCGAAGCTTTTATTCCCCTACTTCGCTCAC-3'
FpQlinkNVc0780	5'-GCGGCCGGATCCTTGTGTCGTTATGCCCGATCC-3'
RpQlinkNVc0780	5'-GCCGCGAAGCTTTCAAGTAGGTAAGGTGGGGAC-3'
FpQlinkNVc2457	5'-GCGGCCGGATCCATGATCGTTGGACTCGGAAC-3'
RpQlinkNVc2457	5'-GCCGCGAAGCTTTTAACGGCGTTCAAGGATCAC-3'
FpQlinkNVc1579	5'-GCGGCCGGATCCATGCCATACAATTCTGCG-3'
RpQlinkNVc1579	5'-GCCGCGAAGCTTTCATCAATGGCTTGTTTCATATG-3'
F-AlmE	5'-ATGCTTGCAGCTAGCATGCCATACAATTCTGCG-3'
R-AlmE	5'-GTACGAACGGAATTCTCATCAATGGCTTGTTTCATATGT-3'
F-AlmE$_{C316S}$	5'-GACTGAGTTTTTTCTcCGGACAAGCCTTGCT-3'
R-AlmE$_{C316S}$	5'-AGCAAGGCTTGTCCGgAGAAAAAACTCAGTC-3'
F-AlmE$_{C316A}$	5'-CGACTGAGTTTTTTCgCCGGACAAGCCTTGC-3'
R-AlmE$_{C316A}$	5'-GCAAGGCTTGTCCGGcGAAAAAACTCAGTCG-3'
F-AlmE$_{L248V}$	5'-GATTTGACTTTTGATgTTTCGGTACATGATA-3'
R-AlmE$_{L248V}$	5'-TATCATGTACCGAAAcATCAAAAGTCAAATC-3'
F-AlmE$_{L248A}$	5'-ATTTGACTTTTGATGcTTCGGTACATGATAT-3'
R-AlmE$_{L248A}$	5'-ATATCATGTACCGAAgCATCAAAAGTCAAAT-3'

Chapter 3: Crystallographic Study of the EptC Phosphoethanolamine Transferase Required for Polymyxin Resistance and Motility in *Campylobacter jejuni*

Reprinted with permission from *Acta Crystallographica Section D: Biological Crystallography*.

ABSTRACT

The food-borne enteric pathogen *Campylobacter jejuni* decorates a variety of its cell-surface structures with phosphoethanolamine (pEtN). Modifying lipid A with pEtN promotes cationic antimicrobial peptide resistance, whereas posttranslationally modifying the flagellar rod protein FlgG with pEtN promotes flagellar assembly and motility—processes important for intestinal colonization. EptC, the pEtN transferase required for all known pEtN cell-surface modifications in *C. jejuni,* is a predicted inner-membrane metalloenzyme with a 5-helix N-terminal transmembrane domain followed by a soluble sulfatase-like catalytic domain in the periplasm. We crystallized and solved the atomic structure of the catalytic domain of EptC (cEptC) to a resolution of 2.40 Å. cEptC adopts the α/β/α fold of the sulfatase protein family and harbors a zinc-binding site. We observed a phosphorylated T266 residue, hypothesized to mimic a covalent pEtN-enzyme intermediate. The requirement of T266 as well as nearby residues N308, S309, H358, and H440 was ascertained via *in vivo* activity assays on mutant strains. Our results establish a basis for the design of pEtN transferase inhibitors.

INTRODUCTION

Bacteria maintain a complex outer surface that gives rise to a variety of biological functions. *O*-linked phospho-form modifications such as phosphoethanolamine (pEtN), phosphoglycerol, and phosphocholine are found on a variety of surface molecules, where they promote biosynthesis and pathogenic functions in certain organisms. For example, pEtN modifications are commonly associated with lipopolysaccharide/lipooligosaccharide (LPS/LOS) in Gram-negative bacteria and have dedicated roles in the recognition and evasion of host innate immune responses (Raetz *et al.*, 2007; Trent *et al.*, 2006; Needham & Trent, 2013). More recently, posttranslational phospho-form modifications of the type IV pili subunits of pathogenic *Neisseria* species and the flagellar rod protein FlgG of *Campylobacter jejuni* have extended these modifications to cell-surface proteins (Cullen & Trent, 2010; Chamot-Rooke *et al.*, 2011; Naessan *et al.*, 2008).

Our laboratory previously identified the EptC protein, a unique pEtN transferase in *C. jejuni* (Cullen & Trent, 2010). *C. jejuni* is a Gram-negative enteric pathogen responsible, in large part, for food-borne illness and bacteria-associated diarrhea in the United States and abroad (Scallan *et al.*, 2011; Havelaar *et al.*, 2013). In terms of burden to health and cost, campylobacteriosis ranks highest among food-borne infections (Batz *et al.*, 2012). Contracted by ingestion of contaminated poultry and beef products, the disease is characterized by gastrointestinal colonization followed by acute diarrhea. In rare instances, secondary neurological symptoms may persist in the form of the debilitating autoimmune disease Guillain-Barré syndrome (Yuki *et al.*, 2012; Israeli *et al.*, 2012).

C. jejuni has evolved a single pEtN transferase, EptC, that modifies an array of cell-surface molecules, including the 1- and 4′-phosphoryl groups of lipid A (Cullen & Trent, 2010), the first heptose residue of the inner-core oligosaccharide of LOS (Cullen *et al.*, 2013), residue T75 of the flagellar rod protein FlgG (Cullen *et al.*, 2012), and *N*-linked glycans of numerous glycoproteins (Scott *et al.*, 2012) (**Figure 3-1**). Conversely, most mucosal pathogens appear to express several pEtN transferases dedicated to the modification of one specific target. For example, pathogenic *Neisseria* species produce the enzymes LptA, Lpt3, Lpt6, and PptA that specifically transfer pEtN to only one of the following: lipid A phosphoryl groups (Cox *et al.*, 2003), position 3 of LOS core heptose II (Mackinnon *et al.*, 2002; Wenzel *et al.*, 2010), position 6 of LOS core heptose II (Wright *et al.*, 2004), or the major pilin protein PilE (Naessan *et al.*, 2008), respectively. Likewise, *E. coli* species possess at least three functional pEtN transferases (EptA, EptB, and CptA) that attach pEtN groups to specific positions of LPS (Raetz *et al.*, 2007). Due to its uniquely promiscuous nature, EptC is linked to multiple and distinct biological functions in *C. jejuni*. Mutational studies demonstrate that loss of pEtN modifications cause drastic reduction in flagellar assembly/motility and significantly higher sensitivity to a range of cationic antimicrobial peptides (Cullen & Trent, 2010; Cullen *et al.*, 2012). Most importantly, EptC expression is required for proper intestinal colonization of *C. jejuni* in avian and mouse infection models (Cullen *et al.*, 2013).

The *eptC* gene (locus tag Cj0256) is clustered in a family of inner-membrane metalloenzymes (COG2194) containing a 5-helix transmembrane domain and a periplasmic catalytic domain that is currently grouped in the sulfatase family (pfam00884) (Marchler-Bauer *et al.*, 2013) (**Figure 3-1**). Placement of the catalytic domain dictates that pEtN modification occurs on the periplasmic side of the inner

membrane, where EptC transfers head groups from the phospholipid phosphatidylethanolamine to target molecules. Indeed, EptC activity is not observed when the enzyme is heterologously expressed in a phosphatidylethanolamine-deficient *E. coli* strain (Cullen & Trent, 2010). To gain insight into the structural basis of enzyme function, we crystallized and determined the atomic coordinates of the catalytic domain of EptC (cEptC) to a resolution of 2.40 Å. Our structural data along with mutational analysis of EptC allowed us to identify zinc-ligand residues, a putative nucleophile, and conserved active-site residues required for *in vivo* activity. Moreover, we juxtaposed the structures of cEptC and the catalytic domains of the related enzymes cLtaS (lipoteichoic acid synthase from *Bacillus subtilis* and *Staphylococcus aureus*) and cLptA (lipid A pEtN transferase from *N. meningitidis*). Observed structural differences around the active sites of cEptC and the lipid A-specific cLptA may provide clues to the promiscuous nature of EptC.

RESULTS AND DISCUSSION

Overall structure of the catalytic domain of EptC

In order to examine the structural basis for enzyme function, we overexpressed the catalytic domain of EptC (cEptC, residues 203-512) fused to a C-terminal Gly_6His_8-tag and determined the X-ray crystal structure (**Figure 3-2**). After testing many gene truncations, we obtained one that expressed enough soluble protein suitable for crystallography, and confirmed the molecular weight of the gene product (**Figure 3-3**). Including the His-tag, cEptC is composed of 325 residues and has a molecular weight of 36.7 kDa. The phases of cEptC were elucidated from a SAD dataset, collected from a SeMet-derived cEptC crystal that diffracted to 2.80 Å in

space group $P2_12_12_1$ (**Table 3-3**). Molecular replacement was carried out using one monomer from the SeMet derivative as a search model for the native dataset. One native crystal diffracted to 2.40 Å in space group $C222_1$. With the exception of residues 213-215 of chains A and C and residues 212-214 of chain B, there is full coverage of the peptide backbone by the electron density map between the initiating M202 and P512 (preceding the C-terminal His-tag). Elevated *B*-factors (overall average = 49.8 $Å^2$), particularly for the lower half of the structure as arranged in **Figure 3-2**, indicate increased mobility in atomic positions—possibly due to the absence of the N-terminal transmembrane region of the enzyme or to the dearth of crystal contacts in this region of each chain in the asymmetric unit.

The cEptC catalytic domain assumes an α/β/α fold with a 6-stranded, central β-sheet that is enveloped by a series of 11 α-helices (**Figure 3-2**), imparting an overall hemispherical shape. β-strands and α-helices are labeled as β1 to β8 and α1 to α11, respectively (**Figure 3-4**). The α/β/α fold is common to proteins in the alkaline phosphatase superfamily (pfam00245), which includes sulfatases. Enzymes of this family cleave ester bonds such as sulfoester and phosphoester linkages, a reaction likely required for EptC function. A zinc-binding site, identified by anomalous diffraction, is located at the flat surface of the hemisphere and is fairly exposed to solvent (**Figure 3-5A**). As the putative nucleophile T266 resides at the N-terminus of α2, a helical dipole may help to stabilize a nucleophilic alkoxide form of its side chain. A phosphoryl group was found attached to T266 to form phosphothreonine (**Figure 3-6**), which we speculate mimics a pEtN-enzyme intermediate. Consistent with the location of cEptC in the oxidizing environment of the periplasm, three disulfide bonds are present (**Figure 3-2**): between residues C262 and C272, C312 and C316, and C377 and C385. C272 is located on the

nucleophilic helix, while C262 is on the loop extending from the N-terminus of that helix. C312 resides on a loop near the nucleophilic T266 and, with C316, may play a role in properly positioning an active-site loop as discussed below. In contrast, the C377-C385 pair is more distant from the active site.

Comparison of cEptC with known phospho-form transferase structures

Phospho-form transferase enzymes transfer phospho-form structures such as pEtN and phosphoglycerol to hydroxyl and phosphoryl groups of target molecules to promote their biological functions. In addition to the cEptC structure presented here, the structures of two other phospho-form transferase proteins have been previously reported: the catalytic domain of *S. aureus* (Lu *et al.*, 2009a) and *B. subtilis* (Schirner *et al.*, 2009) lipoteichoic acid synthase LtaS (cLtaS; PDB 2W5S and 2W8D), and most recently the catalytic domain of the lipid A pEtN transferase LptA (cLptA; PDB 4KAY) of *N. meningitidis* (Wanty *et al.*, 2013). LtaS, a phosphoglycerol transferase distantly related to EptC, synthesizes the polyphosphoglycerol backbone structure of lipoteichoic acid in Gram-positive bacteria. LptA is functionally similar to EptC in that it adds pEtN to the phosphoryl groups of lipid A; however, LptA displays specificity for lipid A alone, while EptC has multiple molecular targets. Like cEptC, the C-terminal domains cLptA and cLtaS adopt the α/β/α alkaline phosphatase-like fold (**Figure 3-7**), which is fused to a predicted N-terminal 5-helix transmembrane domain. Together, these three proteins represent a class of bacterial membrane-associated, biosynthetic enzymes with similar activities (phospho-form transfer), although the products and biological outcomes are widely varied.

The cEptC and cLtaS structures share a number of common core β-strands and α-helices that align with a C_α r.m.s.d. of 3.74 Å over 216 residues (**Figures 3-7A** and **3-7C**; **Figures 3-5A** and **3-5C**). The main differences between the two structures are the positioning of loop regions and the presence of insertions only observed in cLtaS. For instance, cLtaS has an extra 79-residue C-terminal region composed of a 5-stranded β-sheet and a long α-helix. Additionally, a helix-loop-helix structure spanning residues 477-500 corresponds to a loop region between β6 and β7 of cEptC. Interestingly, this loop in the cEptC structure is oriented towards the active site and positions H440-$N_{\varepsilon 2}$ within hydrogen bonding distance (3.0 Å) of a phosphoryl oxygen atom of phospho-T266 (**Figures 3-7A** and **3-8A**). In contrast, the helix-loop-helix structure of cLtaS is oriented in the opposite direction, causing this region of the active site to be more open in cLtaS (**Figures 3-7C** and **3-8C**). Finally, LtaS contains an additional helix positioned at a loop region between the corresponding β2 and β3 of cEptC (**Figures 3-7A** and **3-7C**; **Figures 3-8A** and **3-8C**). Side chains on the N-terminal end of this helix and the preceding loop were shown to bind phosphoglycerol in *S. aureus* cLtaS (PDB 2W5S; Lu *et al.*, 2009a). The major structural differences observed between cEptC and cLtaS presumably contribute to their distinct enzyme-substrate interactions.

The cEptC and cLptA structures are strikingly similar as noted by their 44% amino acid sequence identity and observed C_α r.m.s.d. of 0.59 Å over 211 residues (**Figures 3-7A** and **3-7B**; **Figures 3-5A** and **3-5B**). The disulfide bonding pattern of cLptA is identical to that of cEptC, with the exception of two additional disulfides between residues C348-C353 and C499-C540 of cLptA. Interestingly, these extra disulfides occur in the only regions that differ significantly from cEptC. C499 is part of an α-helix that corresponds to α10 of cEptC. Also, cLptA contains an extra 8-

residue α-helix between the corresponding β8 and α10 of cEptC that, together with the C499-C540 disulfide, causes α10 to take on a different pitch than observed in cEptC. The second additional disulfide between C348 and C353 of cLptA is in a loop spanning residues 339-354 adjacent to the active site. Curiously, this loop is significantly shorter in the cEptC structure (residues 325-329), despite the similarity of sequence and structure with cLptA around this region. It is possible that this extended loop plays a part in substrate specificity in cLptA, and its absence from the cEptC structure may allow for entry of more diverse molecular targets.

Active-site architecture of cEptC

The cEptC active site contains a tetrahedrally coordinated zinc ion and a putative nucleophilic threonine, T266, covalently bound to a phosphoryl group to form phosphothreonine (**Figure 3-8A**). We reason that phospho-T266 mimics an enzyme reaction intermediate representing the first step in a two-step transferase reaction. Phosphothreonine was also observed in the cLptA (Wanty *et al.*, 2013) and *B. subtilis* cLtaS (Schirner *et al.*, 2009) structures (**Figures 3-7B** and **3-7C**), but not the *S. aureus* cLtaS (Lu *et al.*, 2009a) structure (**Figure 3-8C**).

Coordinated divalent metals are important cofactors for enzymes in the alkaline phosphatase protein superfamily, as they are implicated in substrate binding and catalysis (Kim & Wyckoff, 1991; O'Brien & Herschlag, 2001; Stec *et al.*, 2000). The cEptC zinc ion is tetrahedrally coordinated by residues E227, phospho-T266, D427, and H428. By sequence alignment and structural comparison, these residues are highly conserved in the characterized phospho-form transferase members of COG2194 (includes pEtN transferases) and COG1368 (includes phosphoglycerol transferases) (**Figure 3-9**). The orientation of these residues is

well conserved between cEptC, cLptA, and cLtaS, with r.m.s.d. values of 0.26 Å over 30 atoms and 0.18 Å over 21 atoms when comparing cEptC to cLptA and *B. subtilis* cLtaS, respectively (**Figure 3-10**). Of the two cLtaS structures solved, one was shown to bind a magnesium ion (Schirner *et al.*, 2009) and the other a manganese ion (Lu *et al.*, 2009a) (compare **Figures 3-10C** and **Figure 3-8C**). The cLptA structure, on the other hand, contains two zinc ions (**Figure 3-8B**), reminiscent of the *E. coli* alkaline phosphatase PhoA (PDB 1ED8) (Wanty *et al.*, 2013; Stec *et al.*, 2000). It is worth noting that only one zinc ion, corresponding to the zinc and magnesium/manganese ions of cEptC and cLtaS, was observed unless the cLptA crystals were soaked with 1 mM zinc sulfate. Nevertheless, the zinc-coordinating residues of cLptA superimpose well onto corresponding residues of alkaline phosphatase (Wanty *et al.*, 2013) and, importantly, most of these residues are conserved in cEptC, cLtaS, and several diverse pEtN transferases (**Figure 3-9**). More specifically, the first-zinc ligand residues E227/240, T266/280, D427/452, and H428/453 of cEptC/cLptA directly correspond to D51, S102, D369, and H370 of alkaline phosphatase, and the second-zinc ligand residues H358/383 and H440/465 correspond to D327 and H412 of alkaline phosphatase. Another second-zinc ligand residue, H331, is found in alkaline phosphatase, but not cLptA or cEptC. Additional studies will be necessary to determine the extent of this similarity, as well as the capacity of cEptC to bind multiple zinc ions.

Mutational characterization of conserved cEptC active-site residues

Based on a sequence alignment of COG2194 members (**Figure 3-9**), we identified conserved amino acid residues that correspond to cEptC active-site residues T266, N308, S309, H358, and H440 (**Figure 3-8A**). In order to assess the

functional requirement of these active-site residues, we prepared mutant strains of *C. jejuni* 81-176 carrying point mutations for each residue as described in the Methods section. We previously established that a functional *eptC* allele is required for pEtN modification of the flagellar rod protein FlgG and the lipid A in *C. jejuni* (Cullen & Trent, 2010; Cullen *et al.*, 2012). These modifications are unequivocally responsible for proper flagellar assembly, motility, and resistance to polymyxin B. To this end, we monitored the *in vivo* activity of mutant EptC enzymes by comparing the minimal inhibitory concentration (MIC) of polymyxin B and the halo-of-growth diameter in motility assays. All of the active-site mutants were significantly less resistant to polymyxin B when compared with parent strain 81-176 (**Figure 3-11**). The S309A mutant displayed a 6-fold decrease in polymyxin B resistance, while the remaining active-site mutants and Δ*eptC* null mutant displayed a 15.8-fold decrease. All of the mutants were also significantly impaired with respect to motility (**Figure 3-12**). Interestingly, the N308A, S309A, H440A and H358A, and T266S mutants displayed the most severe motility defects, while the T266A mutant was comparable to the Δ*eptC* null strain after 24 h of growth. It was previously determined that roughly 10% of the Δ*eptC* mutant population produces visible flagella, (Cullen & Trent, 2010), giving rise to partial recovery of motility as demonstrated in **Figure 3-12** (compare 81-176, Δ*eptC,* and the non-motile strain H358A as a reference). The more pronounced motility defects observed for most of the active-site mutants presented here may reflect the fact that these strains express inactive protein that precludes partial recovery of motility via unknown mechanisms.

Taken together, the loss of motility and polymyxin B resistance suggest a crucial role for T266, N308, S309, H358, and H440 in EptC function. It is possible that H358 and H440 (1) coordinate a second zinc ion (similar to cLptA and PhoA) or

(2), along with N308 and S309, bind substrate (similar to cLtaS). Notably, N308-$O_{\delta 1}$ and -$N_{\delta 2}$ and S309-O_{γ} are 5.5, 6.7, and 6.4 Å from phospho-T266 O3P (chain A), respectively, which places them within hydrogen bonding distance of a pEtN moiety covalently linked to T266.

To further explore the phosphorylation of T266, we expressed and purified soluble cEptC protein that contained a T266S mutation and investigated the presence of phosphoserine. The molecular weights of native and mutant cEptC proteins were determined by mass spectrometry collected in the positive mode with an accuracy of +/-5 u. The deconvoluted cEptC spectrum showed a major peak at 36750 u, consistent with the calculated $[M+H]^+$ of native cEptC (**Figure 3-3A**). As expected, another major peak was observed with a mass shift of +80 u, demonstrating the presence of phosphothreonine as observed in the cEptC crystal structure. Corresponding peaks with masses 36477 u and 36556 u suggests the loss of two histidine residues (-273 u), likely due to proteolysis of the disordered His-tag during expression and purification. Contrary to native cEptC, the T266S mutant spectrum showed one major peak at 36740 u, consistent with unmodified protein, and a corresponding minor peak indicating the loss of two histidine residues (-272 u) (**Figure 3-3C**). Taken together, these results suggest that serine cannot replace threonine for EptC catalysis and that indeed the phosphothreonine observed in the cEptC crystal structure and mass spectrum is most likely due to the catalytic activity of cEptC.

CONCLUSION

We present the structure of the catalytic domain of EptC, the only known phosphoethanolamine transferase from the Gram-negative pathogen *C. jejuni*. Despite the wide array of molecules modified by EptC, our structure displays extensive similarities to the lipid A-specific cLptA of *N. meningitis* and fewer similarities to the distantly related, phosphoglycerol-specific cLtaS of *B. subtilis* and *S. auerus*. Like cLptA and cLtaS, cEptC assumes an α/β/α fold that coordinates a divalent metal ion and possesses a nucleophilic threonine modified with a phosphoryl group. We suggest that this residue mimics a covalent phosphoethanolamine-enzyme intermediate. Given the similarities among the active sites of these phospho-form transferases and alkaline phosphatase, the enzymes are expected to share mechanistic features. *C. jejuni* strains with point mutations at EptC residues T266, N308, S309, H358, and H440 suffered polymyxin B sensitivity and severe loss of motility, indicating that the active sites of EptC and other lipid A phosphoethanolamine transferases may be effective targets for drug design. *C. jejuni* has been increasingly recognized as a major causative agent of food-borne illness worldwide. Our analysis of the catalytic domain of EptC augments the structural characterization of phosphoethanolamine transferases and sets the foundation for the development of inhibitory drugs to combat infections of *C. jejuni* and related bacteria.

MATERIALS AND METHODS

Strains and plasmid construction

Primers, plasmids, and strains used in this study are listed in **Tables 3-1** and **3-2**. Plasmid pETcEptC was constructed for the overexpression of the catalytic domain of EptC (cEptC, residues 203-512) fused to a Gly_6His_8-tag at the C-terminus, using the Polymerase Incomplete Primer Extension (PIPE) method (Klock & Leslie, 2009). Briefly, the plasmid pET28b(+) (Novagen) was linearized using V-PIPE primers 1 and 2 and subsequently gel-purified and treated with restriction enzyme *Dpn*I (New England Biolabs). The insert was colony PCR-amplified from *C. jejuni* 81-176 using the I-PIPE primers 3 and 4. The insert and linearized vector were mixed and transformed into XL1-blue *E. coli* (Stratagene) and incubated overnight at 37 °C on LB agar plates containing 50 μg/mL kanamycin.

EptC active-site mutants were constructed by site-directed mutagenesis of plasmids pAtsAKO::*cj0256*-Kan^R (Cullen *et al.*, 2012) using primers 5-16 (**Tables 3-1** and **3-2**). The mutant alleles were introduced into the *eptC* mutant strain *C. jejuni* 81-176 Δ*cj0256* as described previously (Cullen & Trent, 2010).

Expression, purification, and crystallization of cEptC

E. coli BL21(DE3) cells expressing cEptC were cultured in LB with 50 mg/L kanamycin at 37 °C. Upon reaching OD_{600} of ~0.6, cultures were cooled to 18 °C and induced with 0.5 mM IPTG. After 18 h, cells were pelleted, resuspended in lysis buffer (500 mM NaCl, 30 mM HEPES pH 7.5, 10% (v/v) glycerol) and sonicated. Lysate was clarified by centrifugation and the soluble portion was poured over a Ni-NTA column (Expedeon) equilibrated with lysis buffer. Beads were washed with 15 mM imidazole in lysis buffer and bound protein was eluted with 150 mM imidazole

in lysis buffer. Eluate was further purified on a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES pH 7.5. SeMet-derived protein was purified similarly, but expressed in minimal media with 50 mg/L kanamycin, supplemented with amino acids for methionine pathway inhibition (Doublié, 1997).

SeMet-derived and native proteins were concentrated to 30 mg/mL in 25 mM NaCl, 10 mM HEPES pH 7.5. Crystals were grown over the course of several days by sitting-drop vapor diffusion at 22 °C by combining 4 µL protein solution with 1 µL 200 mM diammonium phosphate, 15% (w/v) PEG 3350. SeMet-derived protein crystallized in space group $P2_12_12_1$, while native protein crystallized in $C222_1$. Before flash freezing in liquid nitrogen, crystals were cryoprotected in 200 mM diammonium phosphate, 16.5% (w/v) PEG 3350, 10% (v/v) glycerol for 10 s.

X-ray diffraction data collection, processing, phasing, and refinement

Native and SeMet derivative data were collected at Advanced Photon Source beamline 23-ID-B and processed and scaled in *HKL2000* (Otwinowski & Minor, 1997) (**Table 3-3**). Reflection intensities were converted to amplitudes in *TRUNCATE* (French & Wilson, 1978) of the *CCP4* software suite (Winn *et al.*, 2011). A single-wavelength anomalous dispersion (SAD) dataset was collected at the Se K-edge (λ = 0.9795 Å) from one SeMet-derived cEptC crystal that diffracted to 2.80 Å. Initial phases were obtained from *AutoSol* of the *PHENIX* software suite (Adams *et al.*, 2010), which identified 46 Se sites (30 expected for 6 monomers, based on sequence; figure of merit = 0.329) in a $P2_12_12_1$ unit cell and generated six partial cEptC monomers in the asymmetric unit (R_{work}/R_{free} = 32.9%/37.8%; map-model correlation coefficient = 0.747). One of the six monomers was subjected to several

cycles of refinement in *Coot* (Emsley & Cowtan, 2004) and *REFMAC5* (Murshudov *et al.*, 1997) of *CCP4*, and then employed as a molecular replacement search model to phase the native dataset in *Phaser* (McCoy *et al.*, 2007) of *CCP4*. One native cEptC crystal diffracted to 2.50 Å (Zn remote, λ = 1.0332 Å), with a $C222_1$ unit cell containing three monomers. Iterative refinement was carried out in *Coot* and *phenix.refine* of *PHENIX*, with default NCS restraints applied in the latter. Notably, for the $P2_12_12_1$ (SeMet-derived) dataset, *phenix.xtriage* of *PHENIX* identified an off-origin Patterson peak that is 30% of the origin peak, which strongly indicates the presence of translational pseudo-symmetry. No pseudo-symmetry was detected for the $C222_1$ (native) dataset.

X-ray fluorescence of native cEptC crystals was examined at Advanced Light Source beamline 4.2.2 to identify the catalytic metal ion. A dataset collected at the Zn K-edge (λ = 1.2680 Å) from one crystal diffracted to 2.40 Å; this was processed in *XDS* (Kabsch, 2010) and scaled in *SCALA* (Evans, 2006) of *CCP4*. Data were truncated at a resolution in which a compromise was reached between $I/\sigma(I)$ and $CC_{1/2}$ values (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013). Refinement was carried out as described above. The two native datasets were merged in *cad* to generate an anomalous difference map in *FFT* (Read & Schierbeek, 1988), both programs of *CCP4*. The Zn K-edge dataset was selected for deposition due to less apparent radiation damage to disulfides and zinc ions, as evident in the F_o-F_c density (**Figure 3-13**) (Weik *et al.*, 2000; Burmeister, 2000). The structure factor amplitudes and atomic coordinates of cEptC were deposited in the RCSB Protein Data Bank under PDB code 4TN0.

Protein mass spectrometry

LC-MS (AB Sciex 4000 QTrap hybrid triple quadrupole-linear ion trap with Shimadzu Prominence analytical LC) was employed in collection of mass spectra. Protein sample was eluted with a 0.5×50 mm PLRP-S column (8 μm, 4000 Å; Michrom BioResources) with mobile phase A and B (acetonitrile:water:formic acid:trifluoroacetic acid, A = 2:98:0.1:0.01; B = 90:10:0.1:0.01). A linear gradient of 5-65% B was used to elute protein in 20 min at a flow rate of 20 μL/min. Automated acquisition of full-scan MS spectra over 450-2000 Da was executed by *Xcalibur* (Thermo Scientific). The acquired convoluted protein spectra were deconvoluted by *ProMass for Xcalibur* 2.5.0 (Novatia) to yield the $[M+H]^+$ mass-to-charge values.

Motility assays

C. jejuni strains were initially grown from freezer stocks on MH agar with 10 μg/mL trimethoprim for 48 h at 37 °C in microaerobic conditions (5% O_2, 10% CO_2, 85% N_2). Following the initial growth, the strains were re-streaked and grown under the same conditions. After 16 h, the strains were resuspended from agar plates in MH broth and diluted to OD_{600} of ~0.05, stab-inoculated into semisolid MH motility medium (containing 0.4% agar), and incubated for 24 h at 37 °C in microaerobic conditions. Relative motilities were assessed from the diameter of the halos of growth. Experiments were performed in duplicate.

Polymyxin B resistance assays

Strains were grown in the same manner as described for motility assays above and swabbed onto MH agar plates with 10 μg/mL trimethoprim. Polymyxin B Etest strips (bioMérieux) were applied to the swabbed plates and incubated in

microaerobic conditions for 48 h at 37 °C before reading minimal inhibitory concentrations (MICs). Experiments were performed in duplicate.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (grants AI064184 and AI076322 to M.S.T. and grant GM106112 to A.T.K.) and the Army Research Office (grant W911NF-12-1-0390 to M.S.T.) Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank Maria Person and the Proteomics Facility at the University of Texas at Austin and funding sources ES007784 (CRED) and RP110782 (CPRIT) for producing protein mass spectrometry data. We also thank Jay Nix for assistance with

anomalous data collection and processing, and Alex Taylor and Art Monzingo for helpful discussions.

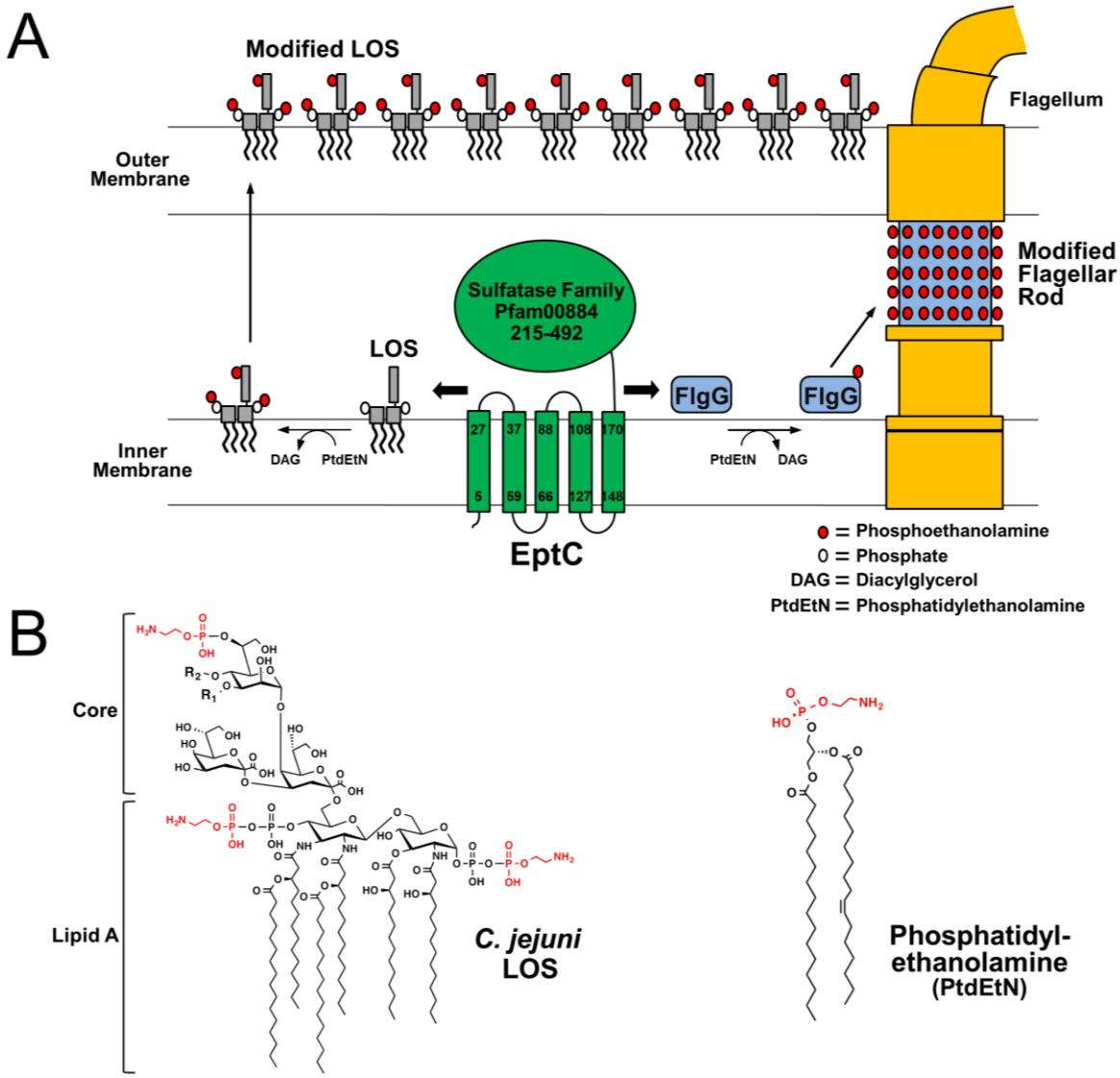


Figure 3-1. Topology and biological functions of *C. jejuni* phosphoethanolamine transferase EptC. **(A)** EptC is predicted to be anchored to the inner membrane by a 5-helix transmembrane domain, with its sulfatase-like catalytic domain positioned in the periplasm. EptC transfers pEtN (red circles) from the head groups of phosphatidylethanolamine onto the LOS core (grey rectangles) and phosphoryl groups (white circles) of lipid A and T75 of FlgG proteins (blue boxes). Modifications of these *C. jejuni* surface structures with pEtN promote flagellar assembly, motility, cationic antimicrobial peptide resistance, and host intestinal colonization (Cullen & Trent, 2010; Cullen *et al.*, 2012; Cullen *et al.*, 2013). The organization of flagellar components has been previously proposed (Chevance & Hughes, 2008). **(B)** Chemical structures of *C. jejuni* LOS and phosphatidylethanolamine, with pEtN moieties in red. R_1, β-D-glucose; R_2, outer core sugars; pEtN, phosphoethanolamine; PtdEtN, phosphatidylethanolamine; DAG, diacylglycerol; LOS, lipooligosaccharide.

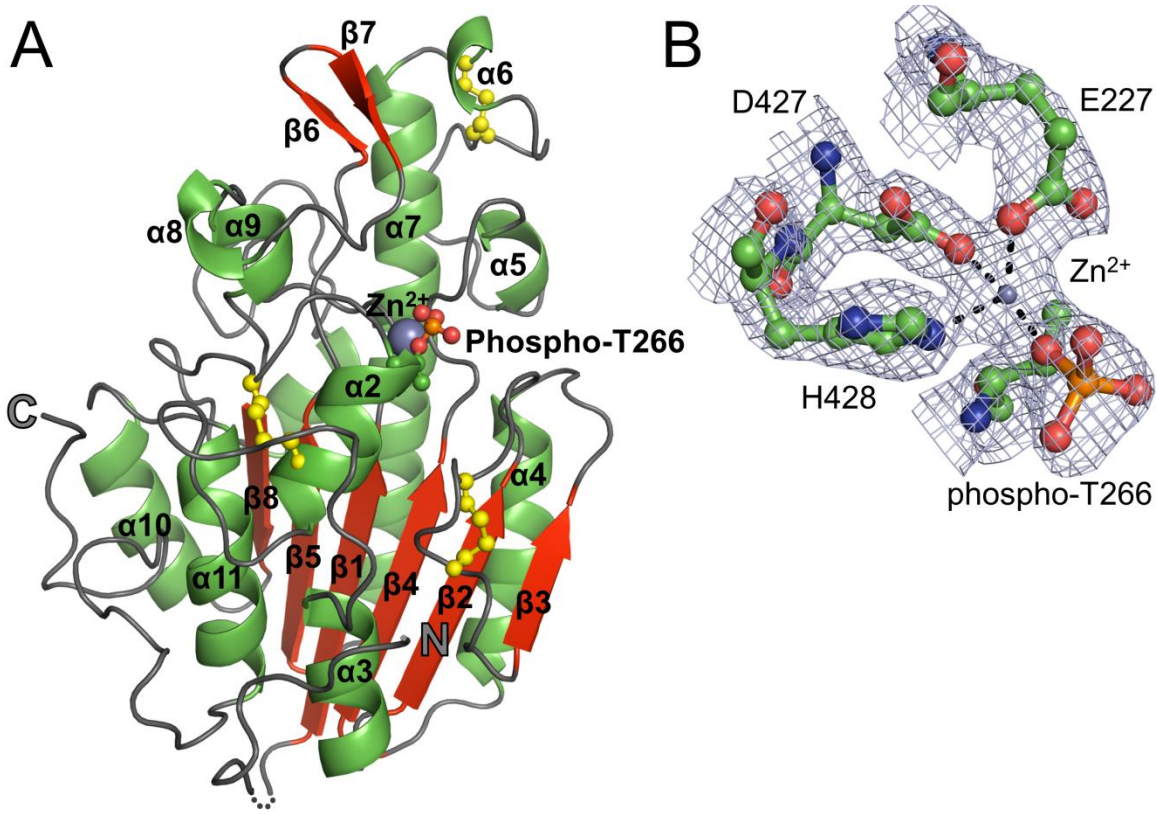


Figure 3-2. Structure of the C-terminal, catalytic domain of the pEtN transferase EptC (cEptC) from *C. jejuni*. **(A)** Secondary structural features of a monomer. α-Helices (green), β-strands (red), and loops (grey) are labeled as in **Figure 3-4**. Disulfides (yellow) and the phosphothreonine residue are displayed as ball-and-sticks; Zn^{2+} (purple) is displayed as a sphere; and the N- and C-termini are labeled. The dotted line indicates residues 213-215, for which there is no observable electron density in monomer A. **(B)** Zinc-binding residues of the phospho-enzyme intermediate. $2F_o$-F_c map density (blue mesh) is contoured to 2.0σ. The metal ion was identified as Zn^{2+} by anomalous diffraction (see **Figure 3-10**). Coordination bonds between side chains and Zn^{2+} are indicated with dashed lines. All molecular images were generated in PyMOL (DeLano).

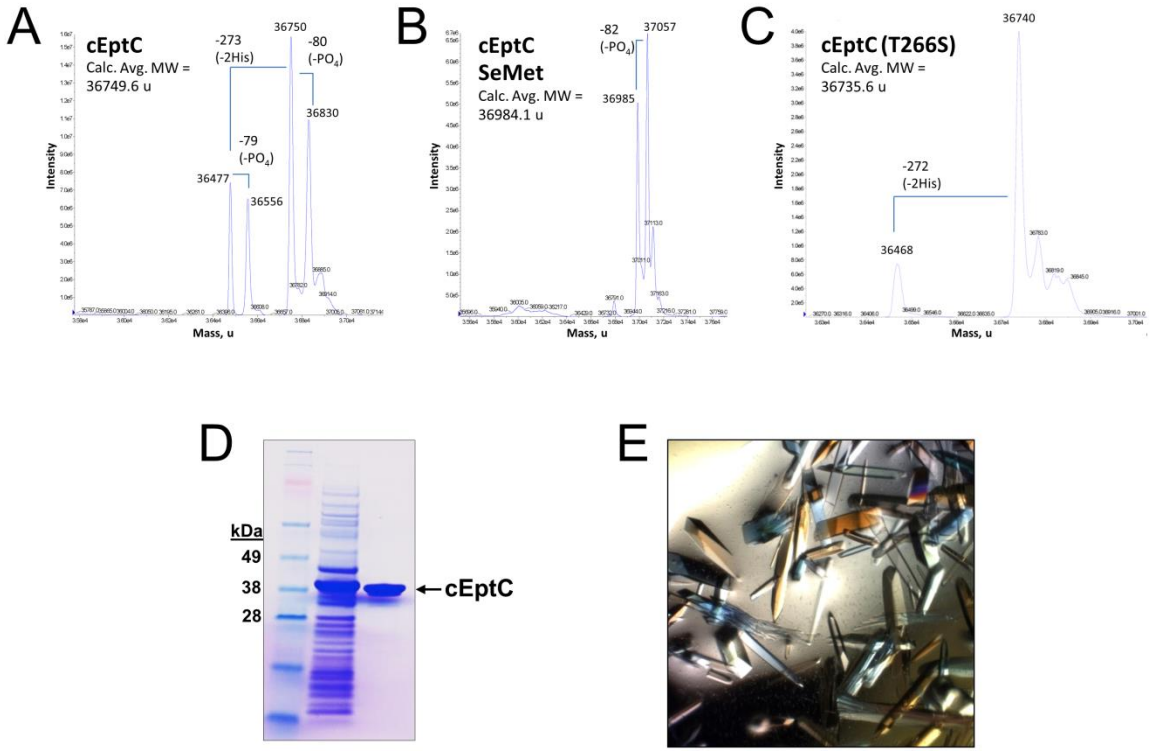


Figure 3-3. Expression and isolation of cEptC proteins. **(A)** Mass spectrum of native cEptC protein used in crystallization experiments (u, atomic mass unit). **(B)** Mass spectrum of selenomethionine-derived cEptC protein used to obtain phases. **(C)** Mass spectrum of the cEptC T266S mutant demonstrating the absence of phosphate modification. **(D)** Coomassie-stained SDS-PAGE gel showing the final purified cEptC protein; lane 1, molecular weight marker; lane 2, cell-free lysate of *E. coli* cells expressing cEptC; lane 3, purified cEptC protein. **(E)** Representative native cEptC protein crystals used in X-ray diffraction experiments.

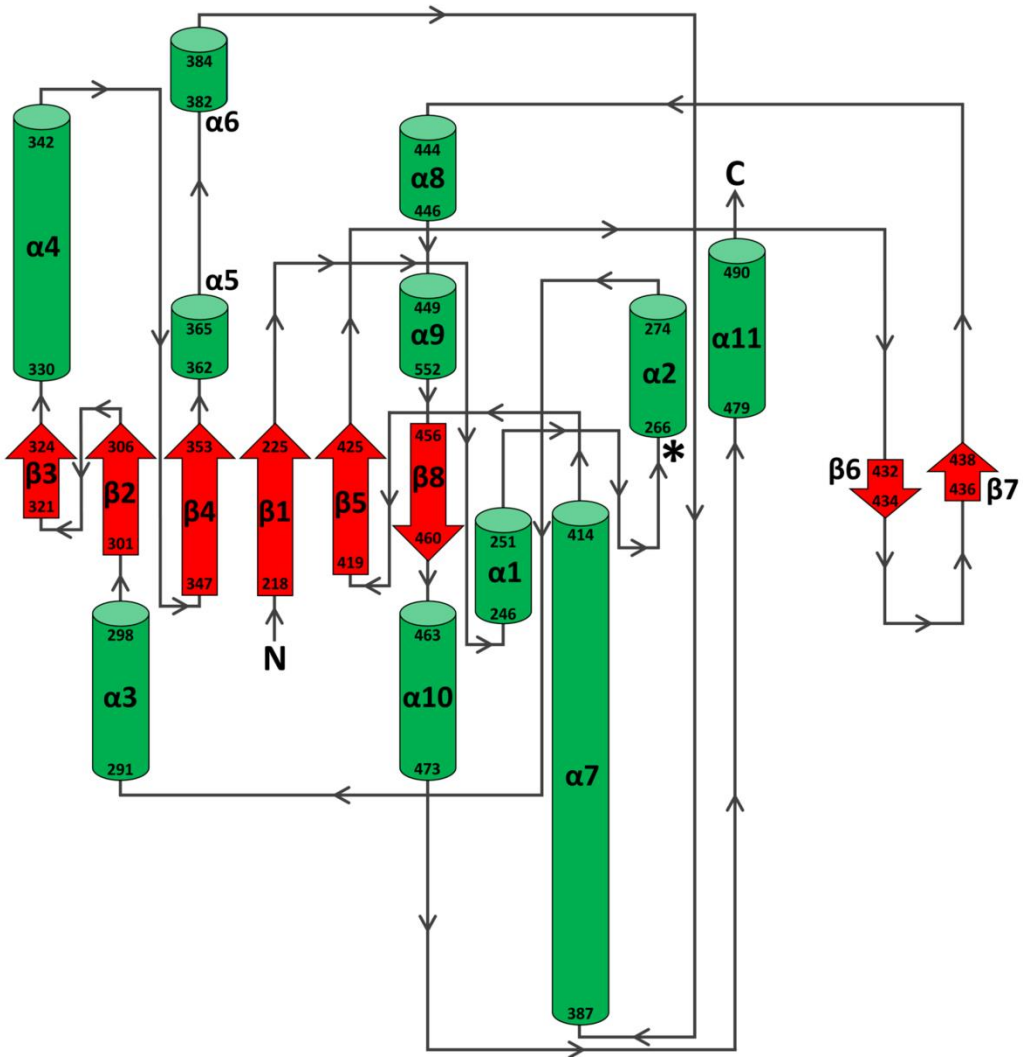


Figure 3-4. Topology map of cEptC. α-Helices (green cylinders) are labeled α1 to α11 and β-strands (red arrows) are labeled β1 to β8, beginning from the N-terminus, as in **Figure 3-2**. The position of the nucleophilic Thr266 on the N-terminus of α2 is indicated with a star. This image is adapted from the topology map generated by uploading the cEptC coordinate file to the PDBsum server (Laskowski, 2009).

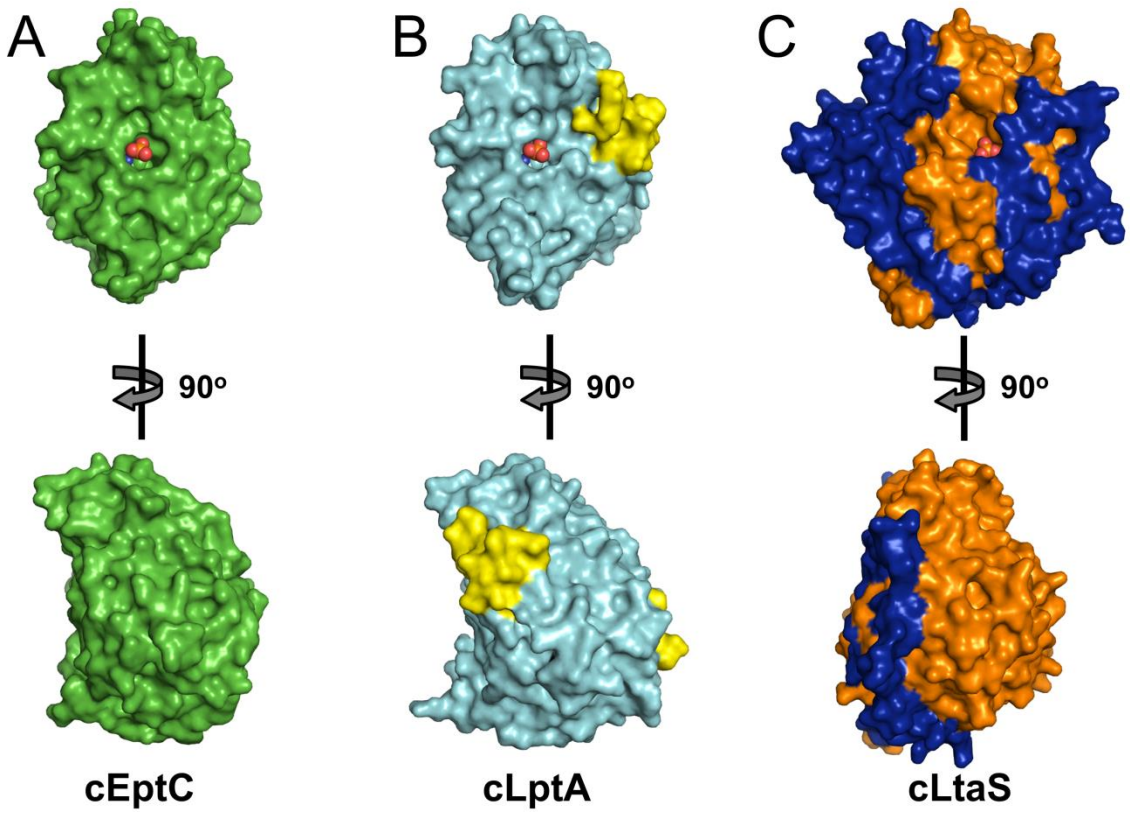


Figure 3-5. Surface rendition of **(A)** cEptC, compared with **(B)** cLptA (PDB 4KAY) and **(C)** *B. subtilis* cLtaS (PDB 2W8D). In the lower panel, each image is rotated 90° to the left. Phosphothreonine residues, represented as spheres, indicate the locations of the active sites. Yellow regions are unique to cLptA and blue regions are unique to cLtaS, as in **Figure 3-7**.

