Structure, Biochemistry, and Substrate Selectivity of the Hydroxymethylglutaryl Synthase of Polyketide β-branching

by

Finn P. Maloney

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2017

Doctoral Committee:

Professor Janet L. Smith, Chair
Professor Carol A. Fierke
Professor David H. Sherman
Professor John J.G. Tesmer

Finn P. Maloney

fpmalone@umich.edu

ORCID ID: 0000-0003-1057-9442

Acknowledgement

Many people contributed to this thesis. First and foremost, I would like to thank my mentor, Professor Janet Smith, who has consistently challenged me to be a better scientist. I would like to thank the members of my committee, including Janet, Professor Carol Fierke, Professor David Sherman, and Professor John Tesmer, who have given crucial guidance and have always pushed me to think outside of the focus of my project. All members of the Smith lab have been tremendously supportive over the years and to whom I owe much of my success. In particular I have to thank Dr. Steffen Bernard, who mentored me as a rotation student and who taught me crystallography, and Rachel Cueny, an undergraduate who contributed to the results presented here. I also acknowledge the beamline staff of GM/CA at APS, for their invaluable contribution to the diffraction experiments presented in this thesis.

I of course need to thank my family for all of their support these past five years, as well as all of my friends in Ann Arbor and from Oberlin who have helped me stay sane. I am also lucky to have had my cat, Arya, for the last three years. Though some others may find her vicious, she shows me every day that there is still good in her. I want to make a special acknowledgment of all those responsible for making Magic: the Gathering (because everyone needs a hobby), which has helped me escape the all-consuming intellectual vortex of graduate school. In the cold, dreary days of Michigan, I

want to thank Buffalo Wild-Wings and Little Bangkok for consistently bringing the heat. Finally, I want to acknowledge the artisans at the Jolly Pumpkin, Bells, and Shorts breweries, whose work has contributed immeasurably to my graduate school career.

Table of Contents

Acknowledgement ii

List of Figures vi

List of Tables vii

Abstract viii

Chapter 1. Background and introduction 1

Significance, enzymology, and engineering of polyketide synthase pathways 1
Polyketide diversification by β-branching 7
Branching ACPs and orchestration of β-branching 11
Thesis summary 14

Chapter 2. HMGS structure and interaction with ACP_D 15

Summary 15
Introduction 16
Results 21
CurD HMGS activity 21
HMGS structure 25
ACP_D/HMGS complex 25
ACP_D/HMGS interface 39
Discussion 43
Methods 48
Protein expression and purification 48
Site directed mutagenesis 49
Activity assay 51
Fluorescence anisotropy binding assay 51
Crystallization 52
Structure solution and refinement 53
Homology modeling of myxovirescin HMGS and ACP_D 54

Chapter 3. Importance of conserved structural motifs to HMGS reactivity and ACP interaction **55**

Introduction 55
Results 61
Dependence of ACP interactions on surface electrostatics 61
Importance of the active site thiol pocket to HMGS activity 64
Role of the HMGS flexible loop region 66
Discussion 69
Methods 74
Protein expression and purification 74
Site directed mutagenesis 75
Activity assay 77
FP affinity assay 78
Appendix 87

Chapter 4. HMGS selectivity for non-natural donor substrates **94**

Introduction 94
Results 98
Structure of acetyl-HMGS intermediate 98
HMGS reaction with propionyl and butyryl substrates 103
Discussion 106
Methods 108
Protein expression and purification 108
Site directed mutagenesis 109
Protein crystallization 110
Structure solution and refinement 111
HMGS activity assay 111
ACP_D acyl-transfer assay 112

Chapter 5. Conclusions and Future Directions **113**

Summary and Conclusions 113
Future Directions 115

Bibliography **123**

List of Figures

Figure 1.1 Architecture of modular PKS 3
Figure 1.2 Reactions catalyzed by PKS domains 6
Figure 1.3 Flowchart of polyketide β-branching 8
Figure 2.1 HMGS reaction 18
Figure 2.2 A. Curacin A β-branching 19
Figure 2.3 Sample HMGS activity data 23
Figure 2.4 Ramachandran analysis of the HMGS structure 27
Figure 2.5 CurD HMGS structure 28
Figure 2.6 Electron density for holo-ACP_D-HMGS 29
Figure 2.7 Alignment of HMGS and HMGCS sequences 30
Figure 2.8 HMGS interaction with the donor ACP 32
Figure 2.9 Ramachandran analysis of apo-ACP_D/HMGS complex structure 33
Figure 2.10 Ramachandran analysis of holo-ACP_D/HMGS complex structure 34
Figure 2.11 Acetylation-dependent position of Ppant 35
Figure 2.12 Alignment of ACP_D and ACP_A sequences 36
Figure 2.13 Ramachandran analysis of acetyl-ACP_D/HMGS complex structure 38
Figure 2.14 Electrostatic surface potentials and interacting surfaces for HMGS and selected ACPs 40
Figure 2.15 Affinity of ACP_D for wild type and variant HMGS 41
Figure 3.1 Structure of conserved structural elements of HMGS 57
Figure 3.2 Electrostatic comparison of HMGS and HMGCS 60
Figure 3.3 ACP competitive binding 63
Figure 3.4 Deacetylation of ACP_D by HMGS 65
Figure 3.5 Fluorescence anisotropy for ACP_D and ACP_A binding experiments 79
Figure 4.1 Polyketide β-branching with non-canonical outcomes 96
Figure 4.2 Role of Met298 in the acetyl-HMGS intermediate 99
Figure 4.3 Ramachandran analysis of $HMGS_{C114Q}$ 101
Figure 4.4 Ramachandran analysis of Holo-ACP_D/$HMGS_{C114Q}$ complex 102
Figure 4.5 HMGS activity with a propionyl donor 103
Figure 5.1 Engineering PKS by addition of β-branches 119

List of Tables

Table 2.1 HMGS activity and ACP_D affinity 22
Table 2.2 Conversion of acetoacetyl-ACP_A to HMG-ACP_A evaluated by Ppant ejection 24
Table 2.3 HMGS-dependent de-acetylation of acetyl-ACP_D 24
Table 2.4 Crystallographic data collection and refinement rtatistics 26
Table 2.5 Cloning and mutagenic primers 50
Table 3.1 Surface electrostatic substitutions 61
Table 3.2 Thiol pocket substitutions 64
Table 3.3 Flexible loop substitutions 68
Table 3.4 Mutagenic primers 75
Table 4.1 Data collection and refinement statistics 100
Table 4.2 Donor acyl-transfer 105
Table 4.3 Mutagenic primers 110

Abstract

Modular polyketide synthase (PKS) pathways generate a diverse array of pharmaceutically significant small molecule natural products, and synthetic PKS biology may facilitate pharmaceutical development, production of industrially important compounds, and discovery of chemoenzymatic reagents. Of interest to these efforts are biosynthetic schemes that install unusual functional groups, including "β-branching" enzymes that generate alkyl substituents found in some polyketides. PKS enzymes act on substrates linked to acyl carrier proteins (ACP) via a phosphopantetheine arm (Ppant), but β-branching enzymes are further selective for an intermediate linked to a specialized, branch-acceptor ACP (ACP_A). ACP-enzyme interactions are a poorly understood facet of PKS biology and selectivity of β-branching enzymes for ACP_A is essential for fidelity of the biosynthetic pathway.

A hydroxymethylglutaryl synthase (HMGS) initiates β-branching using an acetyl nucleophile that is delivered to HMGS by a distinct, branch-donor ACP (ACP_D). This thesis summarizes research into the structural basis for the distinct selectivity of HMGS for its acetyl-ACP_D and polyketide-ACP_A substrates. We solved crystal structures of HMGS and determined features that both distinguish it from its primary metabolism homolog and are involved in ACP interaction. Structures of the ACP_D/HMGS complex revealed that ACP_D recognition is dependent on electrostatic interactions and on unique

structural features of ACP_D. In these first structures of a natively bound PKS ACP/enzyme complex, we discovered that ACP_D Ppant positioning is substrate dependent and identified distinct pre- and post-reaction positions of the Ppant. Furthermore, differences in the ACP_A and ACP_D interactions with HMGS apparently result in different Ppant positions that we selectively disrupted with active site substitutions. Finally, we demonstrated that HMGS is reactive with a non-natural donor-substrate, which has promising implications for the use of HMGS in synthetic biology and chemoenzymatic applications.

Chapter 1. Background and introduction

Significance, enzymology, and engineering of polyketide synthase pathways

From the use of herbal remedies to modern pharmaceuticals, natural products have been at the center of medicine since antiquity. Natural products and their derivatives make up half of approved pharmaceuticals and encompass diverse chemical scaffolds that have found use in virtually every category of therapeutic[1]. These bioactive small molecules often have complex structures, containing several stereocenters and/or unusual functional groups. While natural products can be difficult to produce synthetically, in many cases they can be cheaply produced by fermentation or semi-synthetic means. Due to this difficulty of total synthesis, much effort has been invested in engineering natural product biosynthetic pathways to produce altered molecules that may have improved pharmaceutical properties, however, such engineered pathways are often marred by diminished throughput. By understanding enzymology of the biosynthetic pathway one may facilitate engineering of processive pathways.

Fungi, plants, and bacteria, all contain polyketide synthases (PKS), which have been a major focus of engineering efforts due to their capacity to assemble acyl-CoA building blocks into complex molecules[2,3]. Modular polyketide synthases are a PKS subclass that form enzyme assembly lines[4–6], where each module is a set of fused

enzymes that extend the nascent polyketide by two carbons and perform chemical modifications after each building block is incorporated (Figure 1.1). Different modifications can be made at each point in the pathway, depending on the modification domains present in the corresponding module. An acyl-carrier protein domain (ACP) tethers intermediates via a post-translationally added phosphopantetheine arm (Ppant), shuttles them between different catalytic domains of a module, and transfers the finished product to the next module of the pathway.

The catalytic sequence of each PKS module is loading, extension, modification, and transfer to the next module[4] (coinciding with that module's extension). A summary of these reactions and the intermediates they generate is provided in Figure 1.2. The ACP is loaded by transthioesterification of an acyl group from an acyl-CoA to the Ppant arm, which is catalyzed by an acyltransferase domain (AT). ATs may occur as a domain within each module (*cis*-AT pathways[5]), or as a standalone enzyme that acts in *trans* on each ACP within the pathway (*trans*-AT pathways[6]). During extension, the ketosynthase domain (KS) accepts the nascent polyketide from the ACP of the previous module (upstream ACP) to a conserved cysteine in the active site. The ketosynthase then decarboxylates the malonyl or methylmalonyl-loaded ACP from within the module during a Claisen condensation that results in a β-keto intermediate tethered to the downstream ACP. The β-keto intermediate is either subject to α-methylation and/or reduction, or to β-branching. At the end of the pathway, the polyketide is released from the final ACP by a thioesterase (TE) domain. Many TEs catalyze cyclization of the polyketide to hydrolyze it from the Ppant. Finally, modifications are often made to the polyketide after

offloading, such as the glycosylations that are present in the biosynthesis of macrolide antibiotics.