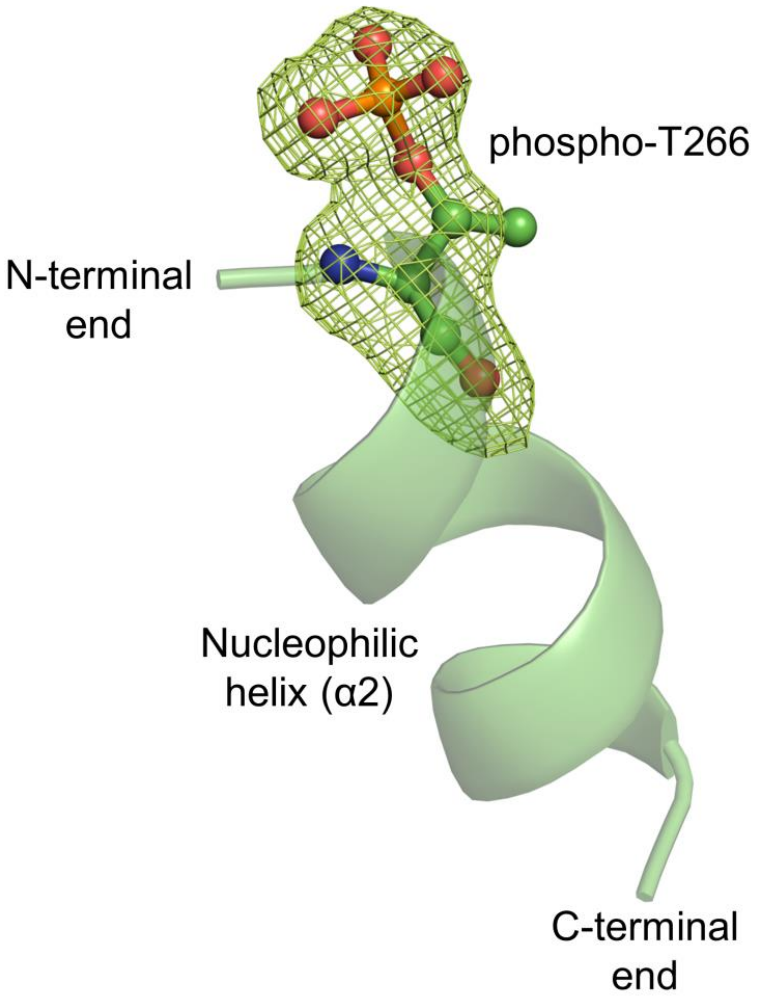


Figure 3-6. Phosphothreonine is located on the N-terminal end of the nucleophilic helix (α2). A helical dipole and neighboring Zn^{2+} (not shown) stabilize the deprotonated side chain of T266, activating the residue as a nucleophile capable of generating a phospho-intermediate. F_o-F_c omit map density for phosphothreonine (green mesh) is contoured to 4.0σ.

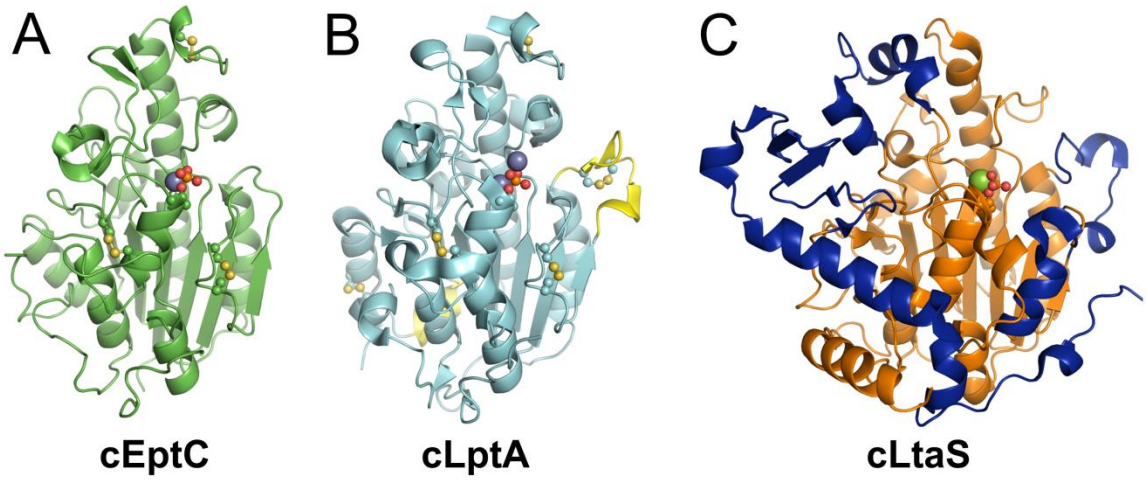


Figure 3-7. Overall structures of pEtN transferases cEptC and cLptA, and *B. subtilis* lipoteichoic acid synthase cLtaS. **(A)** cEptC (green), with phosphothreonine and disulfides shown as ball-and-sticks and Zn^{2+} ion as a purple sphere. **(B)** cLptA (PDB 4KAY; cyan), with phosphothreonine and disulfides shown as ball-and-sticks and Zn^{2+} ions as purple spheres. Structural insertions found in cLptA, but not cEptC, are colored yellow. cLptA and cEptC superpose with C_α r.m.s.d. of 0.59 Å over 211 residues. **(C)** *B. subtilis* cLtaS (PDB 2W8D; orange), with phosphothreonine shown as ball-and-sticks and Mg^{2+} ion as a green sphere. Structural features unique to cLtaS are colored blue. cLtaS and cEptC superpose with C_α r.m.s.d. of 3.74 Å over 216 residues.

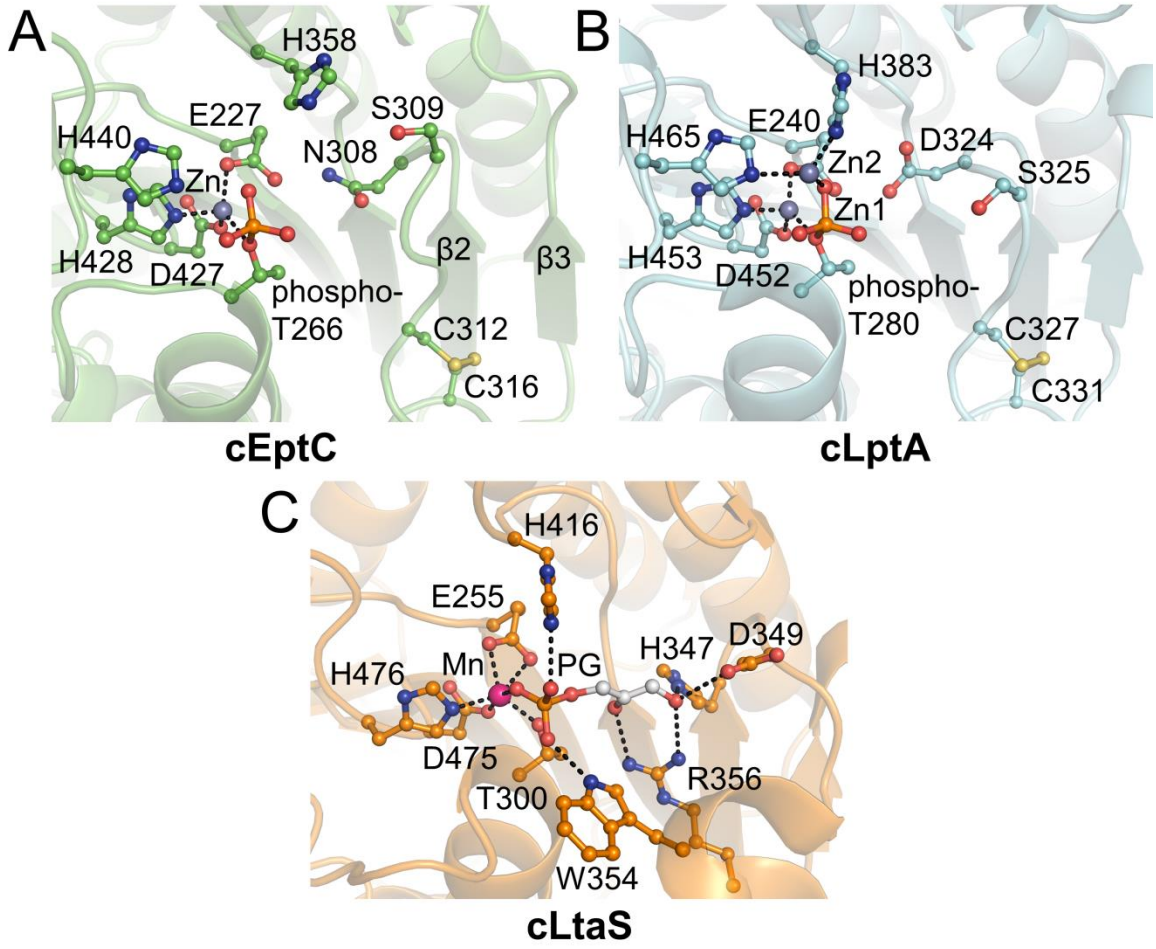


Figure 3-8. Juxtaposed active sites of **(A)** cEptC, **(B)** cLptA (PDB 4KAY), and **(C)** *S. aureus* cLtaS (PDB 2W5S). Active-site architectures and zinc-binding residues are conserved between cEptC and cLptA (**Figure 3-9**), although only one Zn^{2+} associates with cEptC *in crystallo*. H358/H383 and metal-ligand residues of cEptC and cLptA are also conserved in cLtaS. Side chains on an active-site loop/helix of cLtaS, corresponding to a loop between *β*2 and *β*3 of cEptC, bind phosphoglycerol (PG, white carbons; Lu *et al.*, 2009a). In cEptC and cLptA, this loop contains N308/D324 and S309/S325, which may contribute to substrate binding. Notably, T266, H358, H440, N308, and S309 mutant strains of *eptC* in *C. jejuni* displayed polymyxin B sensitivity and severe loss of motility (**Figures 3-11** and **3-12**).

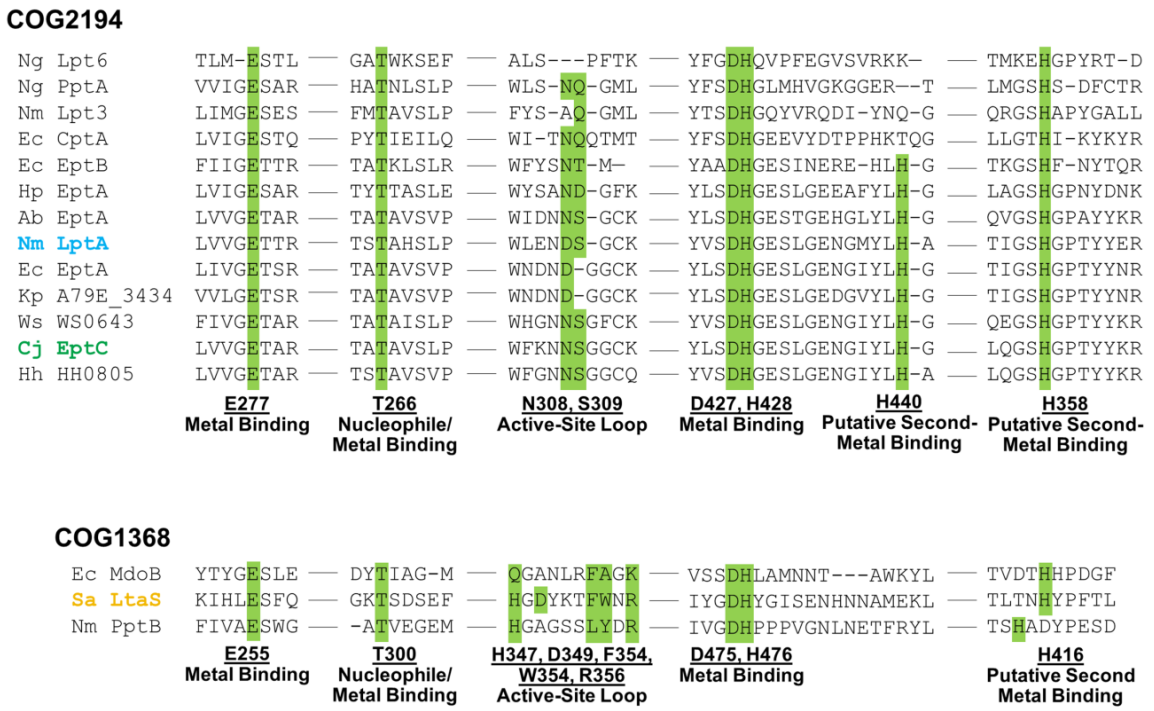


Figure 3-9. Minimized sequence alignment (Clustal Omega; Sievers & Higgins, 2014) of representative phospho-form transferases in the COG2194 and COG1368 groups, demonstrating conservation of active-site residues. COG2194 consists of putative membrane-associated, metal-dependent hydrolases, including pEtN transferases such as EptC and LptA; COG1368 includes phosphoglycerol and related transferases of the alkaline phosphatase superfamily, such as LtaS. EptC and LtaS residues are labeled under alignments for COG2194 and COG1368, respectively. The locus tags used in the sequence alignments are as follows: Ng Lpt6 (*Neisseria gonorrhoea,* NGO2071), Ng PptA (*N. gonorrhoea,* NGO1540), Nm Lpt3 (*N. meningitidis,* NMB2010), Ec CptA (*E. coli,* b3955), Ec EptB (*E. coli,* Y75_p3631), Hp EptA (*Helicobacter pylori,* HP0022), Ab EptA (*Acinetobacter baumannii,* ABAYE0731), Nm LptA (*N. meningitidis* NMB1638), Ec EptA (*E. coli,* Y75_p4001), Kp A79E_3434 (*Klebsiella pneumonia,* A79E_3434), Ws WS0643 (*Wolinella succinogens,* WS0643), Cj EptC (*C. jejuni,* Cj0256), Hh HH0805 (*H. hepaticus,* HH0805), Ec MdoB (*E. coli,* Y75_p4244), Sa LtaS (*S. aureus,* MS7_0770), Nm PptB (*N. meningitidis,* NMV0885).

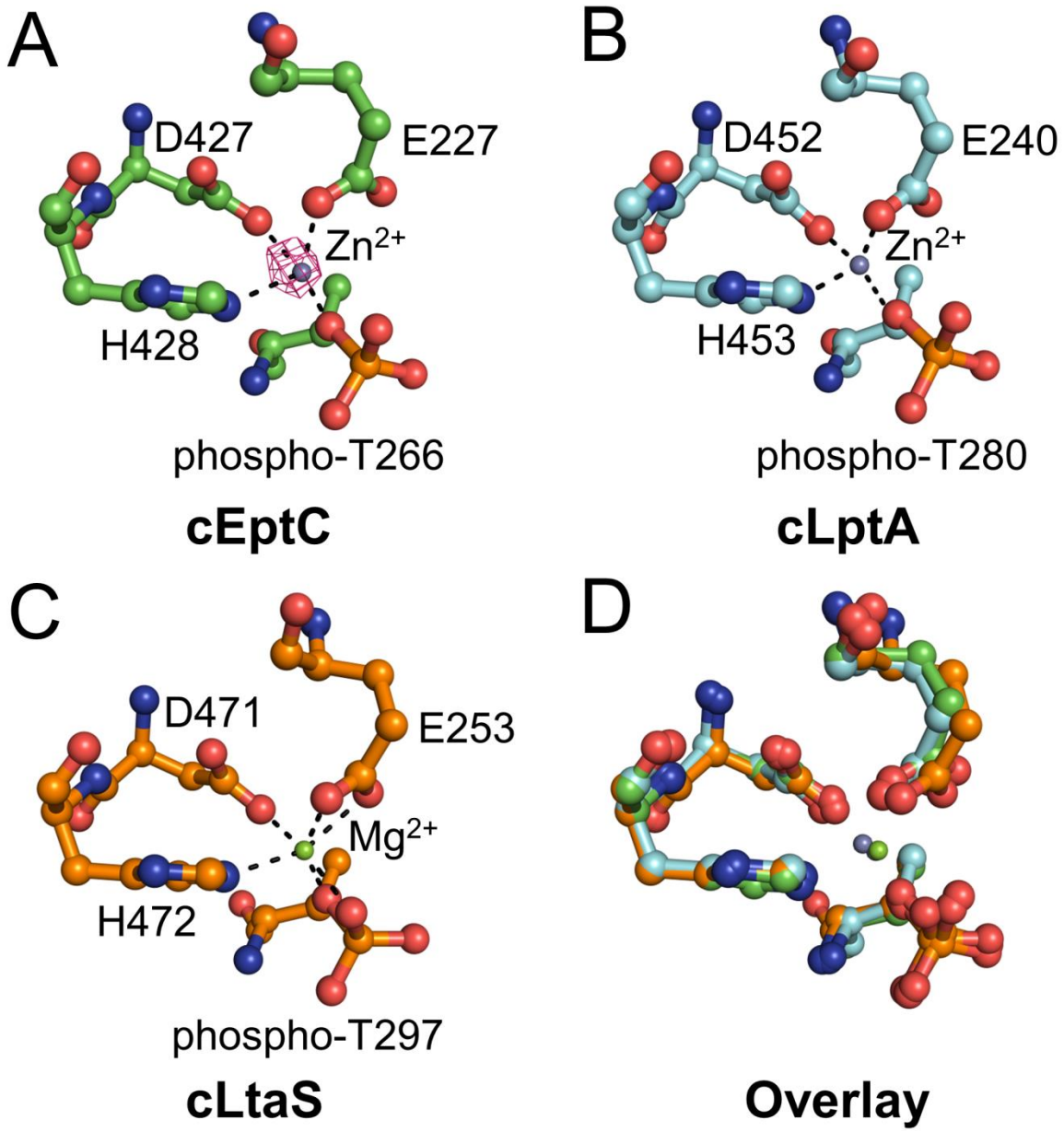


Figure 3-10. Metal-binding sites of **(A)** cEptC, **(B)** cLptA (PDB 4KAY), and **(C)** *B. subtilis* cLtaS (PDB 2W8D) are conserved. Anomalous difference density for Zn^{2+} (magenta mesh) of cEptC is contoured to 6σ. Coordination bonds between side chains and metal ions are indicated with dashed lines. **(D)** Superposed models between **A** and **B** (30 atoms, r.m.s.d. = 0.26 Å) and **A** and **C** (21 atoms, r.m.s.d. = 0.18 Å) are displayed as one image with the coordinates of **A** as a reference.

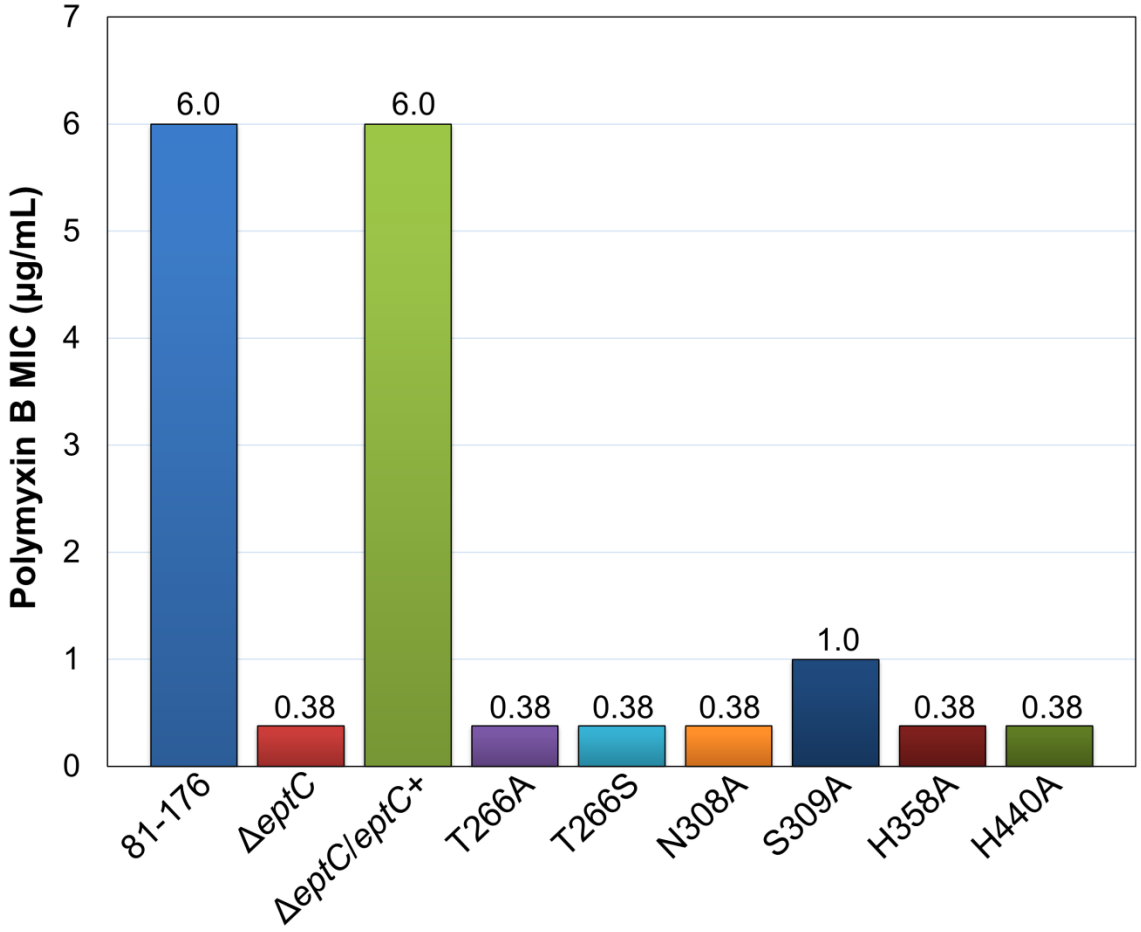


Figure 3-11. Minimal inhibitory concentrations (MICs) of polymyxin B for *C. jejuni* EptC active-site mutants. Parent strain (81-176), *eptC* null mutant (Δ*eptC*), complemented Δ*eptC* strain carrying a copy of the native *eptC* allele (Δ*eptC/eptC*$^{+}$), and mutant strains carrying *eptC* alleles with single amino acid mutations (T266A, T266S, N308A, S309A, H358A, and H440A) were swabbed onto MH agar plates with 10 µg/mL trimethoprim and subjected to the polymyxin B Etest strip assay. MICs were recorded after 48 h. Numbers above each bar in the graph indicate polymyxin B MICs in µg/mL.

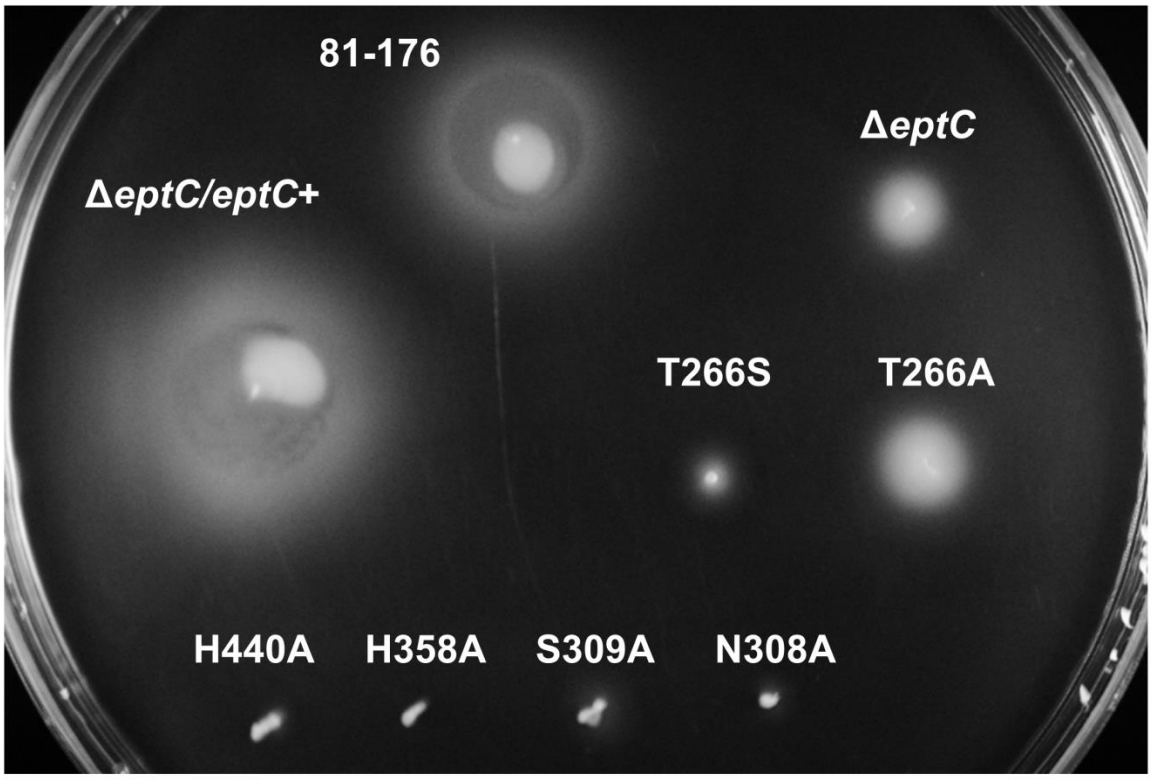


Figure 3-12. EptC active-site mutant motility assays. Parent strain (81-176), *eptC* null mutant (Δ*eptC*), complemented Δ*eptC* strain (Δ*eptC/eptC*$^{+}$), and strains carrying *eptC* active-site mutant alleles (T266A, T266S, N308A, S309A, H358A, and H440A) were stab-inoculated into semisolid MH motility medium and incubated in microaerobic conditions. To assess relative motility, bacterial growth diameters were visually compared 24 h post-inoculation.

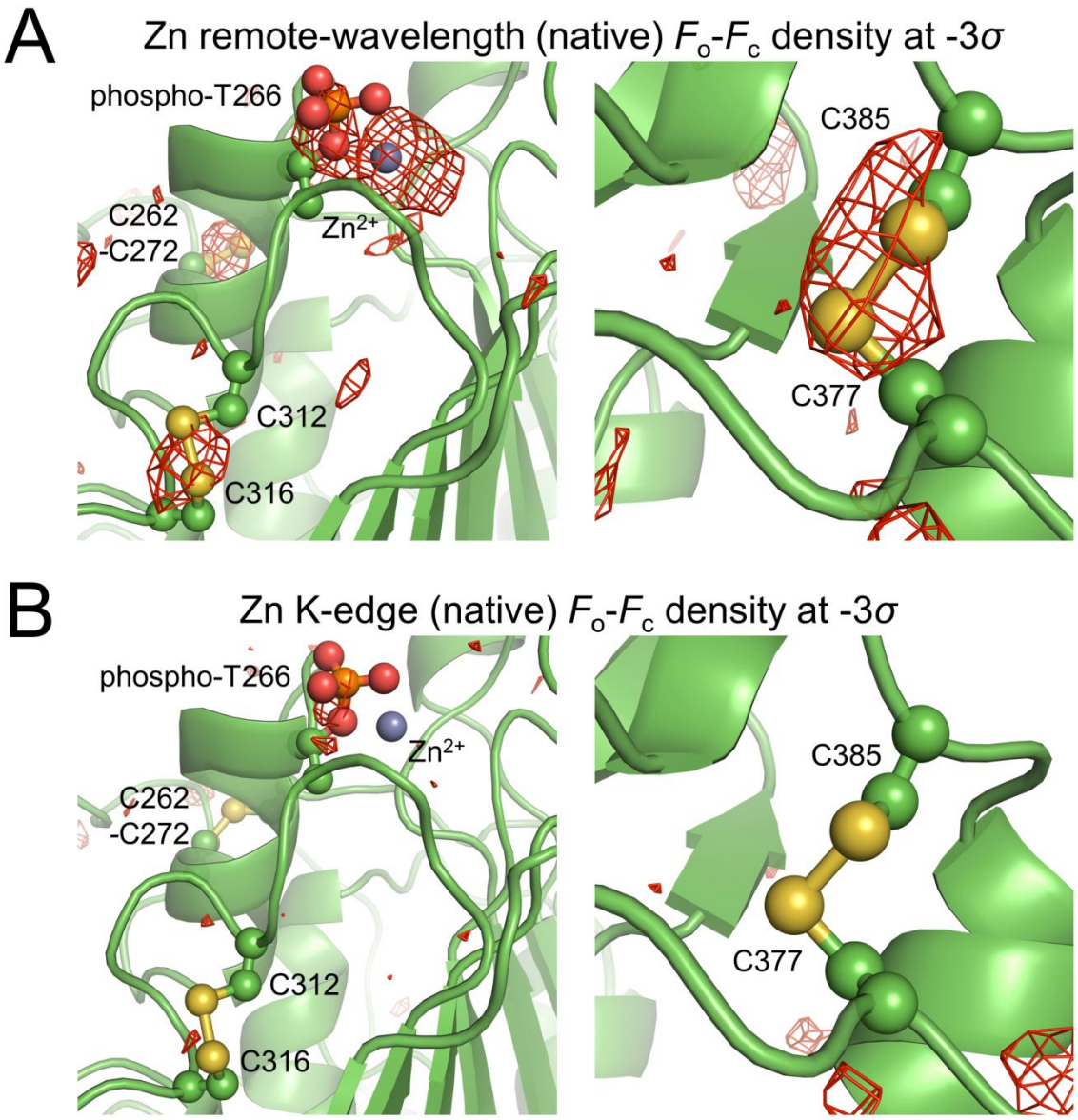


Figure 3-13. Negative F_o-F_c electron density maps reveal radiation damage to the Zn remote-wavelength dataset. Density (red mesh) for phospho-T266, Zn^{2+}, and disulfides (C262-C272, C312-C316, and C377-C385) of chain A from **(A)** Zn remote-wavelength and **(B)** Zn K-edge datasets is contoured at -3σ. These images are representative of the entire asymmetric unit of each dataset. Because the Zn K-edge dataset is of higher quality in these crucial regions, it was selected for deposition over the Zn remote-wavelength dataset.

Table 3-1. Bacterial strains and plasmids.

Strains and plasmids	Genotype or description	Source or reference
Strains		
Escherichia coli		
XL1-Blue	General cloning strain, *recA1 endA1 gyrA96thi-1 hsdR17*	Stratagene
BL21(DE3)	Protein expression strain, F^- *ompT gal dcm lon* $hsdS_B(r_{B^-}\ m_{B^-})$ λ(DE3)	Novagen
Campylobacter jejuni		
81-176	Serotype HS: 23, 26	(Cullen & Trent, 2010)
81-176 A1	81–176 Δ*cj0256* clone1, Cam^R	(Cullen & Trent, 2010)
81-176 B	Complement strain, 81–176 Δcj0256 Δ*astA*::cj0256, Cam^R, Kan^R	(Cullen & Trent, 2010)
81-176 $eptC_{T266A}$	81–176 Δ*cj0256* Δ*astA*::$cj0256_{T266A}$, Cam^R, Kan^R	This Study
81-176 $eptC_{T266S}$	81–176 Δ*cj0256* Δ*astA*::$cj0256_{T266S}$, Cam^R, Kan^R	This Study
81-176 $eptC_{N308A}$	81–176 Δ*cj0256* Δ*astA*::$cj0256_{N308}$, Cam^R, Kan^R	This Study
81-176 $eptC_{S309A}$	81–176 Δ*cj0256* Δ*astA*::$cj0256_{S308A}$, Cam^R, Kan^R	This Study
81-176 $eptC_{H358A}$	81–176 Δ*cj0256* Δ*astA*::$cj0256_{H358A}$, Cam^R, Kan^R	This Study
81-176 $eptC_{H440A}$	81–176 Δ*cj0256* Δ*astA*::$cj0256_{H4440A}$, Cam^R, Kan^R	This Study
Plasmids		
pET28b(+)	Expression vector containing a T7 promoter, Amp^R	Novagen
pETcEptC	pET28b(+)-derived expression plasmid used to express cEptC, Amp^R	This Study
pAtsAKO::*cj0256*:Kan^R	pGEM-atsAKO:Kan^R with *cj0256* inserted	(Cullen & Trent, 2010)
$pETcEptC_{T266S}$	T266S mutant derivative of pETcEptC, Amp^R	This Study
$pAstAEptC_{T266A}$	T266A mutant derivative of pAtsAKO::*cj0256*:Kan^R	This Study
$pAstAEptC_{T266S}$	T266S mutant derivative of pAtsAKO::*cj0256*:Kan^R	This Study
$pAstAEptC_{N308A}$	N308A mutant derivative of pAtsAKO::*cj0256*:Kan^R	This Study
$pAstAEptC_{S309A}$	S309A mutant derivative of pAtsAKO::*cj0256*:Kan^R	This Study
$pAstAEptC_{H358A}$	H358A mutant derivative of pAtsAKO::*cj0256*:Kan^R	This Study
$pAstAEptC_{H4440A}$	H440A mutant derivative of pAtsAKO::*cj0256*:Kan^R	This Study

Table 3-2. Primers used in this study.

1 F-VPIPEcEptC	3'-ACCCCCGCCTCCCCTGGATTTGCCTTAAGTTTAGGATTTAA-5'
2 R-VPIPEcEptC	3'-AGGAGATATACCATGTTTAAAACTATAGCTAATGATGCTTATA-5'
3 F-IPIPEcEptC	3'-GGGGAGGCGGGGGTGGACACCACCACCACCACCACCACTGAGAT-5'
4 R-IPIPEcEptC	3'-CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATT-5'
5 F-EptCT266S	3'-ATTTTTCATCATGTGGAACAGCAAGCGCGGTAAGTTTGCCTTGTATGTTT-5'
6 R-EptCT266S	3'-AAACATACAAGGCAAACTTACCGCGCTTGCTGTTCCACATGATGAAAAAT-5'
7 F-EptCT266A	3'-ATTTTTCATCATGTGGAACAGCAGCGGCGGTAAGTTTGCCTTGTATGTTT-5'
8 R-EptCT266A	3'-AAACATACAAGGCAAACTTACCGCCGCTGCTGTTCCACATGATGAAAAAT-5'
9 F-EptCN308A	3'-GCATGGTTTGATAATGCATCTGGTGGTTGTAAA-5'
10 R-EptCN308A	3'-TTTACAACCACCAGATGCATTATCAAACCATGC-5'
11 F-EptCS309A	3'-TGGTTTGATAATAATGCAGGTGGTTGTAAAGGG-5'
12 R-EptCS309A	3'-CCCTTTACAACCACCTGCATTATTATCAAACCA-5'
13 F-EptCH358A	3'-CATTTGCAAGGTTCTGCAGGGCCAACTTATTAT-5'
14 R-EptCH358A	3'-ATAATAAGTTGGCCCTCGAGAACCTTGCAAATG-5'
15 F-EptCH440A	3'-AATGGTATTTATCTTGCAGGTATGCCTTATGCT-5'
16 R-EptCH440A	3'-AGCATAAGGCATACCTCGAAGATAAATACCATT-5'

Underlined regions indicate introduced mutations.

Table 3-3. Data-collection and refinement statistics for cEptC.

Values for the outer shell are given in parentheses.

	Native, Zn K-edge	**Native, Zn remote-wavelength**	**SeMet-derived, Se K-edge**
Data collection			
Space group	$C222_1$	$C222_1$	$P2_12_12_1$
X-ray source	Advanced Light Source, 4.2.2	Advanced Photon Source, 23-ID-B	Advanced Photon Source, 23-ID-B
Wavelength (Å)	1.2680	1.0332	0.9795
Temperature (K)	100	100	100
Unit cell parameters			
a, *b*, *c* (Å)	121.9, 183.1, 121.5	122.5, 183.6 121.7	119.1, 120.8, 182.7
α = β = γ (°)	90.0	90.0	90.0
Resolution (Å)	55-2.40 (2.46-2.40)	50-2.50 (2.54-2.50)	50-2.80 (2.85-2.80)
R_{merge}† (%)	11.2 (99.6)	11.9 (95.5)	12.1 (82.9)
$CC_{1/2}$‡ (%)	99.6 (67.4)	99.0 (85.4)	99.8 (75.6)
Average $I/\sigma(I)$	10.6 (1.5)	18.2 (2.8)	14.3 (2.0)
No. of observed reflections	392584	349675	729109
No. of unique reflections	53319 (3892)	47144 (2327)	119456 (5831)
Oscillation step (°)	0.2	1.0	1.0
No. of images	900	180	360
Exposure time per image (s)	0.4	1.0	1.0
Completeness (%)	99.7 (99.6)	99.9 (100)	99.7 (97.6)
Redundancy	7.4 (7.4)	7.4 (7.2)	6.1 (4.9)
Monomers in asymmetric unit	3	3	6
Solvent content (%)	61.8	60.6	59.0
Matthews coefficient ($Å^3$ Da^{-1})	3.22	3.12	3.00
Refinement			
Resolution (Å)	54.58-2.40 (2.43-2.40)		
R_{work}/R_{free}# (%)	21.5/25.6 (33.0/38.8)		
No. of atoms			
Overall	7411		
Protein	7378		
Ligand§	15		
Water	18		
Average *B*-factors ($Å^2$)			
Overall	49.8		
Protein	49.8		
Ligand§	44.3		
Water	38.7		
R.m.s.d., bond lengths (Å)	0.009		
R.m.s.d., bond angles (°)	1.28		
Ramachandran favored (%)	96.8		
Ramachandran allowed (%)	3.2		
PDB code	4TN0		

†$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity for all reflections.

‡$CC_{1/2}$ = correlation coefficient between two random half-datasets (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013).

#$R_{work}/R_{free} = \sum_{hkl}\big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

§Includes phosphoryl group and Zn atoms.

Chapter 4: The [4+2]-Cyclase SpnF at 1.50-Å Resolution

ABSTRACT

In the biosynthetic pathway of the spinosyn insecticides, the tailoring enzyme SpnF performs a [4+2]-cycloaddition on a 22-membered macrolactone to forge an embedded cyclohexene ring. To learn more about the details of this reaction, which has been hypothesized to proceed through a Diels-Alder mechanism, the 1.50-Å resolution structure of SpnF bound to *S*-adenosylhomocysteine was determined. SpnF is structurally related to class I *S*-adenosylmethionine-dependent methyltransferases, yet its *S*-adenosyl coenzyme does not participate in methyltransfer and primarily plays a structural role. By analogy with related enzymes, SpnF was observed in a closed conformation in which an 860-Å^3 substrate cavity is shielded from bulk water. All mutations to residues lining this cavity greatly reduced SpnF activity. This structure sets the stage for advanced experimental and computational studies to determine the mechanism of SpnF-mediated cyclization.

INTRODUCTION

Polyketides generated by iterative polyketide synthases (PKSs) commonly contain carbocycles, whereas those generated by processive PKSs typically do not (Crawford & Townsend, 2010; Hertweck *et al.*, 2007; Yu *et al.*, 2012; Keatinge-Clay, 2012). The intramolecular carbon-carbon bond-forming reactions that lead to carbocyclic products are often facilitated by the reactive nature of poly β-keto intermediates generated by iterative Type I PKSs, Type II PKSs, and Type III PKSs

(**Figure 4-1**). These cyclizations are catalyzed by standalone cyclases in Type II pathways and cyclases incorporated within the multidomain synthases of Type I pathways (Ames *et al.*, 2008; Crawford *et al.*, 2009; Korman *et al.*, 2010). Their active sites position reactive substrate moieties for cyclization through regiospecific carbon-carbon bond-forming condensations. Iterative Type I PKSs and associated tailoring enzymes are also known to generate and cyclize highly unsaturated polyene intermediates into such molecules as calicheamicin and ikarugamycin (Ahlert *et al.*, 2002; Zhang *et al.*, 2014).

The spinosyn biosynthetic pathway is unusual in that it contains a processive Type I PKS yet generates a carbocyclic product (Kirst, 2010). In the construction of the tetracyclic spinosyn core, a desaturated, 22-membered macrolactone is cross-bridged through two carbocyclization reactions (**Figure 4-2**). While the tailoring enzyme SpnL is believed to catalyze cyclization through a Rauhut-Currier mechanism (Aroyan *et al.*, 2009), the [4+2]-cyclization catalyzed by SpnF has been hypothesized to proceed through a Diels-Alder mechanism (Diels & Alder, 1929), in which an s-*cis*-configured diene is fused with a dienophile (e.g., an olefin) through an exergonic electrocyclization that converts two π-bonds into two σ-bonds (Kim *et al.*, 2011). Although other biosynthetic pathways (e.g., those of lovastatin, equisetin, solanapyrones, and spirotetronates) are hypothesized to contain [4+2]-cyclases, SpnF is the first enzyme to be reported whose sole function is to catalyze such a reaction (Kelly, 2008).

The order of the spinosyn tailoring reactions was determined through *in vitro* studies. After generation of spilactone by a processive Type I PKS, SpnJ catalyzes oxidation of the C15-hydroxyl group to a ketone functionality (Kim *et al.*, 2007) (**Figure 4-2**). Next, SpnM catalyzes a 1,4-dehydration (Kim *et al.*, 2011)

yielding a second, π-conjugated system that contains the putative dienophile. SpnF then catalyzes cyclohexene ring formation via [4+2]-cycloaddition between the C4-C7 diene and the C11-C12 dienophile (Kim *et al.*, 2011). The tricyclic intermediate is subsequently rhamnosylated in a glycosylation reaction catalyzed by SpnG (Isiorho *et al.*, 2012), after which the rhamnose moiety is permethylated through the combined activities of the *S*-adenosylmethionine (SAM)-dependent methyltransferases SpnH, SpnI, and SpnK (Kim *et al.*, 2010). SpnL facilitates a putative Rauhut-Currier reaction to generate another intramolecular carbon-carbon bond, thereby completing the construction of the tetracyclic skeleton (Kim *et al.*, 2011). Finally, SpnP transfers forosamine to generate the biologically-active spinosyns (Isiorho *et al.*, 2014).

SpnM, SpnF, and SpnL have been hypothesized to participate in the carbocylization reaction since the spinosyn biosynthetic gene cluster was sequenced (Waldron *et al.*, 2001). Their individual roles have now been demonstrated (Kim *et al.*, 2011); however, the mechanisms utilized by these tailoring enzymes remain to be deciphered. SpnM resembles an α/β-hydrolase, but contains an alanine where a catalytic serine is anticipated. SpnF and SpnL resemble *S*-adenosylmethionine (SAM)-dependent methyltransferases (MTs), yet whether SAM participates in these cyclization reactions is unclear. The carbocyclization catalyzed by SpnF is hypothesized to proceed through a Diels-Alder mechanism, known to be facilitated by the participation of an electron-deficient dienophile. If SpnF catalyzes a Diels-Alder reaction a feature that withdraws electron density from the dienophile may be within its active site.

We now present the 1.50-Å resolution structure of SpnF bound to *S*-adenosylhomocysteine (SAH). Through consensus docking of the reaction substrate

and product, their putative binding modes were interrogated. Mutagenesis studies identified residues important to SpnF-mediated cyclization. The structural results reported here set the stage for advanced experimental and computational approaches to determine the exact mechanism of cyclization.

RESULTS AND DISCUSSION

SpnJ and SpnM prepare the macrolactone for [4+2]-cyclization

The spilactone ring produced by the spinosyn Type I PKS is relatively unreactive, with a conjugated polyene on one side of the macrolactone, similar to pimaricin. However, the combined oxidation and 1,4-dehydration catalyzed by SpnJ and SpnM introduces a conjugated π-system on the opposite side of the macrolactone, creates ring strain, and enables a [4+2]-cycloaddition that can be accelerated by SpnF 500-fold (k_{cat} vs. k_{non}) (Kim *et al.*, 2011). All the inter-olefinic bonds of the SpnM product (e.g., the SpnF substrate), including C5-C6 of the reactive diene, have been observed by ^{1}H-NMR in s-*trans* conformations (Kim, 2010). Low-energy conformations of the SpnM product can thus be represented by a macrocycle that resembles an elongated triangle in which the 2 longest sides are constructed of approximately coplanar atoms (7- and 10-long). While conformations of the SpnM product in which C5-C6 is in the s-*cis* conformation are higher in energy, it is this population that is hypothesized to undergo an experimentally-observed nonenzymatic [4+2]-cycloaddition that yields the same product generated by SpnF (Kim *et al.*, 2011). Within the SpnF active site it is unknown what fraction of molecules are in the reactive C5-C6 s-*cis* conformation and whether the s-*trans* and s-*cis* conformations of C5-C6 can interconvert.

Overall structure of SpnF

To learn more about how cyclization is catalyzed, the crystal structure of SpnF was determined (**Figure 4-3, Table 4-1**). The catalytic core of RebM (the rebeccamycin MT, PDB code: 3BUS) was used as a molecular replacement search model (Singh *et al.*, 2008). Two monomers of SpnF were identified in the asymmetric unit within space group $P2_1$. These monomers align well (0.165 Å r.m.s.d. over 1357 atoms), with the pitch of helix αG being the only significant difference in conformation (**Figure 4-4**). Only a few portions of the polypeptide chains are not well-resolved: the histidine tag, the N-terminus (residues 1–2 and 1–3 in monomers A and B, respectively), and two loop regions (9–12, 186–191 in monomer A; 8–12, 184–185 in monomer B) are invisible, while only weak density is present for the C-terminal glycine (residue 282). No Ramachandran outliers are observed, and 97.6% of residues fall within the favored region. Slightly elevated R_{work} and R_{free} values, a common feature of SpnF datasets, may be due to the presence of multiple lattices in the diffraction patterns.

Based on the F_o-F_c omit maps, SAH, not SAM, is bound to each monomer in a 490-Å^3 pocket resembling that of class I SAM-dependent MTs (Martin & McMillan, 2002; Voss & Gerstein, 2010) (**Figure 4-3B**). SAH is likely present from the SpnF purification. The H41 side chain and the I40 backbone NH form charged hydrogen bonds with the SAH carboxylate, while the backbone carbonyls of G80 and M147 form charged hydrogen bonds with the homocysteine amino group of SAH. The T102 and Q107 side chains, as well as the G82 NH, form hydrogen bonds with the SAH ribose hydroxyl groups. The D130 side chain and the A131 NH form hydrogen bonds with the SAH adenine. Hydrogen-bonded to one another, the Y23 and E152

side chains form a lid for the cofactor binding pocket that essentially separates it from the substrate cavity.

The H42 and Q148 side chains bind a molecule of malonate (present in the crystallization buffer) in each monomer, apparently in two alternate conformations (**Figure 4-3B**). As malonate was found to significantly enhance crystal quality and bind in the same cavity in which the SpnF substrate is anticipated to bind, it may stabilize the closed conformation observed for SpnF.

Similar to many natural product MTs (Liscombe *et al.*, 2012), the catalytic core (a 7-stranded β-sheet surrounded by 5 α-helices) of SpnF is embellished with an N-terminal extension (αY) and a "cover" formed by a loop (residues 181-202, inserted between β5 and αE) and 2 α-helices (αF and αG, inserted between β6 and β7). The closest structural homologues identified by the DALI server (Holm & Rosenström, 2010) are a sarcosine dimethylglycine MT, RebM, a geranyl diphosphate MT, and a mycolic acid cyclopropane synthase (PDB codes 2O57, 3BUS, 4F86, and 1KPG with Z-scores 28.9, 28.5, 28.4, and 28.0, respectively) (McCoy *et al.*, 2009; Singh *et al.*, 2008; Ariyawutthiphan *et al.*, 2012; Huang *et al.*, 2002). Most of the top DALI hits are *C*-MTs that catalyze the attack of a carbon-carbon double bond on the electrophilic methyl group of SAM. They also share a less common ribose-binding motif in which the ribose hydroxyl groups are coordinated by a glutamine, a serine/threonine, and a backbone NH, instead of a side chain carboxylate. The solved structures of geranyl diphosphate MT and cyclopropane synthase provide examples of closed, ternary complexes in which both a cofactor and a substrate are completely isolated from solvent (PDB codes 4F86 and 1KPG) (Ariyawutthiphan *et al.*, 2012; Huang *et al.*, 2002), whereas the crystal structure of RebM provides an example of an MT in an open state (PDB Code 3BUS) (Singh *et al.*, 2008).

The 860-Å^3 volume of the substrate cavity located between the SpnF catalytic core, αY, and the cover is large enough to completely enclose the SpnF substrate. BusF is the SpnF homologue in the biosynthetic pathway of butenyl-spinosyns, which differ from the spinosyns in that a butenyl substituent replaces the ethyl substituent at C21 (Hahn *et al.*, 2006). While 23 residues out of 275 differ between the two enzymes, 22 are surface residues. As the one buried substitution (M147I) is distant from the active site, SpnF and BusF apparently possess nearly identical substrate cavities.