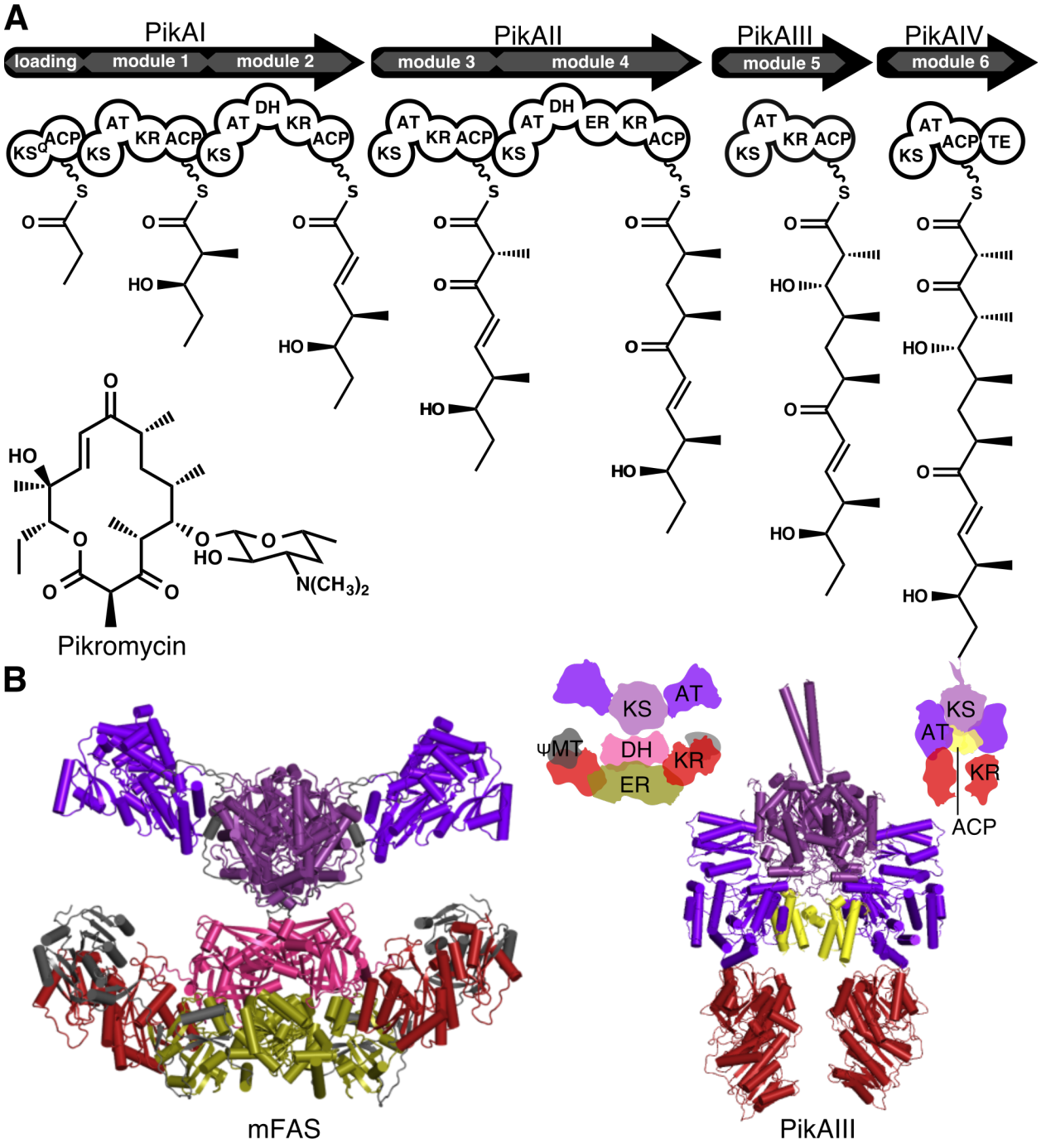


Figure 1.1 Architecture of modular PKS. A. The PKS pathway that generates pikromycin, a macrolide antibiotic, consists of 7 modules outlined in black. The intermediate generated by each module is shown tethered to that module's ACP, and

the final product is generated from cyclization by the TE and glycosylation. **B**. mFAS[7] (2VZ8), for a long time, was a model for the architecture of domains with in a module. A cryo EM structure of PikAIII[8] revealed that the module adopts an an arch-like arrangement and that the ACP (yellow) repositions depending on the intermediate that it tethers. A recent model of a mycocerosic acid PKS module[9] suggests a FAS like architecture for fully reducing modules (containing the full KR-DH-ER reductive sequence). See Figure 1.2 for reactions catalyzed by each enzymatic domain.

While there are structures for many of these domains commonly found in PKS modules[10], information on substrate selectivity and the architecture of the module is still limited. For a long time, the best model for the domain architecture of a PKS module was fatty-acid synthase (FAS, Figure 1.1), which was supported by crystal structures of KS-AT didomains[11–13]. The cryo-EM structure of PikAIII (Figure 1.1) turned this paradigm on its head, and captured a new, arch-like arrangement for a KS-AT-KR-ACP module[8,14]. Crucially, these structures also revealed a new entrance to the KS active site to which the intramodular ACP can dock for the extension reaction. This structure may not be the end of the story though, as indicated by a recent study of a mycocerosic acid synthase-like PKS module (MAS), which contained a full DH-ER-KR reductive suite, unlike PikAIII. Independent structures of the KS-AT didomain and DH-ER-KR tridomain both revealed FAS-like arrangements for each multidomain[9].

Currently, prospects for engineering new PKS pathways with the goal of creating custom molecules or libraries of molecules are grim, though success has been found with more limited goals[2,3,15]. Conservative changes, such as point mutations to individual domains or domain swaps, have been the most successful[2]. Combinatorial approaches sometimes result in compatible sequences of a few modules[16], but the rules determining why some combinations of modules have throughput and others don't are

still unknown. Finally, some studies aimed at engineering PKS dimodules to produce biofuels and molecules of industrial interest have had success as a proof of principle[17]. An additional concern for synthetic biology is that the natural producers of these molecules are often unculturable, necessitating the use of a heterologous host in which to express the pathway.

Some pathways include β-alkylations, referred to a "β-branches", where the β-keto intermediate produced by a module with no reductive modification domains may be substituted with a β-alkyl, most commonly a methyl[18]. β-branching is catalyzed by a set of standalone enzymes that are almost always encoded by a β-branching cassette that follows the gene for the module that accepts the branch. For virtually every step of the PKS catalytic cycle, a variation or non-canonical activity can be found in nature, which increases the vast chemical space accessible by PKS pathways. KR domains can generate either stereochemistry of a β-hydroxyl and DH domains can create cis, trans, and even β-γ double bonds, for example. β-branching may be the most versatile of these modification schemes, as changing the enzymes encoded by the β-branching cassette can lead to incorporation of exotic and seldom-seen functional groups into the polyketide (Figure 1.3).

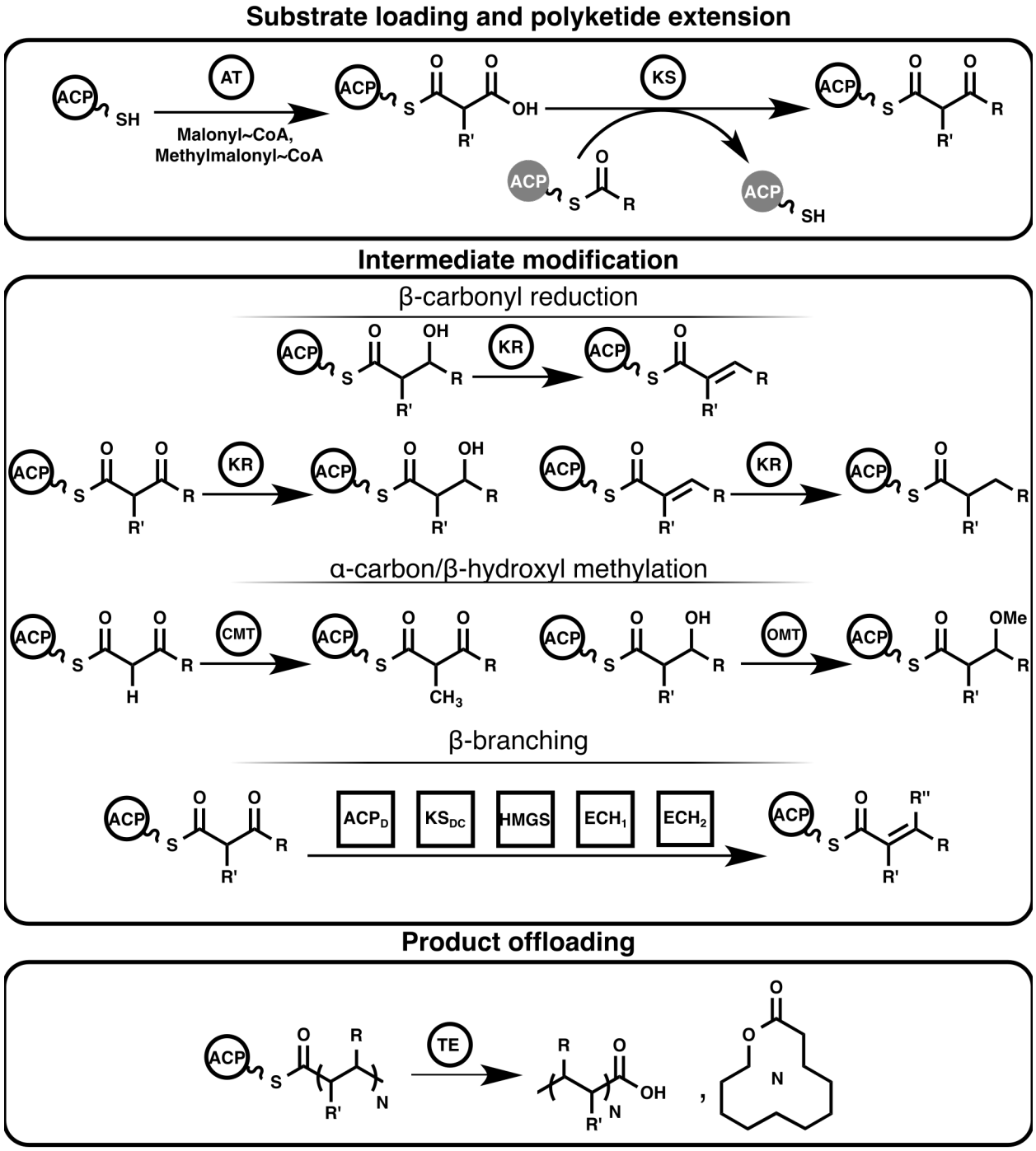


Figure 1.2 Reactions catalyzed by PKS domains. Domains within the same module are depicted as unfilled circles, while the upstream ACP is shown in gray. β-branching enzymes, which act *in trans* are shown as squares. R corresponds to the remainder of the polyketide intermediate, R' to the α-methyl or hydrogen (depending on the use of methylmalonyl- or malonyl-CoA as the extender), and R'' to a β-branch.

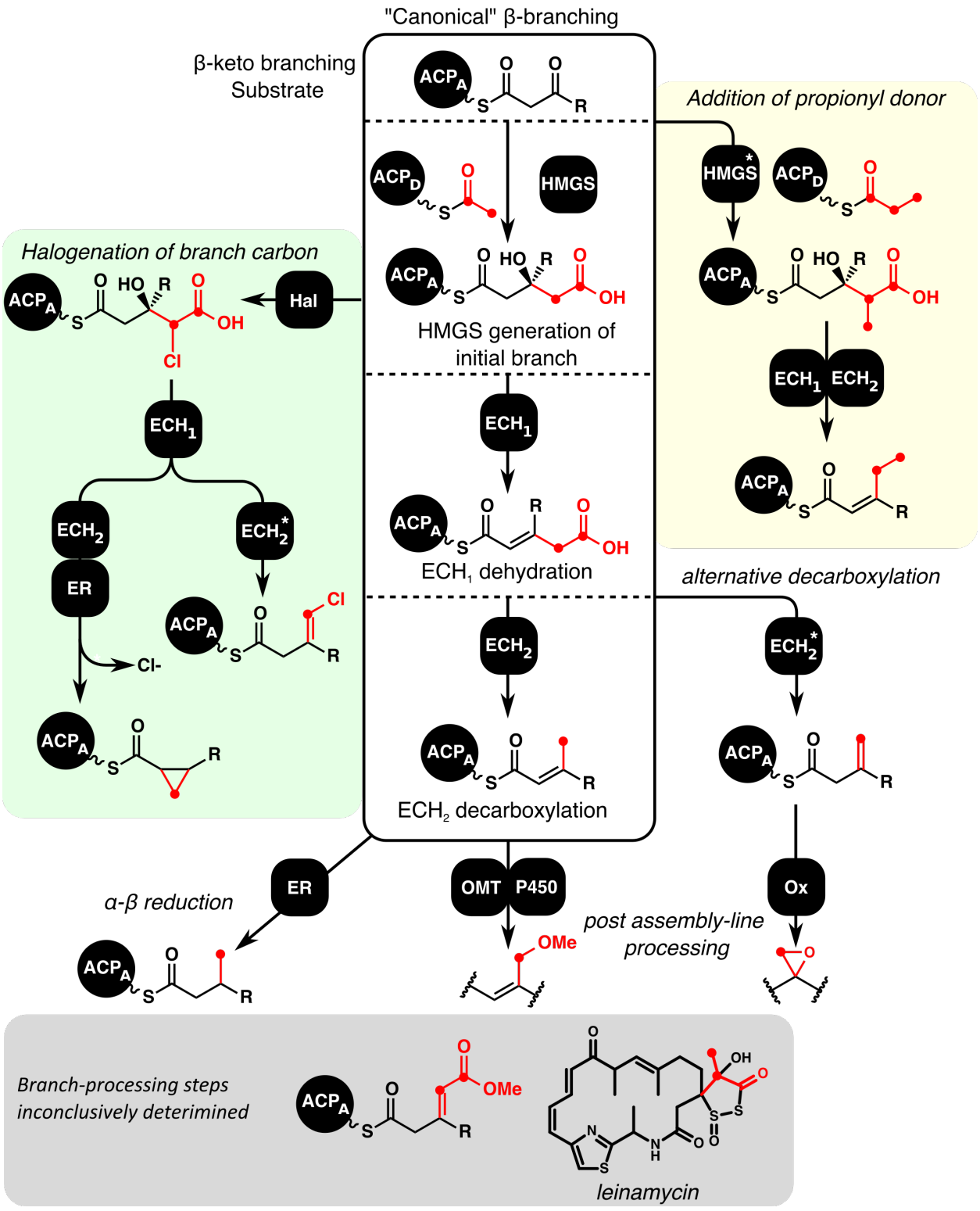


Figure 1.3 Flowchart of polyketide β-branching. The canonical enzymatic sequence that generates a methyl branch is outlined. Branch atoms are shown in red. Alternative branch processing steps are labeled in italics. Alternative branch processing involving halogenation, as occurs in curacin A and jamaicamide branching[19] is boxed in green.

Use of a propionyl donor by HMGS, as occurs in myxovirescin[20] and leinamycin[21] branching is boxed in yellow. The bryostatin and leinamycin β-branches (gray box) are highly unusual and the processing steps that generate them have not been conclusively determined.

Polyketide diversification by β-branching

Just as each module follows the pattern of a conserved extension step followed by variable extension steps to produce the different functional groups that decorate a polyketide natural product, β-branching proceeds by a conserved branch initiation step, followed by variable branch modification steps that tailor the final structure of the β-alkyl substituent[18] (Figure 1.3). The branch carbon originates from an acetyl group, or, rarely, a propionyl, that is tethered by a specialized donor ACP (ACP_D) encoded in the β-branching cassette. The ACP that tethers the polyketide intermediate subject to branching is referred to as the acceptor ACP (ACP_A). Often, β-branching will happen at only one point in the biosynthetic pathway, though some pathways have as many as five branches[22,23], and, very rarely, such pathways may install different branches. A common, hydroxymethylglutaryl-like (HMG) intermediate is the starting point for the branch modification enzymes, which is formed from the activity of a hydroxymethylglutaryl synthase (HMGS)[24,20,25] (Figure 1.3). The following is a discussion of the reactions catalyzed by each enzyme associated with β-branching and how they contribute to the formation of diverse branches.

i. pre-branching enzymes. The acetyl branch unit is generated by decarboxylation of malonyl-ACP_D by a decarboxylating ketosynthase (KS_{DC})[18,24,20]. KS_{DC} sequences are highly divergent compared to non-branching KSs, but they all have a conserved serine

substitution for the KS catalytic cysteine. The KS_{DC} is often encoded with HMGS in a polycistronic gene[26–28]. In the leinamycin pathway, which generates an ethyl branch, a malonyl-CoA decarboxylase-like enzyme replaces KS_{DC} and decarboxylates a methymalonyl-ACP_D, producing a propionyl-ACP_D substrate for HMGS[29,21]. Finally, the question of how ACP_D is loaded with malonyl or methylmalonyl is open in most cases. For some *trans*-AT pathways, the AT has been shown to be capable of loading ACP_D[24,20]. An embedded AT is often proposed to load ACP_D in *cis*-AT pathways, but this hypothesis is untested.

ii. Hydroxymethylglutaryl synthases. HMGS generates the initial branch by the aldol addition of an acetyl donor to a β-keto polyketide acceptor, forming an HMG-like product (Figure 1.3). It is homologous to HMG-CoA synthase (HMGCS), a primary metabolism enzyme of the mevalonate isoprenoid biosynthesis pathway[30], which catalyzes the same reaction as HMGS, but with CoA-tethered substrates[31]. ACP_D delivers the acetyl branch-unit to a cysteine in the HMGS active site, which is subsequently converted to an enolate. After HMGS interacts with the acceptor ACP (ACP_A), the enolate attacks the β-carbon of the β-keto polyketide acceptor, forming the carboxymethyl branch. In rare cases, a specialized HMGS may use a propionyl donor-acyl, resulting in an ethyl branch[20,29,21].

iii. ECH_1 dehydratases. β-branching cassettes often encode two enoyl-CoA hydratase-like enzymes that dehydrate and decarboxylate the HMG-like initial branch[32]. ECH_1 dehydrates the β-carbon to generate an α-β unsaturated 3-R-glutaconyl intermediate. β-branching during bryostatin biosynthesis creates unusual O-methyl-

carboxyvinyl branches that have "β-γ" double bonds, but it is unclear whether these result from a non-canonical activity of the bryostatin ECH_1 or from the activity of another processing enzyme[25,33].

iv. ECH_2 decarboxylases. The second ECH decarboxylates the γ-carbon of the ECH_1 product[32] (this corresponds to the carboxylate of the acetyl added by HMGS) to generate a 3-methylcrotonyl-like intermediate. This is often the final step of β-branching, after which the branched intermediate is delivered to the KS of the next module in the pathway. ECH_2 is proposed to decarboxylate its 3-R-glutaconyl substrate first, resulting in an enolate intermediate that is stabilized by an oxyanion hole[34]. Canonically, the enolate intermediate collapses and is reprotonated at the γ-position, producing and α-β unsaturated product. In several cases[18], ECH_2 instead reprotonates at the α-position to generate a vinyl branch. Unfortunately crotonase enzymes like ECH_2 have notoriously poor active site conservation and the identity of the catalytic acid for these non-canonical ECH_2s is unclear[34,35].

v. Enoylreductases. In some pathways, the methylcrotonyl product of ECH_2 is further reduced by an enoylreductase domain. While most ERs simply saturate the α,β positions, the curacin A pathway encodes an unusual ER that executes a cyclopropanation reaction by eliminating a chloride added to the branch carbon by a halogenase[19]. The structure of the curacin A ER is known and structural motifs in the ER active site were identified that are specific to cyclopropanase activity[36].

vi. Halogenases. The curacin A halogenase chlorinates the branch carbon to allow the cyclopropanation reaction, but the jamaicamide pathway also uses a β-

branching halogenase to create a vinyl-chloride branch[19,37,38]. In both cases, the halogenase enzyme is a domain in the same module as ACP_A. The pathways diverge at the decarboxylation step and the jamaicamide ECH_2 generates a non-canonical β-γ double bond as the final step of branching. Interestingly the jamaicamide ER is active and able to reduce an α-β unsaturated substrate, but cannot reduce the β-γ unsaturated vinyl-chloride branch[19].

vii. Other modification enzymes. The final step of bryostatin branching is catalyzed by an O-methyltransferase, which methylates the carboxylates of both O-methyl-carboxyvinyl branches[33]. A few pathways encode enzymes that tailor the final β-branch structure after the polyketide is offloaded from the assembly line. In myxovirescin biosynthesis, a P450 hydroxylates the methyl branch and that hydroxyl is then methylated by and O-MT[20,39]. Finally, several pathways[40–42] install unusual epoxide branches. This requires a non-canonical ECH_2 to generate vinyl branches, which are then converted to epoxides by a diooxygenase[41].

Branching ACPs and orchestration of β-branching

β-branching requires several enzymes to act *in trans* on an ACP-tethered substrate, whereas most other PKS enzymes act *in cis* on the ACP fused within their respective module. Handoff of the polyketide intermediate from an ACP to the KS of the next module sometimes requires the interaction of those two domains *in trans*, but this is facilitated by docking domains appended to the ACP and KS that enhance affinity and ensure pairing of the ACP with the correct KS[12,43,44]. Docking domains are not present

on the *trans*-acting enzymes to efficiently target them to ACP_A, implying a different mechanism for how β-branching is targeted to a specific module in the assembly line.