Substrate and product docking

To obtain representative structures of the SpnF substrate and product for docking into the SpnF active site, conformations of these molecules were investigated using two force fields (UFF and MMFF94s) (Rappe *et al.*, 1992; Halgren, 1999). The restraints imposed by the two extended conjugated systems of the substrate greatly reduce the available conformational space compared to similarly-sized macrocycles. With the C5-C6 bond fixed in the s-*cis* state, the conformations of other rotatable bonds (e.g., s-*cis* or s-*trans* at C1-C2, C3-C4, C12-13, and C14-C15) were investigated. The conformer in **Figure 4-2**, in which the carbonyls point in the same directions as those of the spinosyns (e.g., with the enone C14-C15 also in the s-*cis* orientation), was predicted to be to be lowest in energy. To obtain structures for the product, the bonds of the substrate structure shown in **Figure 4-2** were adjusted to those of the cyclized product and the resulting molecule was energy-minimized. The UFF and MM94s force fields yielded very similar structures for both the SpnF substrate and product.

Consensus docking of the SpnF substrate and product was performed using the programs AutoDock Vina (Trott & Olson, 2010) and DOCK6 (Brozell *et al.*, 2012) (**Tables 4-2** – **4-9**). In general, the top solutions represent poses that have the best shape complementarity to the active site cavity. This is ideal since natural product MTs are known to enclose their substrates in complementary fashion and SpnF was captured in its closed conformation. An orientation was observed for the substrate structures and the product structures that not only fits well in the substrate cavity but also forms favorable hydrophobic and hydrophilic contacts (**Figure 4-5**).

The triangular substrate enters the cavity shortest-side (C17-C21) first. Hydrophobic contact is made with V19, M22, Y23, V26, L30, C38, I40, W53, E152, V178, T196, G197, L198, F230, M231, F234, W256, V260, and M270. Hydrogen bonds are formed between the C17-OH and the side chains of H42 and Q148 (where the malonate carboxylate binds in the crystal structure). The substrate appears to rest on a platform created on the catalytic core by Y23, I40, H42, Q148, and E152 and to be blanketed by the cover residues T196, G197, L198, F230, F234, and W256. The product makes nearly the same contacts as the substrate. Neither substrate nor product makes any contacts with SAH. Space is available in both the substrate and product complexes to accommodate the C6-methyl group incorporated by the Type I PKS during the biosynthesis of Spinosyn D.

Mutagenesis

To investigate which residues contribute most to SpnF activity Y23F, C38A, H42A, H42Q, Q148A, Q148E, E152A, and E152L mutants were compared to wild-type SpnF through a cyclization assay (**Figure 4-6**) (A51T, R123Q, M147I, T196A, T196S mutants have not yet been assayed for activity) (Kim *et al.*, 2011). The assay

measures tricyclic product formed after equivalent incubations of the SpnM substrate with SpnM and the SpnF variant. All mutants showed a severe loss in activity compared to wild-type SpnF. The SpnF active site may thus be finely tuned, with each residue making comparable contributions to catalysis. Perhaps this could be inferred from the observation that none of the residues lining the substrate cavity are different in BusF.

The cycloaddition reaction

While computational studies support a concerted, highly asynchronous Diels-Alder mechanism for SpnF (Hess & Smentek, 2012), non-Diels-Alder routes (e.g., dipolar, biradical, [6+4]-cyclization followed by Cope rearrangement) are not easily disproven. Because the SpnF substrate undergoes stereospecific [4+2]-cycloaddition in an aqueous environment in the absence of enzyme, SpnF is not thought to conduct covalent, acid/base, or metal-assisted catalysis. SpnF may facilitate cycloaddition through a combination of (1) stripping the substrate of water molecules, (2) stabilizing the C5-C6 s-*cis* conformation, and (3) withdrawing electron density from the dienophile — T196 is in position to form a hydrogen bond with the C15-carbonyl and could withdraw electron density through the C11-C15 π-system (the E152 carboxylate may also affect dienophile electronics).

The structure suggests that SAM, the coenzyme most likely bound *in vivo*, may primarily play a structural role. Indeed, the active site of the sarcosine dimethylglycine MT is only loosely structured in the absence of SAM/SAH (McCoy *et al.*, 2009). Without SAM playing a mechanistic role in SpnF, why this carbocyclase evolved from an MT can only be hypothesized. Perhaps an ancestral enzyme that recognized a spinosyn intermediate for methyltransfer came to catalyze

carbocyclization when binding a more reactive substrate (the same may be true of SpnL). RdmB from the aclacinomycin biosynthetic pathway is an example of an enzyme with a class I MT fold that binds SAM but performs a reaction other than methyltransfer — O_2-dependent oxidation; however, SAM does not play a purely structural role in this enzyme (Jansson *et al.*, 2003).

Carbocycles within polyketides are generally derived from intramolecular cyclization of reactive, π-enriched intermediates. In contrast to iterative Type I, Type II, and Type III PKS pathways, the spinosyn pathway introduces reactive π-bonds after the core polyketide backbone has been assembled by the PKS machinery. This may avoid byproducts generated from side reactions between adventitious nucleophiles and the conjugated π-systems of the reactive intermediates bound to the relatively-slow, processive Type I PKS enzymes ($k_{cat} \approx 1$ min^{-1} for the erythromycin PKS) (Khosla *et al.*, 2007). This unique strategy relies on a previously uncharacterized type of cyclase. Experimental and computational studies can now utilize the structural information reported here to further elucidate the details of this remarkable chemistry.

MATERIALS AND METHODS

Materials

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Carbosynth Limited, SAM was purchased from Ark Parm, Inc., and all other chemicals were purchased from Sigma-Aldrich or Fisher Scientific. All chemicals were used without any further purification. Ni-NTA resin was purchased from

Expedeon. Amicon centrifugal units were purchased from Millipore. PCR primers were obtained from Sigma-Aldrich.

Cloning

Previously, expression and purification of SpnF led to insoluble protein, and the chaperone proteins GroEL/ES were required to obtain soluble protein (Kim *et al.*, 2011). Reexamining the *spnF* sequence led to the observation of a new start codon upstream of the *spnF* ORF annotated by NCBI. The longer *spnF* gene construct was amplified from the genomic DNA of *Saccharopolyspora spinosa* strain NRLL18537 using the primers 5′-CCGGAATTCCATATGCAGAGATCAGGCATACCGGTG-3′ and 5′-CCAAGCTTTTAGCCGACCGGCTTCCGCGCCGTC-3′. The amplicon was then inserted into a pET28b(+) vector (Novagen) between the *Nde*I and *Hind*III sites. *Escherichia* coli TOP10 cells (Invitrogen) were transformed with the ligated plasmid and plated on Luria Broth (LB)-agar supplemented with kanamycin (50 μg/mL), and a desired expression plasmid was obtained from a single colony.

Protein expression and purification

The expression plasmid was transformed into the *E. coli* BL21(DE3) expression host. The transformed host was grown in overnight cultures (4 mL of LB with 50 μg/mL kanamycin) that were used to inoculate 6 × 1 L LB (50 mg/L kanamycin). The cultures were incubated at 37 °C with shaking until OD_{600} of 0.6, at which point the temperature was lowered to 15 °C. When cultures reached 15 °C, IPTG was added to a final concentration of 0.5 mM. The cultures were grown for an additional 16 h. Cells were harvested via centrifugation (5000 × *g* for 10 min) and resuspended in lysis buffer (20% (v/v) glycerol, 500 mM NaCl, 30 mM HEPES, pH

7.5). Cells were lysed with a sonicator and centrifuged (30000 × *g* for 30 min). The lysate was loaded onto a Ni-NTA column equilibrated with lysis buffer. The column was washed with 20 column volumes of 20 mM imidazole in lysis buffer, and protein was eluted with 2 column volumes of 200 mM imidazole in lysis buffer. SpnF was further purified on a gel filtration column (Superdex 200, GE Healthcare Life Sciences) equilibrated with 5% (v/v) glycerol, 150 mM NaCl, and 10 mM HEPES, pH 7.5. A centrifugal concentrator was employed to yield 23 mg/mL SpnF. Aliquots were flash frozen in liquid nitrogen and stored at -80 °C until further use.

Crystallization, data processing, and structure determination

Crystals were grown over 6 months at 25 °C with the sitting-drop vapor-diffusion method. Drops consisted of 2:1 ratio of SpnF (23 mg/mL, 2 mM SAM, 5% (v/v) glycerol, 150 mM NaCl, 10 mM HEPES, pH 7.5) and crystallization buffer (1.8 M ammonium sulfate, 2% (v/v) PEG 400, 10 mM HEPES, pH 7.5, 100 mM sodium malonate, pH 7.0). Crystals were briefly soaked in 35% (v/v) PEG 400, 50 mM HEPES, pH 7.5 before they were flash frozen in liquid nitrogen. Data for SpnF were collected at 100 K at Advanced Photon Source beamline 23-ID-D and processed and scaled in XDS (Kabsch, 2010) (**Table 4-1**). Data truncation was guided by $I/\sigma(I)$ and $CC_{1/2}$ values (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013). Free-R flags were added to 5% of reflections in Uniqueify of the CCP4 software suite (Winn *et al.*, 2011). The phases were solved with molecular replacement using the catalytic core of the rebeccamycin MT, RebM (PDB code: 3BUS), as the search model in the program Phaser (McCoy *et al.*, 2007) of CCP4. The solution was submitted to the ARP/wARP server for automated refinement (Langer *et al.*, 2008) before several rounds of refinement were carried out in Coot (Emsley *et al.*, 2004), Refmac5 of

CCP4 (Murshudov *et al.*, 1997), and phenix.refine of the PHENIX software package (Adams *et al.*, 2010). Default NCS restraints were enabled in Refmac5 and phenix.refine. All protein images were generated in PyMOL. Atomic coordinates for the SpnF/SAH complex have deposited in the Protein Data Bank under PDB code 4PNE.

SpnF activity assay

Assay mixtures of wild-type and mutant SpnF (15 μM) contained 2 mM SpnM substrate and 1.25 μM SpnM in a total volume of 50 μL Tris HCl buffer (50 mM, pH 8) (Kim *et al.*, 2011). The reaction was incubated for 2 h at 30 °C before quenching with 80 μL of cold ethanol. The sample was centrifuged and the supernatant was analyzed by HPLC using a Varian Microsorb-MV 100-5 C18 250 × 4.6 mm column. The following method was used: linear gradient of 5-30% B over 25 min, 30-70% B over 3 min, 70-5% B over 5 min (solvent A, water; solvent B, acetonitrile; flow-rate,1 mL/min; detection at 254 nm).

SpnF mutants

SpnF mutants (Y23F, Y23A, C38A, H42A, H42Q, A51T, R123Q, M147I, Q148A, Q148E, E152A, E152L, E152Q, T196A, T196S) were engineered by site-directed mutagenesis (QuikChange, Stratagene) using primers 5′-GCAGCAGGTTGGGCAGATGTTTGACCTGGTCA-3′ and 5′-TGACCAGGTCAAACATCTGCCCAACCTGCTGC-3′ for Y23F, 5′-CAGGTTGGGCAGATGGCTGACCTGGTCACGCC-3′ and 5′-GGCGTGACCAGGTCAGCCATCTGCCCAACCTG-3′ for Y23A, 5′-CGGGCGGCCCCGCCGCCATCCACC-3′ and 5′-GGTGGATGGCGGCGGGGCCGCCCG-3′ for C38A, 5′-CTGCGCCATCCACGCCGGCTACTGGGAG-3′ and 5′-

CTCCCAGTAGCCGGCGTGGATGGCGCAG-3′ for H42A, 5′-CTGCGCCATCCACCAAGGCTACTGGGA-3′ and 5′-TCCCAGTAGCCTTGGTGGATGGCGCAG-3′ for H42Q, 5′-GAGAACGACGGGCGGACTTCCTGGCAGCAGG-3′ and 5′-CCTGCTGCCAGGAAGTCCGCCCGTCGTTCTC-3′ for A51T, 5′-GCGGACTAAGCCACCAGGTGGACTTCTCGTG-3′ and 5′-CACGAGAAGTCCACCTGGTGGCTTAGTCCGC-3′ for R123Q, 5′-CGACGCCGCCTGGGCCATCCAGTCGCTGTTGGAGATG-3′ and 5′-CATCTCCAACAGCGACTGGATGGCCCAGGCGGCGTCG-3′ for M147I, 5′-GACGCCGCCTGGGCCATGGCGTCGCTGTTGGAGATGTC-3′ and 5′-GACATCTCCAACAGCGACGCCATGGCCCAGGCGGCGTC-3′ for Q148A, 5′-GCCTGGGCCATGGAGTCGCTGTTGG-3′ and 5′-CCAACAGCGACTCCATGGCCCAGGC-3′ for Q148E, 5′-GCAGTCGCTGTTGGCGATGTCCGAACCGG-3′ and 5′-CCGGTTCGGACATCGCCAACAGCGACTGC-3′ for E152A, 5′-CATGCAGTCGCTGTTGCTGATGTCCGAACCGGAC-3′ and 5′-GTCCGGTTCGGACATCAGCAACAGCGACTGCATG-3′ for E152L, and 5′-TGCAGTCGCTGTTGCAGATGTCCGAACCG-3′ and 5′-CGGTTCGGACATCTGCAACAGCGACTGCA-3′ for E152Q, 5′-GGGGACAGGTGGCCGTCCGGCCTTCGGATCT-3′ and 5′-AGATCCGAAGGCCGGACGGCCACCTGTCCCC-3′ for T196A, 5′-GGGGACAGGTGGCCGGCCGGCCTTCGGATCT-3′ and 5′-AGATCCGAAGGCCGGCCGGCCACCTGTCCCC-3′ for T196S. Protein expression and purification followed the protocol described for wild-type SpnF.

Energy minimization

Representative substrate and product structures were obtained using Avogadro 1.1.0 (Hanwell *et al.*, 2012). Energies were minimized using the steepest descent algorithm with the UFF and MMFF94s force fields (Rappe *et al.*, 1992; Halgren, 1999).

Consensus docking

All heteroatoms and solvent atoms were removed from the SpnF coordinates. Docking calculations were performed with AutoDock Vina (Trott & Olson, 2010) and DOCK6.6 (Brozell *et al.*, 2012). For AutoDock Vina, nonpolar H atoms were merged and a cubic grid (15 Å per side) centered on coordinates 6.983, 1.769, 15.230 was drawn around the manually docked SpnF product with AutoDockTools 1.5.6 (Morris *et al.*, 2009). For DOCK6, molecules were processed with Dock Prep, charges were adjusted with Add Charge, and after deleting hydrogen atoms the surface of SpnF was calculated with Write DMS (all features of UCSF Chimera 1.8.1) (Pettersen *et al.*, 2004). Spheres were generated in sphgen of DOCK6 and those within 6 Å of every atom of the manually docked SpnF product were accepted (any spheres outside of the substrate cavity were rejected). A scoring grid with 0.2 Å spacing was generated around the spheres with an addition 1.0 Å on each side, resulting in a box of 18.944 Å × 18.316 Å × 15.454 Å centered on coordinates 8.065, 0.447, 13.745. In each program, docking was performed with the optimized defaults. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6 (**Tables 4-2** – **4-9**).

ACKNOWLEDGEMENTS

We thank Mark Ruszczycky for discussions and advice. We thank Kay Diederichs for assistance in processing diffraction data. We thank the National Institutes of Health (GM106112, A.T.K. and GM35906, H.-w.L.) and the Welch Foundation (F-1712, A.T.K. and F-1511, H.-w.L.) for financial support. Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

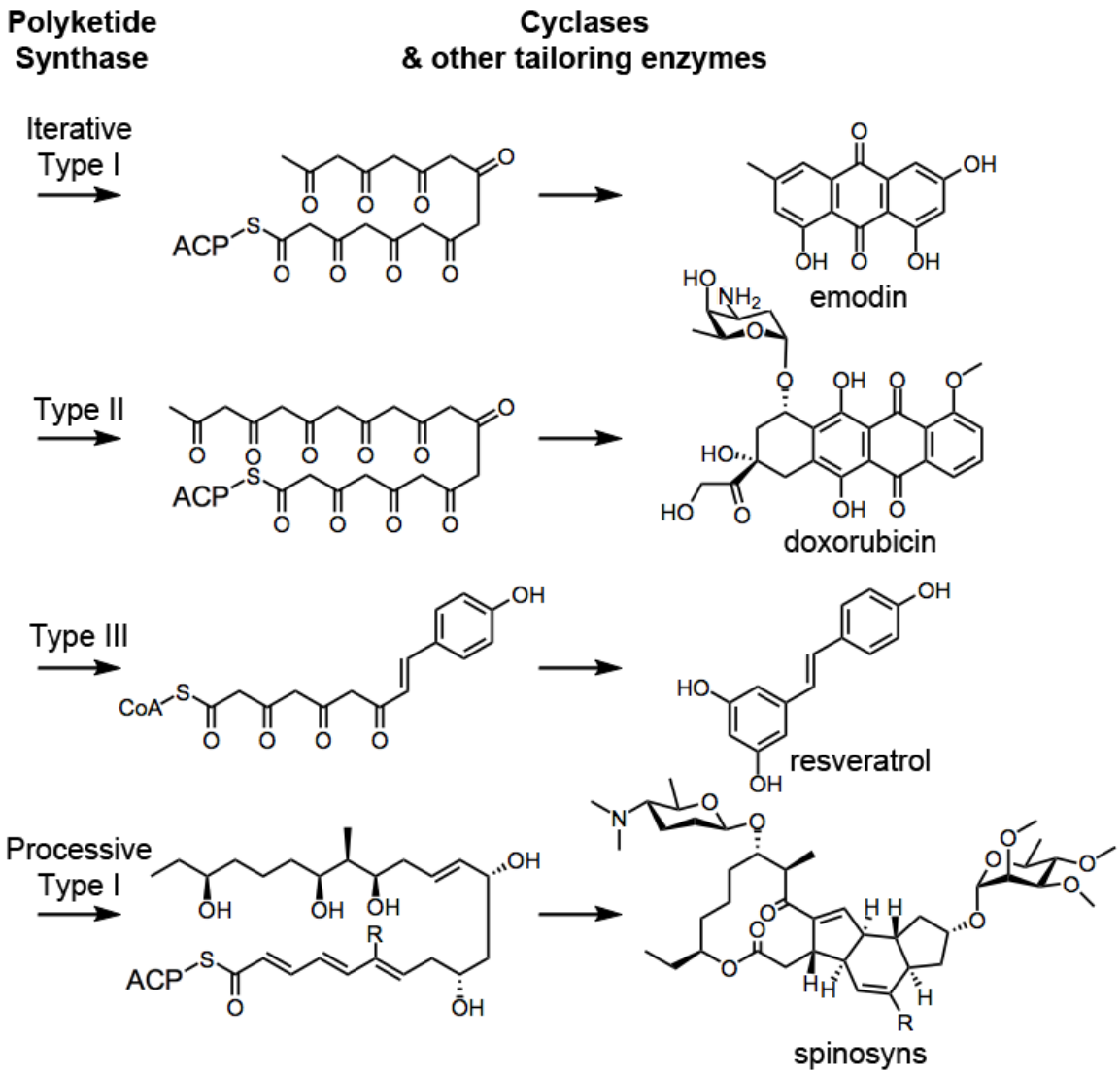


Figure 4-1. Production of polyketides containing carbocyles. Cyclases that accompany iterative PKSs often mediate the formation of polyketides with rings composed entirely of carbon. The spinosyn biosynthetic pathway is rare in that it contains a processive Type I PKS and generates carbocyclic polyketides.

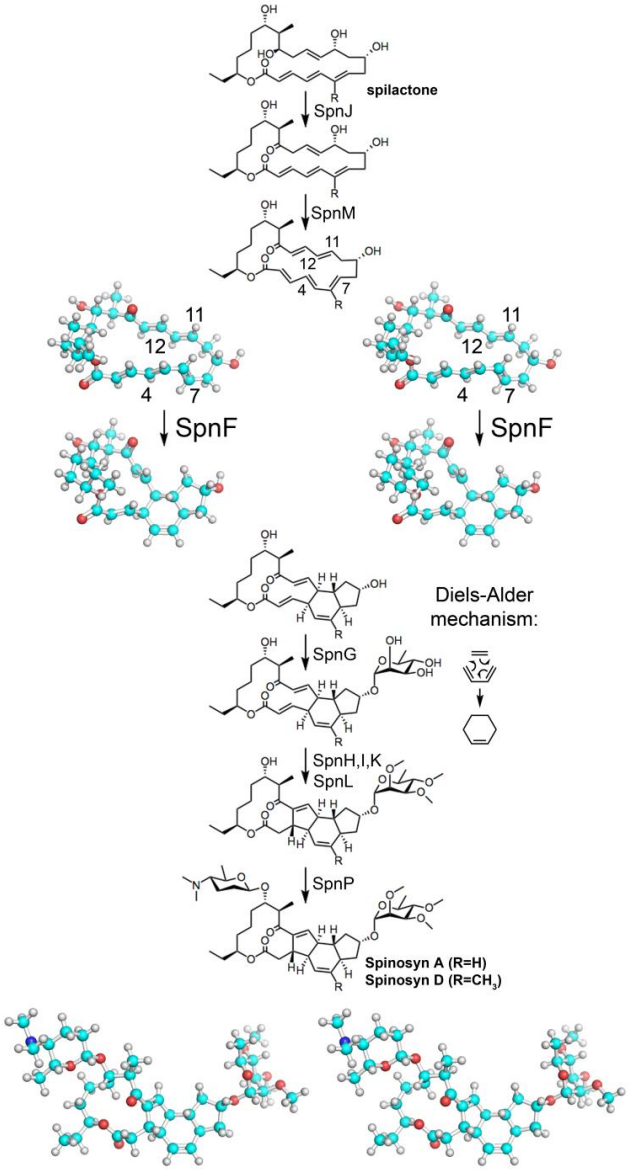


Figure 4-2. Reactions mediated by SpnF and accompanying tailoring enzymes. The spinosyn Type I PKS generates spilactone, which is processed by several tailoring enzymes into spinosad (Spinosyn A and D). The SpnF substrate can spontaneously undergo [4+2]-cycloaddition to yield the same product generated by the enzyme, albeit 500-fold more slowly. The SpnF substrate, with its C11-C12 double bond in s-*cis* geometry, is shown in its lowest energy conformation, as computed by the program Avogadro. Its rigid, 22-membered macrocycle resembles an elongated triangle, with one side constructed of 10 coplanar atoms and another of 7. SpnF has been proposed to catalyze a Diels-Alder reaction to yield the cyclohexene ring embedded in spinosyns. The SpnF substrate and product and Spinosyn A are depicted as stereo ball-and-sticks.

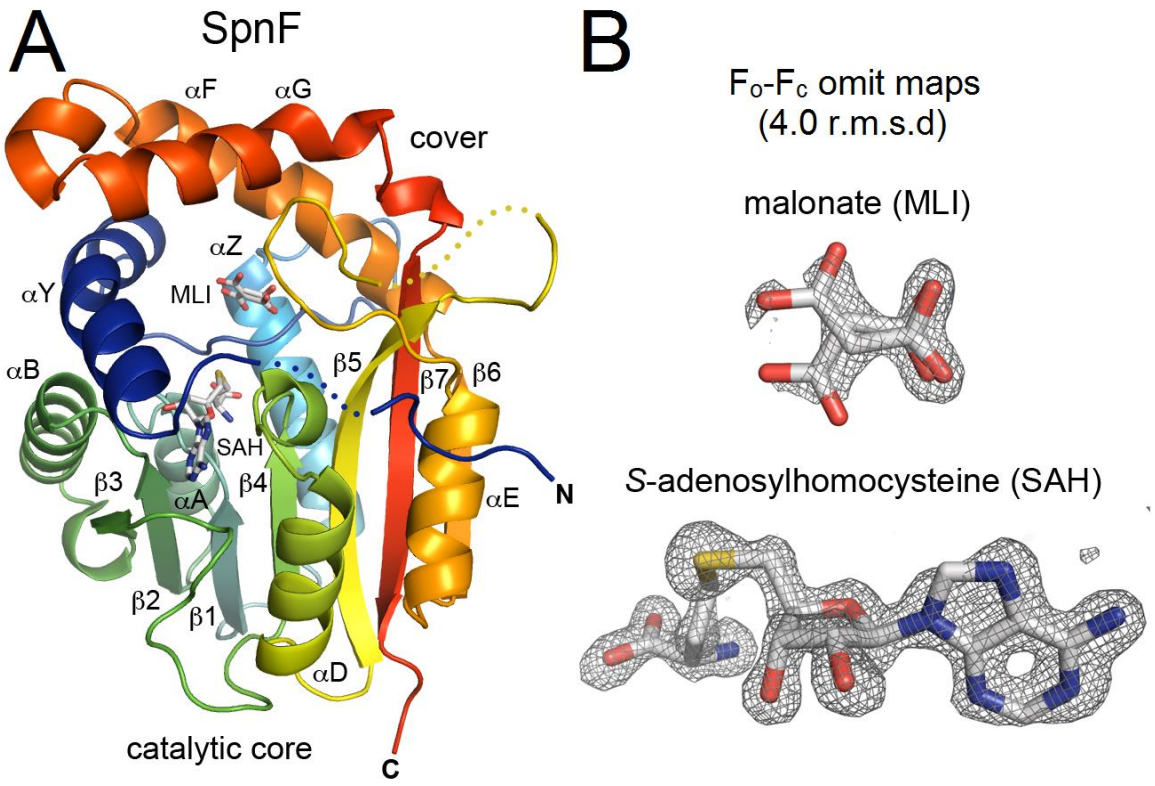


Figure 4-3. SpnF complexed with SAH at 1.50-Å resolution. **(A)** SpnF adopts a SAM-dependent methyltransferase fold, with a catalytic core and a cover like most natural product MTs. **(B)** The electron density maps reveal that two molecules are bound to each SpnF monomer – SAH and what appears to be malonate in two conformations. The F_o-F_c omit maps, contoured at 4.0 r.m.s.d., show the molecules of malonate and SAH bound to one of two SpnF monomers in the asymmetric unit.

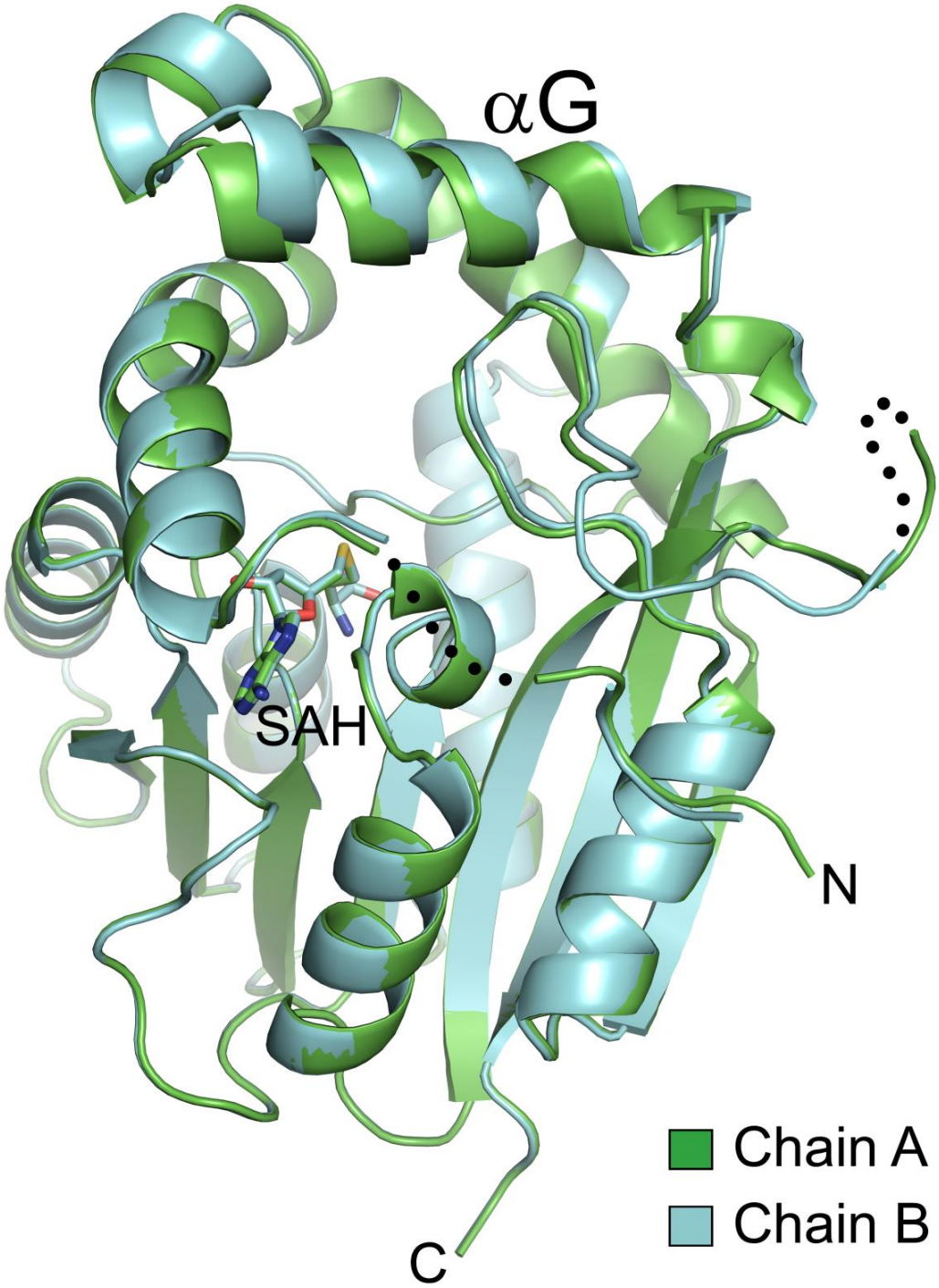


Figure 4-4. Superposition of the two SpnF monomers. The polypetide chains are nearly identical (0.165 Å r.m.s.d. over 1357 atoms), with the only major difference in conformation being the pitch of helix αG. N- and C-termini are labeled. Dotted lines indicate residues for which there is no observable electron density (9–12, 186–191 in chain A; 8–12, 184–185 in chain B).

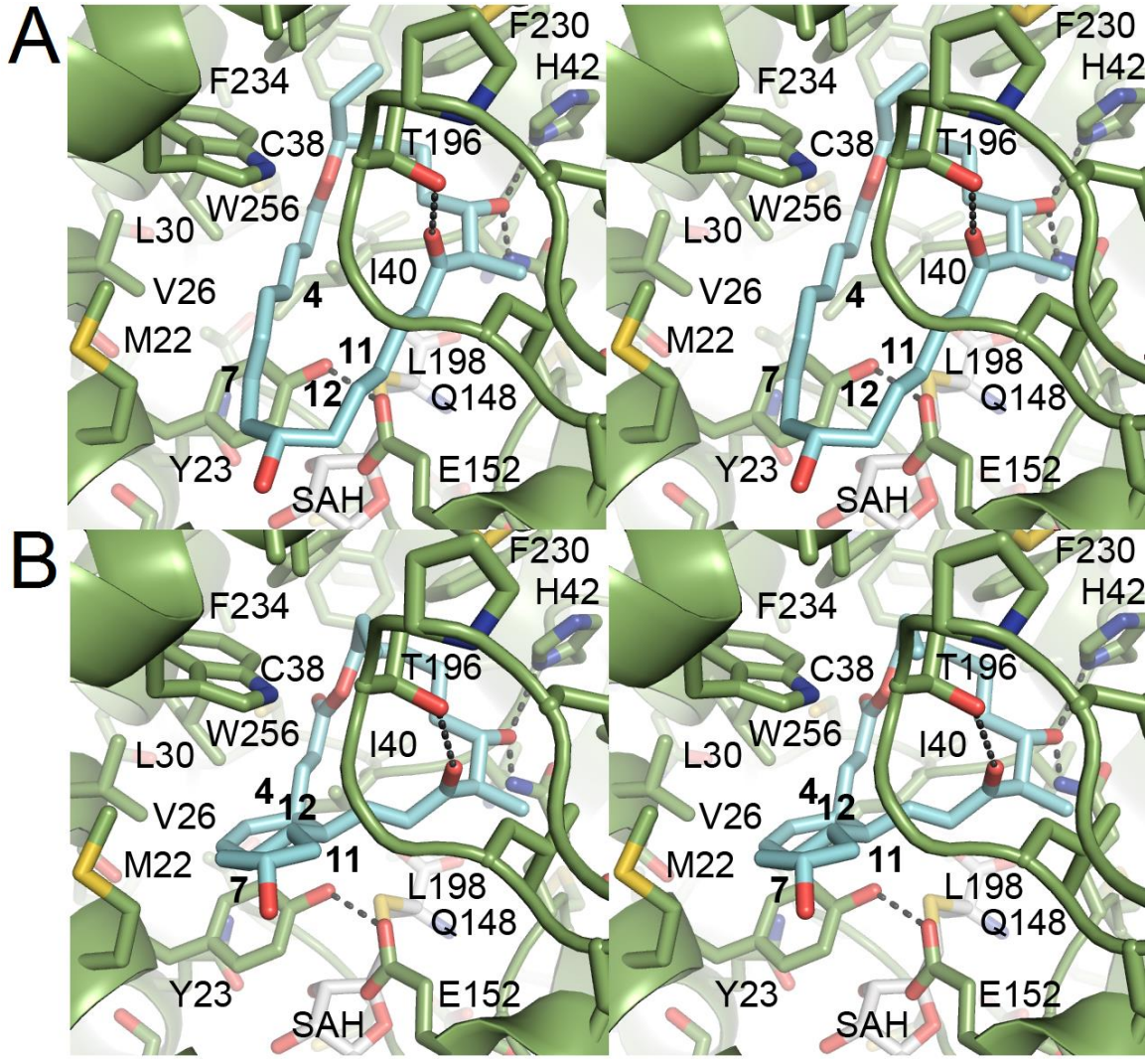


Figure 4-5. Consensus docking of **(A)** substrate and **(B)** product. The programs AutoDock Vina and DOCK6 were used to dock SpnF substrate and product (in the conformations shown in **Figure 4-2**) into the substrate cavity. A top orientation for both molecules in both programs is shown in the stereodiagram. The majority of contacts are hydrophobic although the C17-OH forms hydrogen bonds with both H42 and Q148. The C11-C12 π-bond may be polarized by T196 through its hydrogen bond to the C11-C15 π-system to react with the C4-C7 diene. SAH does not make contact with either the substrate or product. All solutions output by the docking software are illustrated in **Tables 4-2 – 4-9**.

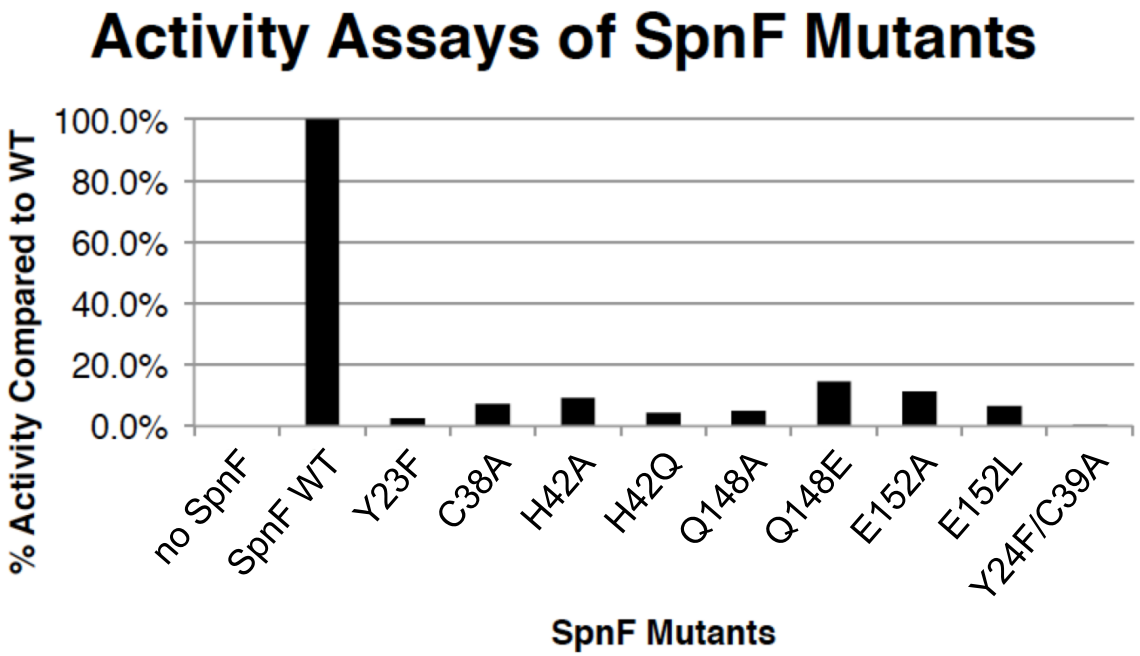


Figure 4-6. Assays of SpnF mutants. Point mutations at key residues shown in **Figure 4-5** were designed. Rates of tricyclic lactone formation were compared to controls with and without wild-type SpnF. All mutants showed severe losses in activity. (A51T, R123Q, M147I, T196A, T196S mutants have not yet been assayed for activity.)

Table 4-1. Crystallographic data and refinement statistics for SpnF.

Data collection and processing	
Space group	$P2_1$
Unit cell parameters	
a, b, c (Å)	41.74, 69.35, 83.63
α, β, γ (°)	90.0, 97.5, 90.0
Resolution (Å)	82.92-1.50 (1.54-1.50)
R_{merge} (%)	11.2 (96.4)
$CC_{1/2}$ (%)	99.6 (61.3)
Average $I/\sigma(I)$	9.12 (1.56)
Completeness (%)	98.0 (97.4)
No. of unique reflections	74158 (5443)
Redundancy	3.7 (3.8)
Refinement	
Resolution	82.92-1.50 (1.52-1.50)
R_{work}/R_{free} (%)	19.7/22.0 (28.8/33.2)
No. of atoms	8506
Protein	8158
Ligand	126
Water	222
Average *B*-factors (Å^2)	19.73
Protein	20.46
Ligand	14.99
Water	23.73
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.37
PDB code	4PNE

Values in parentheses refer to the highest-resolution shell.

Table 4-2. All AutoDock Vina outputs for SpnF substrate energy-minimized with UFF force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

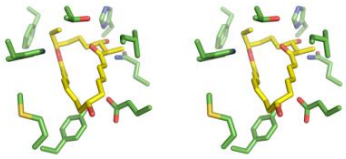
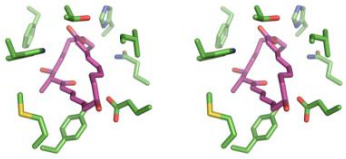
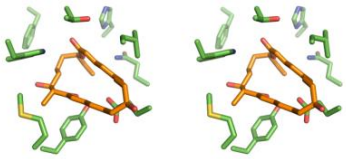
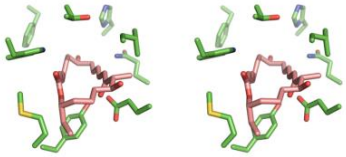
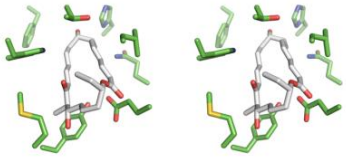
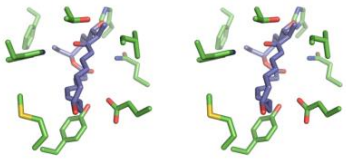
Substrate (UFF), AutoDock Vina		
Mode	Affinity (kcal/mol)	Docked pose
1	-7.5	
2	-7.1	
3	-5.9	
4	-5.3	
5	-5.0	
6	-4.9	

Table 4-3. All DOCK6 outputs for SpnF substrate energy-minimized with UFF force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

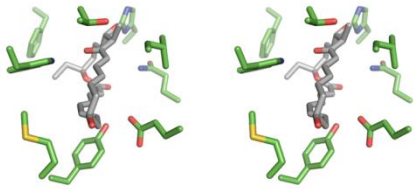
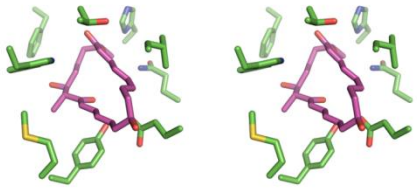
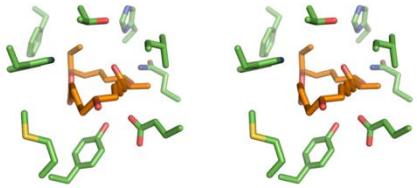
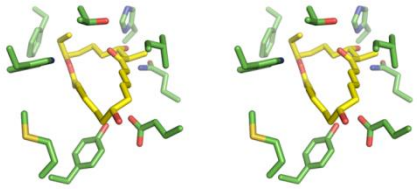
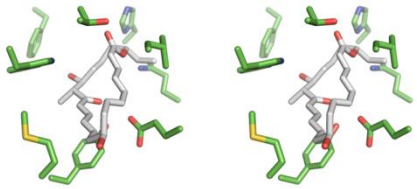
Substrate (UFF), DOCK6		
Mode	Grid score	Docked pose
1	-27.5	
2	-25.7	
3	-10.5	
4	-5.4	
5	-1.2	

Table 4-4. All AutoDock Vina outputs for SpnF substrate energy-minimized with MMFF94s force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

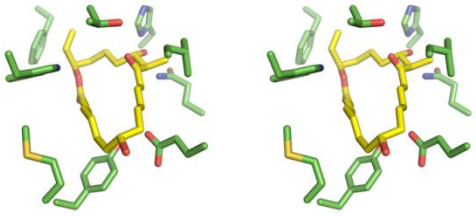
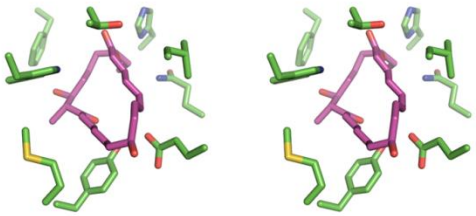
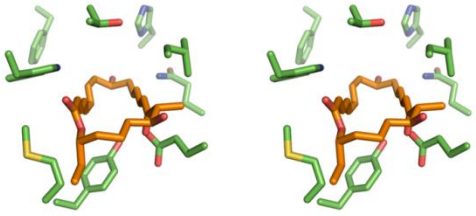
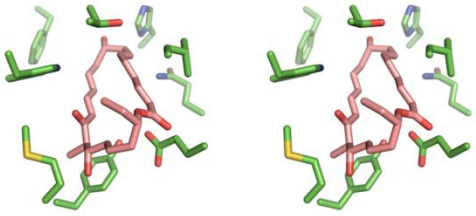
Substrate (MMFF94s), AutoDock Vina		
Mode	Affinity (kcal/mol)	Docked pose
1	-7.6	
2	-5.9	
3	-5.6	
4	-4.6	

Table 4-5. All DOCK6 outputs for SpnF substrate energy-minimized with MMFF94s force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

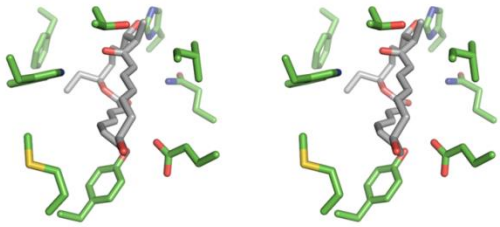
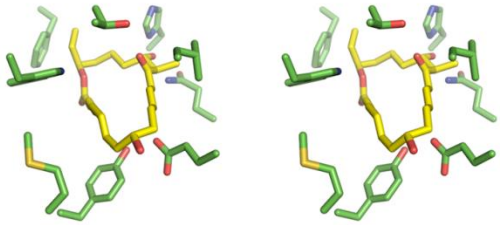
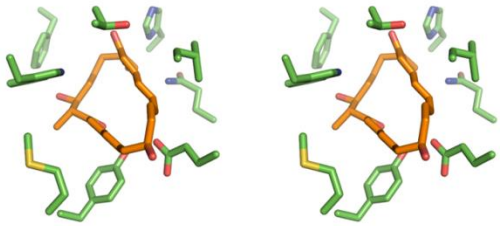
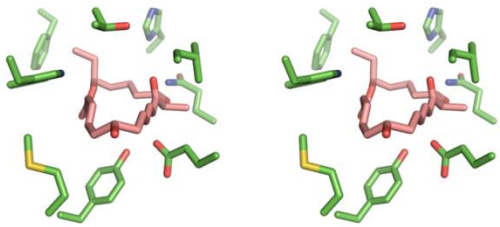
Substrate (MMFF94s), DOCK6		
Mode	Grid score	Docked pose
1	-18.4	
2	-9.1	
3	-3.7	
4	88.2	

Table 4-6. All AutoDock Vina outputs for SpnF product energy-minimized with UFF force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

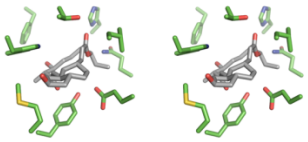
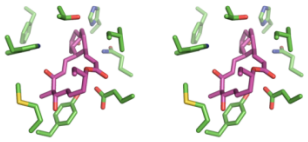
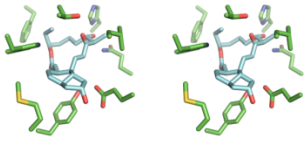
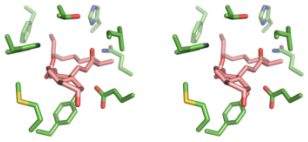
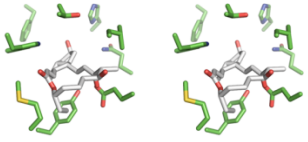
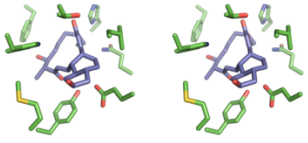
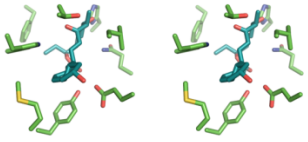
Product (UFF), AutoDock Vina		
Mode	Affinity (kcal/mol)	Docked pose
1	-7.0	
2	-6.6	
3	-6.3	
4	-6.2	
5	-5.7	
6	-4.6	
7	-4.1	

Table 4-7. All DOCK6 outputs for SpnF product energy-minimized with UFF force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

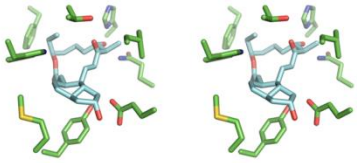
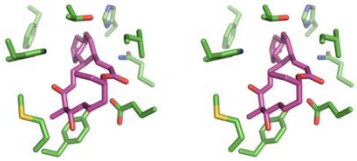
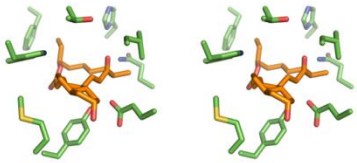
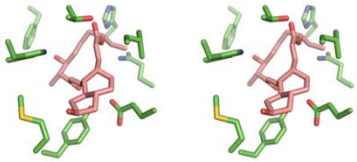
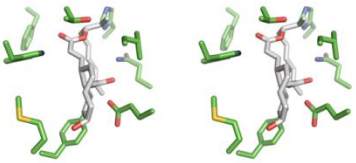
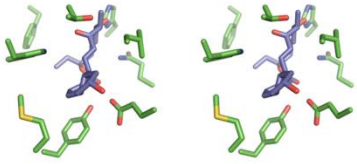
Product (UFF), DOCK6		
Mode	Grid score	Docked pose
1	-24.8	
2	-13.7	
3	-12.1	
4	-8.27	
5	3.6	
6	3.7	

Table 4-8. All AutoDock Vina outputs for SpnF product energy-minimized with MMFF94s force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

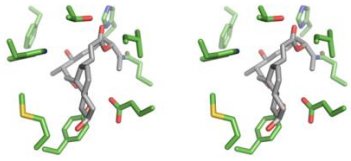
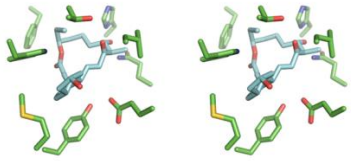
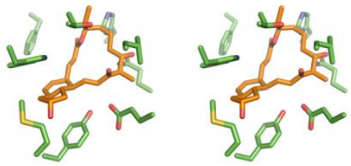
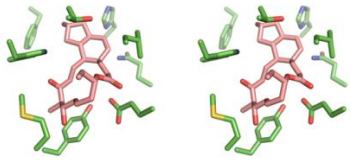
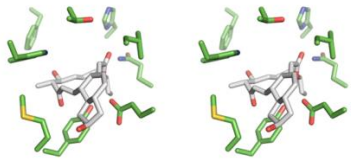
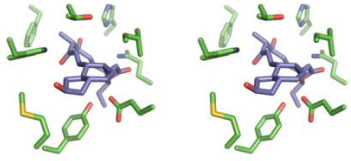
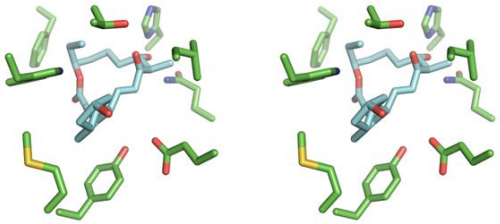
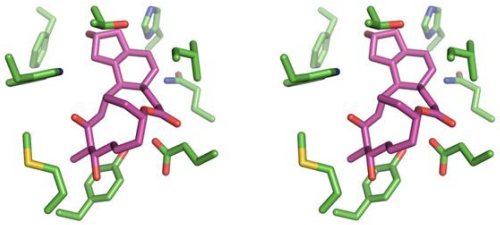
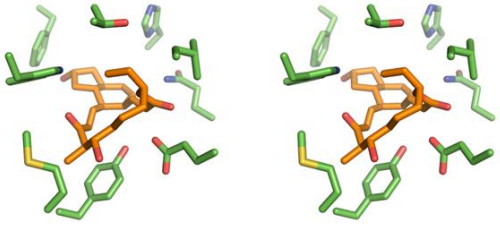
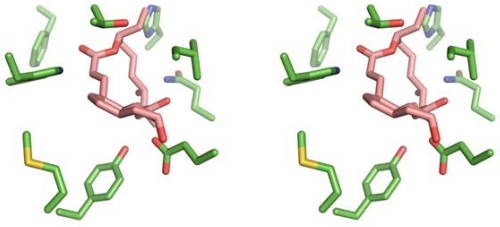
Product (MMFF94s), AutoDock Vina		
Mode	Affinity (kcal/mol)	Docked pose
1	-6.0	
2	-5.4	
3	-4.6	
4	-4.5	
5	-3.3	
6	-3.2	

Table 4-9. All DOCK6 outputs for SpnF product energy-minimized with MMFF94s force field. Docked poses are in stereo, with residues M22, Y23, H42, Q148, E152, T196, L198, F234, and W256 shown for reference. Top docking modes (highlighted) are chemically reasonable, fully contained within the binding pocket, and in agreement between AutoDock Vina and DOCK6.