ACP_A and ACP_D are specialized for their roles in β-branching. ACP_D sequences clade separately from other PKS ACPs, and ACP_D is a standalone domain. ACP_As have a conserved tryptophan in the hydrophobic core that also allows them to be distinguished from other ACPs by sequence[45]. The tryptophan occurs 6 residues after the Ppant serine and packs between helix II and helix III in the ACP. These helices define a surface commonly associated with enzyme interaction[46,47]. ACP_As also often occur as tandem di- or tri-domains, whereas non-branching ACPs are always single domains. Tandem ACP di- and tri-domains enhance the efficiency of β-branching *in vitro* and *in vivo*[48,49]. Inactivation of one or more ACPs in a multidomain protein decreases catalytic throughput, but the presence of the dead ACPs increases throughput relative to a single ACP[49]. Thus, these multidomains may create a high effective concentration to promote the successive reactions of standalone β-branching enzymes on the ACP_A-tethered substrate. Whereas ACP_A is specialized in a way that allows it to interact with both the β-branching enzymes as well as KS and AT domains, ACP_D needs to be specialized such that it interacts only with the β-branching KS_{DC}, HMGS, and whatever domain primes the Ppant arm with a malonyl substrate (or methylmalonyl).

HMGS plays a critical role as the initiating enzyme of β-branching that establishes the chemical platform upon which the other β-branching enzymes act. HMGS is a homolog of KS domains, and similarly it must selectively and sequentially

interact with distinct ACPs. For the KS, selectivity is mediated by discrete active site entrances and docking domains, which localize the upstream ACP to its cognate entrance. HMGS likely achieves its high ACP selectivity exclusively by direct protein-protein interaction with ACP_D and ACP_A. Swapping the donor and acceptor substrates between ACP_D and ACP_A results in diminished or abolished HMGS activity, and HMGS is non-reactive when a non-branching associated ACP from the same biosynthetic pathway delivers the acceptor substrate[24,20,25]. Taken together, these data are strongly suggestive that distinguishing structural motifs of ACP_D and ACP_A enable them to make protein-protein interactions with HMGS that are adapted to their respective roles in the HMGS mechanism. It is unknown whether HMGS, like the KS, has distinct active site entrances to accomplish this.

Due to the extreme versatility of β-branching, understanding the rules that allow it to proceed efficiently should naturally be a goal of synthetic biology. As with PKS modules, individual enzymes and their mechanisms are generally well understood, but architectural details are sparse. In particular, little is known about enzyme-ACP interactions, which are even more crucial for standalone β-branching enzymes than module enzymes that are fused to the ACP. Sequence motifs that acceptor ACPs have been identified but it is not understood how these allow ACP_A, but not other ACPs, to interact with β-branching enzymes. Even more puzzling is HMGS, which needs to interact with the unusual donor ACP and the acceptor ACP, but not the ACP of any other module.

Thesis summary

This work focuses on how structural elements of HMGS, ACP_D, and ACP_A influence substrate selectivity during β-branching initiation. Chapter 2 describes the structure of the curacin A HMGS and ACP_D, and how the two proteins interact both pre- and post-acetyl transfer, as adapted from Maloney et. al. *PNAS.* 2016. Next, chapter 3 describes the role that residues which were observed to be important to the HMGS-ACP_D interaction play in the ACP_A interaction. Conserved residues that surround the HMGS active site, and were not observed to interact with ACP_D, are also studied here to evaluate whether they may be important for ACP_A docking to HMGS. Finally, chapter 4 covers donor-acyl selectivity by HMGS. Several substitutions based on HMGSs that use a propionyl donor were made to the curacin A HMGS to determine whether they confer greater donor-acyl tolerance to the enzyme.

Chapter 2. HMGS structure and interaction with ACP_D

This chapter is published as: Maloney FP, Gerwick L, Gerwick WH, Sherman DH, and Smith JL. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *PNAS*. **113.** 10316-10321 (2016)

Summary

Alkyl branching at the β position of a polyketide intermediate is an important variation on canonical polyketide natural product biosynthesis. The branching enzyme, 3-hydroxy-3-methylglutaryl synthase (HMGS), catalyzes the aldol addition of an acyl donor to a β-keto-polyketide intermediate acceptor. HMGS is highly selective for two specialized acyl carrier proteins (ACP) that deliver the donor and acceptor substrates. The HMGS from the curacin A biosynthetic pathway (CurD) was examined to establish the basis for ACP selectivity. The donor ACP (CurB) had high affinity for the enzyme (K_d 0.5 μM) and could not be substituted by the acceptor ACP. High-resolution crystal structures of HMGS alone and in complex with its donor ACP reveal a tight interaction that depends on exquisite surface shape and charge complementarity between the proteins. Selectivity is explained by HMGS binding to an unusual surface cleft on the donor ACP, in a manner that would exclude the acceptor ACP. Within the active site, HMGS discriminates between pre- and post-reaction states of the donor ACP. The free phosphopantetheine (Ppant) cofactor of ACP occupies a conserved pocket that excludes the acetyl-Ppant substrate. In comparison to HMG-coenzyme A (CoA) synthase, the homologous enzyme from primary metabolism, HMGS has several

differences at the active site entrance, including a flexible-loop insertion, which may account for the specificity of one enzyme for substrates delivered by ACP and the other by CoA.

Introduction

Polyketides are a large and chemically diverse group of natural products that includes many pharmaceuticals with a broad range of biological activities and applications as antibiotics, antifungals, anti-inflammatory drugs, and cancer chemotherapeutic agents[50]. Polyketide synthase (PKS) biosynthetic pathways are subjects of efforts to engineer diversification of natural products in pursuit of pharmaceutical leads and compounds of industrial importance[2]. They are rich sources for development of chemo-enzymatic catalysts based on PKS enzymes with unusual catalytic activities.

Modular type-I PKS pathways, among the most versatile of nature's systems, are biosynthetic assembly lines composed of modules that act in a defined sequence to produce complex products with a variety of functional groups and chiral centers. Each module is a set of fused catalytic domains that extend and modify a polyketide intermediate. Biosynthesis proceeds from intermediates tethered to acyl carrier protein (ACP) domains via a thioester link to a phosphopantetheine (Ppant) cofactor. A ketosynthase (KS) domain catalyzes extension of the intermediate, and subsequent modification domains typically catalyze reduction and/or methylation of the β-keto (3-keto) extension product. Beyond the enzymes for these core reactions, many PKS pathways also include other catalytic functionality. Among the most interesting of these

non-canonical capabilities is alkylation at the β position by a set of β-branching enzymes[18]. A 3-hydroxymethylglutaryl (HMG) synthase enzyme (HMGS) catalyzes the key branch-incorporation step, generating a β-hydroxy, β-carboxyalkyl acyl-ACP (Figure 2.1). The final structure of a β-branch depends on the carboxyalkyl group and the series of enzymatic steps that tailor the initial branch generated by HMGS[18].

The natural product curacin A, produced by the marine cyanobacterium *Moorea producens*[28], has cytotoxic activity and has been explored as an anticancer agent[51]. The hybrid PKS/NRPS (non-ribosomal peptide synthetase) pathway for curacin A contains an abundance of unique enzymes that install distinctive functional groups[52], including a cyclopropane ring formed by a surprising variation of the β-branching process[19]. In curacin A β-branching, the initial HMG-ACP intermediate undergoes chlorination, dehydration, and decarboxylation before a reductive ring-closing reaction generates the final cyclopropane group[19] (Figure 2.2).

HMGS is a homolog of the KS extension enzyme in PKS pathways but it is more closely related to HMG-coenzyme A (CoA) synthase (HMGCS), an enzyme of primary metabolism in the mevalonate-dependent isoprenoid pathway, where it acts directly before HMG-CoA reductase[30]. HMGCS generates HMG-CoA from acetyl-CoA and acetoacetyl-CoA by an aldol addition (Figure 2.2). The acetyl group is transferred to a catalytic cysteine, then deprotonated by a conserved glutamate. The resulting enolate nucleophile attacks the β-carbonyl of the second substrate, acetoacetyl-CoA. The covalent enzyme-product complex is hydrolyzed to release HMG-CoA. HMGCS is well

characterized, including crystal structures of complexes between the enzyme and its product, substrates, and inhibitors[53–58].

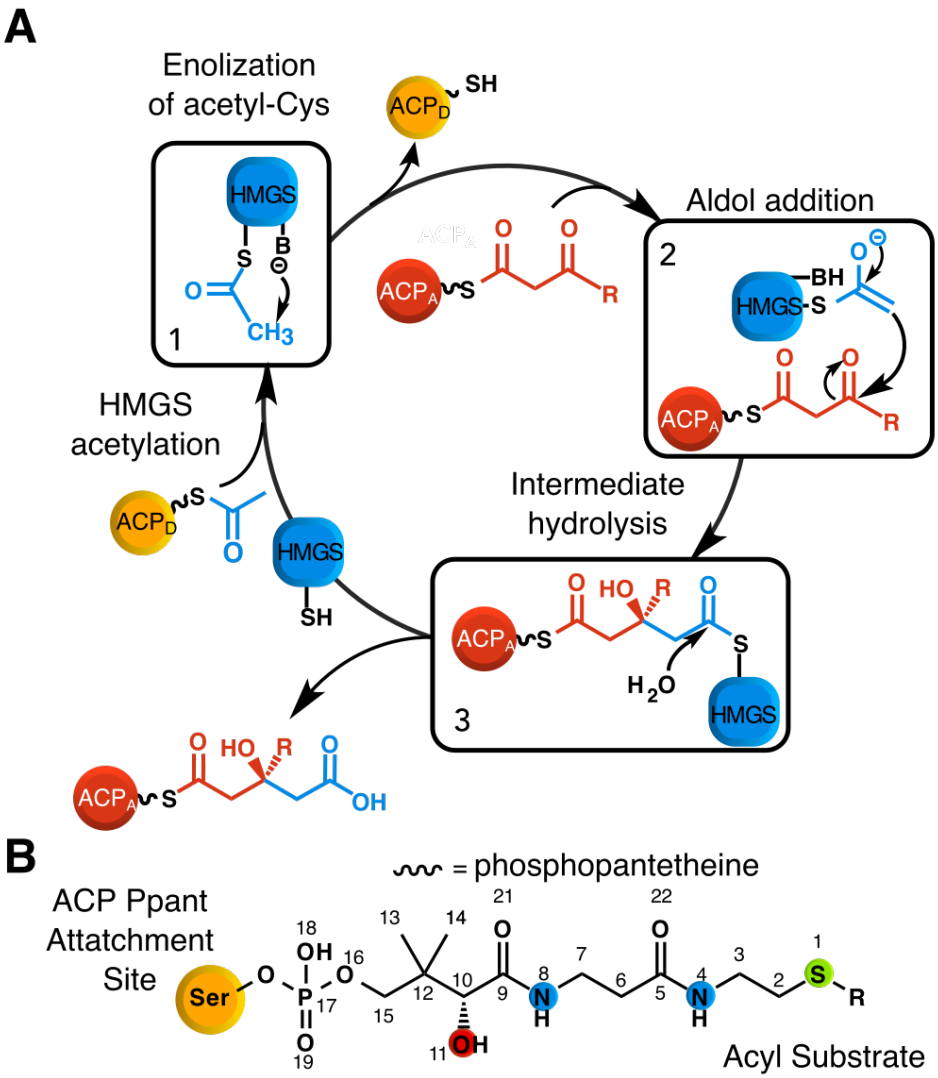


Figure 2.1 HMGS reaction. A. Reaction steps. 1) ACP_D transfers an acetyl group to HMGS Cys114 and Glu82 deprotonates the acetyl group; 2) the resulting enolate nucleophile attacks acetoacetyl-ACP_A; 3) The HMG-ACP_A product is hydrolyzed from Cys114. R indicates the polyketide intermediate (methyl in curacin A biosynthesis). **B.** Structure of the Ppant cofactor (represented as a squiggle symbol in **A**).

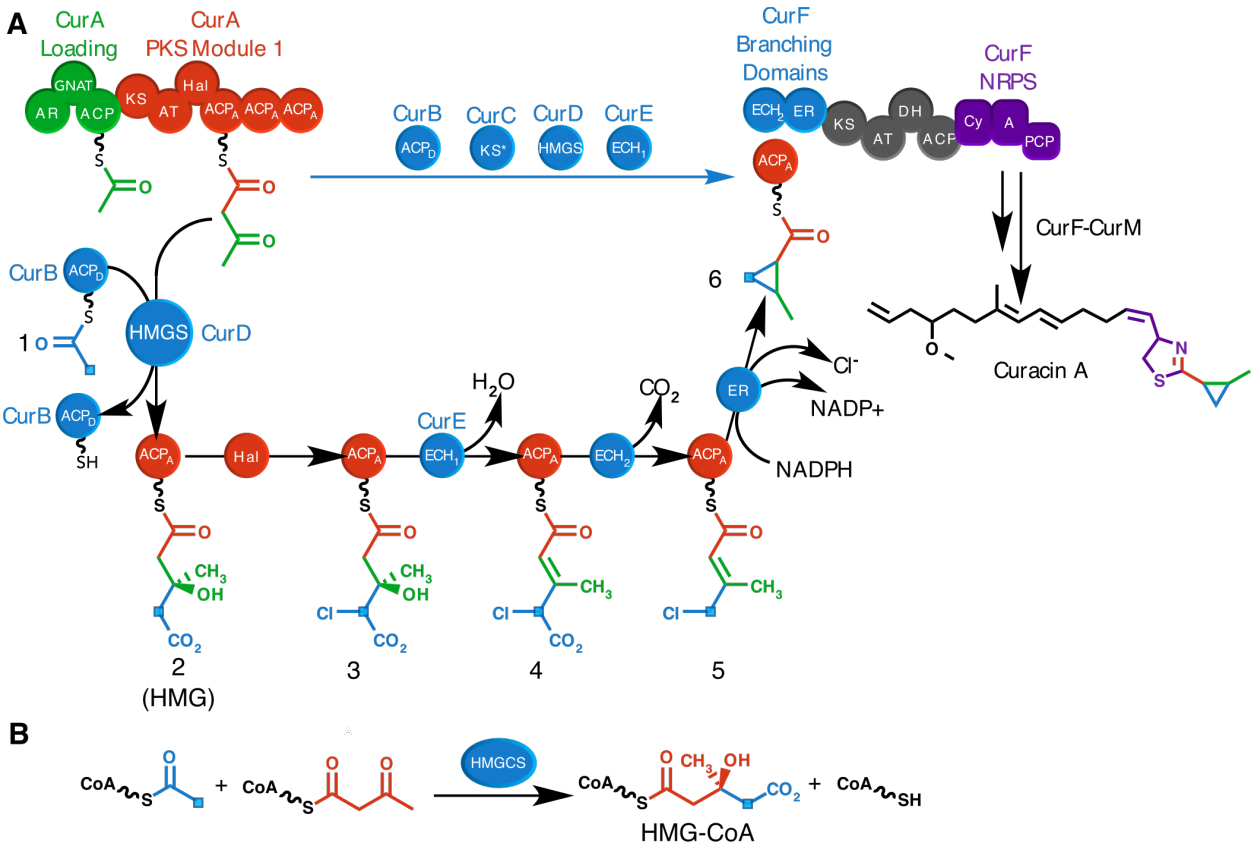


Figure 2.2 A. Curacin A β-branching. CurC KS_{DC} is proposed to decarboxylate malonyl-CurB, generating acetyl-CurB (ACP_D), **1**. CurD HMGS catalyzes formation of HMG-CurA ACP^3 (ACP_A), **2**. CurA Hal chlorinates the γ-carbon of HMG-ACP_A, which is subsequently dehydrated to **3** and decarboxylated to **4** by CurE ECH_1 and CurF ECH_2, respectively. CurF ER catalyzes cyclopropanation to **6** by NADPH-dependent addition of hydride and elimination of chloride. **B. HMGCS Reaction.** HMG-CoA synthase generates HMG-CoA from acetyl- and acetoacetyl-CoA via an aldol addition.

The HMGS and HMGCS reactions (Figure 2.1) are analogous and, in the case of the curacin A HMGS (CurD), have identical acyl substrates that are tethered to distinct ACPs and not to CoA. Whereas HMGCS distinguishes acetyl-CoA and acetoacetyl-CoA based on the acyl group and the acetylation status of the catalytic cysteine, HMGS also displays ACP selectivity: a standalone acetyl "donor" ACP (ACP_D), and an acetoacetyl-like "acceptor" ACP (ACP_A) bearing the polyketide intermediate within a specific module of the PKS pathway[20,25]. HMGS enzymes are highly selective for acetoacetyl-ACP_A and acetyl-ACP_D, and do not react with CoA substrates[24,20,25]. At the sequence level, the donor and acceptor ACPs clade separately from each other and from non-branch ACPs from the same pathways[25,45]. Thus, by selecting for the correct acyl-ACPs, HMGS prevents the formation of aberrantly branched metabolites.

The basis of ACP selectivity by HMGS is unknown, and detailed views of ACP interactions with PKS enzymes are few. The homologous KS extension enzymes distinguish their cognate donor and acceptor ACPs through two active site entrances[8] in interactions facilitated by fusion or non-covalent interaction of appended docking domains[43,44], which are absent in HMGS. HMGS selectivity for ACP_A and ACP_D presumably originates from distinguishing features of each ACP that result in different modes of interaction with HMGS. Here we present the biochemical characterization of the curacin A HMGS (CurD) and the interaction with its cognate ACP_D (CurB). Crystal structures of HMGS and of complexes with ACP_D reveal a striking shape complementarity between the proteins and specific charge interactions that orient the Ppant cofactor in the HMGS active site. Pre- and post-acetyl transfer states of Ppant

capture sequential views of the catalytic cycle. The results are a detailed benchmark of high-affinity enzyme-ACP interactions that advance our understanding of enzyme selectivity for carrier proteins.

Results

CurD HMGS activity. Purified, recombinant ACP_D (CurB) and the second cognate ACP_A from the CurA tandem ACP_A tridomain (Figure 2.2) were acylated *in vitro*. Recombinant CurD HMGS converted 83% of equimolar acetyl-ACP_D and acetoacetyl-ACP_A to HMG-ACP_A in a 10 minute reaction at 25 °C (Table 2.1). Reaction progress was monitored by LC-MS using the acyl-Ppant ejection assay[59] (Figure 2.3, Table 2.2). We detected no conversion of acetoacetyl-ACP_A to HMG-ACP_A when the catalytic Cys114 was substituted with serine (Table 2.2). The C114S substitution also prevented acetyl transfer from acetyl-ACP_D (Table 2.3).

Table 2.1 HMGS activity and ACP_D affinity

	HMGS activity[1] (%)	K_d (apo-ACP_D) (μM)[2]	K_d (holo-ACP_D) (μM)
HMGS and crystallization variant[3]			
Wild type	82.8 ± 0.3	1.1 ± 0.1	0.5 ± 0.2
AAA[3]	87 ± 1	--	--
Mismatched acyl-ACP substrates[5]			
ACP_D as donor and acceptor	29 ± 2	--	--
ACP_A as donor and acceptor	ND	--	--
HMGS active site substitutions			
C114S	ND[4]	--	--
P166A	88.1 ± 0.3	--	--
S167A	95.5 ± 0.3	--	--
ACP_D/HMGS interface substitutions			
ACP_D			
R42A	95.7 ± 0.2	--	--
HMGS			
R33A	ND	6.9 ± 0.7	8 ± 2
R33D	2 ± 3	29 ± 3	25 ± 3
D214A	86 ± 8	10.9 ± 0.9	9 ± 2
D214R	45 ± 4	17 ± 5	11 ± 3
D222A	27 ± 1	28 ± 2	1.5 ± 0.6
D222R	17 ± 1	34 ± 2	5 ± 1
E225A	55 ± 3	90 ± 4	--
E225R	8 ± 1	68 ± 5	12 ± 1
R266A	56 ± 1	110 ± 10	17 ± 2
R266E	26 ± 1	180 ± 10	17 ± 5

[1] Conversion of equimolar acetyl-ACP_D and acetoacetyl-ACP_A to HMG-ACP_A in a 10 minute reaction at 25 °C. Each value corresponds to the average of 3 measurements.
[2] Affinities were measured by fluorescence polarization using BoDIPY-tagged ACP_D. Each value corresponds to the average of 3 measurements.
[3] All crystal structures were of the $HMGS_{AAA}$ variant (K343A/Q344A/Q346A). Cys114 is the catalytic nucleophile.
[4] No product detected.
[5] In each reaction, either ACP_D or ACP_A was loaded with both acyl substrates, and equimolar quantities of acetyl-ACP and acetoacetyl-ACP were used.