Product (MMFF94s), DOCK6		
Mode	Grid score	Docked pose
1	3.0	
2	15.8	
3	58.0	
4	69.3	

Chapter 5: The Missing Linker: A Dimerization Motif Located within Polyketide Synthase Modules

Reprinted with permission from *ACS Chemical Biology*.

ABSTRACT

The dimerization of multimodular polyketide synthases is essential for their function. Motifs that supplement the contacts made by dimeric polyketide synthase enzymes have previously been characterized outside the boundaries of modules, at the N- and C-terminal ends of polyketide synthase subunits. Here we describe a heretofore uncharacterized dimerization motif located within modules. The dimeric state of this dimerization element was elucidated through the 2.6-Å resolution crystal structure of a fragment containing a dimerization element and a ketoreductase. The solution structure of a standalone dimerization element was revealed by nuclear magnetic resonance spectroscopy to be consistent with that of the crystal structure, and its dimerization constant was measured through analytical ultracentrifugation to be ~20 μM. The dimer buries ~990 Å^2 at its interface, and its C-terminal helices rigidly connect to ketoreductase domains to constrain their locations within a module. These structural restraints permitted the construction of a common type of polyketide synthase module.

INTRODUCTION

Modular polyketide synthases (PKSs) are megadalton enzymatic assembly lines that biosynthesize complex polyketides such as the antibacterial erythromycin, the antifungal amphotericin, and the insecticide spinosyn (Khosla *et al.*, 2007; Smith & Tsai, 2007; Meier & Burkart, 2009; Keatinge-Clay, 2012). During the synthesis of a

polyketide, each chain extension reaction is catalyzed by a separate module within the PKS. Thus, the three domains necessary for chain extension are present in all module types: an acyltransferase (AT) that selects an extender unit, a ketosynthase (KS) that catalyzes the condensation of the extender unit with a growing polyketide chain, and an acyl carrier protein (ACP) that shuttles an extender unit or an extended polyketide chain between active sites of the synthase. Within modules, processing enzymes may also be present, such as a ketoreductase (KR) that reduces the β-keto group of the newly extended polyketide chain to a β-hydroxy group, a dehydratase (DH) that removes the β-hydroxy group to form a double bond, and an enoylreductase (ER) that reduces the double bond. Multimodular PKS assembly lines are dimeric, in contrast to the related monomeric nonribosomal peptide synthetase (NRPS) assembly lines (**Figure 5-1**) (Sieber *et al.*, 2002; Hur *et al.*, 2012). However, the only PKS enzyme that requires a dimeric oligomerization state is KS, and the only other enzyme that makes contact across the 2-fold axis of modules is DH (Tang *et al.*, 2006; Keatinge-Clay, 2008; Akey *et al.*, 2010). The other domains commonly within modules, AT, KR, ER, and ACP, are monomeric and loop out from the 2-fold axis of the synthase. Accessory motifs have been identified that aid KS and DH in dimerizing the polypeptides of a PKS into its homodimeric subunits: a coiled coil is frequently located at the N-terminal end of a subunit, a four-helix bundle is often present at the C-terminal end, and a dimeric thioesterase (TE) that cleaves the elongated polyketide from the assembly line is usually positioned at the C-terminal end of the final PKS subunit (Broadhurst *et al.*, 2003; Buchholz *et al.*, 2009; Tsai *et al.*, 2001).

Here we report a dimerization motif located within PKS modules. This ~55-residue dimerization element (DE) is immediately N-terminal to the KR domain in

~50% of modules that contain KR as their sole processing enzyme (β-modules) (Keatinge-Clay, 2012). The 2.6-Å resolution crystal structure of a fragment from the third module of the spinosyn PKS that contains DE and KR, Spn(DE+KR)3, reveals the architecture of the DE dimer as well as a rigid connection between the last helix of DE and the KR domain. NMR analysis of a standalone DE from the first module of the spinosyn PKS is consistent with the crystal structure. Analytical ultracentrifugation revealed its dimerization constant to be 19.9 ± 1.5 μM. By positioning the C-terminal ends of a structurally characterized KS+AT didomain in proximity with the N-terminal ends of the Spn(DE+KR)3 didomain, a model of a β-module was constructed. Thus, the biophysical characterization of DE has also helped to elucidate the domain organization of a common type of PKS module.

RESULTS AND DISCUSSION

Crystal structure of a DE+KR dimer

A region containing the DE and KR domains of the third module of the spinosyn PKS (Spn(DE+KR)3; SpnC residues 913–1471, numbered here 1–559) was heterologously expressed in *E. coli* BL21(DE3) and purified to homogeneity (Waldon *et al.*, 2001). Crystals of Spn(DE+KR)3 diffracted to 2.6 Å resolution, and its structure was solved by molecular replacement using AmpKR2 (PDB code: 3MJS) as a search model (**Figure 5-2**) (Zheng *et al.*, 2010). The asymmetric unit contains two Spn(DE+KR)3 dimers. Well-defined electron density at the N-terminal ends of the KR domains enabled the DE domains to be built. The KR domain is similar to those already reported, particularly AmpKR2 (1.1 Å C_α r.m.s.d.), another A1-type KR. As in that structure, a conserved methionine (Met371) latches over bound $NADP^+$ and

helps order the adjacent lid helix (Zheng *et al.*, 2010; Keatinge-Clay & Stroud, 2006; Keatinge-Clay, 2007). Searches on the Dali and CATH servers indicate that DE possesses a novel fold (Green *et al.*, 2007; Holm & Rosenstrom, 2010).

Each DE monomer contains three helices (α1, α2, and α3; residues 12–21, 26–33, and 41–62, respectively) connected by two loops (loop I and loop II; residues 22–25 and 34–40), nearly as predicted by the PSIPRED server from its sequence (Bryson *et al.*, 2005; Jones, 1999). Within each monomer, α2 lies out of the plane formed by α1 and α3. Through the association of two monomers, a V-shaped dimer is generated with an angle of ~65° between the α3 helices (**Figure 5-2A**). A surface area of ~990 $Å^2$ is buried at the dimer interface (Reynolds *et al.*, 2009). The main dimeric contact is between the N-terminal end of α3 from one monomer and a pocket created by the helices of the other monomer, specifically between residues L41, T44, L45, L48, W51 of α3 from one monomer and residues F17, W18, V21, L26, V29, L33, L35, L41, T44, L48 from the other monomer. The most significant hydrophobic contacts are made by the nearly invariant F17, W18, and W51.

Several residues from α3 mediate the connection between DE and KR (**Figure 5-2B**). Interactions with αG of the KR domain include hydrophobic contacts between E57, R58, A61, and L483, as well as a hydrogen bond between E57 and S480. At the C-terminal end of α3, W64 inserts into a hydrophobic pocket created by M476, P477, S480, I498, and D500, and R65 forms a salt bridge with E493. Rigid residues connect the end of DE to the beginning of the KR domain (α3 and β1, respectively).

The DE dimer possesses several mysterious features (**Figure 5-2C**). Its hydrophobic core appears somewhat loosely packed, perhaps due in part to interactions made by the invariant residue W18. The indole NHs of W18, and hence

their dipoles, point toward one another across the DE 2-fold axis, with the hydrogens separated by only 1.6 Å. Within α3, P46 neither introduces a kink nor interacts with neighboring residues, yet it is completely conserved (**Figure 5-2D**). Both of the SpnDE3 dimers in the asymmetric unit are slightly asymmetric, with the third helix (α3) of one monomer relatively bent by ~53° (**Figure 5-3**). The "straight" and "bent" monomers within each DE dimer show lower structural similarity with one another (0.93 and 1.0 Å Cα r.m.s.d.) than the equivalent "straight" and "bent" DE monomers from the other dimer (0.43 and 0.29 Å Cα r.m.s.d., respectively). Usually all or none of the β-modules in a PKS contain a DE; however, some PKSs (e.g., the oleandomycin PKS) are composed of both β-modules that contain DE as well as β-modules that do not. The DE dimer is apparently unique to cis-AT PKSs.

Other dimerization motifs that supplement the dimeric contacts made by KS, DH, and TE have been characterized (Broadhurst *et al.*, 2003; Buchholz *et al.*, 2009). The N-terminal docking domain (NDD) is a ~25-residue coiled coil commonly at the N-terminal ends of subunits. The C-terminal docking domain (CDD) at the C-terminal ends of subunits is composed of three helices from each polypeptide; the first two helices dimerize to form four-helix bundles, while the last helix specifically binds to the NDD of a downstream subunit to form a second four-helix bundle. Thus, the set of nonenzymatic dimerization motifs in PKS subunits is the NDD/CDD pair and the three-helix DE described here.

NMR analysis of a standalone DE

The many unusual features of DE prompted us to study an isolated DE domain in solution (**Figure 5-4**). Thus, SpnDE1 (SpnA residues 1877–1944, numbered 1–68 here; ~24% identical to SpnDE3) was investigated by NMR

spectroscopy (**Figure 5-4A**). Like SpnDE3, SpnDE1 is composed of three helices (α1, α2, and α3; residues 14–27, 31–37, and 44–67) and two loops (loop I and loop II; residues 28–30 and 38–43) as determined by NOE cross-peak patterns and PREDITOR (of 2037 unique NOEs observed for the SpnDE1 dimer, 548 were intra-residue and 1489 were inter-residue) (**Figure 5-4B**) (Berjanskii *et al.*, 2006). Interestingly, the N-terminal end of α3 in SpnDE1 is six residues longer than the equivalent helix in SpnDE3.

The spectra collected for SpnDE1 are consistent with the conformation of DE observed in the crystal structure. Intermonomer NOEs indicate that the solution structure is indeed dimeric. For example, the V26 methyl protons interact with L54, L57, and A58 side-chain protons at distances of 3–6 Å (V26, L54, and L57 are highly conserved). Such NOEs could not originate from within a monomer unless substantial conformational changes occurred. The atypically high chemical shift of 11.55 ppm for the indole NH of the invariant W23 (equivalent to W18 in SpnDE3) also supports a model in which those protons are in close proximity, with the symmetry-related tryptophan indole rings deshielding one another. The relative stabilities of different regions of the SpnDE1 monomer were examined. Two-dimensional ^{15}N–^{1}H NMR spectra collected at several pH values revealed the relative amide proton exchange rates. The amide protons of the 21 N-terminal and 10 C-terminal residues are the most susceptible to exchange, as are those of residues D42–A46 on the solvent-exposed loop II; none of these residues are highly conserved. In contrast, the amide protons of E27–D30 in loop I are protected from exchange, consistent with their being steadfastly hydrogen-bonded. Loop I is consistently four residues in length with an aspartate in the fourth position that caps α2.

Contribution of DE to dimerization

To our knowledge, no dimerization constant for a PKS dimerization motif has been determined. Therefore, we sought to measure the dimerization constant of an isolated DE. We performed sedimentation velocity analysis of SpnDE1 spanning a concentration range from 1.85 to 79.2 μM. Van Holde–Weischet analysis of the results showed a clear mass action shift of the sedimentation coefficient distribution consistent with a rapid monomer–dimer equilibrium (**Figure 5-5**) (Demeler *et al.*, 1997). To further identify the oligomerization state, the experimental data were fitted by parsimonious regularization using genetic algorithm analysis (Brooks & Demeler, 2007) followed by Monte Carlo analysis (Demeler & Brookes, 2008). For each concentration, a major species was identified whose sedimentation coefficients were globally fitted to a binding isotherm. This approach yielded a dimerization constant of 19.9 ± 1.5 μM for SpnDE1. Molecular weight predictions from the genetic algorithm analysis for the lowest and highest concentrations analyzed resulted in a molecular weight of 8.61 kDa for the lowest concentration (7.8 kDa expected for the monomer) and 13.8 kDa for the highest concentration measured (15.6 kDa expected for the dimer), indicating that the fitted association reaction indeed reflects the monomer–dimer reaction.

We sought to observe how the oligomerization state of DE is affected by the addition of a KR domain. While Spn(DE+KR)3 appeared to be dimeric by gel filtration chromatography (**Figure 5-6**), as had Amp(DE+KR)2 in previous studies (Zheng *et al.*, 2010), we opted for sedimentation velocity analytical ultracentrifugation to more accurately determine oligomerization states. Spn(DE+KR)3 could not be used for technical reasons, and thus measurements were made with Amp(DE+KR)2. Analytical ultracentrifugation showed Amp(DE+KR)2

was not a dimer, but rather a monomer (~56 kDa as determined by genetic algorithm–Monte Carlo analysis, 56 kDa expected) when measured at 3.8 and 11.5 μM (**Figure 5-7**). Since Spn(DE+KR)3 is dimeric within the crystal structure, the dimerization constant for a DE+KR fragment may be closer to crystallographic concentrations (~10 mM). The implication is that as the number of enzymes within a PKS subunit increases, so must the number of dimerization motifs.

Modeling a β-module

The crystal structure of Spn(DE+KR)3 enabled us to construct a model of a β-module (**Figure 5-8**). The relative positions of a KS+AT dimer, a DE+KR dimer, and two ACP domains were restrained by the average lengths of interdomain peptides (from the set of DE-containing β-modules in the MAPSI database): ~14-residue, unconserved linkers connect the KS+AT dimer to DEs (the shortest observed was 9 residues in the third module of the geldanamycin PKS), ~14-residue, unconserved linkers connect KRs to ACPs (the shortest observed was 9 residues in the seventh module of the spinosyn PKS), and ~19-residue, unconserved linkers connect ACPs to KSs of the next module (the shortest observed was 16 residues in the sixth module of the spinosyn PKS) (**Figure 5-9**) (Tang *et al.*, 2006; Alekseyev *et al.*, 2007; Tae *et al.*, 2009). The first two linkers were included in the Spn(DE+KR)3 construct that crystallized; that they did not appear in the electron density maps indicates their inherent flexibility. The low sequence conservation of the third linker also suggests its unstructured nature. Thus, flexible linkers of appropriate length were built between the structured fragments.

The model naturally raises several considerations about the assembly line architecture of a PKS: (1) Modules must be as close to one another as possible along

the 2-fold axis of a PKS subunit in order to provide the ACP domains, bound by ~19-residue linkers to the next module, enough slack to access KSs within the same module. (2) AT domains may need to adopt another conformation in order for bound extender units to be accessible to cognate ACP domains (but not to ACPs from the preceding module). (3) A physical interaction may exist between the KS and KR domains via the KR structural subdomain. A hydrophobic groove (in part created by W85 and W184 in SpnKR3) is conserved in the KR structural subdomain and faces KS in the model. An interaction between KS and KR could explain why DE is not always present in β-modules. (4) KRs may be oriented similarly in DH- and ER-containing modules (γ- and δ-modules). The DH dimer can fit between the KR domains, and ERs have sufficient space to loop out from the KR domains in the orientation observed in the KR+ER didomain structure (Keatinge-Clay, 2008; Akey *et al.*, 2010; Zheng *et al.*, 2010). Through the evolution of processive, multimodular synthases from iterative synthases that may have resembled the modern mammalian fatty acid synthase (FAS), several architectural refinements occurred (Maier *et al.*, 2008; Brignole *et al.*, 2009). While the relative orientations of the chain-elongation enzymes, KS and AT, did not change significantly, those of the processing enzymes diverged substantially (Tang *et al.*, 2006; Tang *et al.*, 2007; Pappenberger *et al.*, 2010).

Through the transition of a dimeric to monomeric ER, space necessary for ACPs to access their cognate enzymes became available (Zheng *et al.*, 2012); however, without the dimeric contacts provided by ER domains, additional interfaces may have been necessary for subunits to remain dimeric. A comparison of multimodular PKSs to FASs shows the loss of the ER dimer interface, but the addition of interfaces at the N-terminal ends of DH and TE, as well as those made by

CDD, NDD, and DE. Thus, enough avidity is generated through these dimerization contacts to keep the polypeptides of PKS subunits dimerized.

As the last common structured element within multimodular PKSs to be structurally elucidated, DE relates the orientations of the processing enzymes to those of the chain-elongation enzymes. In the absence of a crystal structure of a PKS module, this small three-helix element represents a very informative final piece in the quintessential puzzle that is the molecular architecture of PKS assembly lines.

MATERIALS AND METHODS

Cloning, expression, and purification

The Spn(DE+KR)3 gene fragment was amplified from genomic DNA extracted from *Saccharopolyspora spinosa* (DNeasy Blood & Tissue Kit; Qiagen) with primers 5′-ATCGTAATC*CATATG*GCTGCCTCCGACGAGGCTCGT-3′ and 5′-TGATTCGAT*CTCGAG*TCACACCAGTTCGTGGCGTCGTTG-3′ (restriction sites in italics; stop codon underlined), digested with *Nde*I and *Xho*I, and ligated into pET28b (Novagen). *E. coli* BL21(DE3) cells transformed with the expression plasmid were inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD_{600} of ~0.4, and induced with 0.5 mM IPTG. After 12 h at 15 °C, cells were harvested by centrifugation and resuspended in lysis buffer (300 mM NaCl, 10% (v/v) glycerol, 50 mM HEPES pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 × *g* for 45 min). The supernatant was poured over a column of Ni-NTA resin (Qiagen), which was then washed with 50 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 300 mM imidazole. Spn(DE+KR)3 was further purified using a Superdex 200 gel filtration column (GE

Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10% (v/v) glycerol, 10 mM HEPES pH 7.5. Spn(DE+KR)3 was concentrated to 20 mg/mL in 25 mM NaCl, 10% glycerol (v/v), 1 mM DTT, 10 mM HEPES pH 7.5.

The SpnDE1 gene fragment was amplified from *S. spinosa* genomic DNA with primers 5′-ATGCTTGCA*CATATG*GCGCTGCGTGAATCTTCCGCCGGCGA-3′ and 5′-GTACGAACG*GAATTC*TCACTGATCCATAATGGACTCGGT-3′. Amplicons were digested with *Nde*I and *Eco*RI and ligated into pET28b for transformation into *E. coli* BL21(DE3) cells. SpnDE1 was expressed and purified as described above; however, following elution from the Ni-NTA matrix, SpnDE1 was exchanged into thrombin cleavage buffer (20 mM Tris-HCl pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$) and digested with 1 U of thrombin (GE Healthcare Life Sciences) per milligram of protein at 4 °C for 12 h. To remove thrombin and uncleaved protein, the digest was applied to 1 mL Ni-NTA and 0.5 mL benzamidine sepharose 6B (GE Healthcare Life Sciences) and eluted with 15 mL lysis buffer. SpnDE1 was then purified using a Superdex 75 gel filtration column (GE Healthcare Life Sciences) equilibrated with 100 mM sodium phosphate pH 7.0. SpnDE1 samples enriched in ^{15}N and/or ^{13}C were prepared as unlabeled samples were, but in M9 minimal media containing 0.6 g/L [^{15}N]ammonium chloride and/or 2 g/L [^{13}C]D-glucose as the sole nitrogen and/or carbon source. Amp(DE+KR)2 was prepared as previously described (Zheng *et al.*, 2010).

Crystallization and structure determination

Crystals were grown by sitting drop vapor diffusion at 22 °C by mixing 2 µL of protein solution (20 mg/mL in 25 mM NaCl, 1 mM DTT, 10% (v/v) glycerol, 10 mM HEPES pH 7.5) with 2.5 µL of crystallization buffer (15% (w/v) poly(ethylene

glycol) 3350, 0.2 M magnesium acetate, 30% (v/v) glycerol). X-ray diffraction data were collected at 100 K at beamline 5.0.2 of the Advanced Light Source synchrotron. The collected data were processed with HKL2000 (Otwinowski & Minor, 1997). The structure was solved by molecular replacement in Phaser using AmpKR2 (PDB code: 3MJS) as the search model (Potterton *et al.*, 2003; Zheng *et al.*, 2010). The model was iteratively built in Coot and refined in Refmac (**Table 5-1**) (Potterton *et al.*, 2003; Emsley & Cowtan, 2004). Local noncrystallographic symmetry (NCS) restraints, automatically generated by Refmac, were used in the refinement for the SpnKR3 domains. Atomic coordinates of the Spn(DE+KR)3 have been deposited in the Protein Data Bank (PDB code: 4IMP).

NMR spectroscopy

NMR spectra of SpnDE1 were collected at 25 °C using a 500-MHz Varian Inova spectrometer equipped with a triple-resonance probe and a z-axis pulsed-field gradient. NMR samples consisted of 1–7 mM protein in 100 mM sodium phosphate pH 7.0 and 90% H_2O/10% D_2O. Resonances were assigned using standard three-dimensional (3D) triple-resonance methods. ^{1}H, ^{13}C, and ^{15}N nuclei of backbone atoms were assigned using 3D HNCA, HNCACB, HNCO, and HN(CO)CACB spectra (Muhandiram & Kay, 1994); side-chain resonances were assigned using 2D ^{1}H–^{1}H 2QF-COSY and TOCSY spectra as well as 3D ^{15}N–^{1}H–^{1}H HSQC-TOCSY and ^{13}C–^{1}H–^{1}H HCCH-TOCSY spectra (Kay *et al.*, 1993). NOE cross-peaks were identified using 2D ^{1}H–^{1}H NOESY as well as 3D ^{15}N–^{1}H–^{1}H and ^{13}C–^{1}H–^{1}H HSQC-NOESY spectra (Pascal *et al.*, 1994). Data were processed with NMRPipe (Delaglio *et al.*, 1995). ^{1}H, ^{13}C, and ^{15}N chemical shifts were referenced using internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as described (Wishart

et al., 1995), where the methyl ^{1}H reference frequency of DSS was multiplied by 0.251449530 and 0.101329118 to obtain the 0-ppm frequencies for ^{13}C and ^{15}N, respectively. Helices were identified by their characteristic NOE cross-peak patterns and the chemical-shift values of C_α, C_β, C, and H_α nuclei. Backbone torsion angles φ and ψ were calculated from ^{1}H, ^{13}C, and ^{15}N chemical shifts with PREDITOR (Berjanskii *et al.*, 2006).

The rates at which the backbone amide protons of SpnDE1 exchange with solvent were measured using a saturation-transfer method (Lillemoen *et al.*, 1997). Uniformly ^{15}N-labeled samples were equilibrated with 100 mM sodium phosphate pH 5.0, 7.0, or 8.0. At each pH, 2D ^{15}N–^{1}H spectra were obtained with and without a 2-s presaturation of the water resonance. The rate at which the amide protons exchange with solvent was estimated by comparing peak heights in spectra obtained with and without presaturation at each pH. Nuclei in relatively well-ordered and relatively flexible regions of the protein were identified by their positive and negative peaks, respectively, in 2D ^{15}N–^{1}H heteronuclear NOESY spectra. Chemical shifts were deposited at the Biological Magnetic Resonance Bank (accession number 18947).

Gel filtration chromatography

A 0.1-mL protein solution of Spn(DE+KR)3 (~100 μM) was injected onto a Superdex 200 gel filtration column equilibrated with 150 mM NaCl, 10% glycerol (v/v), 10 mM HEPES pH 7.5. The molecular weight was estimated by comparing to standards (Gel Filtration Standard; Bio-Rad), as previously described (Zheng *et al.*, 2010).

Analytical ultracentrifugation

Sedimentation velocity experiments were performed on a Beckman Optima XL-I at the Center for Analytical Ultracentrifugation of Macromolecular Assemblies (CAUMA) at the University of Texas Health Science Center at San Antonio. Calculations were performed with the UltraScan software (Demeler, 2012) at the Texas Advanced Computing Center at the University of Texas at Austin and at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio (Brookes & Demeler, 2008; Demeler, 2010). All samples were measured in a 10 mM phosphate buffer containing 150 mM NaCl. All data were collected at 20 °C and spun at 45 krpm (Amp(KR+DE)2) or 60 krpm (SpnDE1), using titanium 2-channel centerpieces (Nanolytics; Potsdam, Germany). The partial specific volume of Amp(KR+DE)2 and SpnDE1 were determined to be 0.7329 cm^3/g and 0.7237 cm^3/g, respectively, based on protein sequence according to the method by Durchschlag, as implemented in UltraScan. All data were first analyzed by two-dimensional spectrum analysis with simultaneous removal of time-invariant noise (Brookes *et al.*, 2006; Brookes *et al.*, 2010) and then by enhanced van Holde–Weischet analysis (Demeler & van Holde, 2004) and genetic algorithm refinement (Brookes & Demeler, 2006; Brookes & Demeler, 2007) where applicable, followed by Monte Carlo analysis (Demeler & Brookes, 2008).

Modeling a β-module

The Ery(KS+AT)5 dimer (PDB code: 2HG4), the Spn(DE+KR)3 dimer (PDB code: 4IMP), and two EryACP2 monomers (PDB code: 2JU2) were positioned with PyMOL along a common 2-fold axis (Tang *et al.*, 2006; Alekseyev *et al.*, 2007; DeLano). The DE+KR dimer was arranged so that its N-terminal ends were as close as possible to the C-terminal ends of the KS+AT dimer and so that its long axis was

perpendicular to that of the KS+AT dimer to avoid steric clashes. Another Ery(KS+AT)5 dimer was placed along the common 2-fold axis, parallel to the first KS+AT dimer and as close as possible to the DE+KR dimer, to represent the next module in a PKS subunit and to provide an anchor for the C-terminal linker of ACP. The ACP monomers were positioned between the KS dimers. Polyalanine linkers were built using Coot in order to connect the KS+AT dimer with the DE+KR dimer, the DE+KR dimer with the ACP domains, and the ACP domains with the downstream KS+AT dimer (Emsley & Cowtan, 2004). Linker lengths were determined from sequence alignments of appropriate linkers obtained from MAPSI (**Figure 5-9**) (Larkin *et al.*, 2007; Tae *et al.*, 2009). Measurements were made between residues known to be structured; the linker spanning KS+AT and DE was measured between the YWL motif and 5 residues prior to the first conserved DE tryptophan; the linker spanning KR and ACP was measured between 7 residues after a highly conserved proline and helix 0 (Alekseyev *et al.*, 2007); the linker spanning ACP and KS was measured between the last highly conserved ACP glycine and the first highly conserved KS proline.

ACKNOWLEDGEMENTS

Financial support was provided by Welch Foundation Grant F-1712 (A.T.K-C.). Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility at the University of Texas at Austin, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology. The Berkeley Center for Structural Biology is supported in part by the National Institutes

of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The development of the UltraScan software is supported by the National Institutes of Health through grant RR022200 (B.D.). Supercomputer time allocations were provided through National Science Foundation grant TGMCB070038 (B.D.). We acknowledge the support of the San Antonio Cancer Institute grant P30 CA054174 for the Center for Analytical Ultracentrifugation of Macromolecular Assemblies at the University of Texas Health Science Center at San Antonio.

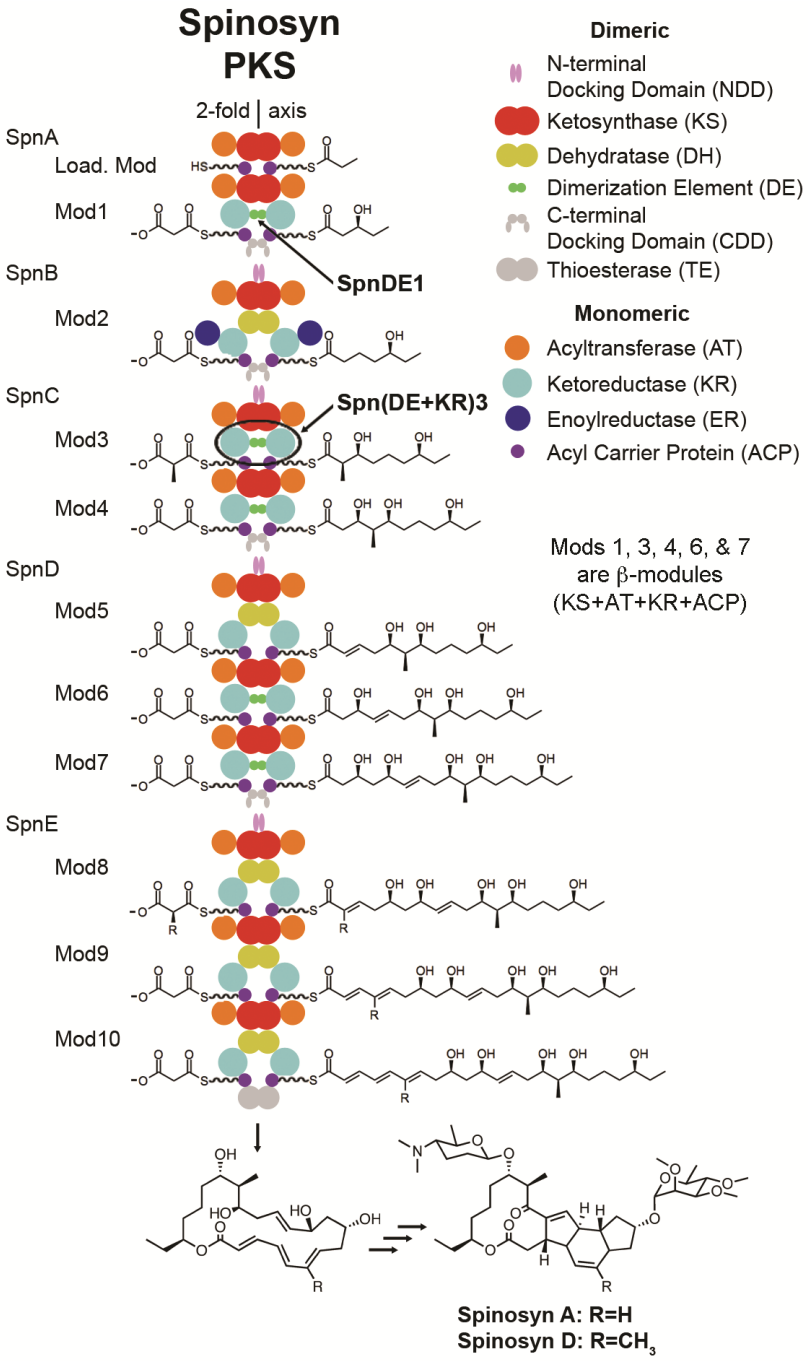


Figure 5-1. Subunit architecture of the spinosyn PKS. Additional dimeric interfaces aid KSs in maintaining the dimeric state of PKS subunits. These interfaces are provided by the enzymes DH and TE, as well as the dimerization motifs of NDD, CDD, and the newly identified DE. The two DE-containing fragments characterized here, SpnDE1 and Spn(DE+KR)3, are from the PKS that helps synthesize the spinosyn insecticides using malonyl and methylmalonyl building blocks.

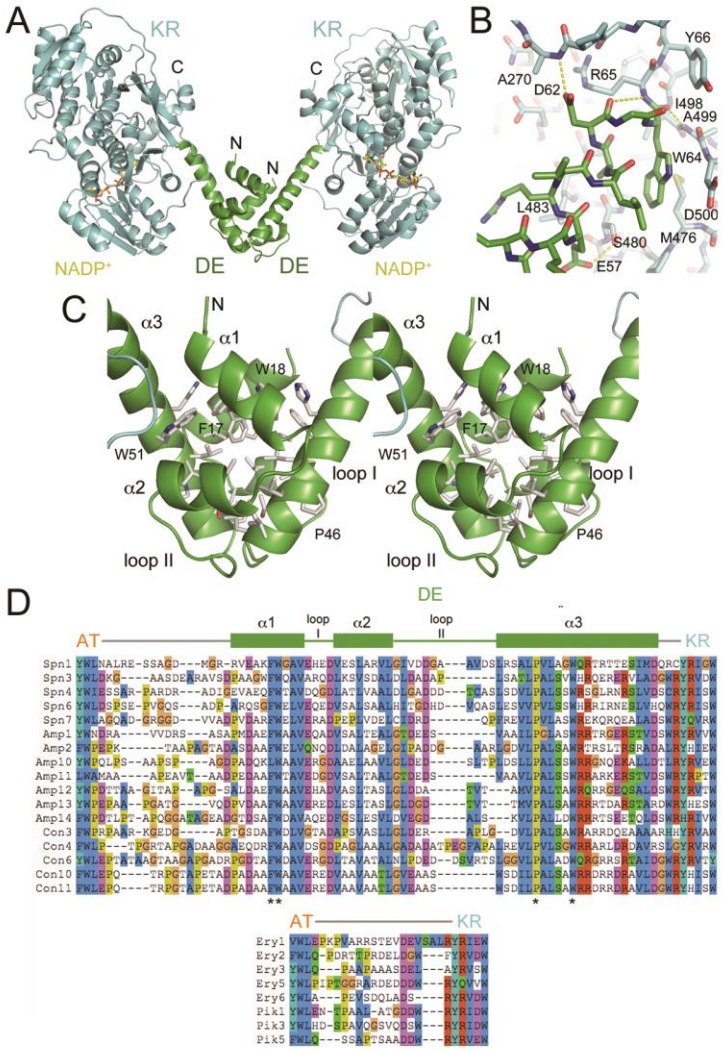


Figure 5-2. DE+KR structure. **(A)** In the crystal structure of Spn(DE+KR)3, the V-shaped DE dimer orients two KR domains bound to NADP+. **(B)** DE rigidly connects to the KR domain. No flexible loop is present between α3 of DE and β1 of KR. **(C)** A stereodiagram of SpnDE3 shows the nearly invariant F17, W18, W51, and other hydrophobic residues that help create the dimer interface of DE. The role of the invariant residue P46 is mysterious. **(D)** A sequence alignment of the ~55-residue DE from DE-containing β-modules shows the boundaries of its three helices, as observed in the Spn(DE+KR)3 crystal structure (residues labeled with * are displayed in the previous panel). Approximately half of β-modules do not contain a DE. Accession codes (GI): Amp1 and Amp2, 14794905; Amp10–Amp14, 14794893; Con3 and Con4, 74026478; Con6, 74026479; Con10 and Con11, 74026481; Ery1 and Ery2, 134097327; Ery3 and Ery4, 134097329; Ery5 and Ery6, 134097330; Pik1, 3800834; Pik3, 3800835; Pik5, 3800836; Spn1, 348173395; Spn3, 348173392; Spn4, 348173391; Spn6 and Spn7, 348173389.

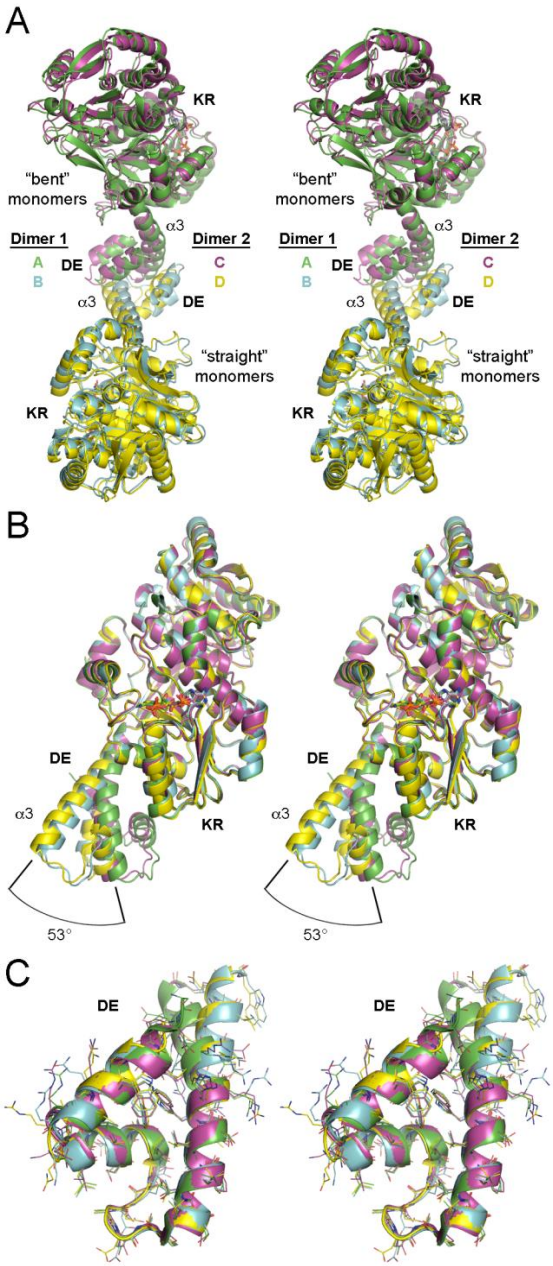


Figure 5-3. The pseudo-twofold symmetry of the DE dimer. **(A)** A stereodiagram shows the superposition of the two DE dimers in the asymmetric unit (as seen looking down their pseudo-twofold axes from the KS dimer). The third helix of DE, α3, bends significantly in the "bent" monomers, resulting in a slight asymmetry within each DE dimer. **(B)** A stereodiagram shows the two principal DE conformations through a superposition of the four DE+KR monomers in the asymmetric unit. **(C)** A stereodiagram reveals that the bend after the third turn of α3 is the principal difference between the four DE monomers in the asymmetric unit.

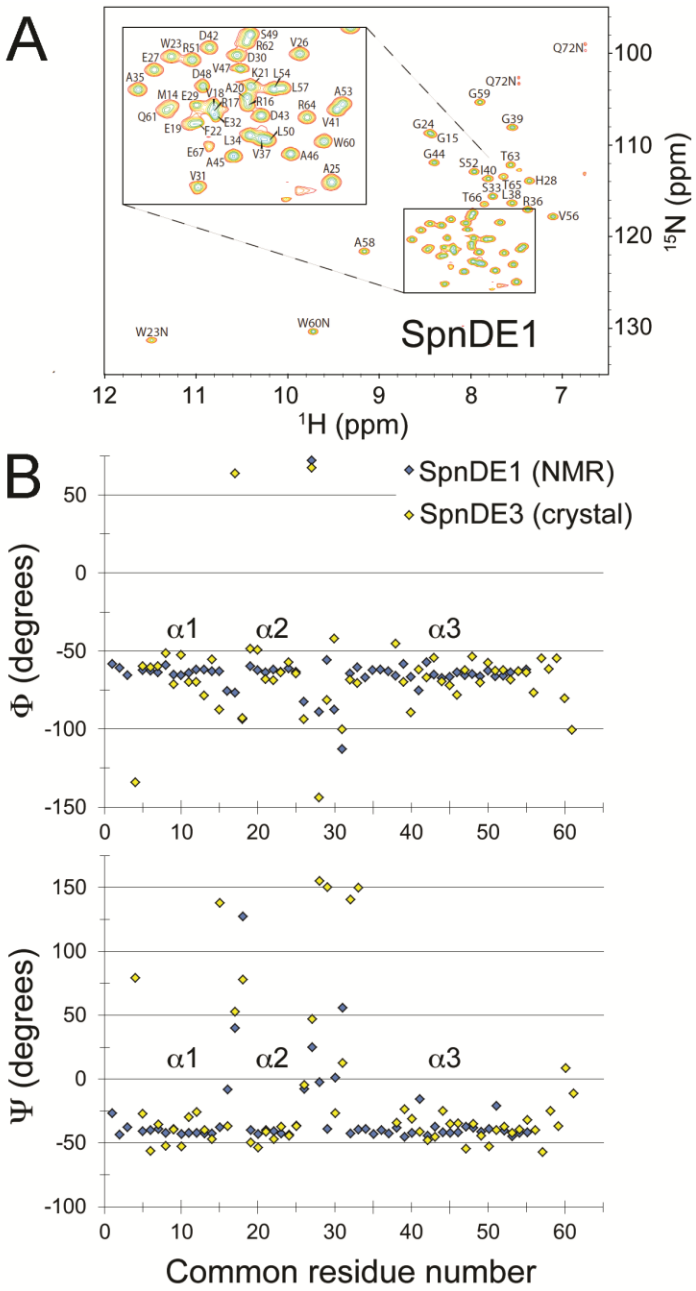


Figure 5-4. NMR characterization of an isolated DE. **(A)** The peaks in the 2D ^{15}N–^{1}H NOESY spectrum of SpnDE1 represent the backbone NHs of structured residues or side-chain NHs. The first 13 and last 5 residues of the SpnDE1 construct are relatively flexible or disordered. **(B)** The Φ and Ψ dihedral angles for SpnDE1 (calculated by PREDITOR from chemical shifts) and SpnDE3 (obtained from monomer A of the Spn(DE+KR)3 crystal structure) display their common secondary structure. The first ordered residues of SpnDE1 and SpnDE3 are Met14 and Asp11, respectively (common residues 1 and 4, respectively, in the plot).

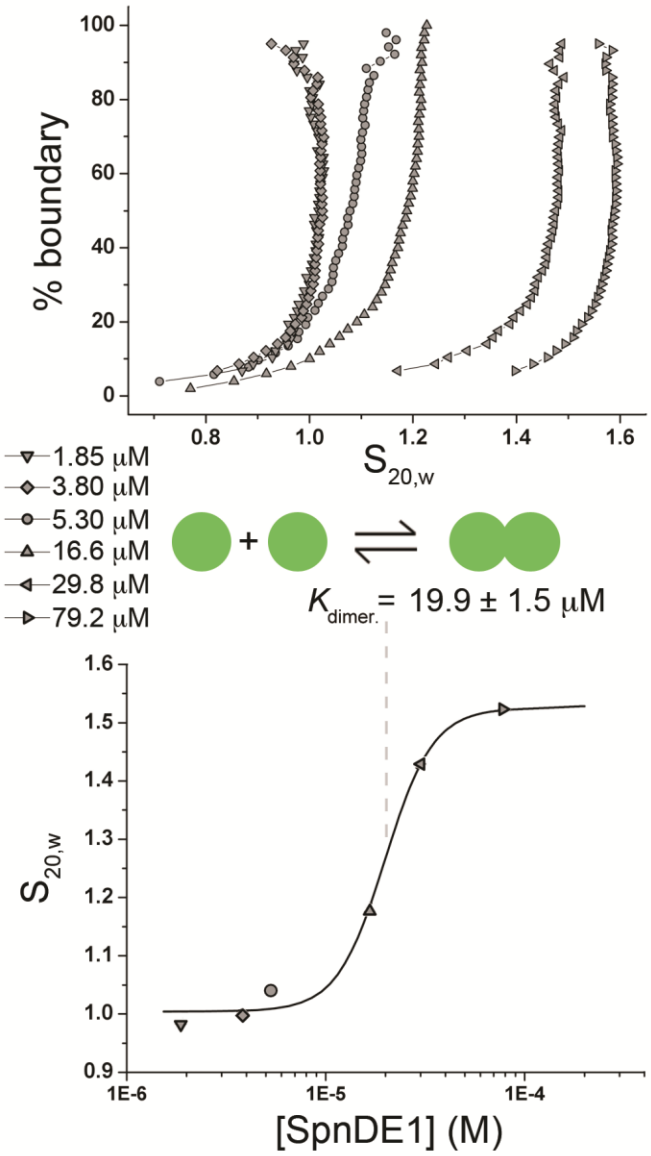


Figure 5-5. Analytical ultracentrifugation. From the van Holde–Weischet integral sedimentation coefficient distributions of SpnDE1 (1.85–79.2 μM) a mass action effect is clearly visible, as distributions shift to higher sedimentation coefficients with increasing concentration. From a global fit of sedimentation coefficients of the major species in each loading concentration to a monomer–dimer binding isotherm, the SpnDE1 dimerization constant was determined to be 19.9 ± 1.5 μM.

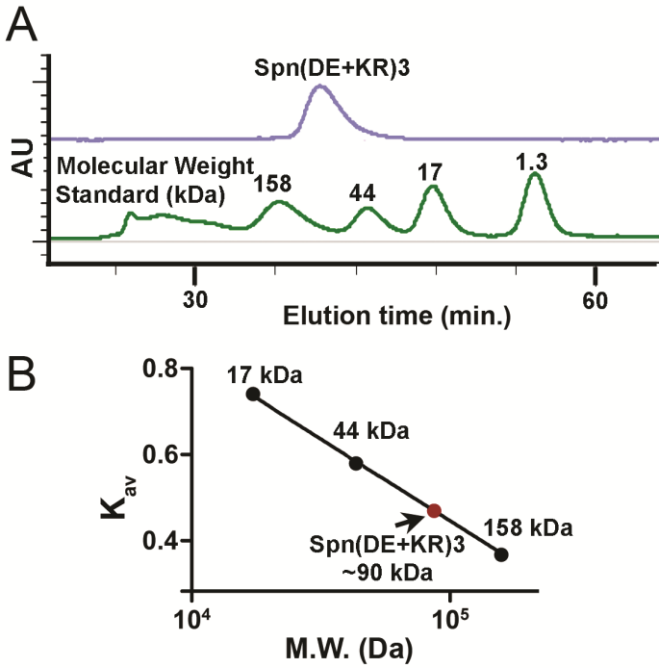


Figure 5-6. Size-exclusion chromatography of a DE+KR fragment. **(A)** The migration of Spn(DE+KR)3 through a Superdex 200 gel filtration column is compared to standards. **(B)** The observed molecular weight was calculated to be ~90 kDa (60 kDa expected for the monomer; 120 kDa expected for the dimer).

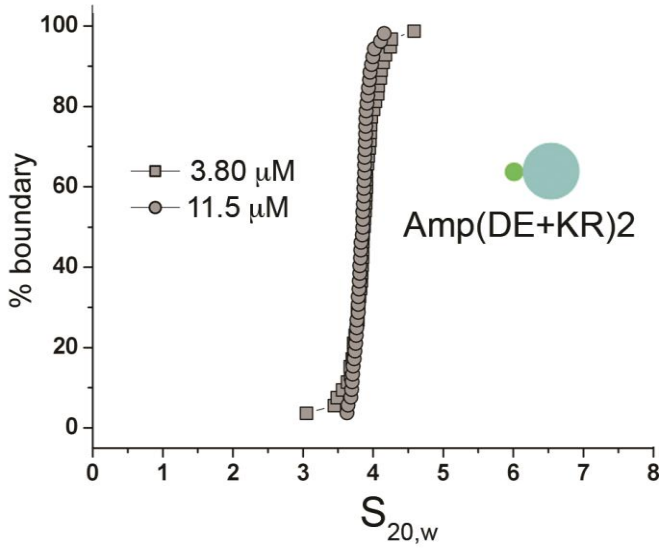


Figure 5-7. Analytical ultracentrifugation of Amp(DE+KR)2. The van Holde–Weischet integral sedimentation coefficient distributions obtained from [Amp(DE+KR)2] = 3.8 μM and 11.5 μM each indicate a homogeneous species with similar sedimentation coefficients. The calculated molecular weight (~56 kDa) is in excellent agreement with predicted monomer molecular weight (56 kDa).

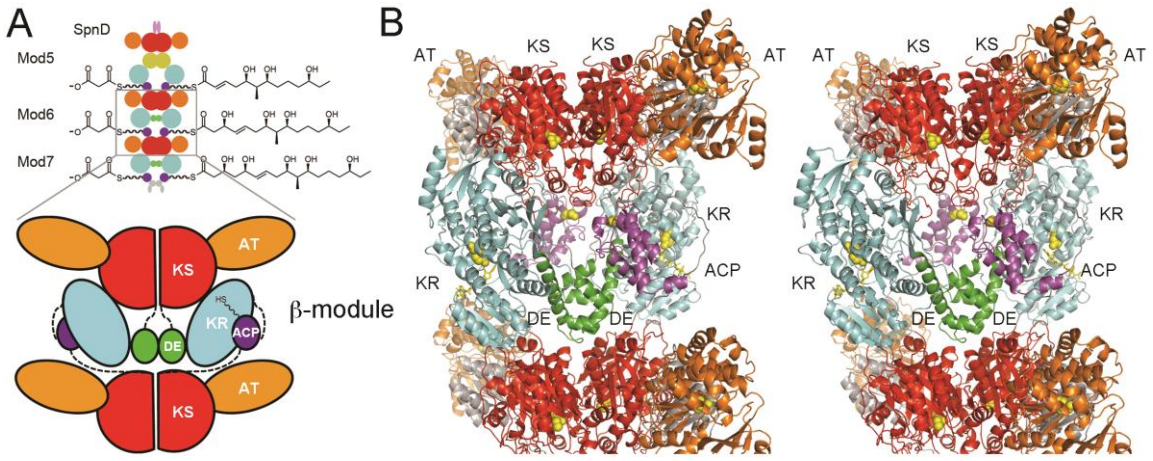


Figure 5-8. Model of a PKS module. **(A)** SpnMod6 is a DE-containing β-module (KS+AT+DE+KR+ACP). A schematic shows the architecture of this module, as well as the first part the downstream SpnMod7, in more detail. **(B)** A stereodiagram shows a β-module, such as SpnMod6, assembled from known structures of a KS+AT dimer, a DE+KR dimer, and two ACPs. Flexible linkers were built between these fragments as well as to the KSs of a subsequent module. The phosphopantetheinyl arm of ACP must interact with the active sites of AT, KS, KR, and the following KS, each shown with a catalytic residue displayed in spheres.