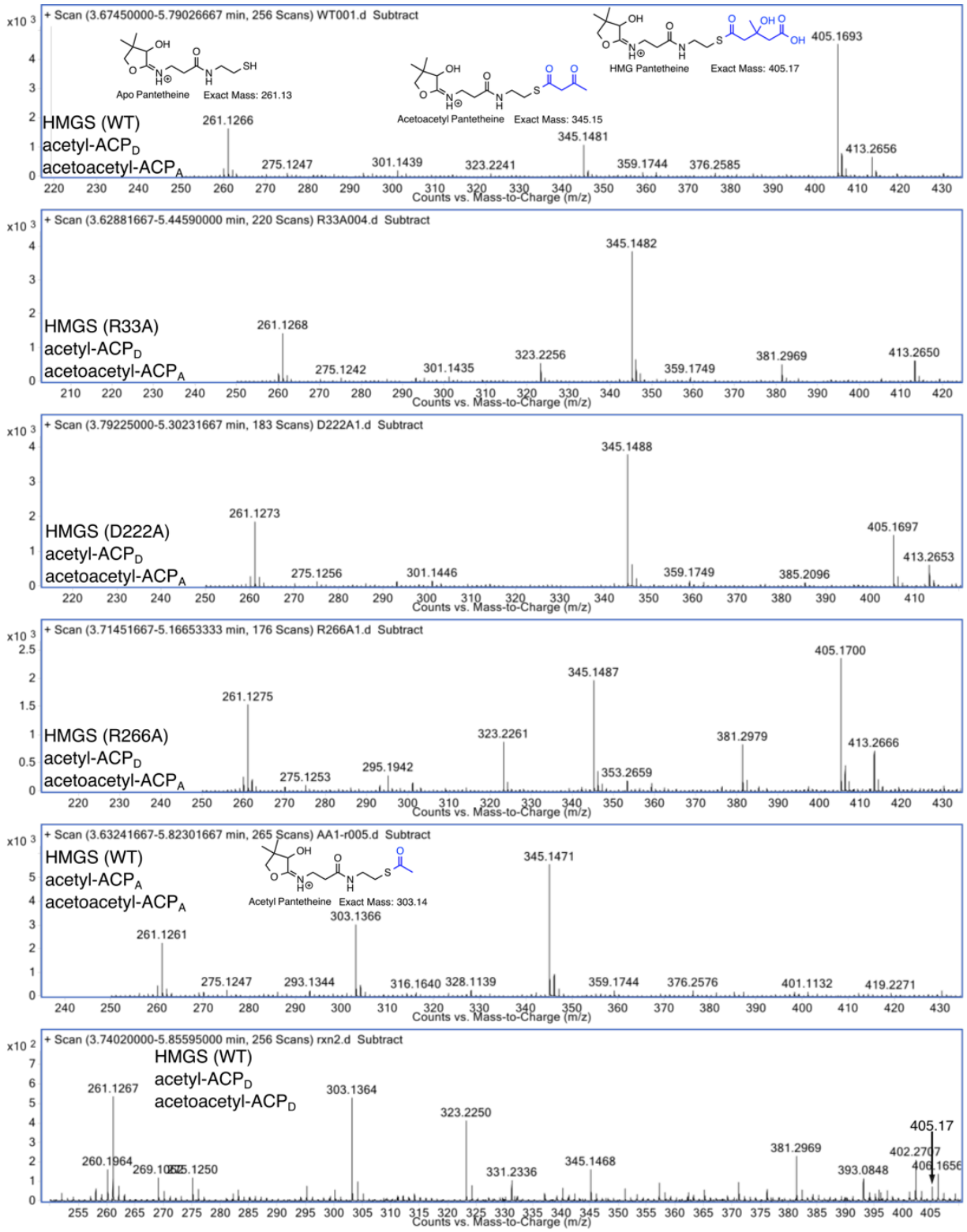


Figure 2.3 Sample HMGS activity data. Ppant ejection mass spectra of the ACP$_A$ elution of an HMGS reaction mix. Ion counts were recorded for the acetoacetyl (calculated m/z=345.15) and HMG (calculated m/z=405.17) peaks. See Table 2.2 for total ion counts for each species.

Table 2.2 Conversion of acetoacetyl-ACP_A to HMG-ACP_A evaluated by pant ejection

	Acetoacetyl (345.15)[1]	HMG (405.17)	Conversion to product (%)
Wild type	1091.33[2]	5258.49	82.8
R33A	6255.19	0	0
D214A	870.19	3527.54	80.2
D222A	7059.05	2753.29	28.1
E225A	2997.98	3324.05	52.6
R266A	3447.38	4178.93	54.8

[1]Predicted masses for each species are given in parentheses.
[2]EIC for spectra in Figure 2.3. One of three replicate experiments for each mutant is shown. Conversion to product was calculated as HMG/(HMG+Acetoacetyl) and averaged for the three trials. Error estimates are given in Table 2.1.

Table 2.3 HMGS-dependent de-acetylation of acetyl-ACP_D

Sample	**% Acetyl-ACP**
acetyl-ACP_D (no incubation)	87
acetyl-ACP_D (18-hr incubation, no enzyme)	88 ± 4
acetyl-ACP_D (18-hr incubation with $HMGS_{WT}$)	13 ± 10
acetyl-ACP_D (18-hr incubation with $HMGS_{C114S}$)	86 ± 0.3

Acetyl-ACP_D was evaluated by mass spec Ppant ejection. Values given are the percent of total ion counts for ejected acetyl-Ppant among all ejected Ppant species. Incubations were at 20 °C in a buffer of 1X MMT (Qiagen) pH 6.5, 20 mM $(NH_4)_2SO_4$. HMGS was present at a 1:10 ratio to ACP_D. Under the conditions of the experiment, wild type HMGS removed the acetyl group from ACP_D by the, but the catalytic variant HMGS C114S did not. Each incubation measurement was made in duplicate on independently prepared samples.

HMGS structure. A triple alanine HMGS variant (K344A/Q345A/Q347A, $HMGS_{AAA}$) was used for all crystal structures and had catalytic activity indistinguishable from the wild type (Table 2.1). The 2.1-Å crystal structure in space group $P3_121$ (Table 2.4, Figure 2.4) was solved by molecular replacement. HMGS is dimeric in solution and in the crystal structure, where the asymmetric unit consists of one subunit (Figure 2.5). The catalytic amino acids Glu82, Cys114, and His250 (Fig 2.6 A) are identically positioned in HMGS and HMGCS, and the overall folds are similar (RMSD of 2.03 Å for 368 Cα atoms). Striking differences occur at the active site entrance and the dimer interface. A disordered loop (HMGS residues 149-163) near the active site encompases a conserved insertion in HMGS (residues 155-164) relative to HMGCS. (Figure 2.7). An adjacent loop at the subunit interface (residues 203-210) has a different conformation than the analogous loop in HMGCS. The HMGS 203-210 loop forms hydrophobic contacts with the partner subunit that involve several residues conserved among HMGS but not HMGCS.

ACP_D/HMGS complex. We tested the ACP selectivity of the curacin A HMGS in experiments where acyl groups were mismatched with ACP_D or ACP_A (Table 2.1, Figure 2.3). ACP_A was not a surrogate acetyl donor, as we detected no conversion of acetyl-ACP_A + acetoacetyl-ACP_A to HMG-ACP_A. In contrast, ACP_D was a weak surrogate acceptor with threefold-reduced conversion of acetyl-ACP_D + acetoacetyl-ACP_D to HMG-ACP_D relative to the natural partners. Thus, HMGS had greater selectivity for the donor ACP than for the acceptor. To investigate the structural basis of ACP selectivity, we pursued structures of HMGS complexes with ACP_A and ACP_D.

Table 2.4 crystallographic data collection and refinement statistics

	HMGS	HMGS/ apo-ACP_D	HMGS/ holo-ACP_D	$HMGS_{C114S}$/ ac-ACP_D	$HMGS_{WT}$/ Ac-ACP_D	$HMGS_{C114S}$/ Holo-ACP_D	$HMGS_{C114S}$/ Ac-ACP_D	$HMGS_{C114S}$
Wavelength (Å)	1.033	1.033	1.033	1.033	1.033	1.033	1.771	1.771
Space Group	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$
a,b (Å)	101.16	101.31	101.12	100.89	100.85	101.94	100.79	101.2
c (Å)	106.23	104.59	104.6	104.46	104.27	104.56	104.32	105.92
Data range (Å)	45.67-2.10	45.59-2.05	45.54-1.60	45.43-1.90	45.39-2.10	45.82-1.89	45.38-2.38	45.66-2.38
	(2.16-2.10)	(2.11-2.05)	(1.63-1.60)	(1.94-1.90)	(2.16-2.10)	(1.93-1.89)	(2.47-2.38)	(2.47-2.38)
Unique Reflections	37159 (3003)	39403 (3028)	81779 (4012)	48869 (3115)	36,067 (2867)	50,923 (3099)	24,958 (2520)	25,488 (2526)
Completeness (%)	100 (100)	100 (99.9)	100 (100)	100.0 (100.0)	99.8 (98.2)	99.7 (94.9)	99.6 (96.7)	99.4 (95.0)
Multiplicity	19.5 (20.0)	19.5 (16.4)	19.8 (20.3)	19.5 (19.5)	10.4 (6.1)	9.9 (9.0)	18.3 (11.2)	18.3 (11.4)
Wilson B-factor	51.1	37	23.9	29.6	37.3	31.1	38.3	46.9
R_{merge} (%)	11.2(222.1)	11.7 (218.1)	7.1 (257.9)	11.6 (152.3)	13.2 (152.4)	7.4 (194.0)	8.9 (123.1)	5.7 (145.6)
Average I/σ_I	15.5 (1.5)	19.1 (1.4)	24.0 (1.4)	16.9 (2.2)	14.8 (1.2)	17.3 (1.2)	25.9 (1.7)	35.3 (1.5)
CC 1/2	1.00 (0.62)	0.99 (0.53)	1.00 (0.51)	1.00 (0.73)	1.00 (0.48)	1.00 (0.49)	1.00 (78.4)	1.00 (0.73)
R_{work} (%)	15.6	15.5	14.8	15.8				
R_{free} (%)	19	19.3	16.5	18.2				
Protein Atoms	3139	3747	3848	3825				
Ppant Atoms	--	--	21	45				
Water/Ion Atoms	161	219	351	219				
RMSD bond length (Å)	0.011	0.011	0.01	0.009				
RMSD bond angle (°)	1.362	1.476	1.437	1.352				
Average B-factors								
Protein	58.5	44.0(HMGS) 76.7(ACP)	30.4 (HMGS) 52.9(ACP)	35.9(HMGS) 53.9(ACP)				
Ppant	--	--	43.6	37.2				
Water	64	55.9	43.9	47				
Ramachandran								
Favored (%)	96.8	97.3	98	97.1				
Generously allowed (%)	99.7	99.8	99.8	99.8				
Outlier (%)	0.3	0.2	0.2	0.2				