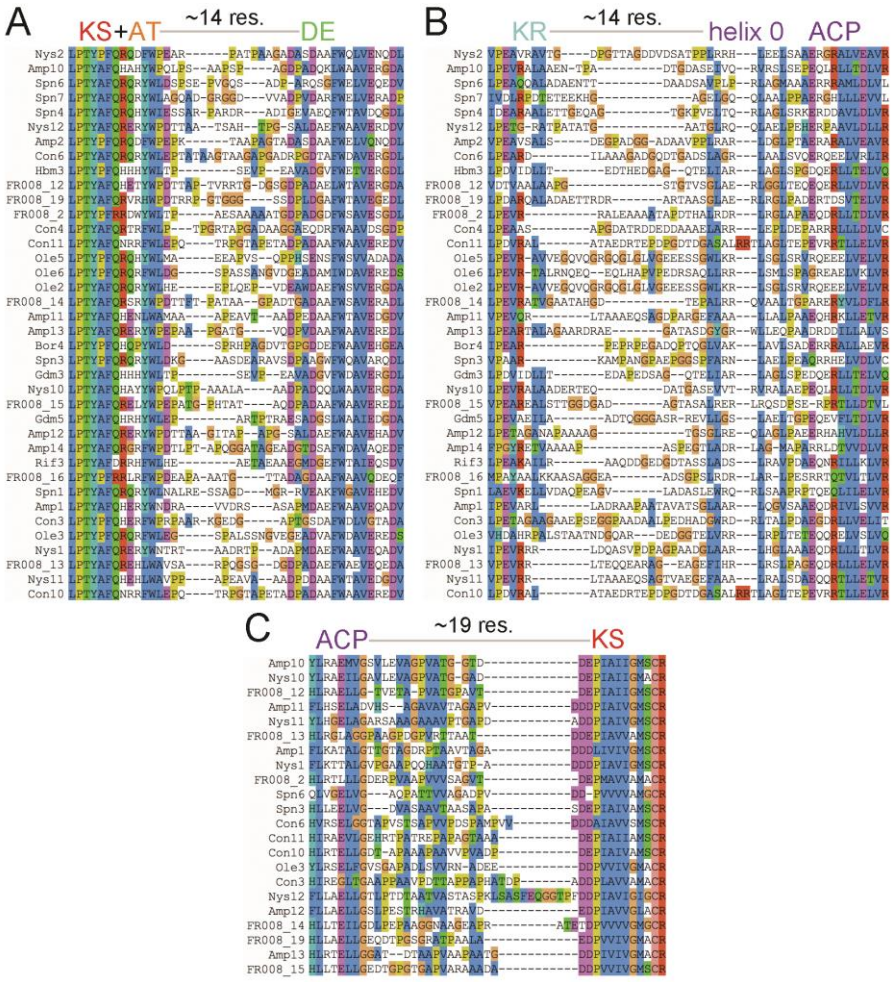


Figure 5-9. Sequence alignment of linkers. **(A)** The average linker between the KS+AT didomain and DE is 14 residues. The shortest linker is 9 residues (Gdm3 and Hbm3). **(B)** The average linker between the KR domain and helix 0 of ACP is 14 residues. The shortest linker is 9 residues (Spn7). **(C)** The average linker between the ACP domain and KS is 19 residues. The shortest linker is 16 residues (Spn6). All sequences were obtained from the MAPSI database and aligned with the program ClustalX. Amp, amphotericin; Bor, borrelidin; Con, concanamycin; FR008, FR008; Gdm, geldanamycin; Hbm, herbimycin; Nys, nystatin; Ole, oleandomycin; Rif, rifamycin; Spn, spinosyn. The number after the PKS acronym indicates from which module the sequence was obtained. Accession codes (GI): Amp1 and Amp2, 14794905; Amp10 - Amp14, 14794893; Bor4, 39725434, Con3 and Con4, 74026478; Con6, 74026479; Con10 and Con11, 74026481; FR008_2, 34766452; FR008_12-FR008_16, 34766455; FR008_19, 34766454; Gdm3, 28192601; Gdm5, 28192602; Hbm3, 63033851; Nys1 and Nys2, 8050849; Nys10 – Nys12, 8050840; Ole2, 9049535; Ole3, 9049536; Ole5, 2492659; Rif3, 300782548; Spn1, 348173395; Spn3, 348173392; Spn4, 348173391; Spn6 and Spn7, 348173389.

Table 5-1. Crystallization data and refinement statistics for Spn(DE+KR)3.

Data Collection	
Space group	$P2_1$
a, b, c (Å)	56.82, 211.69, 101.74
β (°)	95.3
Resolution (Å)	50–2.60 (2.64–2.60)
R_{merge}	0.056 (0.555)
$I/\sigma(I)$	20.3 (2.7)
Completeness (%)	98.2 (97.5)
Redundancy	3.5 (3.5)
Refinement	
Resolution (Å)	50–2.60
No. reflections	69297
R_{work}/R_{free}	0.204/0.245
No. atoms	
Protein	15485
$NADP^+$	192
Water	37
B-factors (Å^2)	
Protein	64.9
$NADP^+$	52.1
Water	39.0
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.149
Ramachandran plot	
Favored/outliers (%)	92.5/1.75
MolProbity	
Clashscore	19.21 (80th percentile)
Overall	3.08
PDB code	4IMP

Values in parentheses refer to the highest-resolution shell.

Chapter 6: The Structure and Activity of a Substrate-permissive β-ketoacyl Thiolase from *Ralstonia eutropha*

ABSTRACT

Thiolases play an integral role in the production of both primary and secondary metabolites in a range of organisms, from microbes to men, by catalyzing the formation and degradation of carbon-carbon bonds. Recently, workers employed β-ketoacyl thiolase B of *Ralstonia eutropha* (*Re*BktB) in an engineered pathway to generate a collection of β-hydroxy acids including the pharmaceutically important 3,4-dihydroxybutyrate, from simple precursors such as acetyl- and glycolyl-CoA (Martin *et al*., 2013). In the present study, we have solved the crystal structure of *Re*BktB in two forms: (1) Untwinned, 2.8-Å and (2) twinned, 2.0-Å. Our structures give insight into the features responsible for the unexpectedly broad substrate specificity of the thiolase and set the groundwork for protein engineering experiments designed to utilize molecules previously inaccessible to thiolases. We also demonstrate *in vitro* thiolytic cleavage of β-keto thioesters of *N*-acetylcysteamine (NAC), a truncated analog of the natural coenzyme. This relatively inexpensive, synthetic handle is expected to substitute for CoA for a variety of *in vitro* reactions, including biosynthetic ones. Interestingly, we found that *Re*BktB prefers 3-oxopentanoyl-*S*-NAC, but also processes 3-oxobutanoyl-, 3-oxohexanoyl-, and 3-oxoheptanoyl-*S*-NAC.

INTRODUCTION

Thiolase enzymes facilitate the synthesis and degradation of carbon-carbon bonds in a number of important biological pathways. Biosynthetic thiolases catalyze carbon-carbon bond formation via the Claisen-like condensation of two acyl-CoA thioesters (such as acetyl-CoA), producing a β-ketoacyl-CoA thioester (such as acetoacetyl-CoA). Such thiolases, specific for short-chain acyl-CoA thioesters, catalyze the formation of HMG-CoA for ketone body, steroid, and isoprenoid biosynthesis (Masamune *et al.*, 1989; Thompson *et al.*, 1989). Degradative thiolases, on the other hand, catalyze carbon-carbon bond cleavage via the thiolysis of a β-ketoacyl-CoA thioester (such as acetoacetyl-CoA), resulting in acetyl-CoA and a shortened acyl-CoA chain (such as acetyl-CoA). These thiolases, generally specific for long-chain β-ketoacyl-CoA thioesters (e.g., 4-16 carbons in length), are found in mitochondria and peroxisomes where they participate in β-oxidation of fatty acids (Masamune *et al.*, 1989; Thompson *et al.*, 1989). Degradative thiolases, found as homodimers, lack the accessory loop that allows biosynthetic thiolases to assemble as tetramers, or dimers of homodimers (Haapalainen *et al*., 2006).

The biosynthetic thiolase reaction proceeds through a two-step ping-pong mechanism (Masamune *et al*., 1989): (1) An active-site cysteine residue, activated as a thiolate by a general base, nucleophilically attacks an acyl-CoA substrate to generate an acyl-enzyme intermediate with concomitant release of CoA; and (2) a second acyl-CoA substrate, activated as an enolate by a general base, nucleophilically attacks the acyl-enzyme intermediate to release an extended β-ketoacyl-CoA product (**Figure 6-1**). The degradative thiolase reaction is similar, except that the enzyme cleaves a β-ketoacyl-CoA substrate as it forms the acyl-enzyme intermediate; it then passes the acyl chain from the cysteine nucleophile to

an incoming CoA thiol. The degradative reaction is thermodynamically favored over the synthetic reaction; however, an excess of acetyl-CoA and the rapid removal of acetoacetyl-CoA favor the biosynthetic route *in vivo* (Modis & Wierenga, 1999; Davis *et al*., 1987).

Bacteria have evolved thiolases, such as β-ketoacyl thiolase B of *Ralstonia eutropha* (*Re*BktB), to assemble the β-ketoacyl-CoA building blocks of biodegradable, thermoplastic polyhydroxyalkanoates (PHAs) (Slater *et al.*, 1998; Valappil *et al*., 2007). One such PHA, poly-3-D-hydroxybutyrate, accumulates in a variety of bacteria as a carbon and energy source and may constitute up to 80% of dry cell weight (Holmes, 1985). The physical properties of these materials are typically comparable with those of petrochemical plastics like polypropylene (Anderson & Dawes, 1990). As such, PHAs are suitable for many applications traditionally fulfilled by petrochemicals, with the advantages of being renewable and biocompatible (Lu *et al*., 2009b). However, unlike their petroleum-based counterparts, PHAs have maintained a prohibitively high production cost (Salehizadeh & van Loosdrecht, 2004). Research to utilize cheaper substrates for fermentation, improve product recovery, and engineer more efficient fermentation strains is ongoing (Verlinden *et al*., 2007).

Recently, *Re*BktB activity was demonstrated on a range of substrates, including acetyl-, propionyl-, butyryl-, isobutyryl-, and glycolyl-CoA (Martin *et al*., 2013). Martin and colleagues incorporated the thiolase into an engineered biosynthetic route to generate a small panel of chiral, organic building blocks, such as the pharmaceutically important 3-hydroxy-γ-butyrolactone and its precursor 3,4-dihydroxybutyrate, in unprecedented titers. The lactone form of this molecule has

served as a starting material in the synthesis of the statin drugs Crestor and Lipitor, the antibiotic Zyvox, and the antihyperlipidemic Zetia (Lee & Park, 2009).

In this report, we demonstrate *in vitro* thiolytic cleavage of β-ketoacyl-*S*-*N*-acetylcysteamine (β-ketoacyl-*S*-NAC) substrates by BktB from *Ralstonia eutropha*. *Re*BktB prefers 3-oxopentanoyl-*S*-NAC, followed by 3-oxobutanoyl-*S*-NAC, and 3-oxohexanoyl- and 3-oxoheptanoyl-*S*-NAC nearly equally (acyl- and β-ketoacyl-*S*-NAC thioesters are short, inexpensive mimics of the natural CoA substrates). In addition, we present two crystal structures of the thiolase: One, untwinned at 2.8-Å resolution and the other, twinned at 2.0-Å resolution. This glimpse into the active site provides a structural basis for the relatively broad substrate selectively of the thiolase. Our work also lays the foundation for protein engineering efforts to further broaden the substrate selectivity of thiolases, ultimately opening new avenues to valuable chiral molecules from common starting materials.

RESULTS AND DISCUSSION

Initial diffraction analysis

Depending on solution conditions, *Re*BktB forms or cleaves carbon-carbon bonds with the assistance of a free CoA thiol. Aside from the inherent utility of these reactions, our interest in the enzyme arose from a desire to understand the molecular basis behind its broad substrate specificity, as demonstrated previously (Martin *et al*., 2013). Therefore, we crystallized and solved the atomic structure of the thiolase in two forms: (1) an untwinned, 2.8-Å dataset in space group $P2_12_12_1$ (4 monomers per asymmetric unit) and (2) a twinned, 2.0-Å dataset in space group $P1$ (16 monomers per asymmetric unit) (**Table 6-1**). Twinning (Yeates & Fam, 1999) of

the latter was suspected after residual factors stalled during refinement in several primitive orthorhombic and primitive monoclinic unit cells, despite the high quality of the electron density map, and was confirmed after closely inspecting the diffraction images (**Figure 6-2**). The dataset was reprocessed in a primitive triclinic unit cell and submitted to *phenix.xtriage* of the *PHENIX* software bundle (Adams *et al.*, 2010), which estimated the twin fraction $\alpha > 0.41$ (where $\alpha = 0.5$ for perfectly twinned and $\alpha = 0$ for untwinned) and identified three pseudo-merohedral twin operators: (1) *h*, -*k*, -*l*; (2) -*h*, *k*, -*l*; and (3) -*h*, -*k*, *l* (**Table 6-2**). The program also performed an *L* Test (Padilla & Yeates, 2003), calculating $\langle |L| \rangle = 0.421$ and $\langle L^2 \rangle = 0.248$ for acentric reflections, which deviate significantly from expected values for normal, untwinned data (**Figure 6-3**). Twin refinement, performed in *REFMAC5* (Murshudov *et al.*, 1997) of the *CCP4* software suite (Winn *et al*., 2011), drastically improved the residual factors after a single cycle (**Table 6-3**). Notably, the unit cell edges for the twinned crystal are up to 9 Å shorter than in the untwinned crystal (**Table 6-1**), and the monomers appear to pack in slightly different modes in the twinned and untwinned unit cells.

In the electron density map of either structure, there is strong coverage of backbone atoms from residues 2-393, except for the loop region consisting of residues ~207-210 (however, these residues are covered in chain A of the 2.0-Å structure). For the 2.8-Å structure, residues R135, S344, and I352 from all four chains fall within the outlier region of the Ramachandran plot (0.26% total). For the 2.0-Å structure, residues 23 (chain O), 56 (chain G), 135 (chain I), 166 (chain N), 170 (chain N), 217 (chains E, H, M), 237 (chains H, J, M, N, P), 239 (chains B, D, G-O), 240 (chain O), 293 (chain J), 295 (chain P) 344 (chain M), 352 (chains C-P), and 374 (chain H) are outliers (0.71% total). Chain A from either structure superposes well

with C_α r.m.s.d. of 0.21 Å (359 atoms), and tetramers (chains A-D) from either structure superpose with C_α r.m.s.d. of 0.68 Å (785 atoms). The remainder of our analysis focuses on the structure derived from the 2.0-Å dataset rather the 2.8-Å one, due to the significantly higher resolution and lower *B*-factors of the former. In particular, we discuss the presumed biological unit comprised of chains A-D because of its superior geometry and electron density map compared to those of the remaining tetramers.

Overall structure of *R. eutropha* β-ketoacyl thiolase B

The tertiary fold of *Re*BktB is nearly identical to that of *Zoogloea ramigera* BktB (*Zr*BktB; PDB 1DLV; Modis & Wierenga, 2000), with C_α r.m.s.d. = 0.68 Å over 349 residues (**Figure 6-4**). Monomers from either enzyme share 51% sequence identity (Sievers *et al.*, 2011) and assume the five-layered αβαβα fold of the thiolase superfamily in which two β-sheets are sandwiched between three clusters of α-helices (Haapalainen *et al.*, 2006) (**Figures 6-5**, **6-6**, and **6-7**; α-helices labeled α1-15 and β-sheets labeled β1-13). Interestingly, each monomer can be divided into equivalent N- and C-terminal halves that share βαβαβαββ topology, indicative of gene duplication (Haapalainen *et al*., 2006).

In addition, the enzymes possess similar tetrameric organization in which residues 120-143 (strands β5 and β6 and their N- and C-terminal loops; *Re*BktB numbering) of one monomer interact with the corresponding residues of three other monomers in a compact tetramerization motif consisting of a 6-stranded, antiparallel β-sheet (**Figures 6-4** and **6-5**). However, the relative orientation of monomers comprising a tetramer differs slightly between the two enzymes, with C_α r.m.s.d. = 1.27 Å over 773 residues (about half of C_α atoms were rejected during

superposition in PyMOL) (**Figure 6-5**). Each *Re*BktB monomer supplies a number of aliphatic side chains, namely those of Y125, A127, A129, A130, A134, A139, and L141, which pack into a hydrophobic core. Other residues, such as P124, L126, P128, R131, W132, and V142 extend away from the core, forming electrostatic and van der Waals interactions with neighboring subunits. This configuration results in an average interfacial area of 291 $Å^2$ and 642 $Å^2$ between distal monomers A/C and A/D, respectively, as calculated by the PDBePISA server (Krissinel & Henrick, 2007). Proximal monomers A/B in this arrangement interact more extensively, with an average interfacial area of 2732 $Å^2$. Beyond the tetramerization motif interactions, proximal monomers A/B further dimerize via the following: β3 on the edge of the central β-sheet forms hydrogen bonds with the same strand of the opposing monomer (residues 85-87) to give one broad, antiparallel β-sheet; α3 associates with α5 of the opposing monomer through hydrophobic interactions, and the C-terminal portion of α4 associates with the C-terminus of the same helix from the opposing monomer, also through hydrophobic interactions.

In all four monomers, α10 displays elevated *B*-factors relative to the remainder of the protein. *B*-factors are also elevated at the mouth of the CoA channel (e.g. α5-α9 and adjacent loops) and the region encompassing β7 and β8, presumably because the phosphoadenosyl moiety of CoA is not present to stabilize the relevant residues (**Figure 6-8**). Furthermore, only in chain A is the β-loop-β motif containing β7 and β8 resolved. R208 and K209 of this motif, relatively conserved in biosynthetic thiolases, are predicted to interact with the phosphoryl groups of CoA when bound to monomer C.

Comparison of the binding pockets of *Re*BktB and other thiolases

In *R. eutropha* BktB, the catalytic residues C90, H350, and C380 (**Figure 6-5C**) are in conformations nearly equivalent to those of the *Z. ramigera* homolog (**Figure 6-9**). Placement of the nucleophilic C90 near the N-terminus of its helix enhances its nucleophilic character by stabilizing the thiolate form of its side chain. H350 and C380 are hypothesized to act as general acid/base residues, exchanging protons with the C90 and CoA-derived nucleophiles, respectively (**Figure 6-1**) (Haapalainen *et al.*, 2006).

An array of conserved hydrophobic side chains lines the narrow CoA channel of *Re*BktB, including M136 (from monomer D), M120, L149, M158, V235, F236, A245, A248, G250, L251, M290, A320, F321, and I352 (from monomer A) as shown in **Figure 6-9**. Save V235 (A234 in *Zr*BktB), these residues are identical to the ones that fulfill this role in *Zr*BktB. In both enzymes, only two hydrophilic side chains, other than those of the catalytic triad, enter the substrate channel for hydrogen bonding interactions: S249 and H157. Given the strict conservation of residues lining this channel, the binding mode of CoA is presumably indistinguishable between these and other biosynthetic thiolases.

As with *Zr*BktB, the active site of *Re*BktB is sequestered from solvent by α5, loop residues 150-157 (C-terminal to α5), and Y66 (Q64 in *Zr*BktB) of the proximal monomer (Kursula *et al*., 2002). Q64, highly conserved in biosynthetic thiolases (**Figure 6-10**), caps the active site of *Zr*BktB, forming hydrogen bonds to the backbone CO of G147 and backbone NH of G381 of the neighboring monomer (Kursula *et al*., 2002). This Gln is equivalently positioned in biosynthetic thiolases from *Clostridium difficile* (Q68; PDB 4DD5) and human mitochondria (Q101; PDB 2IBY). In *Re*BktB, however, this Gln is absent and its space is occupied by Y66, with

the phenol ring packed between L89, P152, and G382 of the neighboring monomer and P62, R63, and M65 of the same monomer (**Figures 6-10** and **6-11**). The phenol hydroxyl group is within hydrogen bonding distance of the backbone carbonyl of A148 (equivalent to G147 in *Zr*BktB) and H150 on the neighboring subunit. In addition, while Q64 is located N-terminal to a helix corresponding to α3 of *Re*BktB, Y66 is the fifth residue of α3 in *Re*BktB and causes the helix to kink sharply (~110°). The positioning of Y66 on α3 appears to restrain its side chain in a position slightly farther from the neighboring monomer than observed for Q64 (**Figures 6-9** and **6-11**). Neighboring helix α5, on which A148 of *Re*BktB resides, is also narrower than the corresponding helix of *Zr*BktB due to deletion of a residue in the former (**Figure 6-10**). These unusual features may be responsible, in part, for the broad substrate specificity observed for the *R. eutropha* enzyme, particularly toward substrates with polar functional groups, such as glycolyl-CoA. Notably, *Re*BktB does not share the sequence conservation of most biosynthetic thiolases along helices α3 and α5 (**Figure 6-10**). Curiously, after superposing proximal monomers of *Re*BktB and *Zr*BktB (PDB 1DLV), the active-site volumes bounded by the NP4 atom of CoA were calculated as 132 $Å^3$ and 157 $Å^3$, respectively, by the 3V server (Voss & Gerstein, 2010). If *Re*BktB indeed accepts more diverse substrates than *Zr*BktB, perhaps variable dynamics at the dimer interface (near Y66/Q64) play a role. Consistent with this, monomers of an *Re*BktB tetramer appear to be oriented relative to one another slightly differently than those of a *Zr*BktB tetramer (**Figure 6-4**).

Interestingly, *R. eutropha* encodes another β-ketoacyl thiolase, *Re*PhbA, which displays narrower substrate preference than *Re*BktB (Slater *et al.*, 1998). When challenged with 3-oxobutanoyl-CoA (acetoacetyl-CoA), 3-oxopentanoyl-CoA, or 3-oxohexanoyl-CoA *in vitro*, *Re*BktB favors cleavage of 3-oxopentanoyl-CoA, but

cleaves all three efficiently; conversely, *Re*PhbA only efficiently cleaves 3-oxobutanoyl-CoA. *Re*BktB and *Re*PhbA share 52% identity (Sievers *et al.*, 2011), with the most divergent active-site residues being those on helices α3 and α5 (**Figure 6-10**). In fact, *Re*PhbA more closely resembles *Zr*BktB (63% identity; Sievers *et al.*, 2011) and typical biosynthetic β-ketoacyl thiolases, suggesting that *Zr*BktB, like *Re*PhbA, may possess a relatively narrow substrate preference.

In degradative thiolases, the residue in place of the Gln is less conserved. For example, it is substituted with S114, T101, and F66 in *Arabidopsis thaliana* peroxisomal (PDB 2WU9), *Saccharomyces cerevisiae* peroxisomal (PDB 1AFW), and *Salmonella typhimurium* (PDB 3GOA) degradative thiolases, respectively. Furthermore, the absence of the tetramerization motif and α5 that follows it opens the "ceiling" of the substrate cavity in such thiolases considerably. That is, the loop preceding α6 is elongated and extends away from the active site, favoring interactions with α1 and α3 of the neighboring monomer. These differences undoubtedly account for the permissiveness of degradative thiolases toward substrates with longer chains. Indeed, in the *S. cerevisiae* (PDB 1AFW) and *A. thaliana* peroxisomal (PDB 2WU9) degradative thiolases, 2-methyl-2,4-pentanediol and ethylene glycol, respectively, were found to occupy the space near α5 and the conserved Gln of biosynthetic enzymes (Modis & Wierenga, 1999; Mathieu *et al.*, 1997; Pye *et al.*, 2010).

Modelling of substrates and reaction coordinate

Although soaking of crystals with CoA and acetyl-CoA was attempted, samples rapidly deteriorated in the presence of either molecule. Therefore, CoA

substrates and acyl-enzyme intermediates were manually modelled into the active site of *Re*BktB as guided by existing structures (**Figures 6-9**, **6-11**, and **6-12**).

Whether *Zr*BktB shares the substrate preference of *Re*BktB demonstrated by Martin *et al*. (2013) is uncertain, although a covalent butyryl-*Zr*BktB intermediate has been observed (Kursula *et al*., 2002; PDB 1M4T). Based on the products generated with *Re*BktB (Martin *et al*., 2013), the bulkier acyl-CoA substrate (e.g. longer than acetyl-CoA) must first be transferred to the enzyme, followed by entry of acetyl-CoA for condensation (glycolyl-CoA is exceptional in that it can enter in the second phase of the reaction). This is reasonable given the limited space available in the active site of the enzyme. As determined previously, an oxyanion hole formed by H350 and a water molecule (hydrogen bonded to N318) may be important in stabilizing the enolate intermediate following deprotonation of the acyl-CoA substrate (Meriläinen *et al*., 2009). For bulky acyl-CoA substrates, M158 and M290 are the likely "gatekeeper" residues as they are located on the opposite side of the enolate oxygen, where the aliphatic group of acetyl-CoA has been shown to reside (**Figure 6-11**; PDBs 1DM3 and 1M3Z). Furthermore, these residues are located at the termini of helices that interface the active site and solvent, in regions displaying elevated *B*-factors; thus, their side chains may be quite mobile (**Figure 6-8**). Mutating M158 and M290, and potentially the nearby L149 and F321, to residues with smaller hydrophobic side chains is hypothesized to further expand the active site, permitting the enzyme to transform even bulkier acyl-CoA substrates (**Figure 6-11**).

Once covalently linked to C90, the bulky substrates are expected to assume conformations similar to that observed in the butyryl-*Zr*BktB complex (see valeryl-*Re*BktB model, **Figure 6-12A**). That is, the carbonyl oxygen fits in an oxyanion hole

formed by the backbone NHs of G382 and C90, while the aliphatic group rests in a pocket defined by the side chains of L149, M158, M290, A320, F321, H350, I352, and C380. As described above, mutation of the two methionine residues, and perhaps others lining this pocket, may increase the affinity of the enzyme toward larger, nonnatural substrates. Unlike its nonpolar counterparts, the glycolyl chain is predicted to point toward the "top" of the cavity, forming a hydrogen bond with either the side chain O of Y66 or the backbone CO of A148 (near the C-terminus of α5) (see glycolyl-*Re*BktB model, **Figure 6-12B**). Unless Y66 undergoes a conformational change, the latter interaction appears more likely as the oxygen-oxygen distance is ~2.9 Å, compared to ~3.7 Å for Y66. If the geometry demonstrated for acyl-enzyme intermediates (PDBs 1DM3 and 1M4T) is preserved in glycolyl-*Re*BtkB, no other free, polar atoms are in the vicinity to interact with the glycolyl hydroxyl group.

In the final step of the reaction, acetyl-CoA enters the active site to condense with the bulky acyl-enzyme intermediate. There is likely some tolerance in the positioning of CoA, given the largely hydrophobic nature of its binding pocket. In *Zr*BktB, only one direct hydrogen bond was observed between the enzyme and pantetheine portion of the coenzyme, with the remaining interactions bridged by water molecules (Modis & Wierenga, 1999). In addition, the *B*-factors of the coenzyme atoms were found to be higher than those of surrounding residues. If acetyl-CoA were to bind without entering the tunnel as deeply as observed in the acetyl-*Zr*BktB complex (PDB 1DM3), the side chains of the only two conserved polar residues, S249 and H157, would be poised for direct hydrogen bonding interactions with the coenzyme.

In vitro thiolytic cleavage of β-ketoacyl-*S*-NACs

*Re*BkB (1 μM) was incubated with 5 mM NAC and 1 mM β-ketoacyl-*S*-NAC substrate: 3-oxobutanoyl-*S*-NAC (**1**), 3-oxopentanoyl-*S*-NAC (**2**), 3-oxohexanoyl-*S*-NAC (**3**), 3-oxoheptanoyl-*S*-NAC (**4**), or (2*RS*)-2-methyl-3-oxopentanoyl-*S*-NAC (**5**). After 1 h or 24 h at 22 °C, reactions were acid-quenched and analyzed by HPLC. The α-branched **5** was found to be an unsuitable substrate for the enzyme, even after 24-h incubation; however, the remaining reactions yielded acetyl-*S*-NAC (**6**) as well as another molecule of **6**, propionyl-*S*-NAC (**7**), butyryl-*S*-NAC (**8**), or valeryl-*S*-NAC (**9**) (**Figure 6-13**) as confirmed by LC/MS (*m/z* of 162.2, 176.2, 190.2, and 204.4 for **7**-**9**, respectively). Peaks for product **6** were integrated and their areas were compared to a standard curve to determine the concentration of **6** formed (**Figure 6-14**). *Re*BktB exhibited preference for the β-ketoacyl-*S*-NAC thioesters as follows: **2** > **1** > **3** ~ **4** (0.873 ± 0.031 mM, 0.225 ± 0.004 mM, 0.089 ± 0.006 mM, and 0.082 ± 0.001 mM **6** formed, respectively) (**Figure 6-15**). Interestingly, these results contrast with the previously determined specificity toward β-ketoacyl-CoA thioesters: 3-oxopentanoyl-CoA > 3-oxohexanoyl-CoA > 3-oxobutanoyl-CoA (Slater *et al.*, 1998). CoA is expected to contribute greater binding energy than its truncated counterpart NAC, which may affect its ability to deliver a β-ketoacyl group to the thiolase active site. Additionally, the lower affinity of NAC may diminish its capacity to degrade bulkier acyl-enzyme intermediates such as butyryl-*Re*BktB. To our knowledge, utilization of β-ketoacyl-*S*-NAC substrates by a β-ketoacyl thiolase has not yet been reported, although ketosynthase domains (also members of the thiolase superfamily) from polyketide synthases have been shown to process such substrates (Tsukamoto *et al.*, 1996).

CONCLUSION

In conclusion, our structures of the biosynthetic thiolase BktB from *Ralstonia eutropha* reveal features which may be important in determining the enzyme's broad specificity toward substrates. Furthermore, our structures set the stage for engineering experiments designed to expand the substrate repertoire of thiolases. Finally, we demonstrate thiolase activity on β-ketoacyl-*S*-*N*-acetylcysteamine substrates, which are inexpensive substitutes for their CoA counterparts. We show that *Re*BktB prefers 3-oxopentanoyl-*S*-NAC over 3-oxobutanoyl-, 3-oxohexanoyl-, and 3-oxoheptanoyl-*S*-NAC thioesters.

MATERIALS AND METHODS

Cloning, expression, and purification of *R. eutropha* β-ketoacyl thiolase B

The *R. eutropha* BktB gene was amplified from the pCDF/pct/tesB plasmid (Martin *et al*., 2003), with primers 5'-ATGCTTGCAcatatgACGCGTGAAGTGGTAGTGGTA-3' and 5'-GTACGAACGgaattcTCAGATACGCTCGAAGATGGCGGC-3' (restriction sites in lower case; stop codon underlined), digested with *Nde*I and *Eco*RI, and ligated into pET28b (Novagen). *E. coli* BL21(DE3) cells carrying this plasmid were grown in LB media with 50 mg/L kanamycin at 37 °C. Upon reaching OD_{600} = 0.4, cultures were cooled to 15 °C and induced with 0.5 mM IPTG. After 17 h, cells were pelleted, resuspended in lysis buffer (500 mM NaCl, 30 mM HEPES pH 7.5, 10% (v/v) glycerol) and sonicated. Cell lysate was centrifuged at 30,000 × *g* for 1 h, and the supernatant was poured over a column of Ni-NTA beads (Expedeon) equilibrated with lysis buffer. Beads were washed with 10 column volumes of 15 mM imidazole in lysis buffer and

protein was eluted with 1.5 column volumes of 150 mM imidazole in lysis buffer. Eluate was further purified on a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES pH 7.5, 10% (v/v) glycerol. Column fractions were exchanged in 25 mM NaCl, 10 mM HEPES pH 7.5, 10% (v/v) glycerol and brought to 37 mg/mL in a centrifugal concentrator. Aliquots were flash frozen in liquid nitrogen and stored at -80 °C until further use.

Crystallization, data collection, processing, phasing, and refinement

Enzyme aliquots were rapidly thawed and placed on ice. Crystals were grown by sitting-drop vapor diffusion by adding 1 µL protein solution to 0.5 µL 22% (w/v) PEG 4000, 100 mM HEPES pH 7.0, 12% (v/v) glycerol at 22 °C. Crystals were harvested from the mother liquor and immediately flash frozen in liquid nitrogen without addition of cryoprotectants. Interestingly, crystals dissolved unless removed from the mother liquor within 48 h of drop preparation.

Datasets for the twinned and untwinned crystals were collected at 100 K at Advanced Light Source beamlines 5.0.2 and 5.0.3, respectively, and indexed, integrated, and scaled in *HKL2000* (Otwinowski & Minor, 1997) (**Table 6-1**). Intensities were converted to amplitudes, and 5% of scaled reflections were reserved for calculating R_{free} in *TRUNCATE* (French & Wilson, 1978) of the *CCP4* software suite (Winn *et al*., 2011). Phases were solved via molecular replacement by searching for eight and two copies of a *Clostridium difficile* thiolase (PDB 4DD5) dimer in space groups $P1$ and $P2_12_12_1$, respectively, in *Phaser* (McCoy *et al.*, 2007) of *CCP4* (Winn *et al*., 2011). To convert side chains to those of the *R. eutropha* enzyme, a SCWRL model (Krivov *et al*., 2009), obtained from the Fold & Function Assignment System (FFAS) server (Jaroszewski *et al*., 2005), was superposed onto each

monomer of the *Phaser* solution. For the 2.8-Å, untwinned structure, iterative refinement was carried out in *Coot* (Emsley & Cowtan, 2004), *phenix.refine* (with default NCS and X-ray/stereochemistry weight = 0.4 to improve geometry) of the *PHENIX* software bundle (Adams *et al*., 2010), and *REFMAC5* (with default NCS) (Murshudov *et al.*, 1997) of *CCP4*. Pseudo-merohedral twinning and associated operators were identified for the 2.0-Å dataset (**Tables 6-2** and **6-3**) in *phenix.xtriage* of *PHENIX*. Iterative refinement was performed in *Coot* and *REFMAC5* (with intensity-based twin refinement and X-ray weighting term = 0.3 to improve geometry). During early stages of twin refinement, medium non-crystallographic symmetry restraints were applied (chains B-O equivalent to A); these were later removed due to worsening of geometry and clashes. Throughout refinement, the MolProbity server was consulted for structure validation (Chen *et al*., 2010).

Syntheses of NAC and β-ketoacyl-*S*-NACs

Syntheses of NAC, **1**, **2**, **3**, and **5** were carried out as described (Piasecki *et al*., 2011). Syntheses of **4** and **6** were carried out as described (Hughes *et al*., 2011a).

***In vitro* thiolytic cleavage of β-ketoacyl-*S*-NACs**

Enzyme aliquots were rapidly thawed and placed on ice. *Re*BktB (1 μM final) was added to 100 mM NaCl, 100 mM HEPES pH 7.5, 10% (v/v) glycerol, 5 mM NAC, and 1 mM **1**, **2**, **3**, **4**, or **5** in a total volume of 1 mL in triplicate. Negative controls lacking either *Re*BktB or NAC were also prepared. After incubating for 1 or 24 h at 22 °C, quenching was effected by mixing 190 μL reaction solution with 10 μL 7 M HCl. Negative controls lacking both *Re*BktB and HCl were also prepared to ensure that quenching did not promote acid-catalyzed hydrolysis of thioesters. Samples were centrifuged and their supernatants were analyzed at 235 nm by reverse-phase

HPLC on a Varian Microsorb-MV 300-5 C18 250 × 4.6 mm column (Agilent Technologies; 50-µL injection loop) and Waters 2998 PDA detector. Mobile phases consisted of 0.1% TFA in water (solution A) and 0.1% TFA in methanol (solution B). Reactions containing **1** were injected (20 µL) on a 30-min linear gradient of 0-20% solution B with a flow rate of 1 mL/min. Reactions containing **2**, **3**, **4**, or **5** were injected (20 µL) on a 30-min linear gradient of 5-50% solution B with a flow rate of 1 mL/min. Peaks for product **6**, monitored at 235 nm, eluted at 17.3 min for reactions containing **1** or 10.2 min for reactions containing **2**, **3**, or **4** (**Figure 6-13**). Peaks for product **6** were integrated in Breeze 2 software (Waters) and their areas were compared to those of a standard curve of 0.1-1 mM **6** to determine concentrations (**Figure 6-14**). For reactions containing **5**, peaks for product **7**, which eluted at 15.3 min, were monitored at 235 nm and integrated. Background from reaction solutions lacking enzyme was subtracted from peak areas, and final concentrations were adjusted for dilution factor. For reactions containing **1**, peaks for product **6** were divided by 2 for direct comparison with other reactions. Product fractions were collected and further analyzed by positive-ESI LC/MS (Agilent Technologies 1200 Series HPLC with a Gemini C18 5 micron 2.1 × 50 mm column coupled to an Agilent Technologies 6130 Series Quadrupole MS). Mobile phases consisted of 0.1% formic acid in water (solution C) and 0.1% formic acid in acetonitrile (solution D). Samples were run on a 12-min linear gradient of 5-95% solution D with a flow rate of 1 mL/min.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (grant GM106112 to A.T.K.). Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

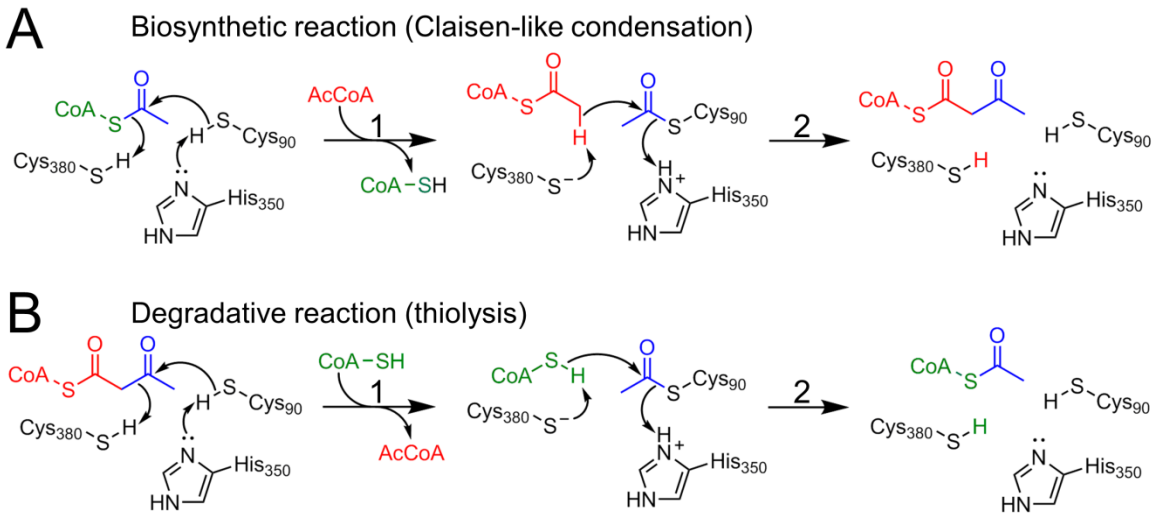


Figure 6-1. Thiolase reaction mechanisms. **(A)** Biosynthetic reaction. **1.** After the first substrate (acetyl-CoA; green and blue) binds, the nucleophilic Cys90, activated via deprotonation by His350, attacks the carbonyl carbon to generate an acyl-enzyme intermediate. The leaving group, CoA (green), abstracts a proton from Cys380 and dissociates from the active site. **2.** The second substrate (acetyl-CoA; AcCoA, red) binds and is deprotonated by Cys380 to form an enolic nucleophile (not shown), which attacks the acyl-enzyme intermediate in a Claisen-like condensation reaction. Cys90 abstracts a proton from His350, and the β-ketoacyl-CoA product (acetoacetyl-CoA; red and blue) dissociates from the active site. **(B)** Degradative reaction. **1.** The first substrate, a β-ketoacyl thioester (acetoacetyl-CoA; red and blue), binds and the nucleophilic Cys90, activated via deprotonation by His350, attacks the β-carbonyl carbon to produce an acyl-enzyme intermediate. The leaving group, acetyl-CoA (AcCoA, red), abstracts a proton from Cys380 and dissociates from the active site. **2.** The second substrate (CoA; green) binds and is deprotonated by Cys380; it then nucleophilically attacks the acyl-enzyme intermediate. Cys90 abstracts a proton from His350, and the product (acetyl-CoA; green and blue) dissociates from the active site.

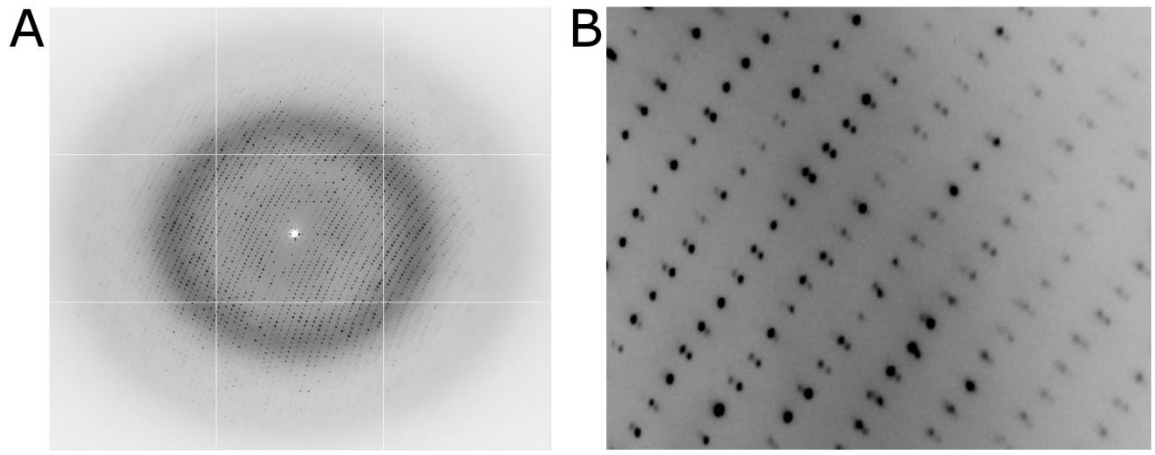


Figure 6-2. Representative diffraction image for the twinned, 2.0-Å dataset. **(A)** Full diffraction image. **(B)** Magnified portion of the image in **A**, revealing twinned reflections. There is comparably little overlap of reflections in the 2.8-Å dataset.

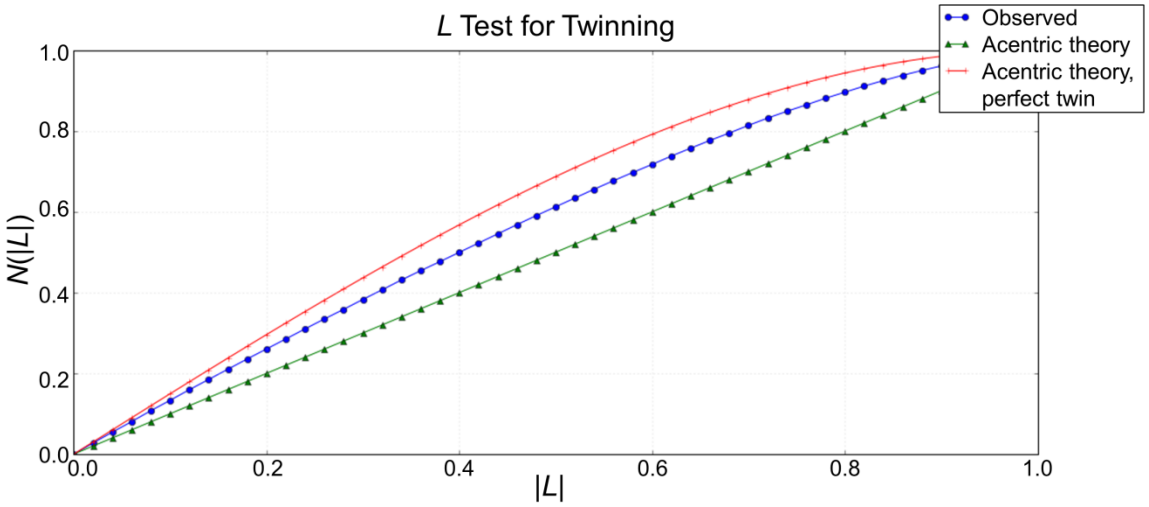


Figure 6-3. L Test for acentric data, adapted from *phenix.xtriage* output. *PHENIX* calculated $\langle|L|\rangle$ = 0.421 and $\langle L^2\rangle$ = 0.248, indicating significant twinning. Padilla and Yeates (2003) defined $L = \frac{I_A - I_B}{I_A + I_B}$, where I_A and I_B are the intensities of local, untwinned reflections A and B. For normal data, the cumulative probability distribution $N(|L|) = |L|$ and the mean values $\langle|L|\rangle$ = ½ and $\langle L^2\rangle$ = 1/3 for acentric reflections. For perfectly twinned data, $N(|L|)$ = ½$|L|(3\text{-}L^2)$, $\langle|L|\rangle$ = 3/8, and $\langle L^2\rangle$ = 1/5 for acentric reflections.

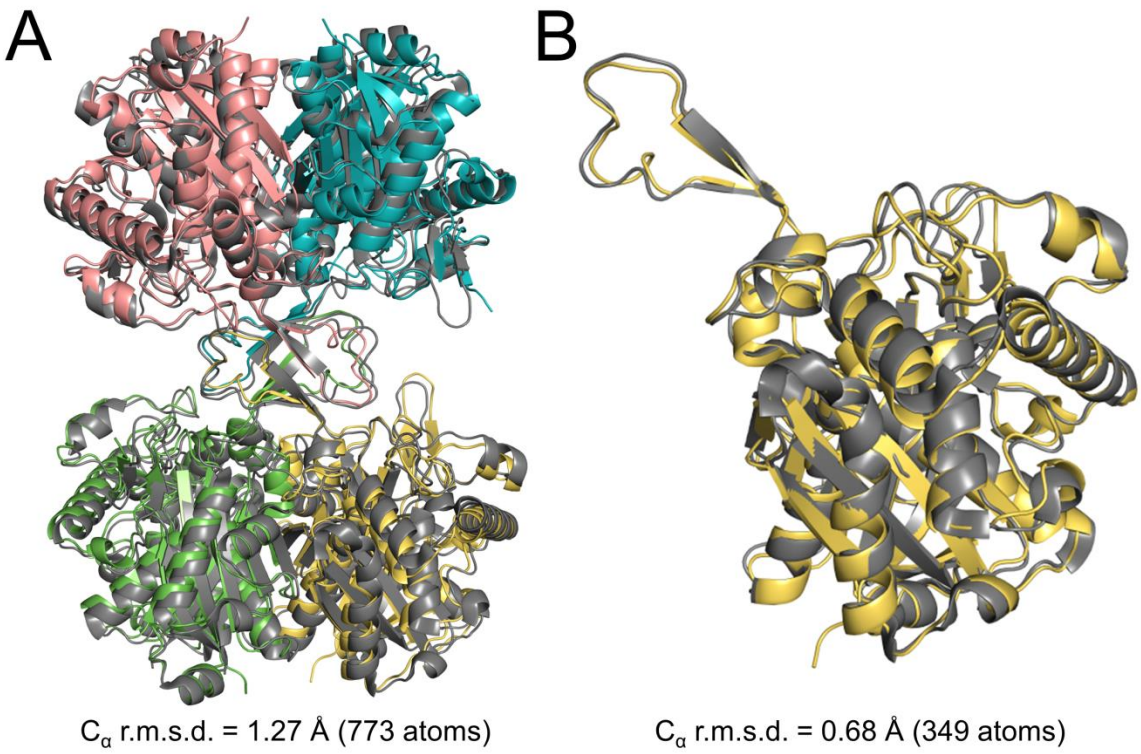


Figure 6-4. Superposed **(A)** tetramers and **(B)** monomers from *Re*BktB and *Zr*BktB. *Re*BktB is colored by chain (as in **Figure 6-5**) and *Zr*BktB is colored grey.

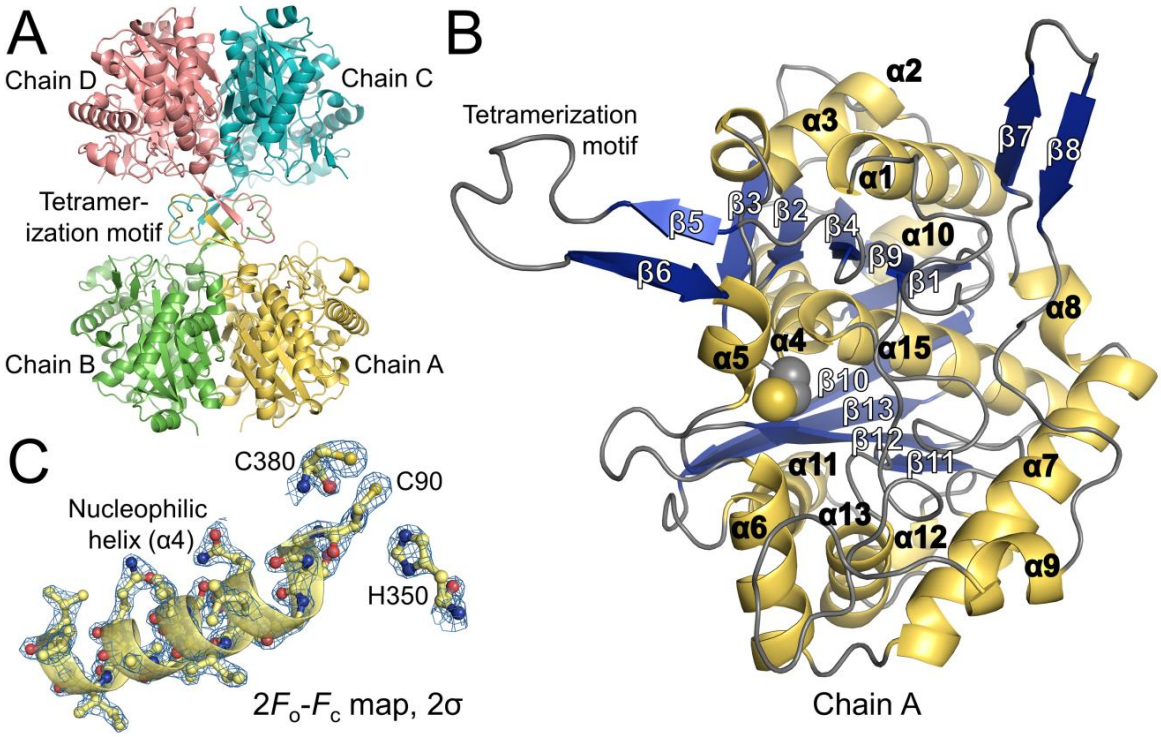


Figure 6-5. Structure of *Re*BktB. **(A)** Presumed biological architecture of the enzyme. Monomers are distinguished by color: A (maize), B (green), C (teal), and D (salmon). The tetramerization motif is indicated in the center of the four monomers. **(B)** Secondary structural features of a monomer (chain A). The five-layered αβαβα fold is evident here: two β-sheets (blue) are sandwiched between three clusters of α-helices (maize). β5, β6, and their N- and C-terminal loops comprise the tetramerization motif. α-Helices and β-sheets are numbered as in the topology map in **Figure 6-7**. The side chain of the nucleophilic C90 is shown as spheres. **(C)** Representative electron density of the active site of the 2.0-Å, twinned structure. 2F_o-F_c map, contoured at 2σ, of the nucleophilic helix (residues 91-106) and catalytic residues C90, H350, and C380. All molecular images were generated in PyMOL.

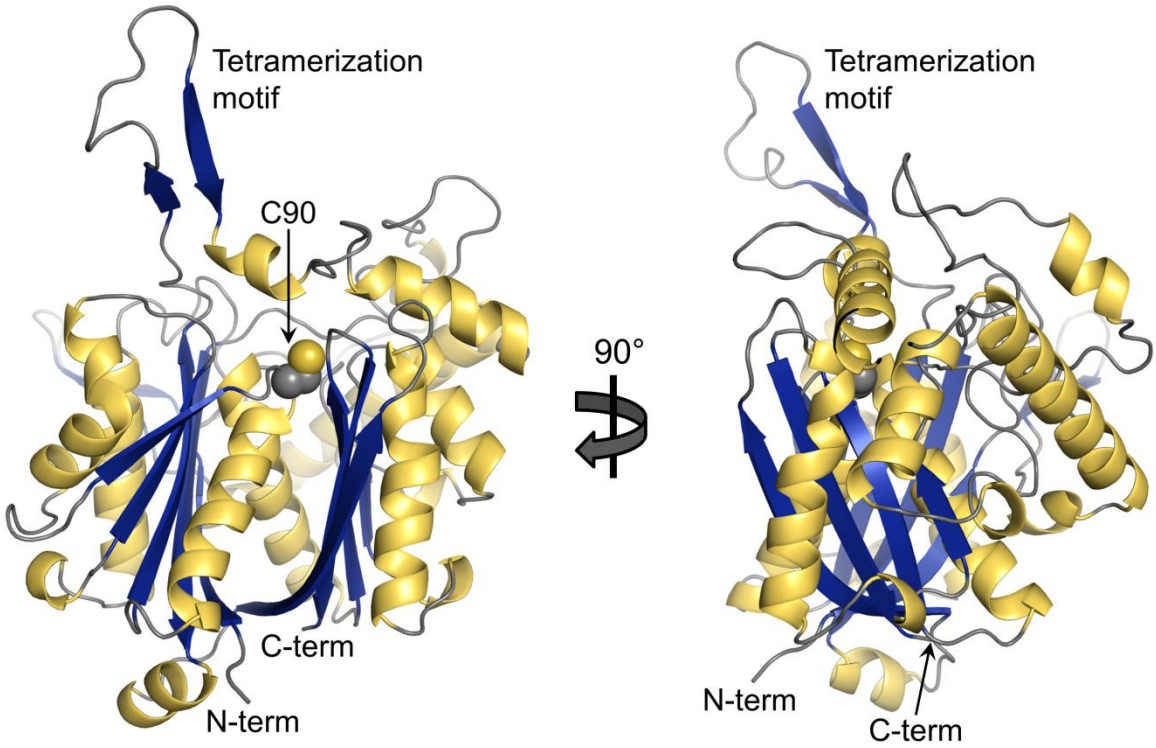


Figure 6-6. Two views of the βαβαβ motif of an *Re*BktB monomer. Loops (grey), α-helices (maize), β-strands (blue), and the nucleophilic C90 (spheres) are highlighted. The five layers of the thiolase superfamily motif are clearly distinguishable in the left panel. At the top of the monomer, the tetramerization motif consisting of residues 120-143 can be seen. N- and C-termini of the chain are labeled.

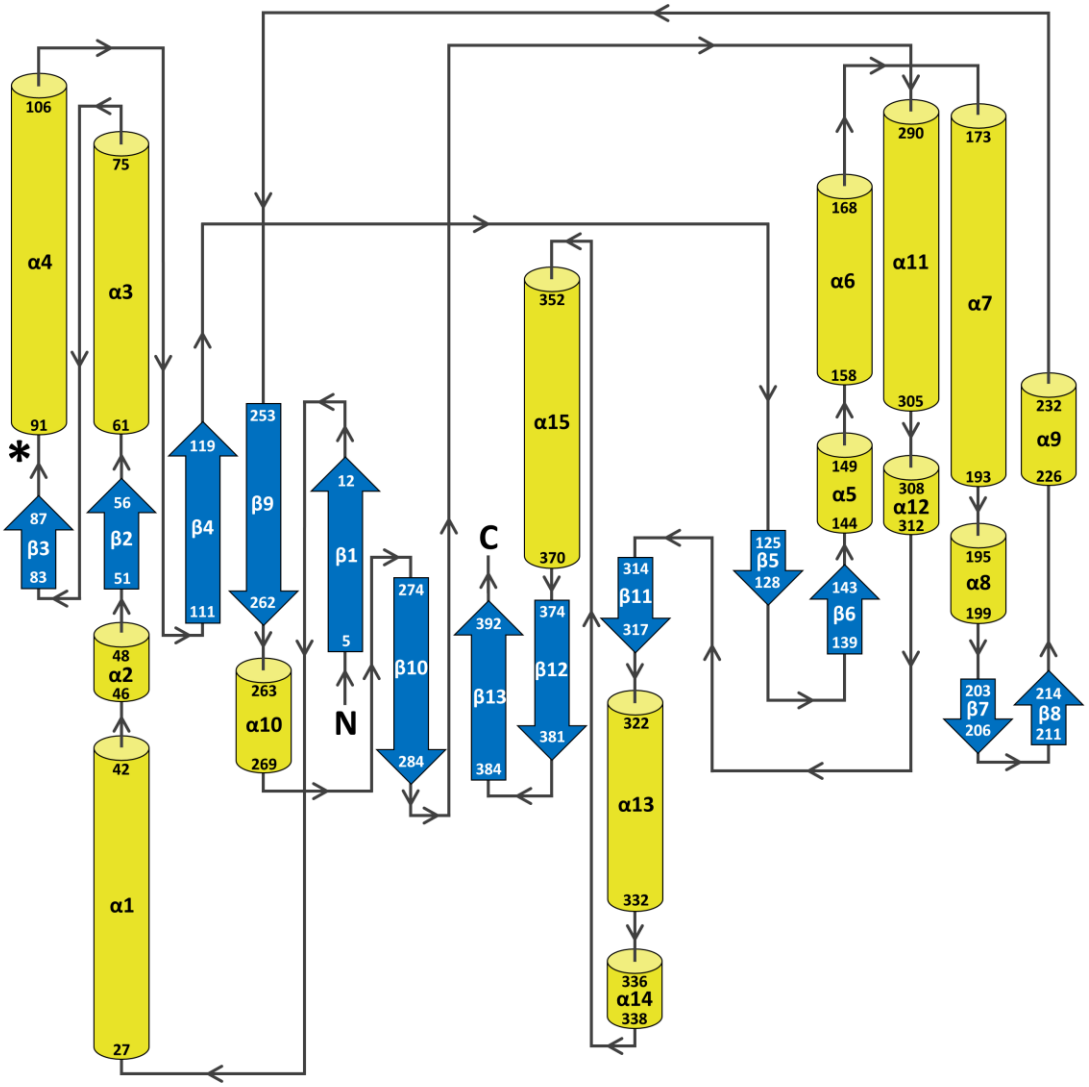


Figure 6-7. Topology of an *Re*BktB monomer, with α-helices (maize) and β-strands (blue) labeled α1-15 and β1-13, respectively. Numbers indicate residues at the boundaries of these secondary structural elements. The nucleophilic C90, located at the N-terminus of α4, is indicated by an asterisk.

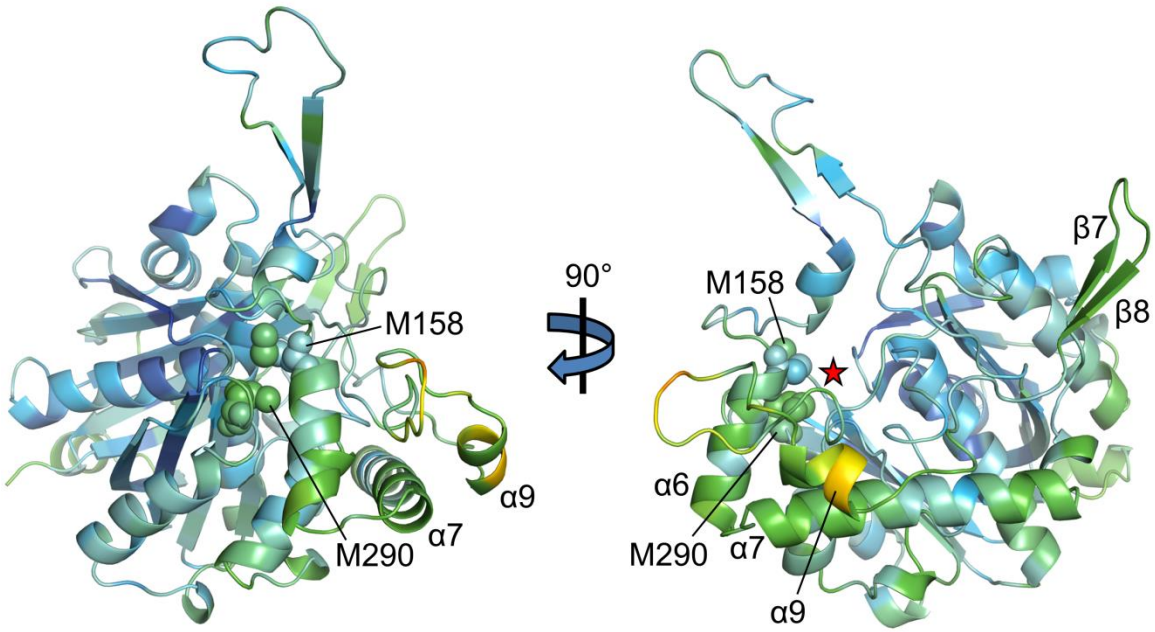


Figure 6-8. *B*-factor heat map of chain A. In this figure, *B*-factors range from ~2 (dark blue) to ~40 (orange). M158 and M290 are labeled and shown as spheres. The mouth of the channel that binds the pantetheinyl arm of CoA is indicated with a red star (right panel). Helix α9 is also labeled; along with its N- and C-terminal loops, this helix binds the adenosyl portion of CoA.

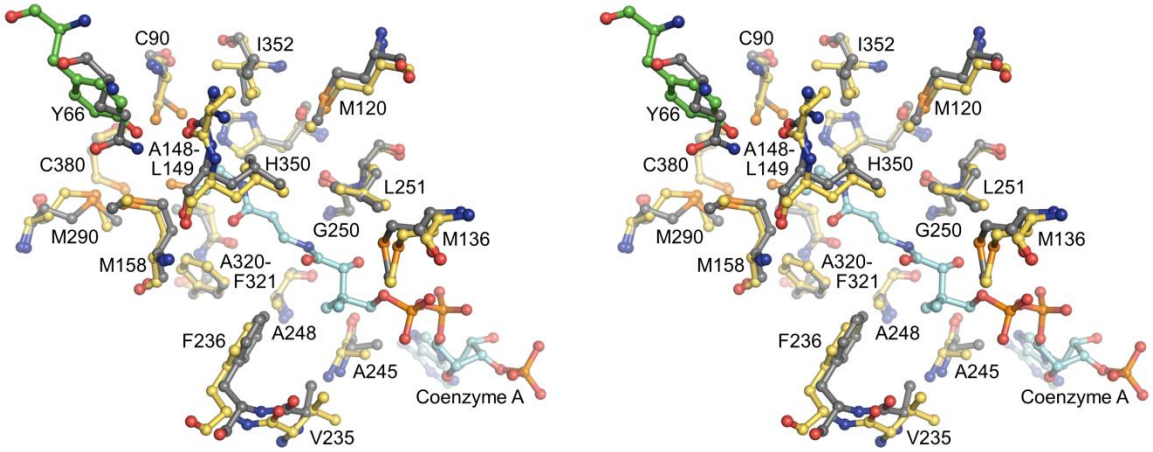


Figure 6-9. Conserved CoA binding pocket of *Re*BktB and *Zr*BktB (in stereo). Hydrophobic residues lining the CoA channel of *Re*BktB (maize) and *Zr*BktB (grey) are superposed and numbered according to *Re*BktB sequence. The pantetheine arm of CoA (cyan), modeled from *Zr*BktB, reaches into this pocket. Highlighted residues from *Re*BktB/*Zr*BktB include C90/89, M120/119, M136/134, A148/G147, L149/148, M158/157, V235/A234, F236/235, A245/243, A248/246, G250/248, L251/249, M290/288, A320/318, F321/319, H350/348, I352/350, and C380/789 of chain A and Y66/Q64 of chain B (green for *Re*BktB, grey for *Zr*BktB).

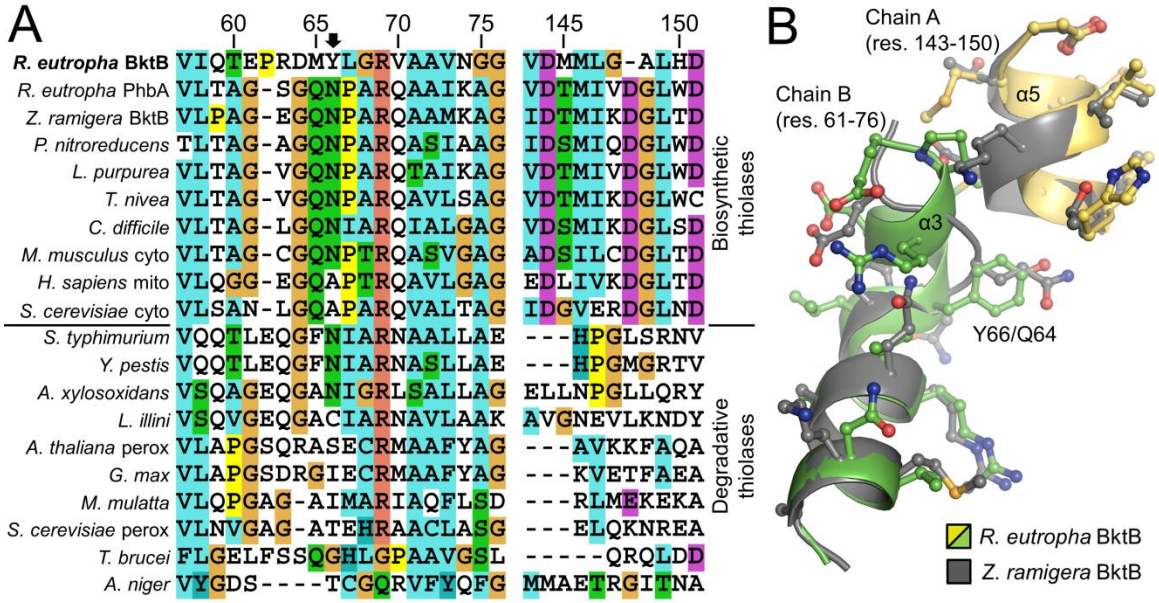


Figure 6-10. *R. eutropha* BktB possesses unique sequences on helices α3 and α5. **(A)** Alignment of thiolase sequences corresponding to *Re*BktB helices α3 (residues 62-75) and α5 (residues 144-149), numbered according to *Re*BktB. Unlike most biosynthetic thiolases, there is little conservation of residues near the active site-capping Y66 (black arrow) of *Re*BktB. In *Re*BktB, Y66 fills the space occupied by Q64, Q68, and Q101 from the *Z. ramigera* (PDB 1DLV), *C. difficile* (PDB 4DD5), and *Homo sapiens* mitochondrial (PDB 2IBY) biosynthetic thiolases. A variety of residues may replace Y66 in the degradative thiolases; for example, T101, F66, and S114 in *S. cerevisiae* peroxisomal (PDB 1AFW), *S. typhimurium* (PDB 3GOA) and *A. thaliana* peroxisomal (PDB 2WU9) degradative thiolases. Biosynthetic thiolases: *R. eutropha* BktB (NCBI YP_725948); *R. eutropha* PhbA (P14611); *Z. ramigera* BktB (P07097); *Pseudomonas nitroreducens* (WP_017518929); *Lamprocystis purpurea* (WP_020506577); *Thiothrix nivea* (WP_002708073); *C. difficile* (Q18AR0); *Mus musculus* cytosolic (NP_033364); *H. sapiens* mitochondrial (P24752); and *S. cerevisiae* cytosolic (XXBYAC). Degradative thiolases: *Salmonella typhimurium* (P0A2H7); *Yersinia pestis* (A4TR28); *Achromobacter xylosoxidans* (WP_020928315); *Leptonema illini* (WP_002774022); *A. thaliana* peroxisomal (Q56WD9); *Glycine max* (NP_001237076); *Macaca mulatta* peroxisomal (AFH32148); *S. cerevisiae* peroxisomal (P27796); *Trypanosoma brucei* (Q57XD5); and *Aspergillus niger* (EHA20372). Sequences were aligned with *ClustalX2* (Larkin *et al.*, 2007). **(B)** Superposed helices α3 (chain B) and α5 (chain A) of *Re*BktB and *Zr*BktB. Backbone atoms are hidden for clarity.

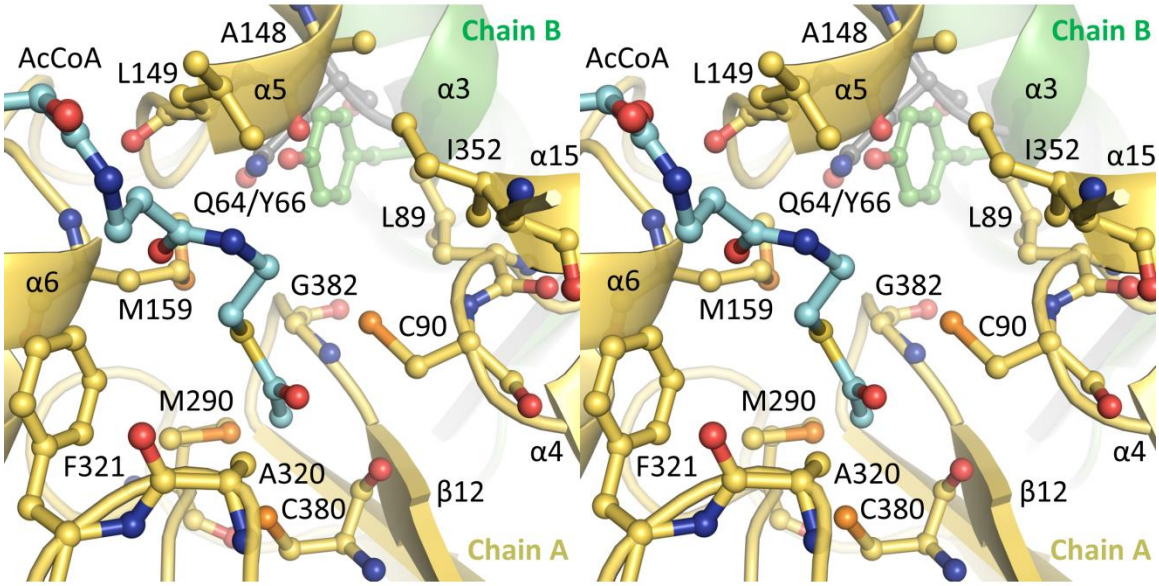


Figure 6-11. Active-site features of *Re*BktB (in stereo). Acetyl-CoA (AcCoA; cyan) is modeled from *Zr*BktB (PDB 1DM3), which was superposed onto *Re*BktB (maize). Q64 (grey) of *Zr*BktB and Y66 (green) of *Re*BktB, supplied by neighboring chain B, cap the active-site pocket. These and neighboring residues on α3 as well as α5 may be responsible for the observed disparities in substrate specificities of thiolases. C90, at the N-terminus of α4, is the nucleophilic residue capable of forming an acyl-enzyme intermediate. Of the catalytic acid/base residues, only C380 is depicted; H350, which stabilizes the enolate oxygen of the substrate, resides on a flexible loop in the foreground (hidden for clarity). The long loop between α9 and β9 is also hidden for clarity.

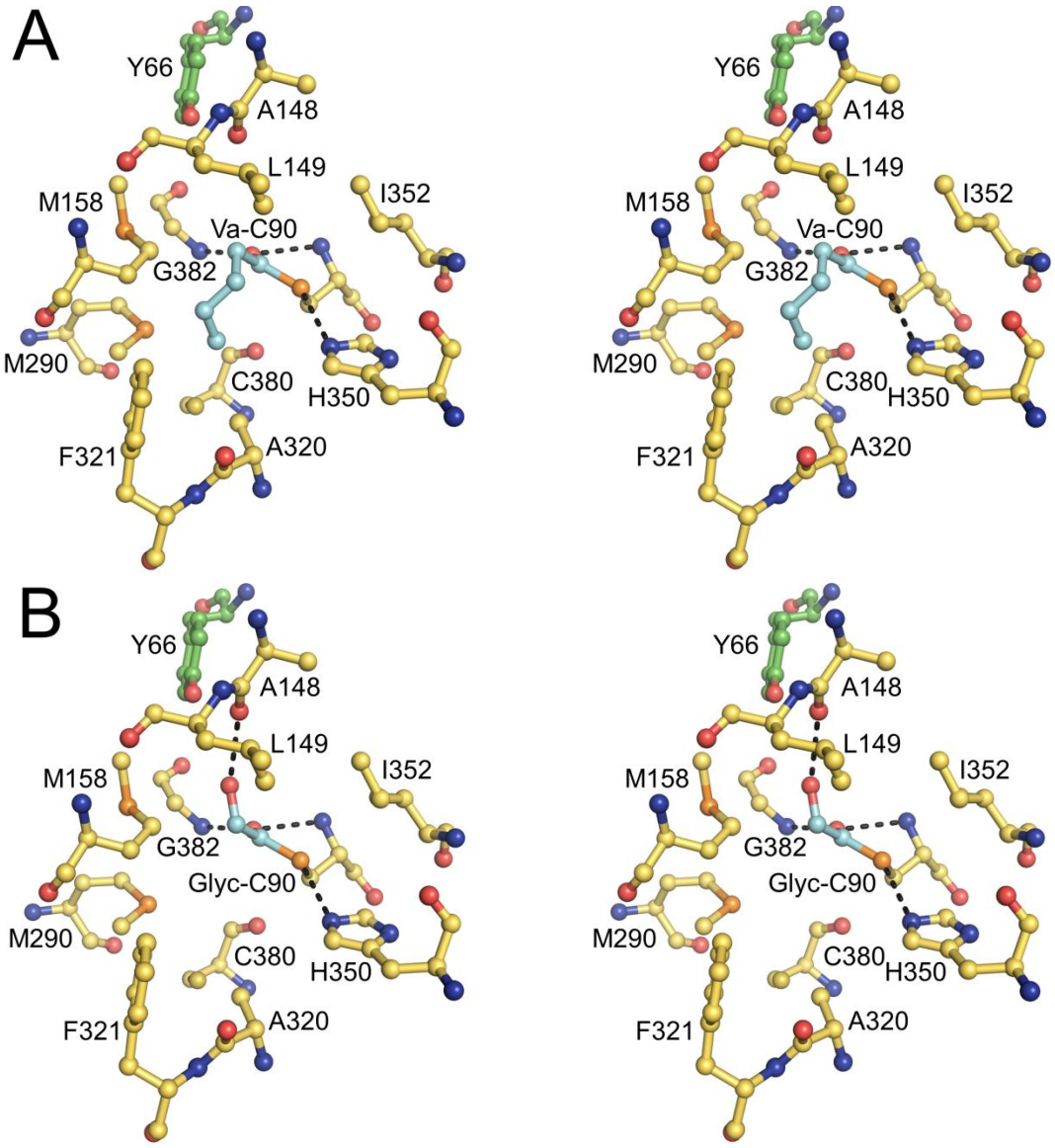


Figure 6-12. Models of acyl-*Re*BktB intermediates, guided by PDBs 1M4T and 1DM3 (in stereo). **(A)** Valeryl-enzyme intermediate (Va-C90). The terminal methyl group appears to fit within the hydrophobic pocket with proper geometry and no serious atomic clashes. **(B)** Glycolyl-enzyme intermediate (Glyc-C90). In contrast to the hydrophobic group of the valeryl-enzyme intermediate, the hydrophilic group of the glycolyl-enzyme intermediate is predicted to point toward the "top" of the cavity, forming a hydrogen bond with the backbone CO of A148, also with proper geometry and no serious atomic clashes. In both models, the oxyanion hole is formed by the backbone NHs of G382 and C90. H350 may interact with the thioester sulfur atom through its $H_{\varepsilon 2}$ atom (the S_{γ}-to-$N_{\varepsilon 2}$ distance is ~2.8 Å).

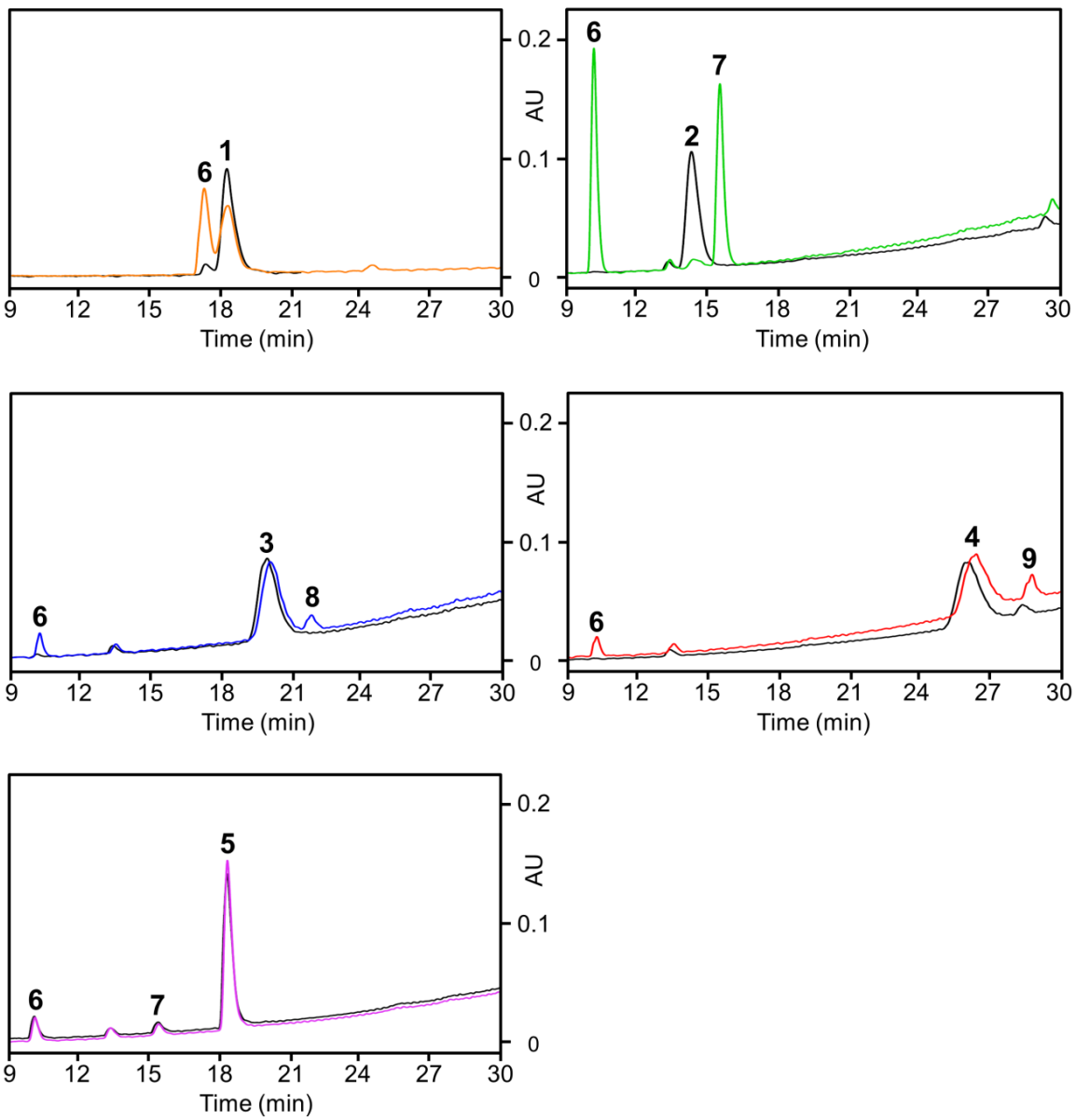


Figure 6-13. Representative reverse-phase HPLC chromatograms (235 nm) for β-ketoacyl-*S*-NAC thiolysis reactions. Colored traces: 1 μM *Re*BktB was incubated with 5 mM NAC and 1 mM substrate **1**, **2**, **3**, **4**, or **5** for 1 h at 22 °C. Grey traces: negative reaction mixtures (–*Re*BktB). Reactions were run in triplicate. Substrate stocks may be contaminated with minute amounts of **6**, a byproduct of NAC synthesis (see upper left trace); this was accounted for by subtracting the average area of peak **6** of the –*Re*BktB control from that of the corresponding +*Re*BktB reaction. Compounds are colored and labeled as in **Figure 6-15**.

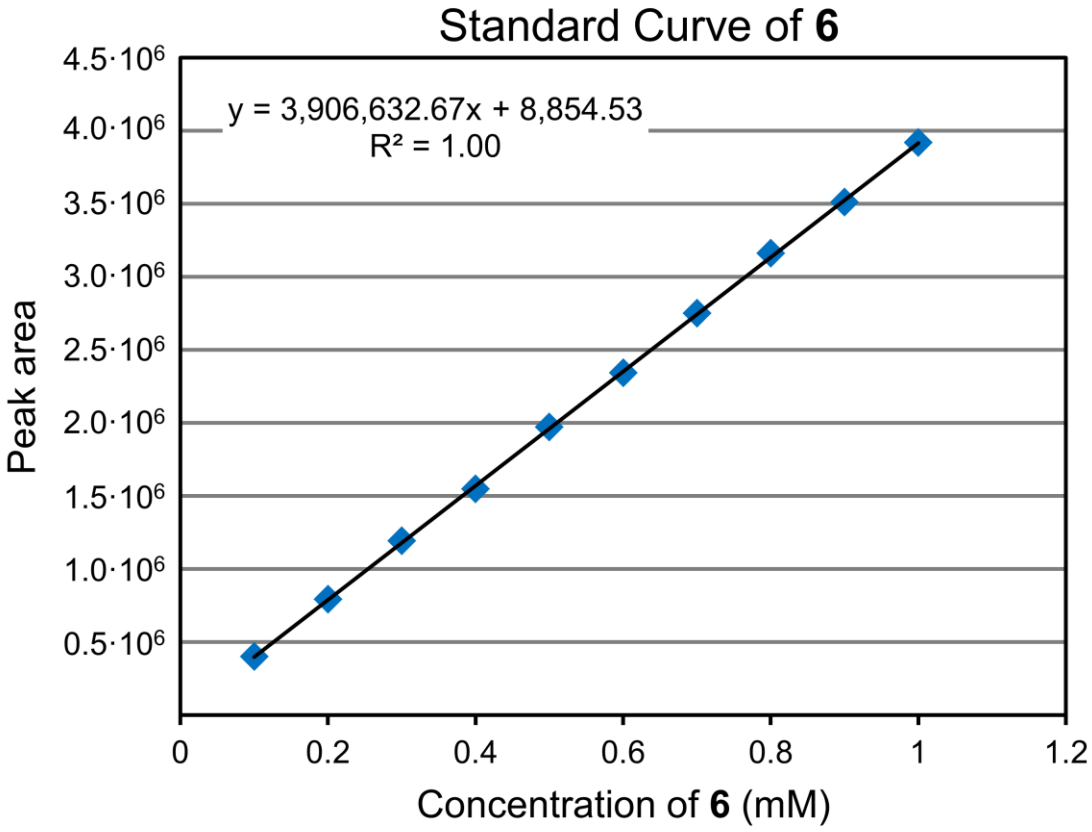


Figure 6-14. Standard curve of **6**. The curve was generated by injecting known concentrations of acetyl-*S*-NAC (0.1-1.0 mM) onto the reverse-phase HPLC column and integrating the peak area.

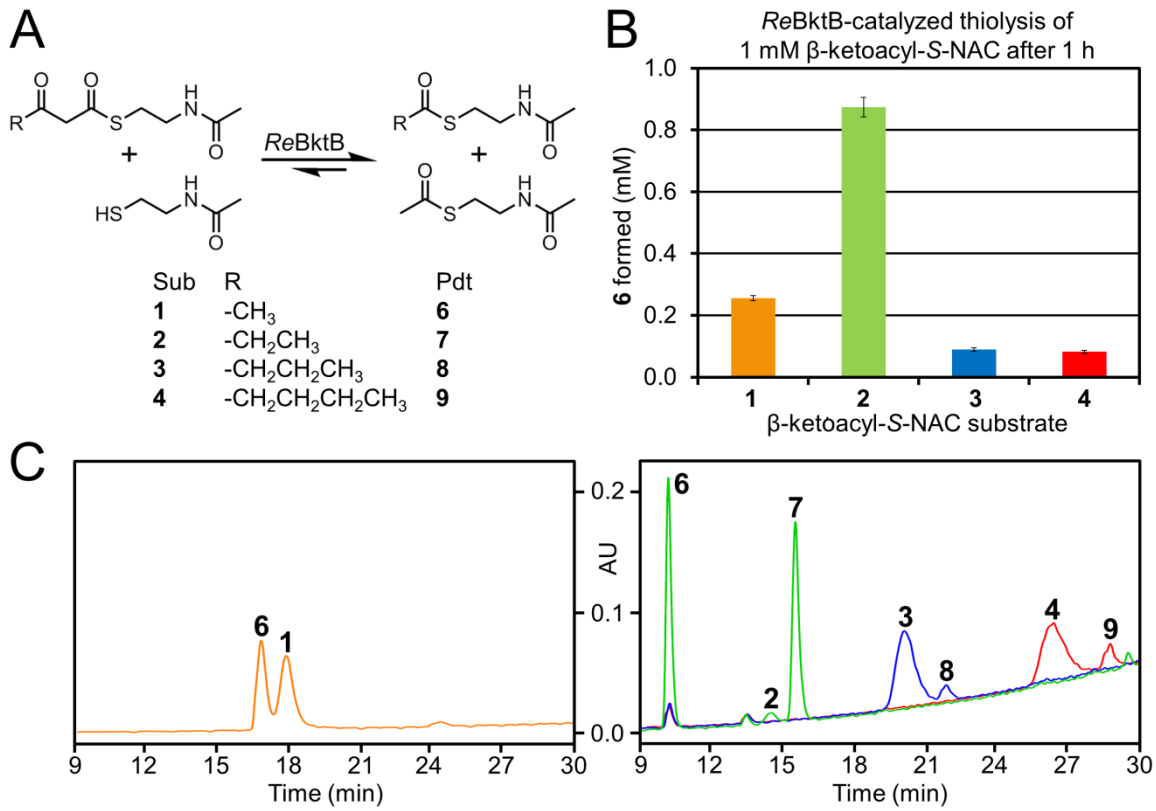


Figure 6-15. *In vitro* thiolytic cleavage of β-ketoacyl-*S*-NAC thioesters by *Re*BktB. **(A)** Reactions assayed with *Re*BktB. R-groups of β-ketoacyl-*S*-NAC substrates (Sub) and acyl-*S*-NAC products (Pdt) are indicated. Compound **5**, an α-methyl variant of **2**, was not a viable substrate for the thiolase, even after 24 h incubation. **(B)** *Re*BktB prefers substrates as follows: **2** > **1** > **3** ~ **4**. *Re*BktB (1 μM) was incubated with 1 mM β-ketoacyl-*S*-NAC and 5 mM NAC for 1 h at 22 °C, then acid-quenched and injected onto a reverse-phase HPLC column. The concentration of product **6** was determined by integrating the area of peak **6** in HPLC chromatograms, subtracting that of corresponding –*Re*BktB controls, and accounting for dilution factor: 0.873 ± 0.031 mM for **2**, 0.225 ± 0.004 mM for **1**, 0.089 ± 0.006 mM for **3**, and 0.082 ± 0.001 mM for **4**. Error bars indicate standard deviation of triplicate measurements. **(C)** Representative reverse-phase HPLC chromatograms (235 nm) for reactions containing **1** (left panel) and **2**, **3**, or **4** (right panel; overlay) after 1 h. A different gradient was employed for reactions containing **1** due to direct overlap substrate and product peaks.

Table 6-1. Data-collection and refinement statistics for *Re*BktB.

Values for the outer shell are given in parentheses.

	Twinned	**Untwinned**
Data collection		
Space group	*P*1	$P2_12_12_1$
Resolution (Å)	50-2.00 (2.03-2.00)	50-2.80 (2.85-2.80)
X-ray source	Advanced Light Source, 5.0.2	Advanced Light Source, 5.0.3
Wavelength (Å)	0.9795	0.9765
Temperature (K)	100	100
Unit cell parameters		
a, *b*, *c* (Å)	72.05, 105.99, 201.14	75.46, 115.15, 202.02
α, β, γ (°)	89.97, 89.98, 89.94	90.00, 90.00, 90.00
R_{merge}† (%)	5.6 (44.4)	7.3 (59.6)
Average $I/\sigma(I)$	10.4 (2.0)	18.7 (2.6)
No. of unique reflections	369,679 (18,535)	42,713 (2,077)
Completeness (%)	94.0 (93.7)	97.8 (96.6)
Redundancy	1.7 (1.7)	6.4 (6.3)
Monomers in asymmetric unit	16	4
Solvent content (%)	47.7	54.2
Matthews coefficient ($Å^3$ Da^{-1})	2.35	2.69
Refinement		
Resolution (Å)	41.8-2.01 (2.06-2.01)	43.8-2.80 (2.91-2.80)
R_{work}/R_{free}‡ (%)	21.9/23.6 (20.2/23.5)	20.6/25.8 (33.1/39.6)
No. of atoms		
Overall	45,477	11,346
Protein	45,238	11,342
Water	239	4
Average *B*-factors ($Å^2$)		
Overall	13.9	89.0
Protein	13.9	89.0
Water	9.6	65.6
R.m.s.d., bond lengths (Å)	0.02	0.012
R.m.s.d., bond angles (°)	2.26	1.57
Ramachandran favored (%)	95.0	95.0
Ramachandran allowed (%)	4.3	4.7
Ramachandran outliers (%)	0.7	0.3

†$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity for all reflections.

‡$R_{work}/R_{free} = \sum_{hkl}\big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

Table 6-2. Estimation of the twin fraction, α, as calculated by *phenix.xtriage* of *PHENIX*. The program identified three pseudo-merohedral twin operators in space group *P*1 and determined $\alpha > 0.41$ for each operator using three different methods. (Note that for perfectly twinned data, $\alpha = 0.5$, and for untwinned data, $\alpha = 0$). More information about Britton and *H* Test methods can be found in the references (Britton, 1972; Yeates, 1988). Details on the maximum likelihood method have not yet been published by the *PHENIX* developers.

Twin operator	*h*, -*k*, -*l*	-*h*, *k*, -*l*	-*h*, -*k*, *l*
Estimated α, by method			
Britton	0.424	0.434	0.424
H Test	0.430	0.438	0.430
Maximum likelihood	0.413	0.422	0.413

Table 6-3. Improvement of residual factors after intensity-based twin refinement of the 2.0-Å dataset in *REFMAC5*. Twin refinement was first attempted in space group $P2_1$ followed by $P2_12_12_1$.

Space group	$P2_12_12_1$	$P2_1$	$P1$
Monomers in asymmetric unit	4	8	16
Twin operator(s)	-	*h*, -*k*, *l*	*h*, -*k*, -*l*; -*h*, *k*, -*l*; -*h*, -*k*, *l*
Twin refinement			
R_{work}/R_{free} before (%)	38.4/43.5	39.0/43.1	26.5/32.3
R_{work}/R_{free} after (%)	-	27.4/34.4	17.8/22.0

Chapter 7: Conclusions and Future Perspectives

This document unveiled the architectures of five bacterial biomolecular machines involved in either cell-surface modification or polyketide biosynthesis. **Chapters 2** and **3** examined the enzymes AlmE and EptC, which decorate the cell surfaces of pathogens *Vibrio cholerae* and *Campylobacter jejuni*, respectively. We solved the atomic coordinates of both proteins to identify amino acid residues that confer substrate specificity. AlmE was engineered to adenylate L-alanine, and its inherent propensity to adenylate D-alanine was ascertained. Additionally, the role of AlmF as a cognate carrier protein to AlmE was substantiated. Mutagenesis of crucial EptC residues abolished motility and polymyxin resistance in *C. jejuni*, supporting the function of these residues in nucleophilic attack as well as metal ion and substrate coordination. Our analyses pave the way for the design of inhibitors for potential treatment of diseases caused by *V. cholerae*, *C. jejuni*, and related organisms. Future studies of these systems should be directed at AlmG, the most mysterious of the AlmEFG trio. AlmG shares little sequence identity with annotated enzymes and executes the final transfer of glycine or diglycine to lipid A, endowing polymyxin resistance. The role of the N-terminal transmembrane domain of EptC as well as the capacity of the enzyme to bind multiple metal ions, like alkaline phosphatase, should also be researched.

In **Chapter 4**, we explored the SAM methyltransferase-like SpnF, which seals three rings in the 22-membered macrolactone precursor to spinosad, an important insecticide. We combined results from X-ray crystallography, mutant enzyme assays, and *in silico* docking to hypothesize which residues bind and polarize the substrate for ring closure. Additional studies will be required to elucidate the exact

cyclization mechanism of SpnF and confirm whether it catalyzes a *bona fide* Diels-Alder reaction. Structural evaluation of SpnL, another SAM methyltransferase homologue in the spinosyn pathway and a putative Rauhut-Currier cyclase, may further illuminate the role of the methyltransferase fold in these biologically unusual, non-transfer reactions.

We investigated dimerization elements (DEs) via X-ray crystallography, NMR, and analytical ultracentrifugation in **Chapter 5**. These ~55-residue domains are scaffolds for about half of polyketide synthase modules containing ketoreductases as the sole β-processing unit. With this "missing linker", we have concluded efforts to structurally characterize the common polyketide synthase domains, which permitted us to assemble a model of a prototypical module. Analyses of interdomain and intermodule contacts are burgeoning (see **Chapter 5**; Bretschneider *et al.*, 2013; Dutta *et al.*, 2014), and the next step is to draft the blueprints of an entire synthase. This will improve our understanding of domain and module boundaries, easing the generation of chimeric synthases and, thus, development of novel polyketides.

Finally, in **Chapter 6**, our structural work on β-ketoacyl thiolase B from *Ralstonia eutropha* (*Re*BktB) allowed us to model intermediates in the active site and revealed an unconserved region that may be implicated in substrate permissiveness. We also demonstrated activity on thioesters of *S*-*N*-acetylcysteamine, including the unexpected 3-oxoheptanoyl-*S*-NAC, and established a preference for 3-oxopentanoyl-*S*-NAC. Future experiments should test even bulkier β-ketoacyl-*S*-NAC derivatives such as 3-oxooctanoyl-*S*-NAC, branched isomers such as 4-methyl-3-oxopentanoyl-*S*-NAC, and charged moieties like 4-hydroxy-3-oxobutanoyl-*S*-NAC to expand the functional repertoire. In addition,

attempts should be made to reroute *Re*BktB catalysis in the biosynthetic direction in order to generate valuable chemical building blocks.

Appendix A: Introduction to X-ray Crystallography

This appendix serves as a brief introduction to the principles of X-ray crystallography as understood by the author. A number of references, citations of which are omitted for clarity, were utilized in arranging this section. The outstanding Crystallography Made Crystal Clear by Gale Rhodes (Rhodes, 2006) served as the major source of inspiration, and the bulk of the figures herein are adapted from that text. Other informative resources include Principles of Protein X-Ray Crystallography by Jan Drenth (Drenth, 2007) and Outline of Crystallography for Biologists by David Blow (Blow, 2010).

OVERVIEW OF X-RAY CRYSTALLOGRAPHY

When viewing an object under an optical microscope, light diffracted by the object is captured by a lens and reassembled into an image on the other side. Essentially, the wavelength of light shone on the object cannot be larger than the object itself. Thus, if one hopes to "view" chemical bonds, X-rays are of the appropriate wavelength (~0.1-100 Å; a typical carbon-carbon single bond is 1.54 Å in length). However, a single molecule cannot be observed with X-rays because they are not focusable with such a lens. Crystallographers therefore substitute a computer for the image-producing lens. In addition, a single molecule is a poor diffractor of X-rays. Crystallographers rely on regularly repeating crystal lattices that constructively recombine the scattered radiation, amplifying the signal to detectable levels.

Growing adequately diffracting crystals is the major barrier of X-ray crystallography. Protein crystals are obtained through gradual precipitation of the

protein from aqueous solution (e.g., vapor diffusion). Crystallization conditions are impossible to predict *a priori* and are thus attained via sparse-matrix screening. A multitude of factors can affect the crystallization process: concentration of protein, precipitant, and salt; type of precipitant and salt; purity and dispersity of protein and buffer components; source organism of protein; proteolysis; temperature; pressure; pH; presence of small molecules (e.g., substrates, coenzymes, detergents, stabilizers like glycerol); vibrations; and duration of crystallization. In addition, the termini of the protein should be carefully designed from multiple sequence alignments prior to cloning to yield a relatively stable, compact molecule free of excessive flexibility.

That a Fourier transform can be applied to a set of diffraction images to visualize electron density is the greatest advantage associated with X-ray crystallography. Leucine, a typical amino acid residue, contains 8 non-H atoms with 51 electrons (hydrogen is a weak X-ray diffractor). If the average protein contains hundreds of residues (say, 350), stretches 50 Å in each dimension, and forms cubic crystals with 0.1 mm per side, then the average protein crystal consists of $\sim 8 \cdot 10^{12}$ protein molecules and $\sim 1.4 \cdot 10^{17}$ electrons (ignoring the $\sim 0.3 \cdot 10^{17}$ contributed by H). Bombardment of this repeating electron lattice with X-rays results in enough constructive interference to produce accurately detectable reflections. After the phases of the reflections are elucidated (the detector only records intensities and positions), the Fourier transform is applied and an atomic model can be constructed within the confines of an electron density map.

CRYSTAL LATTICES

The asymmetric unit is the smallest repeating element of the crystal. One or more asymmetric units comprise a unit cell, which stacks on identical unit cells in a three-dimensional array. Within the unit cell, the asymmetric unit(s) must fulfill any symmetry operations defined by the space group in which the protein crystallized (e.g., a crystal with one two-fold symmetry axis contains two asymmetric units, related by rotation of 180°, per unit cell). While symmetry may exist within the asymmetric unit itself, known as non-crystallographic symmetry, these operations are not designated by the space group.