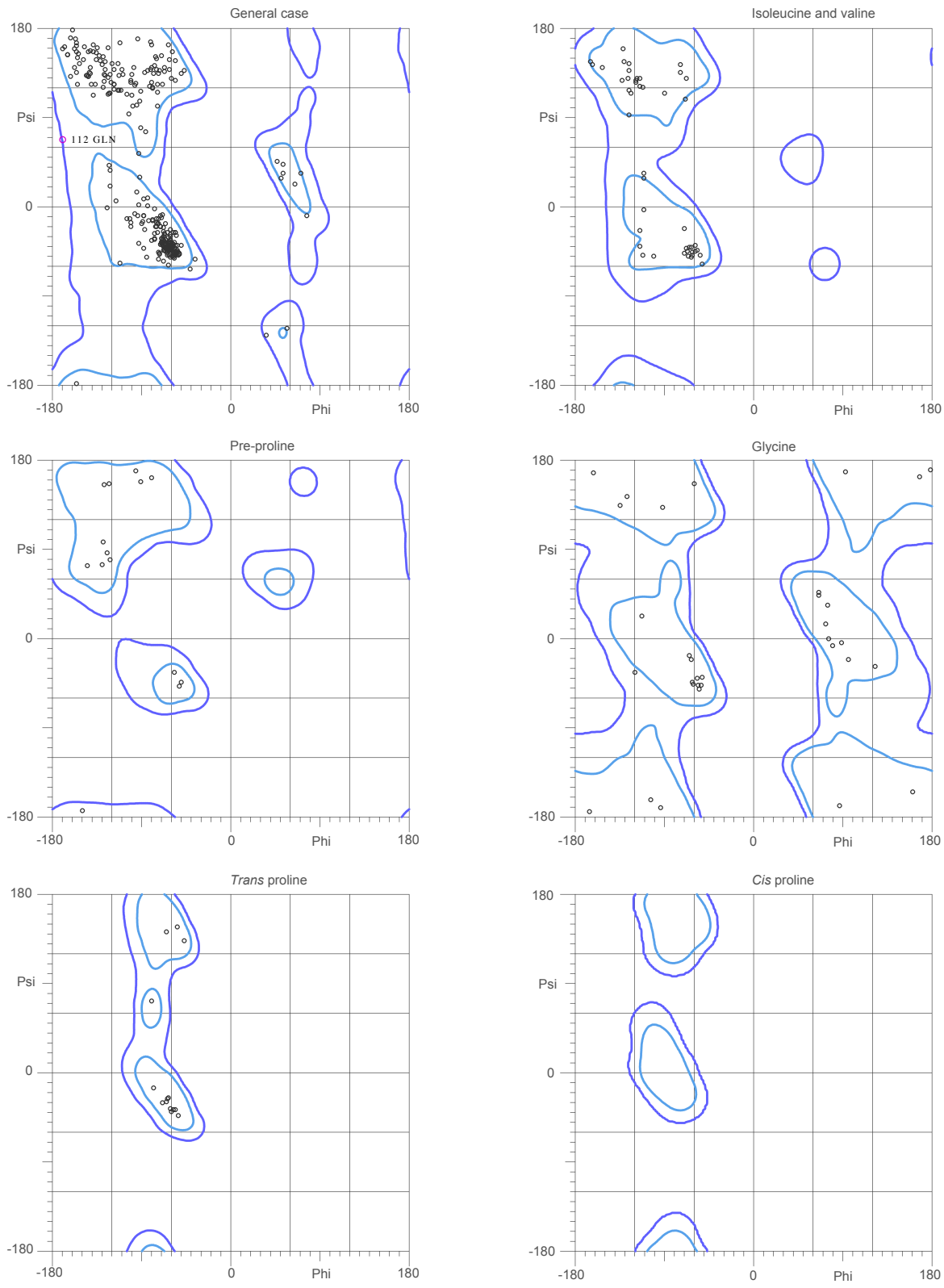


Figure 2.4 Ramachandran analysis of the HMGS structure.

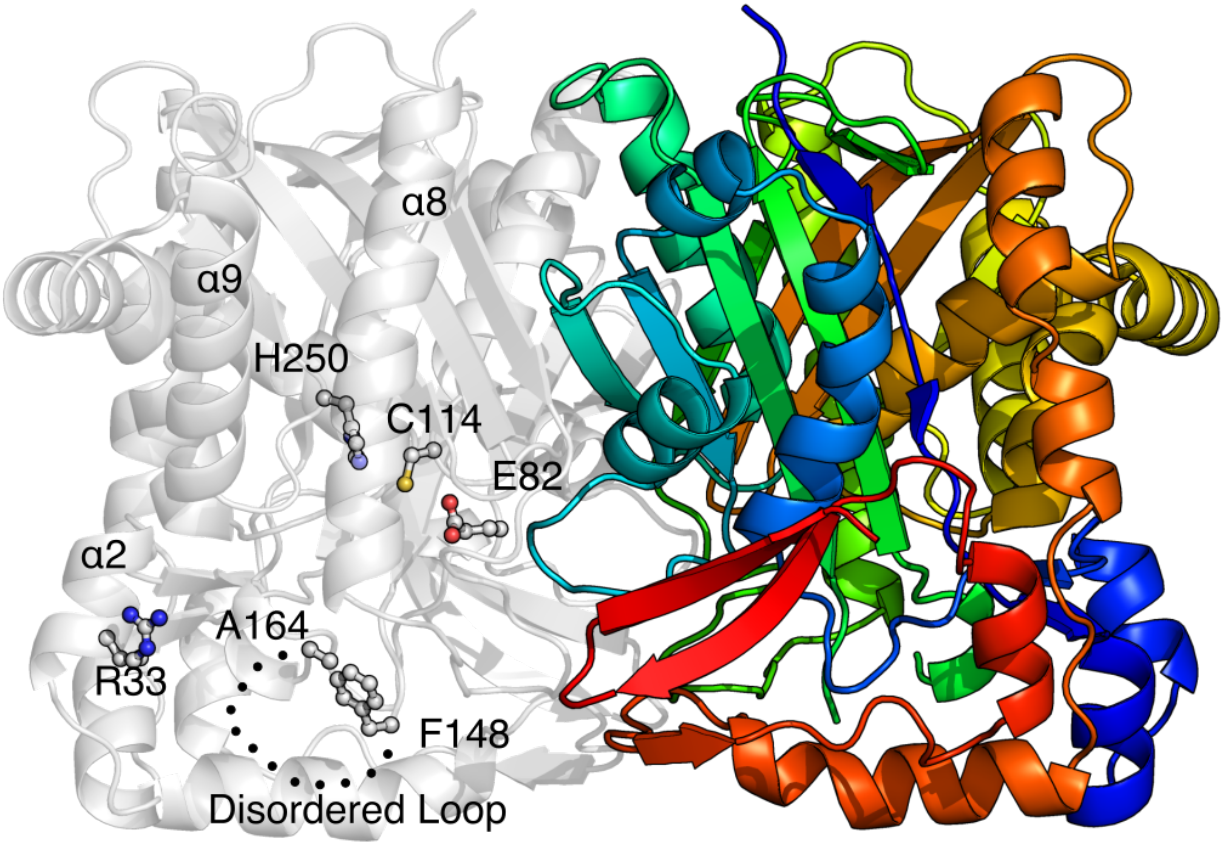


Figure 2.5 CurD HMGS structure. Within the dimer, the right-hand subunit is colored by sequence from the blue N-terminus to the red C-terminus. Key residues are shown in ball-and-stick on the gray left monomer, including the catalytic Cys114, Glu82 and His250. Phe148 and Ala164 are the boundaries of the 15-residue disordered loop at the active site entrance. The basic side chain of Arg33 interacts with the Ppant phosphate and is conserved in HMGCS sequences. A dotted line denotes the disordered loop region connecting Phe148 to Ala164.

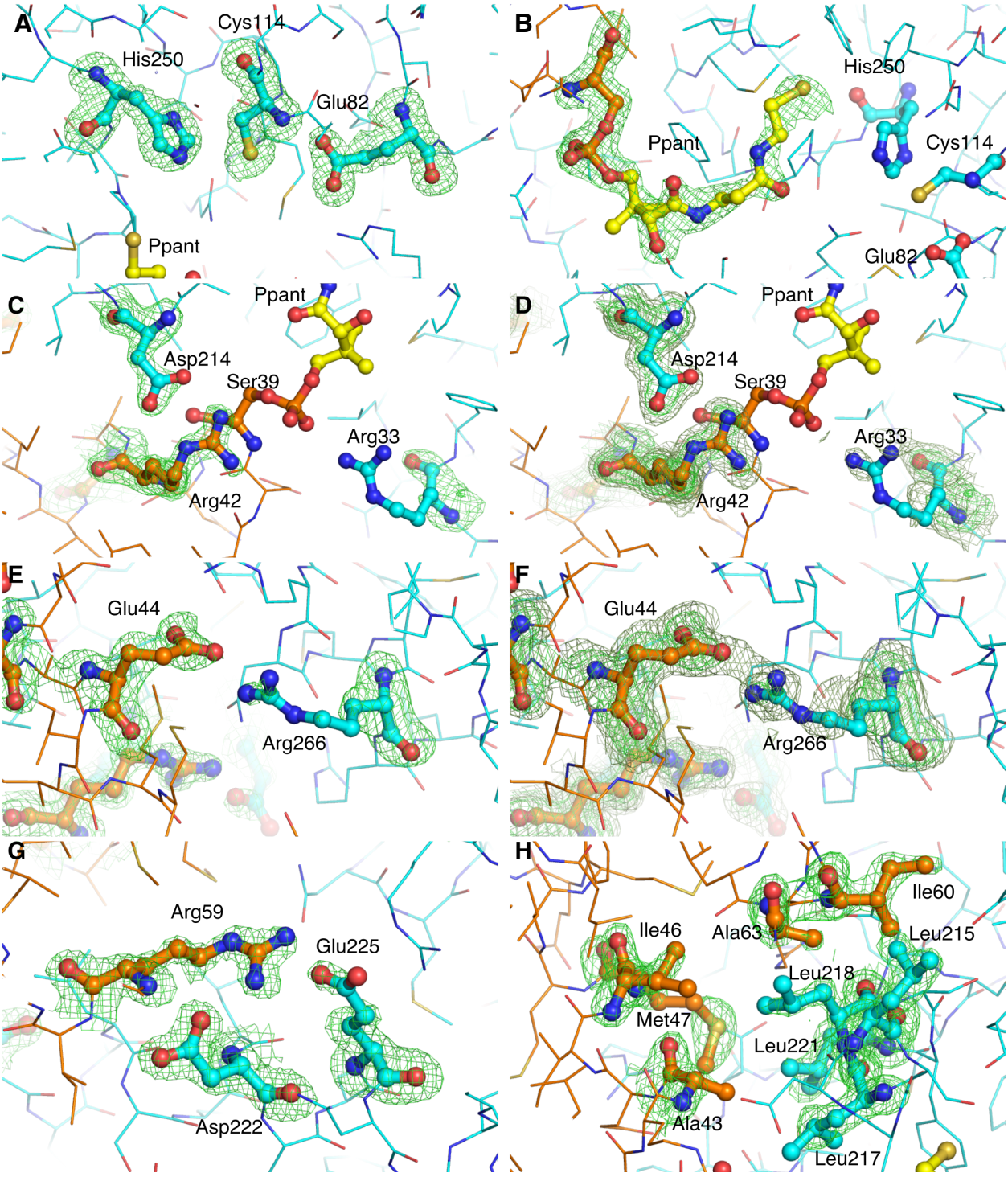


Figure 2.6 Electron density for holo-ACP$_D$-HMGS. Key residues (ball-&-stick) were omitted, and models were refined in phenix.refine with simulated annealing. SA-omit density is contoured at 3.0σ in green (A-H) and at 1.5σ in pale green (D,F). HMGS residues are in cyan, ACP$_D$ in orange, and Ppant in yellow in each panal: **A.** Catalytic HMGS residues. **B.** Ppant. **C-G.** Ionic contacts (Figure 2.7) **H.** Nonpolar contacts (Figure 2.7 D). Weaker omit density for Arg33 and Arg266 is shown in D and F.