The axes of the unit cell are denoted **a**, **b**, and **c**, and the angles are denoted α, β, and γ. As shown by W.L. Bragg (assisted by his father W.H. Bragg) in 1913, diffracted X-rays can be imagined as reflections from regularly repeating, parallel planes of atoms which intersect edges **a**, **b**, and **c** of each unit cell. In this model, explained below in further detail, each series of parallel planes produces a single reflection, or spot in the diffraction pattern. These planes are labeled with Miller indices *hkl*; for example, index 130 describes the (130) planes, which intersect edges **a**, **b**, and **c** one, three, and zero times, respectively (**Figure A-1**). It is worth noting that the (130) planes are identical to the (-1 -3 0), or ($\bar{1}$ $\bar{3}$ 0), planes.

Although the unit cell of a crystal can often be defined under several lattice types, it is standard to select the unit cell which maximizes symmetry. First, one of 14 possible Bravais lattices is identified (**Table A-1**). Each Bravais lattice is described by one of 7 crystal systems (**Figure A-2**) as well as one of 4 lattice types (**Figure A-3**). After choosing the Bravais lattice, the crystallographer determines the space group which defines the symmetry of the unit cell. In contrast to lattices, there are 230 possible space groups; however, only 65 apply to protein crystals due to the

chiral nature of their constituent amino acids (**Table A-1**). Symmetry operations in these 65 space groups are restricted to rotations, translations, and screw axes (combined rotations and translations). The most common space group, $P2_12_12_1$, contains three twofold screw axes normal to one another; that is, along each unit cell axis, the asymmetric unit can be rotated 180° and translated one-half the length of the unit cell edge. In other words, an atom located at (x, y, z) is also located at ($-x$ + ½, $-y$, z + ½), ($-x$, y + ½, $-z$ + ½), and (x + ½, $-y$ + ½, $-z$). In addition, reflections on the $h00$, $0k0$, and $00l$ planes will appear at $h = 2n$, $k = 2n$, and $l = 2n$, respectively. Accordingly, the systematic absence of odd-numbered reflections along these axes can be used to diagnose the space group as $P2_12_12_1$ from diffraction images (**Figure A-4**). Such information for all space groups is available online in Volume A of the International Tables for Crystallography (International Union of Crystallography, 2006).

Symmetry elements are also present within the diffraction pattern itself. Each reciprocal lattice contains an inversion center at its origin; that is, the intensity of reflection hkl is equivalent to that of reflection $\overline{hkl}$, since they arise from opposing sides of the same set of parallel planes (see Friedel's law, below). Thus, to measure all unique reflections from a crystal in the lowest-symmetry space group, $P1$, it is only necessary to collect 180° of diffraction data (½ of the reflections are unique). If other symmetry elements are present, a smaller angle will suffice. For example, in space group $P23$, which possesses two- and three-fold rotation axes, 1/12 of the reflections are unique. Nevertheless, by measuring redundant reflections the crystallographer increases signal-to-noise and corrects for variations in X-ray absorption. The reciprocal-lattice symmetry for all crystals can be described by just 11 Laue groups.

BRAGG'S LAW

Bragg demonstrated that diffraction occurs only when

$$2d_{hkl}sin\theta = n\lambda, \qquad \textbf{(A-1)}$$

where n is an integer, d_{hkl} is the space between parallel planes with index (frequency) hkl, and θ is the angle of the incoming and outgoing X-rays of wavelength λ (**Figure A-5**).

Using $sin\theta = BC/AC$, one can obtain $BC = AC\ sin\theta = d_{hkl}sin\theta$. Thus, ray X_2 travels an additional distance of $2BC$ or $2d_{hkl}sin\theta$ in comparison to ray X_1. If this path length difference is equal to an integral number of wavelengths ($n\lambda$), the reflected waves will interfere constructively to yield an outgoing beam; otherwise, the reflected waves will interfere destructively (they will be out of phase) and there will be little or no outgoing beam. The intensity of the outgoing beam depends on the amount of electron density along the series of parallel planes (hkl).

Bragg's law can also be considered in reciprocal space. For every real space lattice with edges **a**, **b**, and **c**, and respective lengths a, b, and c, there is a reciprocal lattice with edges **a***, **b***, and **c***, and respective lengths $1/a$, $1/b$, and $1/c$. In **Figure A-6**, the X-ray beam IO strikes the crystal in plane **a*b*** of the reciprocal lattice. A circle intersecting origin O with radius $1/\lambda$ and center C on IO represents the wavelength of the beam in reciprocal space. BAO is a right triangle, so

$$sin\theta = \frac{OB}{OA} = \frac{OB}{2/\lambda}. \qquad \textbf{(A-2)}$$

Therefore,

$$\frac{2}{OB} sin\theta = \lambda. \quad \textbf{(A-3)}$$

Also, $OB = 1/d_{hkl}$ because B is a reciprocal lattice point. Therefore,

$$2d_{hkl}\ sin\theta = \lambda. \quad \textbf{(A-4)}$$

This is Bragg's law, where $n = 1$. The criteria for diffraction are fulfilled when point B intersects the circle; that is, ray X is produced in the direction of CB. Rotating the crystal (and therefore the reciprocal lattice) about origin O will bring another point B' in contact with the circle, producing a new ray X' in the direction of CB'. Rotating the blue circle with radius $1/\lambda$ and center C around the X-ray beam IO generates a sphere of reflection (also known as the Ewald sphere). As the sphere of reflection rotates about the origin O, it intersects a reciprocal lattice point B_{hkl}, producing reflection hkl in the direction of B_{hkl} from C. The sphere of reflection may intersect any reciprocal lattice point within $2/\lambda$ of the origin as it rotates, generating yet another sphere known as the limiting sphere. This larger limiting sphere, with its center on origin O and radius $2/\lambda$, contains reciprocal lattice points equal to the number of reflections producible by rotating the crystal through every available orientation in the X-ray beam. Thus, the number of possible reflections is based on the wavelength of the radiation and the volume of the unit cell: shorter wavelengths allow for a larger sphere of reflection and larger unit cells allow for smaller

reciprocal unit cells (giving a denser sphere of reflection); both of these examples increase the number of possible reflections.

FOURIER SUMS AND TRANSFORMS

Simple waves like X-rays can be described by periodic functions such as

$$f(x) = F\ cos2\pi(hx + \alpha) \qquad \textbf{(A-5)}$$

or

$$f(x) = F\ sin2\pi(hx + \alpha), \qquad \textbf{(A-6)}$$

where $f(x)$ is the vertical height, x is the horizontal position, F is the amplitude, h is the frequency, and α is the phase of the wave.

Very complex waves like those scattered from crystals can be described as a sum (Fourier sum) of such periodic functions. For example, a Fourier sum can be represented as

$$\begin{aligned} f(x) = {} & F_0 cos2\pi(0x + \alpha_0) + F_1 cos2\pi(1x + \alpha_1) \\ & + F_2 cos2\pi(2x + \alpha_2) + \ldots + F_n cos2\pi(nx + \alpha_n) \end{aligned} \qquad \textbf{(A-7)}$$

or

$$f(x) = \sum_{h=0}^{n} F_h cos2\pi(hx + \alpha_h), \qquad \textbf{(A-8)}$$

with amplitude F_h, frequency h, and phase α. Crystallographers work with the simple wave equation $cos2\pi(hx) + isin2\pi(hx)$ and, hence, the Fourier sum

$$f(x) = \sum_{h=0}^{n} F_h[cos2\pi(hx) + isin2\pi(hx)], \quad \textbf{(A-9)}$$

where α is implicit and i is imaginary number $\sqrt{-1}$. Using $cos\theta + isin\theta = e^{i\theta}$, equation (**A-9**) can be rewritten as

$$f(x) = \sum_{h=0}^{n} F_h e^{2\pi i(hx)}$$
$$\text{or } f(x) = \sum_{h} F_h e^{2\pi i(hx)}. \quad \textbf{(A-10)}$$

In three dimensions, equation (**A-10**) becomes

$$f(x,y,z) = \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{2\pi i(hx+ky+lz)}. \quad \textbf{(A-11)}$$

Ultimately, the crystallographer desires a real-space electron density map from a reciprocal-space diffraction pattern. Electron density can be obtained using the Fourier sum

$$\rho(x,y,z) = \frac{1}{V}\sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{-2\pi i(hx+ky+lz)}, \quad \textbf{(A-12)}$$

where $\rho(x,y,z)$ is the electron density at coordinates (x, y, z); V is the unit cell volume; h, k, and l are frequencies (indices); and F_{hkl} is the structure factor for a diffracted X-ray (e.g., a spot in the diffraction pattern). Because every spot in a

diffraction pattern is influenced by every atom in the unit cell, F_{hkl} is the sum of the scattering factor, f_{hkl}, for every atom in the unit cell:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad \textbf{(A-13)}$$

for reflection *hkl* and atom *j*. Every element has a different value for f_j based on its number of electrons. The types of atoms determine the amplitude of the reflection, while the positions of the atoms determine the phase. F_{hkl} can also be calculated as an integral of tiny volume elements:

$$F_{hkl} = \int_x \int_y \int_z \rho(x,y,z) e^{2\pi i(hx+ky+lz)}\, dx\, dy\, dz. \quad \textbf{(A-14)}$$

F_{hkl} is the Fourier transform of $\rho(x,y,z)$. This transform converts real-space distances to reciprocal-space distances, and is the back-transform of equation (**A-12**). Notably, equation (**A-12**) is a sum instead of an integral because the spots in a diffraction pattern are described by discrete values.

In order to solve equation (**A-12**) and obtain an electron density map, three pieces of information are necessary: amplitude, frequency, and phase. The amplitude of a structure factor is proportional to reflection intensity, I_{hkl}:

$$|F_{hkl}| \, \alpha \, \sqrt{I_{hkl}}. \quad \textbf{(A-15)}$$

The frequency of a structure factor is equal to $1/d_{hkl}$, the reciprocal spacing between the parallel planes (*hkl*) that yield the reflection interest (see **Figure A-5**). (Low-frequency terms, which appear as spots near the center of the diffraction pattern,

contribute to the coarse details of the electron density map; high-frequency terms, which are farther from the center, contribute to the fine details.) The phase of a structure factor, however, cannot be calculated from the diffraction image alone.

Solving equation (**A-12**) for each reflection in a complete diffraction pattern generates an electron density map; however, equations (**A-13**) and (**A-14**) are equally important to the crystallographer. By comparing the calculated structure factors F_c, obtained from equation (**A-13**) or (**A-14**), with observed structure factors F_o, obtained from the diffraction pattern, the crystallographer is able to evaluate the progress of phasing and, thus, decide when model building is complete. In addition, F_{hkl} values can be calculated from a known structure with equation (**A-13**) or (**A-14**) and fed into equation (**A-12**) (the phases, in particular) to approximate the initial $\rho(x, y, z)$ values of an unknown structure.

CALCULATING PHASES

As mentioned above, the goal of crystallography is to obtain a real-space electron density map from a reciprocal-space diffraction pattern, which can be achieved through equation (**A-12**).

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}. \qquad \textbf{(A-12)}$$

This equation can also be written as

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}| e^{i\alpha_{hkl}} e^{-2\pi i(hx+ky+lz)}, \qquad \textbf{(A-16)}$$

where the phase, α, is included explicitly. The structure factor for reflection *hkl*, F_{hkl}, describes the frequency (indices *h*, *k*, and *l*), amplitude ($|F_{hkl}|$, which is proportional to $\sqrt{I_{hkl}}$), and phase (α). Because X-ray detectors record only the first two parameters, the phase of each reflection must elucidated before an electron density map is obtainable. Phases can be calculated by one of three major methods: isomorphous replacement, anomalous dispersion, or molecular replacement (in some cases, combining multiple methods can be fruitful.)

Isomorphous replacement requires at least two datasets: a native dataset and a heavy-atom derivative dataset. A "heavy" atom is one with a relatively high atomic number and thus electron count, such as platinum, mercury, or uranium. When such an atom is incorporated into the crystal, typically by soaking for hours to months, it will slightly alter the intensities of the diffraction pattern (recall that every atom contributes to every reflection). Compared to the native diffraction pattern, the altered intensities from the derivative arise from the heavy atoms alone. The term "isomorphous" refers to equivalent asymmetric unit and unit cell parameters between the native and derivative crystals, a prerequisite for measuring intensity differences between otherwise identical reflections. To solve the phases, we can look to the vector equation

$$\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H, \qquad \textbf{(A-17)}$$

which involves the structure-factor vectors of the protein ($\mathbf{F}_P$), heavy-atom derivative ($\mathbf{F}_{PH}$), and heavy atoms alone ($\mathbf{F}_H$) (**Figure A-7**). From the diffraction patterns, we know the amplitudes of $\mathbf{F}_{PH}$ and $\mathbf{F}_P$ (vector lengths $|\mathbf{F}_{PH}|$ and $|\mathbf{F}_P|$), but we do not know the phases. If we can determine the location of the heavy atoms

alone, we will know both the amplitude (vector length) and phase (vector angle) of $\mathbf{F}_H$, allowing us to arrive at a geometric solution (**Figure A-8**). However, both $\mathbf{F}_P$ and $\mathbf{F}_P$* of **Figure A-8** can satisfy equation (**A-17**). If an average of the two solutions is not adequate for initial phasing, it may be necessary to prepare a second or even a third or a fourth heavy-atom derivative. Use of more than one derivative is referred to as multiple isomorphous replacement, or MIR, as opposed to single isomorphous replacement, or SIR.

We can determine the location of the heavy atoms alone ($\mathbf{F}_H$) with Patterson maps—maps of interatomic vectors with peaks (atoms) at the head and tail of each vector (**Figure A-9**). The Patterson function (**A-18**), a Fourier sum lacking phases, resembles equation (**A-12**):

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}, \qquad \textbf{(A-18)}$$

where (u, v, w) is a point on the Patterson map and $|\mathbf{F}_{hkl}|^2$ is the amplitude (length of a vector, proportional to intensity). A difference Patterson function is convenient for locating the heavy atoms, as there are significantly fewer peaks. For such a function, the amplitude is $(\Delta\mathbf{F})^2 = (|\mathbf{F}_{PH}| - |\mathbf{F}_P|)^2$, or that of the heavy atom alone:

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_l \Delta\mathbf{F}_{hkl}^2 e^{-2\pi i(hu+kv+lw)}. \qquad \textbf{(A-19)}$$

Searching for symmetry-related Patterson peaks in Harker sections further simplifies the task of locating heavy atoms, allowing one to calculate the structure factor $\mathbf{F}_H$ and therefore $\mathbf{F}_P$ (using $\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$). For example, as described above, an atom located at (x, y, z) in space group $P2_12_12_1$ is also located at (-x + ½, -y, z + ½), (-

$x, y + ½, -z + ½$), and ($x + ½, -y + ½, -z$). Vectors relating atoms at (x, y, z) and ($-x + ½, -y, z + ½$) by symmetry can be found on the (u, v, w) = [$x - (-x + ½), y - -y, z - (z + ½)$] = ($2x + -½, 2y, -½$) = ($2x + ½, 2y, ½$) or $z = ½$ plane. After locating the Patterson atom (peak) at coordinates ($u, v, ½$), we can solve $u = 2x + ½$ for x and $v = 2y$ for y to determine the real coordinates of the atom. Note that z can take any value: We can simply assign $z = 0$, allowing this heavy atom to serve as a reference point for others.

Like isomorphous replacement, anomalous dispersion relies on at least two datasets: A native dataset and a heavy-atom derivative dataset (selenomethionine derivatives are routinely expressed and purified for this purpose). Anomalous dispersion exploits small variations in the intensities of Friedel pairs that arise in the presence of heavy atoms. Irradiation of derivative crystals with X-rays at energies corresponding to inner-shell electron transitions in the heavy atom (e.g., near the absorption edge; **Figure A-10**) alters the phases of some of the outgoing rays; this, in turn, affects the intensities of Friedel pairs. In the absence of anomalous dispersion Friedel pairs are equivalent in amplitude, in accordance with Friedel's law:

$$|\mathbf{F}_{hkl}| = |\mathbf{F}_{\overline{hkl}}|. \qquad \textbf{(A-20)}$$

These reflections have opposite phases—that is, if reflection hkl has phase α, then reflection $\overline{hkl}$ has phase $-\alpha$. As illustrated in **Figure A-11**, anomalous scattering alters a structure factor $\mathbf{F}_{PH}^{0}$ by introducing two perpendicular vectors, $\Delta\mathbf{F}'_{H}$ (real) and $\Delta\mathbf{F}''_{H}$ (imaginary), to yield $\mathbf{F}_{PH}^{anom}$:

$$\mathbf{F}_{PH}^{anom} = \mathbf{F}_{PH}^{0} + \Delta\mathbf{F}'_{H} + \Delta\mathbf{F}''_{H}. \qquad \textbf{(A-21)}$$

Figure A-11 also demonstrates the effect of anomalous scattering on Friedel pairs, explained below. As defined by Friedel's law, the phases of $\mathbf{F}^{0}_{\mathrm{PH}+}$ and its mate $\mathbf{F}^{\mathbf{0}}_{\mathrm{PH}-}$ are α and $-\alpha$, respectively; likewise, the real components $\Delta\mathbf{F}'_{\mathrm{H}+}$ and $\Delta\mathbf{F}'_{\mathrm{H}-}$ are related by reflection over the real axis. The imaginary components $\Delta\mathbf{F}''_{\mathrm{H}+}$ and $\Delta\mathbf{F}''_{\mathrm{H}-}$, on the contrary, are inverted relative to one another (e.g., by reflection over the real axis and change of sign). Because of this, the pairs $\mathbf{F}^{\mathrm{anom}}_{\mathrm{PH}+}$ and $\mathbf{F}^{\mathrm{anom}}_{\mathrm{PH}-}$ do not have opposite phases like their non-anomalous counterparts. As with isomorphous replacement, we can solve the phases geometrically (**Figure A-12**). The magnitudes of $\Delta\mathbf{F}'_{\mathrm{H}}$ and $\Delta\mathbf{F}''_{\mathrm{H}}$ are approximately invariant for the element of interest, and their phases can be calculated with Patterson maps. Rearranging equation (**A-21**) and accounting for one Friedel mate:

$$\mathbf{F}^{0}_{\mathrm{PH}+} = \mathbf{F}^{\mathrm{anom}}_{\mathrm{PH}+} - \Delta\mathbf{F}'_{\mathrm{H}+} - \Delta\mathbf{F}''_{\mathrm{H}+}. \qquad (\mathbf{A\text{-}22})$$

To look at the other Friedel mate, we reflect $\mathbf{F}^{\mathrm{anom}}_{\mathrm{PH}-}$ and its components across the real axis to get

$$\mathbf{F}^{0}_{\mathrm{PH}+} = |\mathbf{F}^{\mathrm{anom}}_{\mathrm{PH}-}| - \Delta\mathbf{F}'_{\mathrm{H}+} - (-\Delta\mathbf{F}''_{\mathrm{H}+}). \qquad (\mathbf{A\text{-}23})$$

Note that reflection does not change the amplitude of $\mathbf{F}^{\mathrm{anom}}_{\mathrm{PH}-}$, only its phase. Furthermore, $\Delta\mathbf{F}''_{\mathrm{H}+}$ is now being added to the other vectors. Two of the solutions for $\mathbf{F}^{0}_{\mathrm{PH}+}$ (circle intersections in **Figure A-12**) will agree at the correct phase for the heavy-atom derivative. If variation in the Friedel pair intensities is not adequate to elucidate phases, data collection at a third wavelength may be necessary (this is

known as multiple-wavelength anomalous dispersion, or MAD, in contrast to single-wavelength anomalous dispersion, or SAD) (**Figure A-10**). Now that phases are known for $\mathbf{F}_{PH}$ and $\mathbf{F}_{H}$, equation (**A-17**) can be solved to acquire phases for $\mathbf{F}_{P}$.

In molecular replacement, a previously solved, homologous structure (search model) provides the phases and the structure of interest provides the frequencies and amplitudes in the form of a diffraction pattern (from which the space group and unit cell parameters can also be computed). The search model is rotated and then translated throughout the unit cell until it aligns with the structure of interest, bringing the observed and calculated diffraction patterns roughly into agreement. This is accomplished by superposing the peaks of the Patterson maps. Given the number of structures in the Protein Data Bank (>100,000 as of this year), molecular replacement is by far the most frequently used and often the most convenient method of phasing.

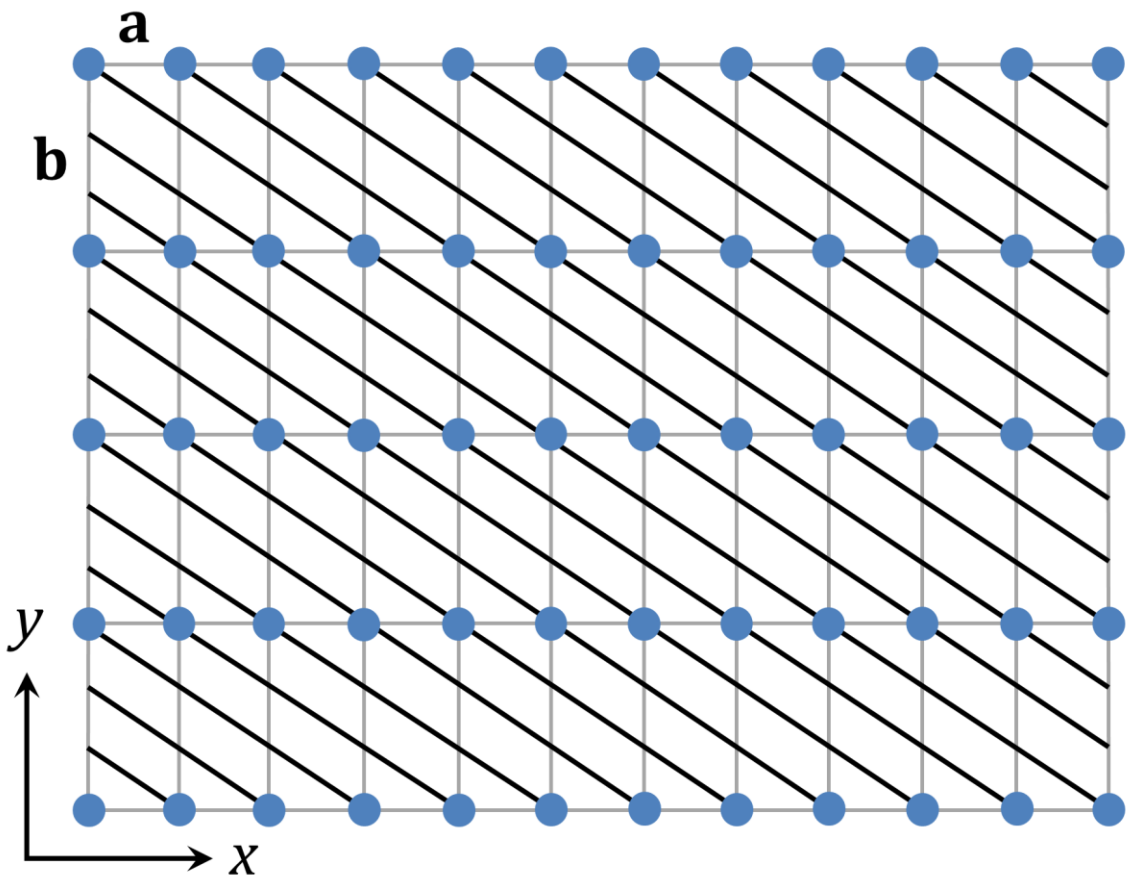


Figure A-1. (130) planes in a crystal lattice. The planes are shown as oblique black lines, lattice points as blue circles. Notice how the (130) planes intersect edges **a**, **b**, and **c** one, three, and zero times, respectively. The vertical and horizontal grey lines can be thought of as (100) planes and (010) planes, respectively.

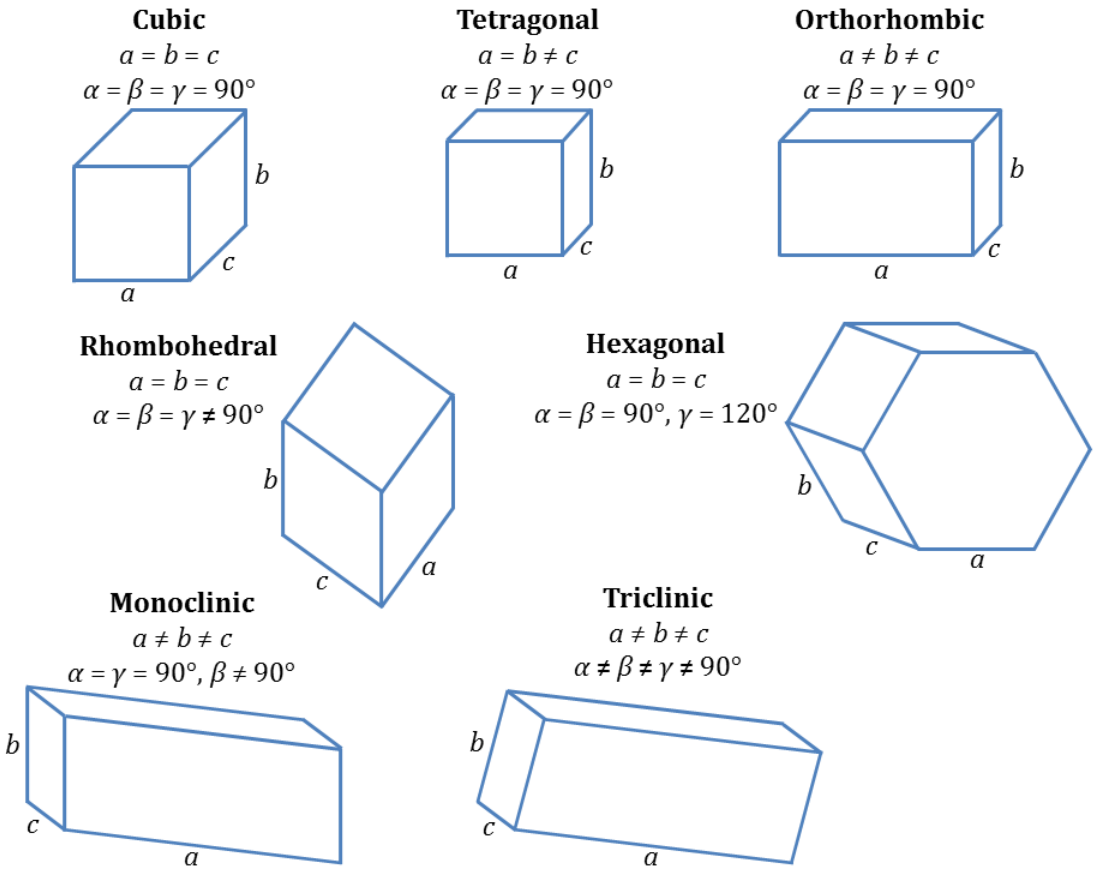


Figure A-2. The 7 possible crystal systems, with unit-cell angle and edge definitions. The systems are ordered from highest symmetry (cubic; upper left) to lowest symmetry (triclinic; lower right).

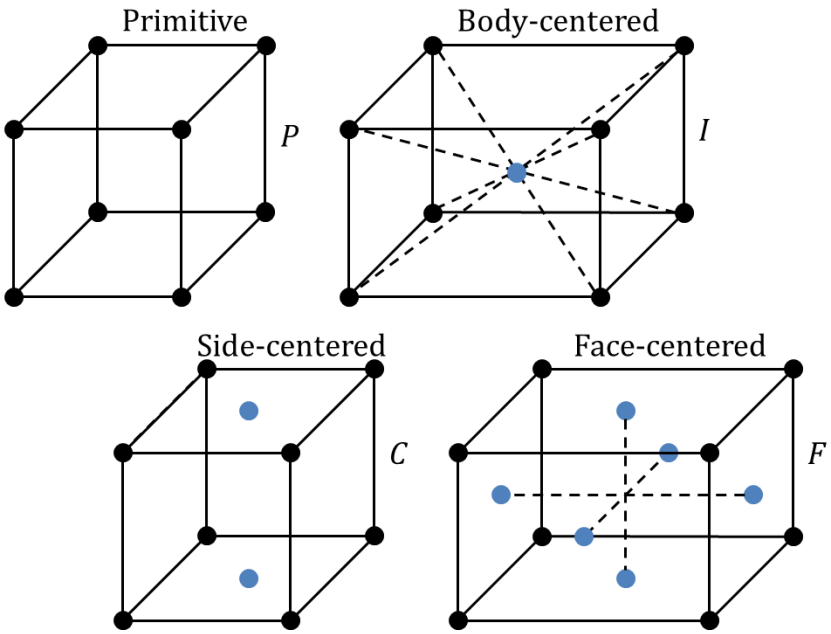


Figure A-3. The 4 lattice types: Primitive (*P*), body-centered or internal (*I*), side-centered (*C*), and face-centered (*F*). The *P* lattice type contains one lattice point per unit cell, as each point is shared with eight adjacent unit cells. Similarly, the *I*, *C*, and *P* lattice types contain two, two, and four lattice points per unit cell.

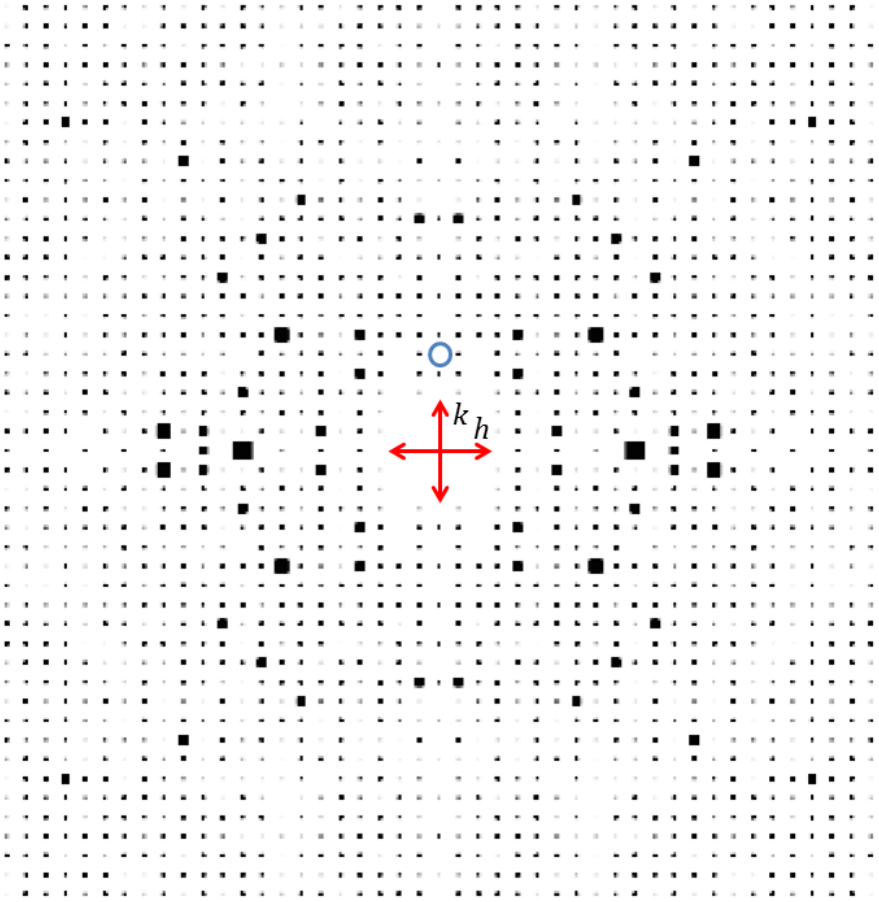


Figure A-4. Section of the zero-level plane (*hk*0) from the diffraction data of selenomethionine-derived EptC, which crystallized in space group $P2_12_12_1$ (see **Chapter 3**). All odd-numbered reflections are missing (systematically absent) along axes *h* and *k*, indicating the presence of two 2_1 screw axes. A third screw axes, along *l*, is similarly observable in the (*h*0*l*) and (0*kl*) planes (not shown). The blue circle exemplifies a systematic absence at $k = 3$.

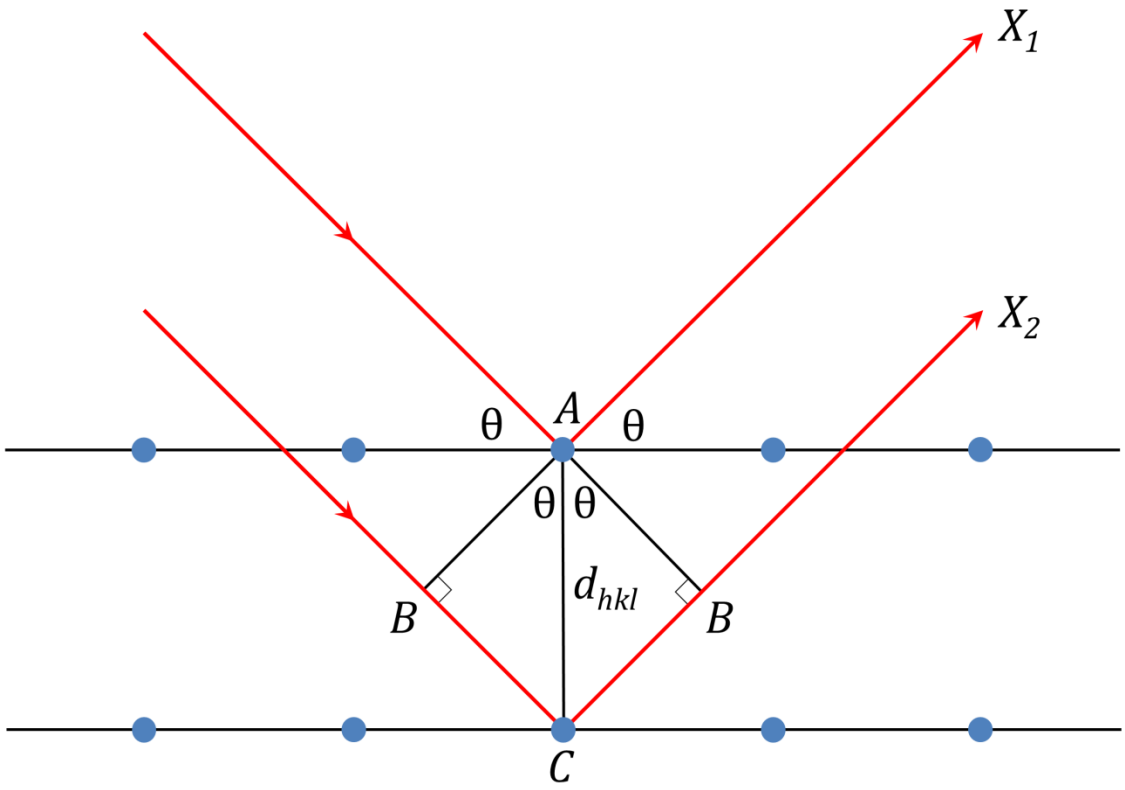


Figure A-5. Illustration of Bragg's law, $2d_{hkl}sin\theta = n\lambda$, which describes the requirements for diffraction. X_1 and X_2 are rays of wavelength λ that have impinged on parallel planes separated by distance d_{hkl}. Lattice points along the planes are indicated by blue circles.

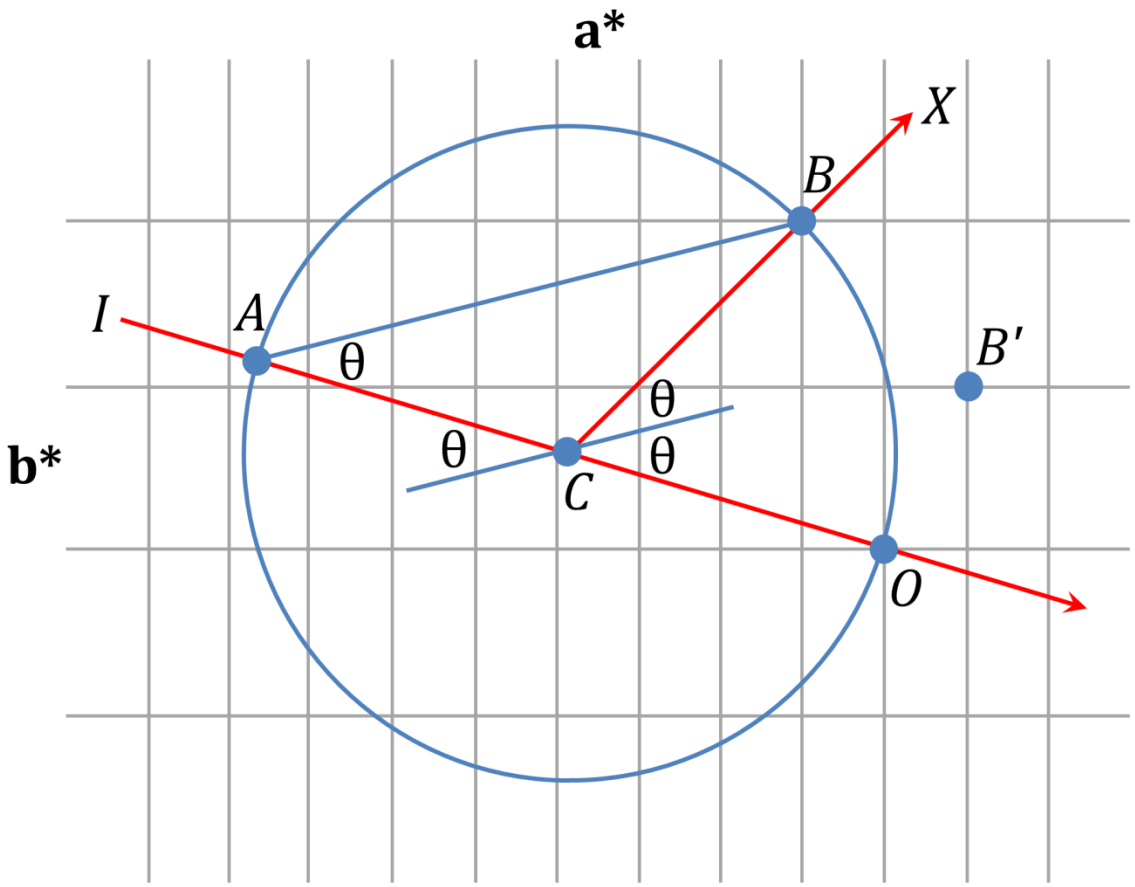


Figure A-6. Bragg's law in reciprocal space. Ray *X* is produced because reciprocal lattice point *B* is on the circle with radius $1/\lambda$. The blue line parallel to *AB* and crossing *C* is the real-space, reflective plane at this angle, as are other parallel lines spaced an integral number of d_{hkl} away. Fully rotating the circle around the X-ray beam *IO* generates the Ewald sphere.

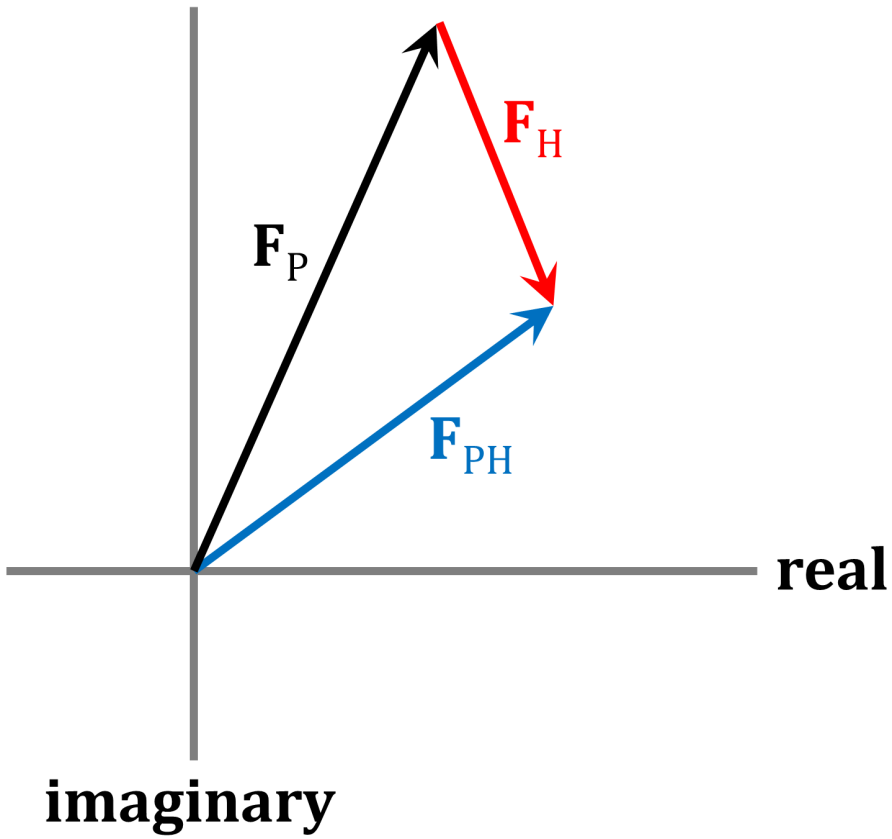


Figure A-7. Vector representation of structure factors. A solution is shown for vector equation (**A-17**), $\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$, which includes structure-factor vectors of a native protein ($\mathbf{F}_P$), heavy-atom derivative ($\mathbf{F}_{PH}$), and heavy atoms alone ($\mathbf{F}_H$). For each vector, the length indicates its amplitude and the angle formed with the real axis indicates its phase. The *x*-axis contains real numbers and the *y*-axis contains imaginary numbers.

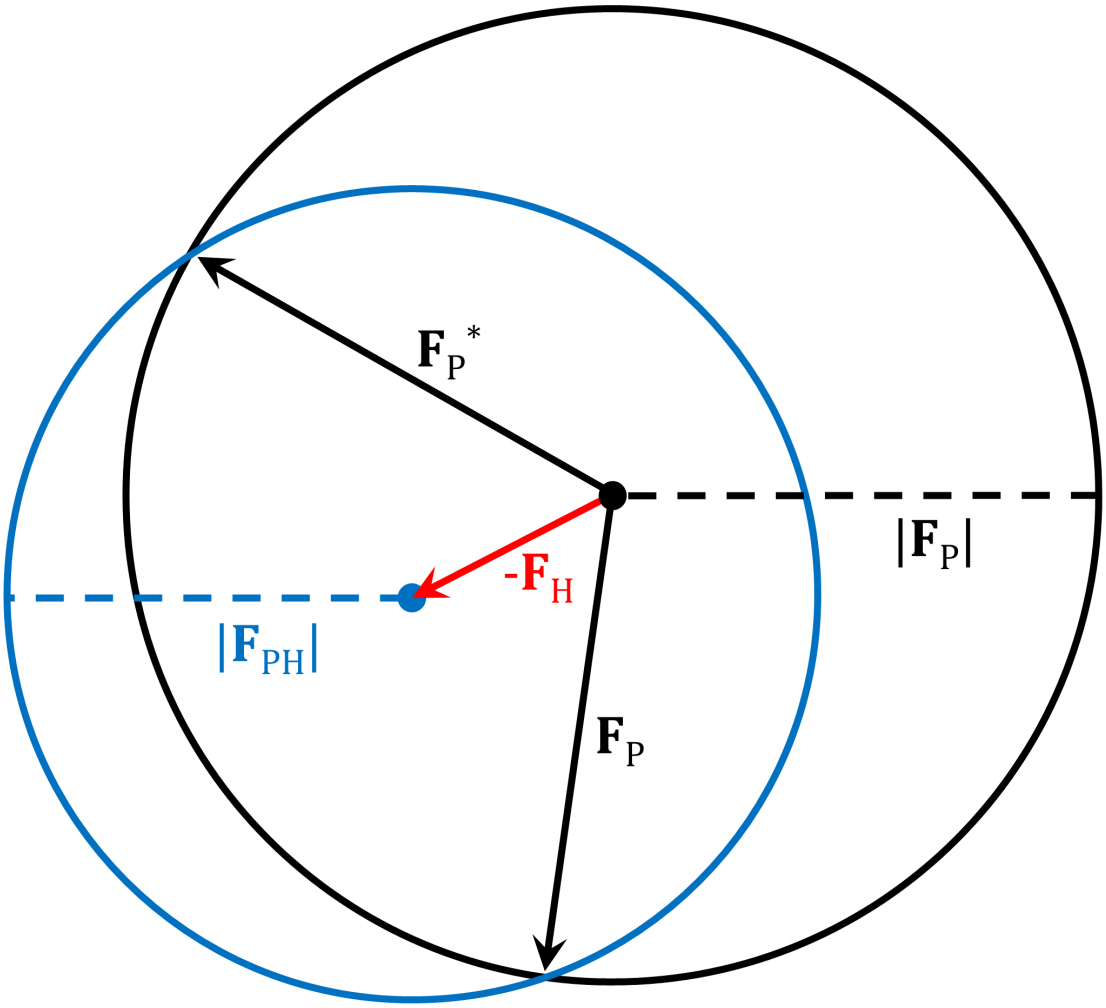


Figure A-8. Harker diagram for isomorphous replacement. The black circle with radius $|\mathbf{F}_P|$ represents a native structure factor and the blue circle with radius $|\mathbf{F}_{PH}|$ represents a heavy-atom structure factor. Vectors $\mathbf{F}_P$ and $\mathbf{F}_P^*$, which point to circle intersections, indicate two possible phase solutions: $\mathbf{F}_P = \mathbf{F}_P^* = \mathbf{F}_{PH} - \mathbf{F}_H$. Additional heavy-atom derivatives may be necessary to determine whether $\mathbf{F}_P$ or $\mathbf{F}_P^*$ represents the correct phase for the native structure factor.

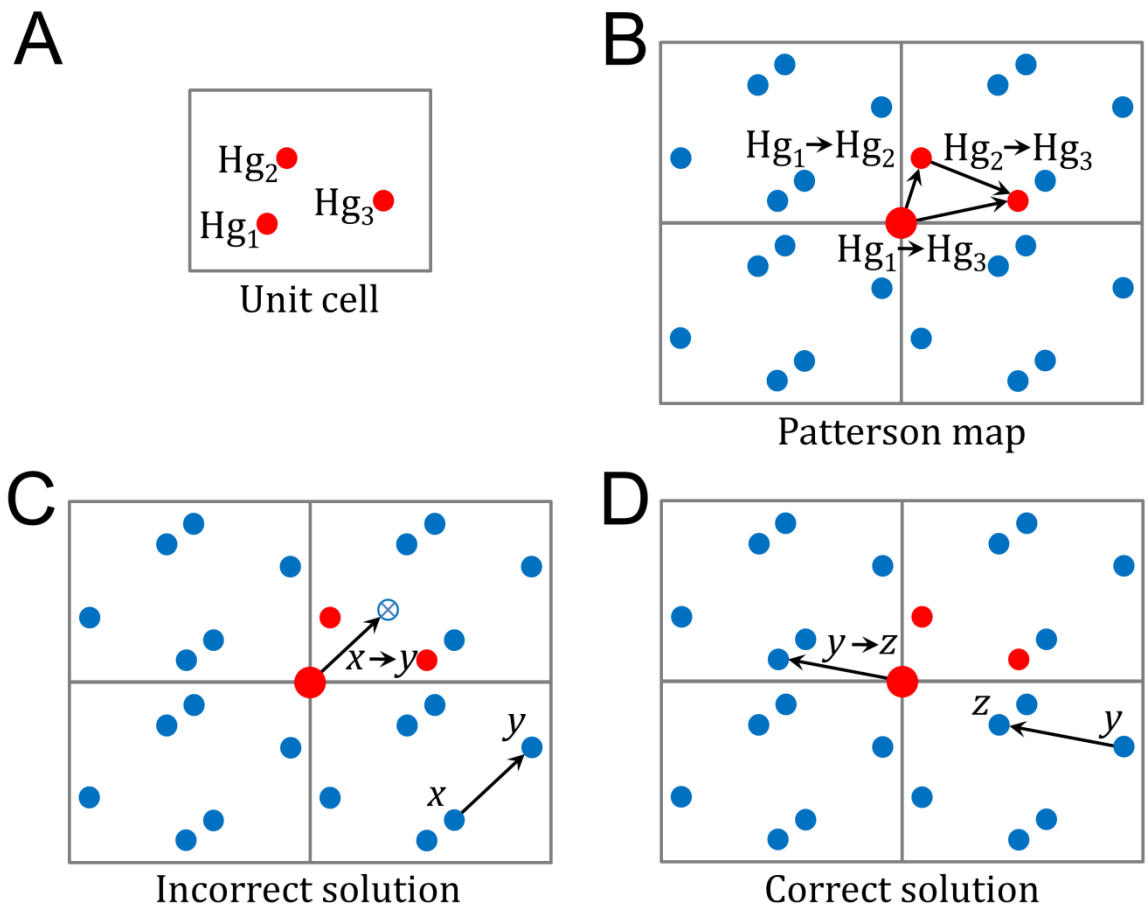


Figure A-9. Two-dimensional Patterson maps. **(A)** Unit cell comprised of 3 mercury atoms (red circles). **(B)** Patterson map of **A** for four adjacent Patterson unit cells. A Patterson unit cell contains $n(n-1)$ Patterson atoms (peaks; blue circles) for a real unit cell containing n atoms. Interatomic vectors point from one Patterson atom to another. Vectors $Hg_2 \rightarrow Hg_1$, $Hg_3 \rightarrow Hg_1$, and $Hg_3 \rightarrow Hg_2$ are not shown for clarity. The intense origin peak arises from vectors to and from the same atom. The atoms of **A** are shown as red peaks in one Patterson unit cell. **(C)** Not knowing **A**, we search for the positions of 3 atoms in the Patterson map. Drawing a vector between two Patterson atoms x and y that do not represent real atoms leads to an incorrect solution. Note how $x \rightarrow y$, when translated to the origin, does not point to a Patterson atom. **(D)** Drawing a vector between two Patterson atoms y and z that represent real atoms leads to a correct solution. Note how $y \rightarrow z$, when translated to the origin, points to a Patterson atom (this is the $Hg_1 \rightarrow Hg_3$ vector).

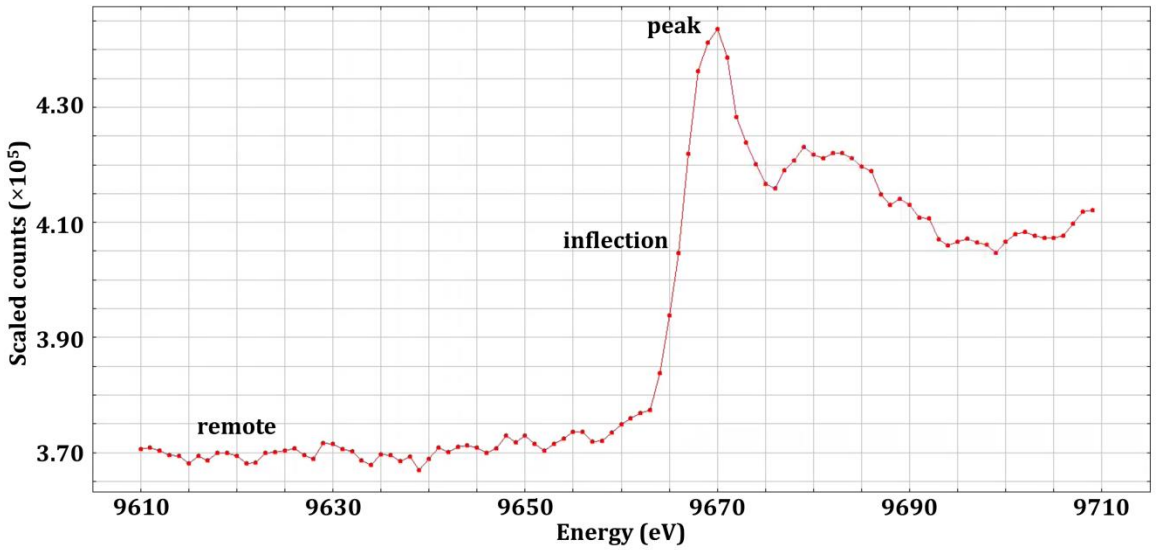


Figure A-10. K-absorption edge of zinc. To obtain this plot, which revealed the presence of zinc, a fluorescence scan was performed at Advanced Light Source beamline 8.2.2 on a cEptC crystal (**see Chapter 3**). Three common phasing wavelengths are highlighted: Peak (or edge; gives maximum value of $\Delta\mathbf{F}''_H$), inflection (half-way down the peak; gives minimum value of $\Delta\mathbf{F}'_H$), and remote. Diffraction at these wavelengths maximizes the difference between $\Delta\mathbf{F}'_H$ and $\Delta\mathbf{F}''_H$. A full 360° of diffraction are collected at each wavelength for a MAD dataset, or the peak wavelength for a SAD dataset.

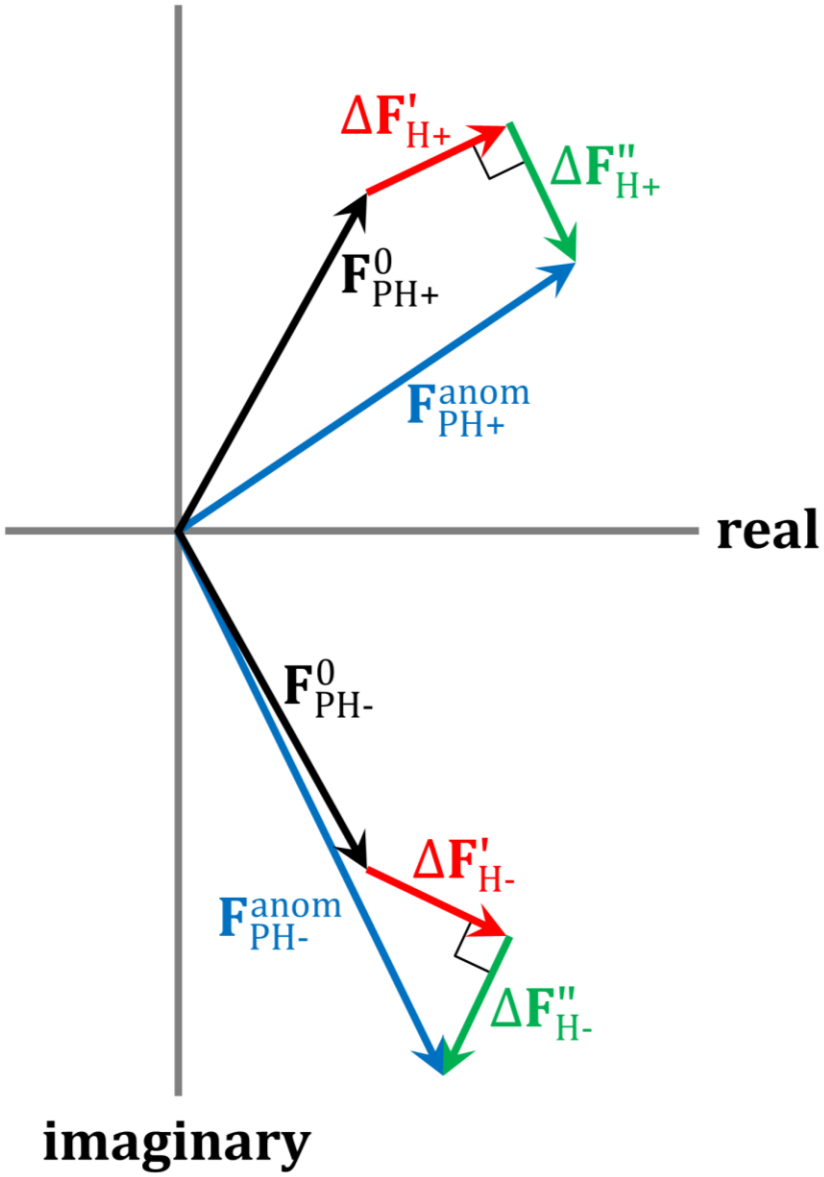


Figure A-11. Vector representation of structure factors for Friedel pairs under anomalous conditions. Equation (**A-21**), $\mathbf{F}_{PH}^{anom} = \mathbf{F}_{PH}^{0} + \Delta\mathbf{F'}_{H} + \Delta\mathbf{F''}_{H}$, describes the effect of anomalous dispersion on the structure factor $\mathbf{F}_{PH}^{0}$ to arrive at $\mathbf{F}_{PH}^{anom}$. Because imaginary component $\Delta\mathbf{F''}_{H}$ is always rotated +90° relative to real component $\Delta\mathbf{F'}_{H}$, Friedel's law is broken; that is, $|\mathbf{F}_{PH+}^{anom}| \neq |\mathbf{F}_{PH-}^{anom}|$. Friedel mates are distinguished here by + and – signs.

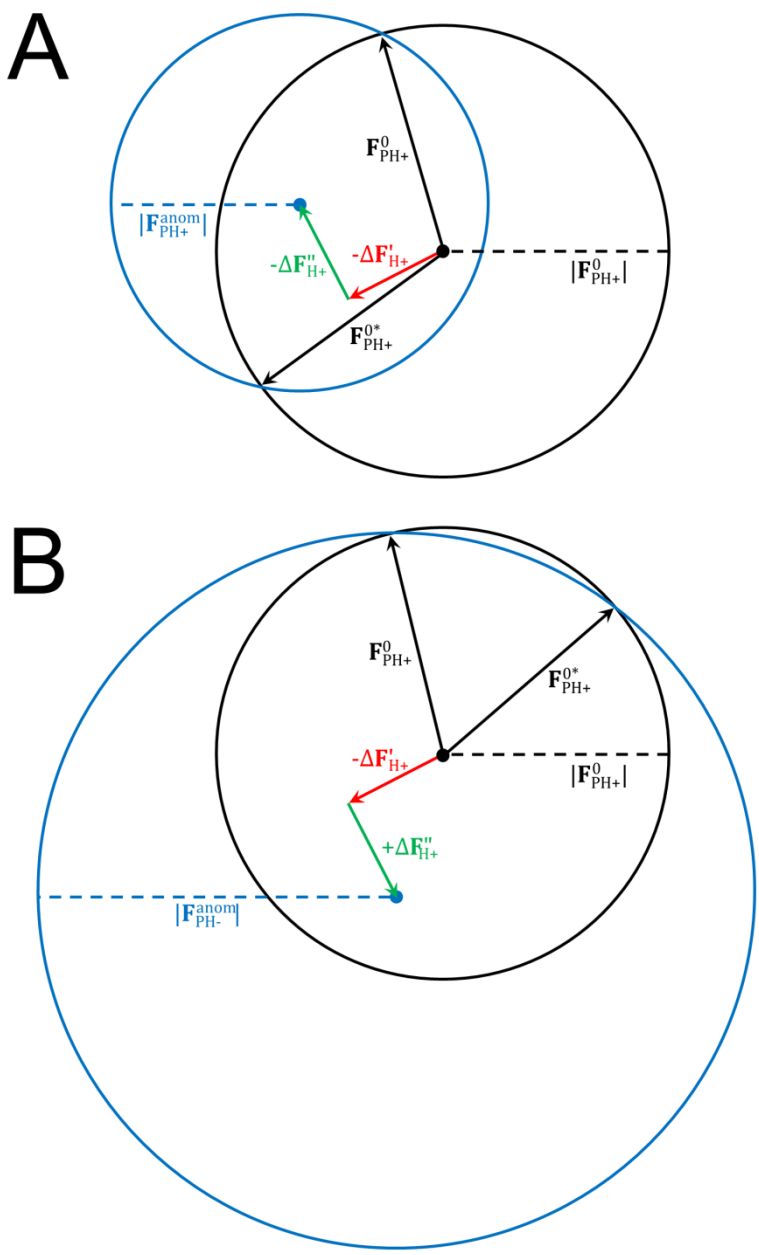


Figure A-12. Harker diagrams for anomalous dispersion of Friedel mates. **(A)** addresses the + mate **(B)** addresses the – mate. Here, we solve equation (**A-22**), $\mathbf{F}^{0}_{PH+} = \mathbf{F}^{anom}_{PH+} - \Delta\mathbf{F}'_{H+} - \Delta\mathbf{F}''_{H+}$, and equation (**A-23**), $\mathbf{F}^{0}_{PH+} = |\mathbf{F}^{anom}_{PH-}| - \Delta\mathbf{F}'_{H+} - (-\Delta\mathbf{F}''_{H+})$, for $\mathbf{F}^{0}_{PH+}$. Amplitudes $|\Delta\mathbf{F}'_{H}|$ and $|\Delta\mathbf{F}''_{H}|$ are known for the heavy atom of interest, and their phases are calculable from Patterson maps. We also know the amplitudes $|\mathbf{F}^{0}_{PH+}|$ and $|\mathbf{F}^{anom}_{PH}|$ from the diffraction patterns. $\mathbf{F}^{0}_{PH+}$ and $\mathbf{F}^{0*}_{PH+}$ represent possible solutions; one of these should agree between Friedel mates. If phases cannot be solved, an additional wavelength (e.g., MAD dataset) can be collected to provide additional information.

Table A-1. The 65 possible space groups in macromolecular crystallography. The 14 Bravais lattices are ordered according to increasing symmetry from bottom to top. %PDB gives the percentage of entries (of 83,959) in the Protein Data Bank representing each space group, as of Sept 25, 2013. (This column does not sum to 100.00% due to space group misannotations and racemate crystallography. For example, there are 15 entries in the PDB with space group *P* -1, which is not normally accessible to chiral, biological molecules.)

Bravais lattice	Space group used to identify Laue class	Space group number	Space group	%PDB
Primitive cubic	*P* 21 3	195	*P* 2 3	0.06
		198	*P* 21 3	0.53
	P 41 3 2	207	*P* 4 3 2	0.05
		208	*P* 42 3 2	0.03
		212	*P* 43 3 2	0.10
		213	*P* 41 3 2	0.14
I centered cubic	*I* 21 3	197	*I* 2 3	0.39
		199	*I* 21 3	0.27
	I 41 3 2	211	*I* 4 3 2	0.11
		214	*I* 41 3 2	0.09
F centered cubic	*F* 2 3	196	*F* 2 3	0.08
	F 41 3 2	209	*F* 4 3 2	0.21
		210	*F* 41 3 2	0.07
Primitive rhombohedral	*R* 3	146	*R* 3	1.44
	R 3 2	155	*R* 3 2	1.42
Primitive hexagonal	*P* 31	143	*P* 3	0.18
		144	*P* 31	0.56
		145	*P* 32	0.63
	P 31 1 2	149	*P* 3 1 2	0.02
		151	*P* 31 1 2	0.10
		153	*P* 32 1 2	0.12
	P 31 2 1	150	*P* 3 2 1	0.39
		152	*P* 31 2 1	3.22
		154	*P* 32 2 1	3.38
	P 61	168	*P* 6	0.26
		169	*P* 61	1.14
		170	*P* 65	1.00
		171	*P* 62	0.21
		172	*P* 64	0.20
		173	*P* 63	0.72
	P 61 2 2	177	*P* 6 2 2	0.08
		178	*P* 61 2 2	1.73
		179	*P* 65 2 2	1.27
		180	*P* 62 2 2	0.37
		181	*P* 64 2 2	0.31
		182	*P* 63 2 2	0.58
Primitive tetragonal	*P* 4	75	*P* 4	0.10
		76	*P* 41	0.75
		77	*P* 42	0.09
		78	*P* 43	0.64
	P 41 21 2	89	*P* 4 2 2	0.05
		90	*P* 4 21 2	0.41
		91	*P* 41 2 2	0.23
		92	*P* 41 21 2	3.23
		93	*P* 42 2 2	0.06
		94	*P* 42 21 2	0.62
		95	*P* 43 2 2	0.30
		96	*P* 43 21 2	3.71
I centered tetragonal	*I* 41	79	*I* 4	0.52
		80	*I* 41	0.29
	I 41 2 2	97	*I* 4 2 2	0.66
		98	*I* 41 2 2	0.60
Primitive orthorhombic	*P* 21 21 21	16	*P* 2 2 2	0.01
		17	*P* 2 2 21	0.14
		18	*P* 21 21 2	5.37
		19	*P* 21 21 21	23.07
C centered orthorhombic	*C* 2 2 21	20	*C* 2 2 21	4.86
		21	*C* 2 2 2	0.22
I centered orthorhombic	*I* 21 21 21	23	*I* 2 2 2	2.13
		24	*I* 21 21 21	0.24
F centered orthorhombic	*F* 2 2 2	22	*F* 2 2 2	0.13
Primitive monoclinic	*P* 1 21 1	3	*P* 1 2 1	0.19
		4	*P* 1 21 1	15.39
C centered monoclinic	*C* 1 2 1	5	*C* 1 2 1	9.58
Primitive triclinic	*P* 1	1	*P* 1	4.42

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst* D**66**, 213-221.

Agarwal V, Lin S, Lukk T, Nair SK, Cronan JE (2012) Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc Natl Acad Sci* **109**, 17406-17411.

Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM, Thorson JS (2002) The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173-1176.

Akey DL, Razelun JR, Tehranisa J, Sherman DH, Gerwick WH, Smith JL (2010) Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18**, 94-105.

Alberts AW & Vagelos PR (1966) Acyl carrier protein: VIII. Studies of acyl carrier protein and coenzyme A in Escherichia coli pantothenate or β-alanine auxotrophs. *J Biol Chem* **241**, 5201-5204.

Alekseyev VY, Liu CW, Cane DE, Puglisi JD, Khosla C (2007) Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci* **16**, 2093-2107.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Ames BD, Korman TP, Zhang W, Smith P, Vu T, Tang Y, Tsai SC (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *Proc Natl Acad Sci* **105**, 5349-5354.

Anderson AJ & Dawes EA (1990) Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol Rev* **54**, 450-472.

Ariyawutthiphan O, Ose T, Minami A, Shinde S, Tsuda M, Gao YG, Yao M, Oikawa H, Tanaka I (2012) Structure analysis of geranyl pyrophosphate methyltransferase and the proposed reaction mechanism of SAM-dependent C-methylation. *Acta Cryst* D**68**, 1558-1569.

Aroyan C, Dermenci A, Miller S (2009) The Rauhut-Currier reaction: a history and its synthetic application. *Tetrahedron* **65**, 4069-4087.

Austin MB & Noel JP (2003) The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* **20**, 79-110.

Batz MB, Hoffmann S, Morris JG Jr (2012) Ranking the disease burden of 14 pathogens in food sources in the United States using attribution data from outbreak investigations and expert elicitation. *J Food Protect* **75**, 1278-1291.

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

Berjanskii MV, Neal S, Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res* **34**, W63-W69.

Blow D (2010) Outline of crystallography for biologists. Oxford, New York, NY.

Bretschneider T, Heim JB, Heine D, Winkler R, Busch B, Kusebauch B, Stehle T, Zocher G, Hertweck C (2013) Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124-128.

Brignole EJ, Smith S, Asturias FJ (2009) Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat Struct Mol Biol* **16**, 190-197.

Britton D (1972) Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. *Acta Cryst* A**28**, 296-297.

Broadhurst RW, Nietlispach D, Wheatcroft MP, Leadlay PF, Weissman KJ (2003) The structure of docking domains in modular polyketide synthases. *Chem Biol* **10**, 723-731.

Brookes E, Boppana RV, Demeler B (2006) Computing large sparse multivariate optimization problems with an application in biophysics. *SC '06 Proceedings of the 2006 ACM/IEEE conference on supercomputing, 0-7695-2700-0/06*.

Brookes E & Demeler B (2006) Genetic algorithm optimization for obtaining accurate molecular weight distributions from sedimentation velocity experiments. *Progr Colloid Polym Sci* **131**, 33-40.

Brookes E & Demeler B (2007) Parsimonious regularization using genetic algorithms applied to the analysis of analytical ultracentrifugation experiments. *GECCO '07 Proceedings of the 9th annual conference on genetic and evolutionary computation, 978-1-59593-697-4/07/0007*.

Brookes E & Demeler B (2008) Parallel computational techniques for the analysis of sedimentation velocity experiments in UltraScan. *Colloid Polym Sci* **286**, 138-148.

Brookes E, Cao W, Demeler B (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *Eur Biophys J* **39**, 423-435.

Brozell SR, Mukherjee S, Balius TE, Roe DR, Case DA, Rizzo RC (2012) Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comp-Aided Mol Des* **26**, 749-773.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res* **33**, W36-W38.

Buchholz TJ, Geders TW, Bartley FE 3rd, Reynolds KA, Smith JL, Sherman DH (2009) Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem Biol* **4**, 41-52.

Bulkley D, Innis CA, Blaha G, Steitz TA (2010) Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proc Natl Acad Sci* **107**, 17158-17163.

Burmeister WP (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Cryst* D**56**, 328-341.

Chamot-Rooke J, Mikaty G, Malosse C, Soyer M, Dumont A, Gault J, Imhaus AF, Martin P, Trellet M, Clary G, Chafey P, Camoin L, Nilges M, Nassif X, Duménil G (2011) Posttranslational modification of pili upon cell contact triggers N. meningitis dissemination. *Science* **331**, 778-782.

Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst* D**66**, 12-21.

Chevance FF & Hughes KT (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat Rev Microbiol* **6**, 455-465.

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

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

Cox AD, Wright JC, Li J, Hood DW, Moxon ER, Richards JC (2003) Phosphorylation of the lipid A region of meningococcal lipopolysaccharide: identification of a family of transferases that add phosphoethanolamine to lipopolysaccharide. *J Bacteriol* **185**, 3270-3277.

Crawford JM, Korman TP, Labonte JW, Vagstad AL, Hill EA, Kamari-Bidkorpeh O, Tsai SC, Townsend CA (2009) Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **461**, 1139-1143.

Crawford JM & Townsend CA (2010) New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* **8**, 879-889.

Crosby J & Crump MP (2012) The structural role of the carrier protein–active controller or passive carrier. *Nat Prod Rep* **29**, 1111-1137.

Cullen TW & Trent MS (2010) A link between the assembly of flagella and lipooligosaccharide of the Gram-negative bacterium Campylobacter jejuni. *Proc Natl Acad Sci* **107**, 5160-5165.

Cullen TW, Madsen JA, Ivanov PL, Brodbelt JS, Trent MS (2012) Characterization of unique modification of flagellar rod protein FlgG by Campylobacter jejuni lipid A phosphoethanolamine transferase, linking bacterial locomotion and antimicrobial peptide resistance. *J Biol Chem* **287**, 3326-3362.

Cullen TW, O'Brien JP, Hendrixson DR, Giles DK, Hobb RI, Thompson SA, Brodbelt JS, Trent MS (2013) EptC of Campylobacter jejuni mediates phenotypes involved in host interactions and virulence. *Infect Immun* **81**, 430-440.

Davis JT, Moore RN, Imperiali B, Pratt AJ, Kobayashi K, Masamune S, Sinskey AJ, Walsh CT, Fukui T, Tomita K (1987) Biosynthetic thiolase from Zoogloea ramigera. I. Preliminary characterization and analysis of proton transfer reaction. *J Biol Chem* **262**, 82-89.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293.

DeLano WL. The PyMOL molecular graphics system. DeLano Scientific LLC, San Carlos, CA, USA; http://www.pymol.org.

Demeler B, Saber H, Hansen JC (1997) Identification and interpretation of complexity in sedimentation velocity boundaries. *Biophys* **72**, 397-407.

Demeler B & van Holde KE (2004) Sedimentation velocity analysis of highly heterogeneous systems. *Anal Biochem* **335**, 279-288.

Demeler B & Brookes E (2008) Monte Carlo analysis of sedimentation experiments. *Colloid Polym Sci* **286**, 129-137.

Demeler B (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. *Curr Protoc Protein Sci* **60**, 7.13.1-7.13.24.

Demeler B (2012) UltraScan III, version 1473: An integrated framework for the analysis of sedimentation velocity experiments. http://www.ultrascan.uthscsa.edu/.

Diederichs K & Karplus PA (2013) Better models by discarding data? *Acta Cryst* D**69**, 1215-1222.

Diels O & Alder K (1929) Synthesis in the hydroaromatic series, IV. Announcement: the rearrangement of malein acid anhydride on arylated diene, triene, and fulvene. *Chemische Berichte* **62**, 2081-2087.

Doublié S (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* **276**, 523-530.

Drenth J (2007) Principles of protein X-ray crystallography, 3rd edition. Springer, New York, NY.

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

Dutta S, Whicher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, Narayan AR, Håkansson K, Sherman DH, Smith JL, Skiniotis G (2014) Structure of a modular polyketide synthase. *Nature* **510**, 512-517.

Dzieciatkowska M, Brochu D, van Belkum A, Heikema AP, Yuki N, Houliston RS, Richards JC, Gilbert M, Li J (2007) Mass spectrometric analysis of intact lipooligosaccharide: direct evidence for O-acetylated sialic acids and discovery of O-linked glycine expressed by Campylobacter jejuni. *Biochemistry* **46**, 14704-14714.

Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst* D**60**, 2126-2132.

Engskog MK, Deadman M, Li J, Hood DW, Schweda EK (2011) Detailed structural features of lipopolysaccharide glycoforms in nontypeable Haemophilus influenzae strain 2019. *Carbohydr Res* **346**, 1241-1249.

Evans PR (2006) Scaling and assessment of data quality. *Acta Cryst* D**62**, 72-82.

Falagas ME, Rafailidis PI, Ioannidou E, Alexiou VG, Matthaiou DK, Karageorgopoulos DE, Kapaskelis A, Nikita D, Michalopoulos A (2010a) Colistin therapy for microbiologically documented multidrug-resistant Gram-negative bacterial infections: a retrospective cohort study of 258 patients. *Int J Antimicrob Agents* **35**, 194-199.

Falagas ME, Rafailidis PI, Matthaiou DK (2010b) Resistance to polymyxins: mechanisms, frequency and treatment options. *Drug Resist Updat* **13**, 132-138.

Fischbach MA & Walsh CT (2009) Antibiotics for emerging pathogens. *Science* **325**, 1089-1093.

French S & Wilson K (1978) On the treatment of negative intensity observations. *Acta Cryst* A**34**, 517-525.

Gangarosa EJ, Bennett JV, Boring JR (1967) Differentiation between vibrio cholerae and Vibrio cholerae biotype El Tor by the polymyxin B disc test: comparative

results with TCBS, Monsur's, Mueller-Hinton and nutrient agar media. *Bull World Health Organ* **36**, 987-990.

Geladopoulos TP, Sotiroudis TG, Evangelopoulos AE (1991) A Malachite Green Colorimetric Assay for Protein Phosphatase Activity. *Anal Biochem* **192**, 112-116.

Goudenège D, Labreuche Y, Krin E, Ansquer D, Mangenot S, Calteau A, Médigue C, Mazel D, Polz MF, Le Roux F (2013) Comparative genomics of pathogenic lineages of Vibrio nigripulchritudo identifies virulence-associated traits. *ISME J* **7**, 1985-1996.

Greene LH, Lewis TE, Addou S, Cuff A, Dallman T, Dibley M, Redfern O, Pearl F, Nambudiry R, Reid A, Sillitoe I, Yeats C, Thornton JM, Orengo CA (2007) The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res* **35**, D291-D297.

Guilhelmelli F, Vilela N, Albuquerque P, Derengowski LD, Silva-Pereira I, Kyaw CM (2013) Antibiotic development challenges: the various mechanisms of action of antimicrobial peptides and of bacterial resistance. *Front Microbiol* **4**, 1-12 353-364.

Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC (2003) The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* **42**, 2866-2873.

Haapalainen AM, Meriläinen G, Wierenga RK (2006) The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem Sci* **31**, 64-71.

Hahn DR, Gustafson G, Waldron C, Bullard B, Jackson JD, Mitchell J (2006) Butenyl-spinosyns, a natural example of genetic engineering of antibiotic biosynthetic genes. *J Ind Microbiol Biotechnol* **33**, 94-104.

Halgren T (1999) MMFF VI. MMFF94s option for energy minimization studies. *J Comput Chem* **20**, 720–729.

Hankins JV, Madsen JA, Giles DK, Childers BM, Klose KE, Brodbelt JS, Trent MS (2011) Elucidation of a novel Vibrio cholerae lipid A secondary hydroxy-acyltransferase and its role in innate immune recognition. *Mol Microbiol* **81**, 1313-1329.

Hankins JV, Madsen JA, Giles DK, Brodbelt JS, Trent MS (2012) Amino acid addition to Vibrio cholerae LPS establishes a link between surface remodeling in gram-positive and gram-negative bacteria. *Proc Natl Acad Sci* **109**, 8722-8727.

Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **4**, 17.

Havelaar AH, Ivarsson S, Löfdahl M, Nauta MJ (2013) Estimating the true incidence of campylobacteriosis and salmonellosis in the European Union, 2009. *Epidemiol Infect* **141**, 293-302.

Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24**, 162-90.

Hertweck C (2009) The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **48**, 4688-4716.

Hess BA Jr & Smentek L (2012) Concerted, highly asynchronous, enzyme-catalyzed [4 + 2] cycloaddition in the biosynthesis of spinosyn A; computational evidence. *Org Biomol Chem* **10**, 7503-7509.

Holm L & Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-W549.

Holmes, PA (1985) Applications of PHB - a microbially produced biodegradable thermoplastic. *Phys Technol* **16**, 32-36.

Huang CC, Smith CV, Glickman MS, Jacobs WR Jr, Sacchettini JC (2002) Crystal structures of mycolic acid cyclopropane synthases from Mycobacterium tuberculosis. *J Biol Chem* **277**, 11559-11569.

Hughes AJ, Detelich JF, Keatinge-Clay AT (2011a) Employing a polyketide synthase module and thioesterase in the semipreparative biocatalysis of diverse triketide pyrones. *Med Chem Comm* **3**, 956-959.

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

Hur GH, Vickery CR, Burkart MD (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* **29**, 1074-1098.

Hutchinson CR (1997) Biosynthetic studies of daunorubicin and tetracenomycin C. *Chem Rev* **97**, 2525-2535.

International Union of Crystallography (2006) International tables for crystallography, volume A: space-group symmetry, 1st online edition; Hahn T, editor. http://it.iucr.org/A/.

Isiorho E, Liu HW, Keatinge-Clay AT (2012) Structural studies of the spinosyn rhamnosyltransferase, SpnG. *Biochemistry* **51**, 1213-1222.

Isiorho E, Jeon BS, Liu HW, Keatinge-Clay AT (2014) Structural studies of the spinosyn forosaminyltransferase, SpnP. *Biochemistry* (In Press).

Israeli E, Agmon-Levin N, Blank M, Chapman J, Shoenfeld Y (2012) Guillain-Barré syndrome--a classical autoimmune disease triggered by infection or vaccination. *Clin Rev Allerg Immun* **42**, 121-130.

Jackson BJ & Kennedy EP (1983) The biosynthesis of membrane-derived oligosaccharides. A membrane-bound phosphoglycerol transferase. *J Biol Chem* **258**, 2394-2398.

Jansson A, Niemi J, Lindqvist Y, Mäntsälä P, Schneider G (2003) Crystal structure of aclacinomycin-10-hydroxylase, a S-adenosyl-L-methionine-dependent methyltransferase homolog involved in anthracycline biosynthesis in Streptomyces purpurascens. *J Mol Biol* **334**, 269-280.

Jaroszewski L, Li Z, Cai XH, Weber C, Godzik A (2005) FFAS server: novel features and applications. *Nucleic Acids Res* **33**, W284-W288.

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

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202.

Kabsch W (2010) XDS. *Acta Cryst* D**66**, 125-132.

Karplus PA & Diederichs K (2012) Linking Crystallographic Model and Data Quality. *Science* **336**, 1030-1033.

Kay LE, Xu GY, Singer AU, Muhandiram DR, Formankay JD (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. *J Magn Reson* B**101**, 333-337.

Keatinge-Clay AT & Stroud RM (2006) The Structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737-748.

Keatinge-Clay AT (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem Biol* **14**, 898-908.

Keatinge-Clay AT (2008) Crystal structure of the erythromycin polyketide synthase dehydratase. *J Mol Biol* **384**, 941-953.

Keatinge-Clay AT (2012) The structures of type I polyketide synthases. *Nat Prod Rep* **29**, 1050-1073.

Kelley LA & Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371.

Kelly WL, Boyne MT 2nd, Yeh E, Vosburg DA, Galonić DP, Kelleher NL, Walsh CT (2007) Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. *Biochemistry* **46**, 359-368.

Kelly WL (2008) Intramolecular cyclizations of polyketide biosynthesis: mining for a "Diels-Alderase"*? Org Biomol Chem* **6**, 4483-4493.

Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR (1999) Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368-1372.

Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu Rev Biochem* **76**, 195-221.

Kim EE & Wyckoff HW (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. *J Mol Biol* **218**, 449-464.

Kim HJ, Pongdee R, Wu Q, Hong L, Liu HW (2007) The biosynthesis of spinosyn in Saccharopolyspora spinosa: synthesis of the cross-bridging precursor and identification of the function of SpnJ. *J Am Chem Soc* **129**, 14582-14584.

Kim HJ (2010) Investigation of post-polyketide synthase (PKS) modifications during spinosyn A biosynthesis in Saccharopolyspora spinosa. PhD Dissertation, University of Texas at Austin.

Kim HJ, White-Phillip JA, Ogasawara Y, Shin N, Isiorho EA, Liu HW (2010) Biosynthesis of spinosyn in Saccharopolyspora spinosa: synthesis of permethylated rhamnose and characterization of the functions of SpnH, SpnI, and SpnK. *J Am Chem Soc* **132**, 2901-2903.

Kim HJ, Ruszczycky MW, Choi SH, Liu YN, Liu HW (2011) Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature* **473**, 109-112.

Kirst H (2010) The spinosyn family of insecticides: realizing the potential of natural products research. *J Antibiot* **62**, 101-111.

Klock HE & Lesley SA (2009) The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol Biol* **498**, 91-103.

Korman TP, Crawford JM, Labonte JW, Newman AG, Wong J, Townsend CA, Tsai SC (2010) Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci* **107**, 6246-6251.

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

Krivov GG, Shapovalov MV, Dunbrack RL Jr (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778-795.

Kursula P, Ojala J, Lambeir AM, Wierenga RK (2002) The catalytic cycle of biosynthetic thiolase: a conformational journey of an acetyl group through four binding modes and two oxyanion holes. *Biochemistry* **41**, 15543-15556.

Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated molecular building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-1179.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

LaRocque RC, Harris JB, Dziejman M, Li X, Khan AI, Faruque AS, Faruque SM, Nair GB, Ryan ET, Qadri F, Mekalanos JJ, Calderwood SB (2005) Transcriptional profiling of Vibrio cholerae recovered directly from patient specimens during early and late stages of human infection. *Infect Immun* **73**, 4488-4493.

Laskowski RA (2009) PDBsum new things. *Nucleic Acids Res* **37**, D355-D359.

Lee SH & Park OJ (2009) Uses and production of chiral 3-hydroxy-γ-butyrolactones and structurally related chemicals. *Appl Microbiol Biotechnol* **84**, 817-828.

Li J, Nation RL, Turnridge JD, Milne RW, Coulthard K, Rayner CR, Paterson DL (2006) Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infect Dis* **6**, 589-601.

Lillemoen J, Cameron CS, Hoffman DW (1997) The stability and dynamics of ribosomal protein L9: investigations of a molecular strut by amide proton exchange and circular dichroism. *J Mol Biol* **268**, 482-493.

Liscombe DK, Louie GV, Noel JP (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat Prod Rep* **29**, 1238-1250.

Lu D, Wörmann ME, Zhang X, Schneewind O, Gründling A, Freemont PS (2009a) Structure-based mechanism of lipoteichoic acid synthesis by Staphylococcus aureus LtaS. *Proc Natl Acad Sci* **106**, 1584-1589.

Lu J, Tappel RC, Nomura CT (2009b) Mini-review: biosynthesis of poly(hydroxyalkanoates). *J Macromolecular Science* C**49**, 226-248.

Mackinnon FG, Cox AD, Plested JS, Tang CM, Makepeace K, Coull PA, Wright JC, Chalmers R, Hood DW, Richards JC, Moxon ER (2002) Identification of a gene (*Ipt-3*) required for the addition of phosphoethanolamine to the lipopolysaccharide inner core of Neisseria meningitidis and its role in mediating susceptibility to bactericidal killing and opsonophagocytosis. *Mol Microbiol* **43**, 931-943.

Maier T, Leibundgut M, Ban N (2008) The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322.

Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Waldor MK (2011) RNA-Seq-Based Monitoring of Infection-Linked Changes in Vibrio cholerae Gene Expression. *Cell Host Microbe* **10**, 165-174.

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH (2013) CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* **41**, D348-D352.

Martin CH, Dhamankar H, Tseng HC, Sheppard MJ, Reisch CR, Prather KL (2013) A platform pathway for production of 3-hydroxyacids provides a biosynthetic route to 3-hydroxy-γ-butyrolactone. *Nat Commun* **4**, 1-9.

Martin J & McMillan F (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opinion Struct Biol* **12**, 783-793.

Masamune S, Walsh CT, Sinskey AT, Peoples OP (1989) Poly-(R)-3-hydroxybutyrate (PHB) biosynthesis: mechanistic studies on the biological Claisen condensation catalyzed by β-ketoacyl thiolase. *Pure Appl Chem* **16**, 303-312.

Masoudi A, Raetz CR, Zhou P, Pemble CW 4th (2013) Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505**, 422-426.

Mathieu M, Modis Y, Zeelen JP, Engel CK, Abagyan RA, Ahlberg A, Rasmussen B, Lamzin VS, Kunau WH, Wierenga RK (1997) The 1.8 Å crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of Saccharomyces cerevisiae: implications for substrate binding and reaction mechanism. *J Mol Biol* **273**, 714-728.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40**, 658-674.

McCoy JG, Bailey LJ, Ng YH, Bingman CA, Wrobel R, Weber AP, Fox BG, Phillips GN Jr (2009) Discovery of sarcosine dimethylglycine methyltransferase from Galdieria sulphuraria. *Proteins* **74**, 368-377.

McQuade TJ, Shallop AD, Sheoran A, Delproposto JE, Tsodikov OV, Garneau-Tsodikova S (2009) A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal Biochem* **386**, 244-250.

Meier JL & Burkart MD (2009) The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev* **38**, 2012-2045.

Meriläinen G, Schmitz W, Wierenga RK, Kursula P (2008) The sulfur atoms of the substrate CoA and the catalytic cysteine are required for a productive mode of substrate binding in bacterial biosynthetic thiolase, a thioester-dependent enzyme. *FEBS J* **275**, 6136-6148.

Meriläinen G, Poikela V, Kursula P, Wierenga RK (2009) The thiolase reaction mechanism: the importance of Asn316 and His348 for stabilizing the enolate intermediate of the Claisen condensation. *Biochemistry* **48**, 11011-11025.

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

Modis Y & Wierenga RK (1999) A biosynthetic thiolase in complex with a reaction intermediate: the crystal structure provides new insights into the catalytic mechanism. *Structure* **7**, 1279-1290.

Modis Y & Wierenga RK (2000) Crystallographic analysis of the reaction pathway of Zoogloea ramigera biosynthetic thiolase. *J Mol Biol* **297**, 1171-1182.

Molinaro A, Silipo A, Castro CD, Sturiale L, Nigro G, Garozzo D, Bernardini ML, Lanzetta R, Parrilli M (2007) Full structural characterization of Shigella flexneri M90T serotype 5 wild-type R-LPS and its ΔgalU mutant: glycine residue location in the inner core of the lipopolysaccharide. *Glycobiology* **18**, 260-269.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* **16**, 2785-2791.

Muhandiram DR & Kay LE (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J Magn Reson* B**103**, 203-216.

Müller-Loennies S, Lindner B, Brade H (2003) Structural analysis of oligosaccharides from lipopolysaccharide (LPS) of Escherichia coli K12 strain W3100 reveals a link between inner and outer core LPS biosynthesis. *J Biol Chem* **278**, 34090-34101.

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst* D**53**, 240-255.

Naessan CL, Egge-Jacobsen W, Heiniger RW, Wolfgang MC, Aas FE, Røhr A, Winther-Larsen HC, Koomey M (2008) Genetic and functional analyses of PptA, a phospho-form transferase targeting type IV pili in Neisseria gonorrhoeae. *J Bacteriol* **190**, 387-400.

Needham BD & Trent MS (2013) Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat Rev Microbiol* **11**, 467-481.

Neuhaus FC & Baddiley J (2003) A Continuum of Anionic Charge: Structures and Functions of D-Alanyl-Teichoic Acids in Gram-Positive Bacteria. *Microbiol Mol Biol Rev* **67**, 686-723.

O'Brien PJ & Herschlag D (2001) Alkaline phosphatase revisited: hydrolysis of alkyl phosphates. *Biochemistry* **40**, 5691-5699.

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

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

Padilla JE & Yeates TO (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Cryst* D**59**, 1124-1130.

Pappenberger G, Benz J, Gsell B, Hennig M, Ruf A, Stihle M, Thoma R, Rudolph MG (2010) Structure of the human fatty acid synthase KS–MAT didomain as a framework for inhibitor design. *J Mol Biol* **397**, 508-519.

Pascal SM, Muhandiram DR, Yamazaki T, Formankay JD, Kay LE (1994) Simultaneous acquisition of 15N- and 13C-edited NOE spectra of proteins dissolved in H2O. *J Magn Reson* B**103**, 197-201.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.

Piasecki SK, Taylor CA, Detelich JF, Liu J, Zheng J, Komsoukaniants A, Siegel DR, Keatinge-Clay AT (2011) Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem Biol* **18**, 1331-1340.

Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Cryst* D**59**, 1131-1137.

Pye VE, Christensen CE, Dyer JH, Arent S, Henriksen A (2010) Peroxisomal plant 3-ketoacyl-CoA thiolase structure and activity are regulated by a sensitive redox switch. *J Biol Chem* **285**, 24078-24088.

Raetz CR, Reynolds CM, Trent MS, Bishop RE (2007) Lipid A modification systems in Gram-negative bacteria. *Annu Rev Biochem* **76**, 295-329.

Rappe AK, Casewit CJ, Colwell KS, Goddard WA 3rd, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* **114**, 10024–10035.

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

Read RJ & Schierbeek AJ (1988) A phased translation function. *J Appl Cryst* **21**, 490-495.

Reichmann NT, Cassona CP, Grundling A (2013) Revised mechanism of D-alanine incorporation into cell wall polymers in Gram-positive bacteria. *Microbiology* **159**, 1868-1877.

Reynolds C, Damerell D, Jones S (2009) ProtorP: a protein-protein interaction analysis server. *Bioinformatics* **25**, 413-414.

Rhodes G (2006) Crystallography made crystal clear, 3rd edition. Elsevier, San Diego, CA.

Rix U, Fischer C, Remsing LL, Rohr J (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat Prod Rep* **49**, 542-580.

Rock CO, Cronan JE, Armitage IM (1981) Molecular properties of acyl carrier protein derivatives. *J Biol Chem* **256**, 2669-2674.

Roy H (2009) Tuning the properties of the bacterial membrane with aminoacylated phosphatidylglycerol. *IUBMB Life* **61**, 940-953.

Salehizadeh H, van Loosdrecht MC (2004) Production of polyhydroxyalkanoates by mixed culture: recent trends and biotechnological importance. *Biotechnol Adv* **22**, 261-279.

Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM (2011) Foodborne illness acquired in the United States--major pathogens. *Emerg Infect Dis* **17**, 7-15.

Scheich C, Kümmel D, Soumailakakis D, Heinemann U, Büssow K (2007) Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res* **35**, e43.

Schirner K, Marles-Wright J, Lewis RJ, Errington J (2009) Distinct and essential morphological functions for wall- and lipo-teichoic acids in Bacillus subtilis. *EMBO J* **28**, 830-842.

Scott NE, Nothaft H, Edwards AV, Labbate M, Djordjevic SP, Larsen MR, Szymanski CM, Cordwell SJ (2012) Modification of the Campylobacter jejuni N-Linked Glycan by EptC Protein-mediated Addition of Phosphoethanolamine. *J Biol Chem* **287**, 29384-29396.

Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS (2013) Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc* **135**, 12646-12651.

Sieber SA, Linne U, Hillson NJ, Roche E, Walsh CT, Marahiel MA (2002) Evidence for a monomeric structure of nonribosomal peptide synthetases. *Chem Biol* **9**, 997-1008.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539-544.

Sievers F & Higgins DG (2014) Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079**, 105-116.

Singh S, McCoy JG, Zhang C, Bingman CA, Phillips GN Jr, Thorson JS (2008) Structure and mechanism of the rebeccamycin sugar 4'-O-methyltransferase RebM. *J Biol Chem* **283**, 22628-22636.

Slater S, Houmiel KL, Tran M, Mitsky TA, Taylor NB, Padgette SR, Gruys KJ (1998) Multiple β-ketothiolases mediate poly(β-hydroxyalkanoate) copolymer synthesis in Ralstonia eutropha. *J Bacteriol* **180**, 1979-1987.

Smith S & Tsai SC (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **24**, 1041-1072.

Stec B, Holtz KM, Kantrowitz ER (2000) A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *J Mol Biol* **299**, 1303-1311.

Stevenson G, Neal B, Liu D, Hobbs M, Packer NH, Batley M, Redmond JW, Lindquist L, Reeves P (1994) Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster. *J Bacteriol* **176**, 4144-4156.

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

Tae H, Sohng JK, Park K (2009) MapsiDB: an integrated web database for type I polyketide synthases. *Bioprocess Biosyst Eng* **32**, 723-727.

Tang Y, Kim CY, Mathews II, Cane DE, Khosla C (2006) The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci* **103**, 11124-11129.

Tang Y, Chen AY, Kim CY, Cane DE, Khosla C (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem Biol* **14**, 931-943.

Thompson S, Mayerl F, Peoples OP, Masamune S, Sinskey AJ, Walsh CT (1989) Mechanistic studies on β-ketoacyl thiolase from Zoogloea ramigera: identification of the active-site nucleophile as Cys89, its mutation to Ser89, and kinetic and thermodynamic characterization of wild-type and mutant enzymes. *Biochemistry* **28**, 5735-5742.

Trauger JW & Walsh CT (2000) Heterologous expression in Escherichia coli of the first module of the nonribosomal peptide synthetase for chloroeremomycin, a vancomycin-type glycopeptide antibiotic. *Proc Natl Acad Sci* **97**, 3112-3117.

Trent MS, Stead CM, Tran AX, Hankins JV (2006) Diversity of endotoxin and its impact on pathogenesis. *J Endotoxin Res* **12**, 205-223.

Trott O & Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* **31**, 455-461.

Tsai SC, Miercke LJ, Krucinski J, Gokhale R, Chen JC, Foster PG, Cane DE, Khosla C, Stroud RM (2001) Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc Natl Acad Sci* **98**, 14808-14813.

Tsukamoto N, Chuck JA, Luo G, Kao CM, Khosla C, Cane DE (1996) 6-deoxyerythronolide B synthase 1 is specifically acylated by a diketide intermediate at the β-ketoacyl-acyl carrier protein synthase domain of module 2. *Biochemistry* **35**, 15244-15248.

Valappil SP, Boccaccini AR, Bucke C, Roy I (2007) Polyhydroxyalkanoates in Gram-positive bacteria: insights from the genera Bacillus and Streptomyces. *Antonie van Leeuwenhoek* **91**, 1-17.

Verlinden RA, Hill DJ, Kenward MA, Williams CD, Radecka I (2007) Bacterial synthesis of biodegradable polyhydroxyalkanoates. *J Appl Microbiol* **102**, 1437-1449.

Voss NR & Gerstein M (2010) 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res* **38**, W555-W562.

Waldron C, Madduri K, Crawford K, Merlo DJ, Treadway P, Broughton MC, Baltz RH (2000) A cluster of genes for the biosynthesis of spinosyns, novel macrolide insect control agents produced by Saccharopolyspora spinosa. *Antonie Van Leeuwenhoek* **78**, 385-390.

Waldron C, Matsushima P, Rosteck PR Jr, Broughton MC, Turner J, Madduri K, Crawford KP, Merlo DJ, Baltz RH (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of Saccharopolyspora spinosa. *Chem Biol* **8**, 487-499.

Wanty C, Anandan A, Piek S, Walshe J, Ganguly J, Carlson RW, Stubbs KA, Kahler CM, Vrielink A (2013) The structure of the neisserial lipooligosaccharide

phosphoethanolamine transferase A (LptA) required for resistance to polymyxin. *J Mol Biol* **425**, 3389-3402.

Weik M, Ravelli RB, Kryger G, McSweeney S, Raves ML, Harel M, Gros P, Silman I, Kroon J, Sussman JL (2000) Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc Natl Acad Sci* **97**, 623-628.

Wenzel CQ, St Michael F, Stupak J, Li J, Cox AD, Richards JC (2010) Functional characterization of Lpt3 and Lpt6, the inner-core lipooligosaccharide phosphoethanolamine transferases from Neisseria meningitidis. *J Bacteriol* **92**, 208-216.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Cryst* D**67**, 235-242.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6**, 135-140.

Wright JC, Hood DW, Randle GA, Makepeace K, Cox AD, Li J, Chalmers R, Richards JC, Moxon ER (2004) lpt6, a gene required for addition of phosphoethanolamine to inner-core lipopolysaccharide of Neisseria meningitidis and Haemophilus influenzae. *J Bacteriol* **186**, 6970-6982.

Wu R, Reger AS, Lu X, Gulick AM, Dunaway-Mariano D (2009) The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase. *Biochemistry* **48**, 4115-4125.

Yeates TO (1988) Simple statistics for intensity data from twinned specimens. *Acta Cryst* A**44**, 142-144.

Yeates TO & Fam BC (1999) Protein crystals and their evil twins. *Structure* **7**, R25-R29.

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

Yu D, Xu F, Zeng J, Zhan J (2012) Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285-295.

Yuki N & Hartung HP (2012) Guillain-Barré syndrome. *N Engl J Med* **366**, 2294-2304.

Zhang G, Zhang W, Zhang Q, Shi T, Ma L, Zhu Y, Li S, Zhang H, Zhao YL, Shi R, Zhang C (2014) Mechanistic insights into polycycle formation by reductive cyclization in ikarugamycin biosynthesis. *Angew Chem Int Ed Engl* **53**, 4840-4844.

Zheng J, Taylor CA, Piasecki SK, Keatinge-Clay AT (2010) Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **18**, 913-922.

Zheng J, Gay DC, Demeler B, White MA, Keatinge-Clay AT (2012) Divergent evolution of multimodular polyketide synthases revealed by a didomain structure. *Nat Chem Biol* **8**, 615-621.