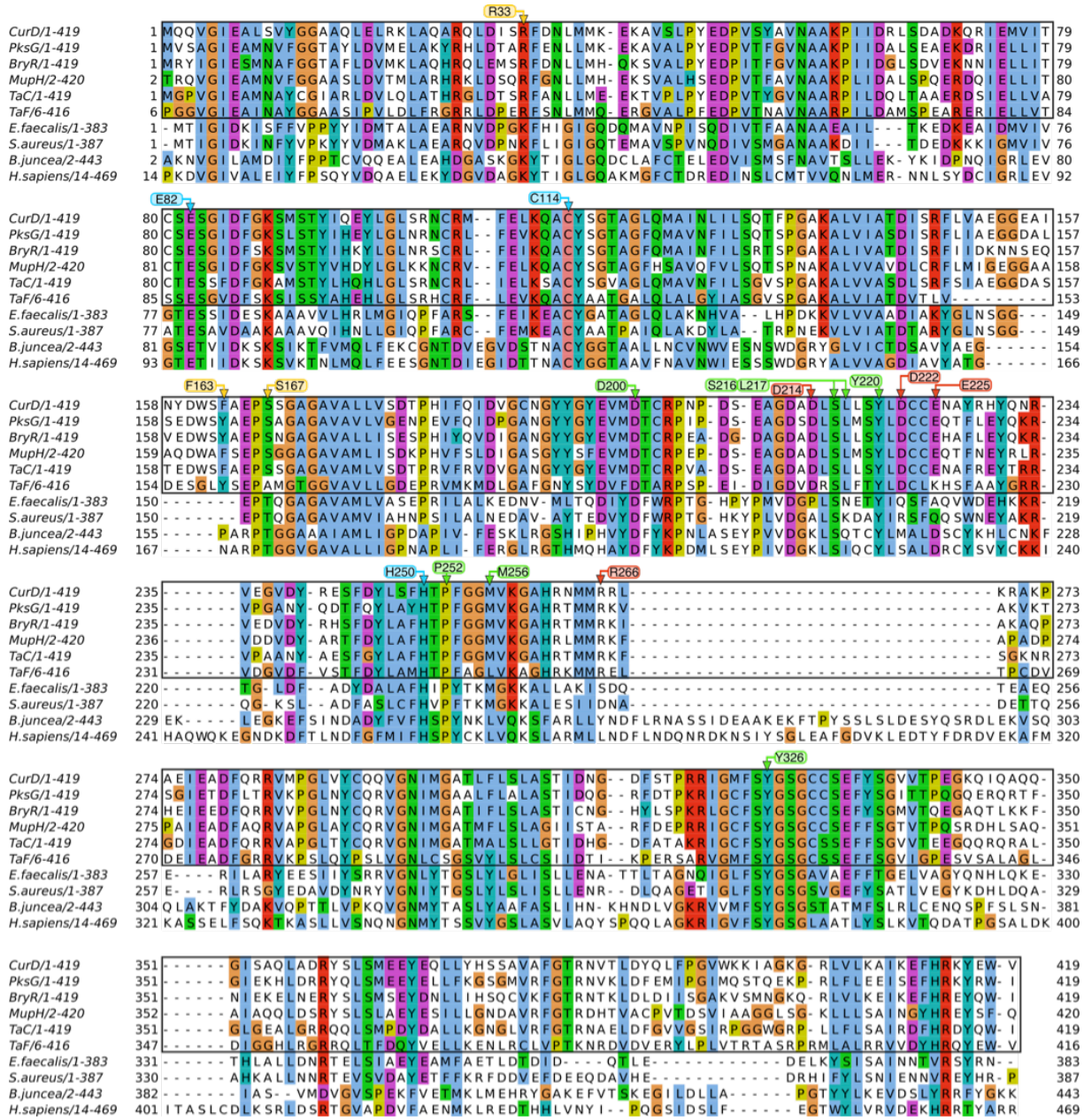


Figure 2.7 Alignment of HMGS and HMGCS sequences. Alignments were generated using Clustal Omega[60] and analyzed in Jalview[61]. The boxed sequences, identified by protein name, are for the β-branching HMGS and the bottom sequences are for the primary metabolism HMGCS, identified by species of origin. Important residues are numbered by the CurD HMGS sequence. Ppant interacting residues are identified in yellow, catalytic residues in blue, thiol pocket residues in green, and ACP_D interacting residues in red. Species of origin and accession codes for HMGS sequences are as follows: CurD from *Moorea producens*, Q6DNE9; PksG from *Bacillus subtilis*, P40830; MupH from *Pseudomonas fluorescens*, Q8RL63; TaC from *Myxococcus xanthus*, Q1D5E5; TaF from *Myxococcus xanthus*, Q1D5E8; BryR from *Ca. Endobugula sertula*, A2CLL9. Accession codes for HMGCS sequences: *Enterococcus faecalis*, Q9FD71; *Staphylococcus aureus*, Q9FD87; *Brassica juncea*, Q9M6U3; *Homo sapiens*, Q01581.

The HMGS active site entrance and surrounding surface were unhindered by crystal contacts, and we captured complexes of HMGS with apo-ACP_D, holo-ACP_D, and acetyl-ACP_D by cocrystallization (Figure 2.8-2.10, Table 2.4). The ACP_D was well ordered at the HMGS active site entrance, made no lattice contacts, and contacted neither the partner subunit of the HMGS dimer nor its bound ACP_D. The holo-ACP_D Ppant had clear electron density in the active site (Figure 2.11, 2.6 B) yet formed only a few interactions, including the phosphate to HMGS Arg33 and ACP_D Arg42, which was also salt-bridged to HMGS Asp214 (Figure 2.8). The Ppant-Arg42-Asp214 network is specific to ACP_D-HMGS pairs (Figure 2.7, 2.12), but the Arg33-Ppant interaction also occurred in HMGCS-CoA complexes[53,55]. Ppant binding induced ordering of amino acids 159-163 in the HMGS disordered loop, forming a 3_{10} helix with hydrogen bonds of Phe163 and Ser167 to the Ppant (Figure 2.8). The Ppant thiol occupies a relatively hydrophobic "thiol pocket" (conserved amino acids Ser216, Leu217, Tyr220, Pro252, Met256 and Tyr326) and is hydrogen bonded to the Ser216 hydroxy group (Figure 2.8). An extensive network of hydrogen bonds involving Ser216, Tyr220, Tyr326 and conserved Asp200 maintains the structure of the thiol pocket.

In the holo-ACP_D/HMGS structure, the Ppant and Cys114 thiol groups are too far apart (7.9 Å) for the acetyl-transfer step of HMGS catalysis (Figure 2.11). This "distal" Ppant position was also occupied in crystals grown from acetyl-ACP_D + HMGS, with no density for an acetyl at either the Ppant or Cys114 (Figure 2.11, Table 2.4). The acetyl group of acetyl-ACP_D was apparently transferred to Cys114 and subsequently hydrolyzed during crystallization (2-3 days) (Table 2.3). We propose that the distal

Ppant position represents a post acetyl-transfer state. Nonproductive loss of the acetyl-CoA donor in absence of the acetoacetyl-CoA acceptor has also been reported for HMGCS[62].

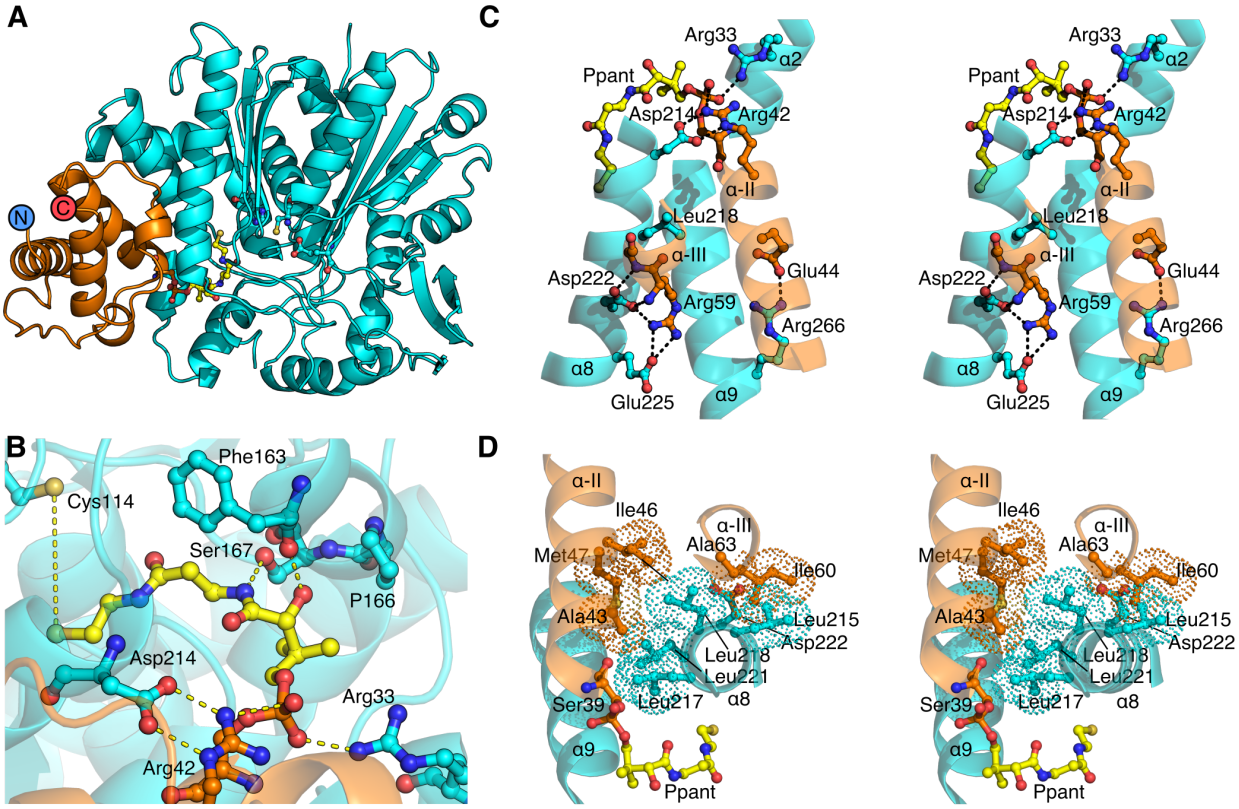


Figure 2.8 HMGS interaction with the donor ACP. A. ACP_D (orange) / HMGS (cyan) complex. Ppant (yellow) and catalytic residues shown in ball-and-stick form. **B.** Ppant in the HMGS active site. Dashed lines represent hydrogen bonds (yellow) and the long separation of Ppant and Cys114 thiol groups (black). **C.** Stereo view of charged contacts in the HMGS - ACP_D interface. **D.** Stereo view of hydrophobic contacts between ACP_D and HMGS. HMGS helices are numbered as in Figure 2.5, and ACP_D helices are labeled by Roman numeral. Helices in C and D are transparent for clarity.

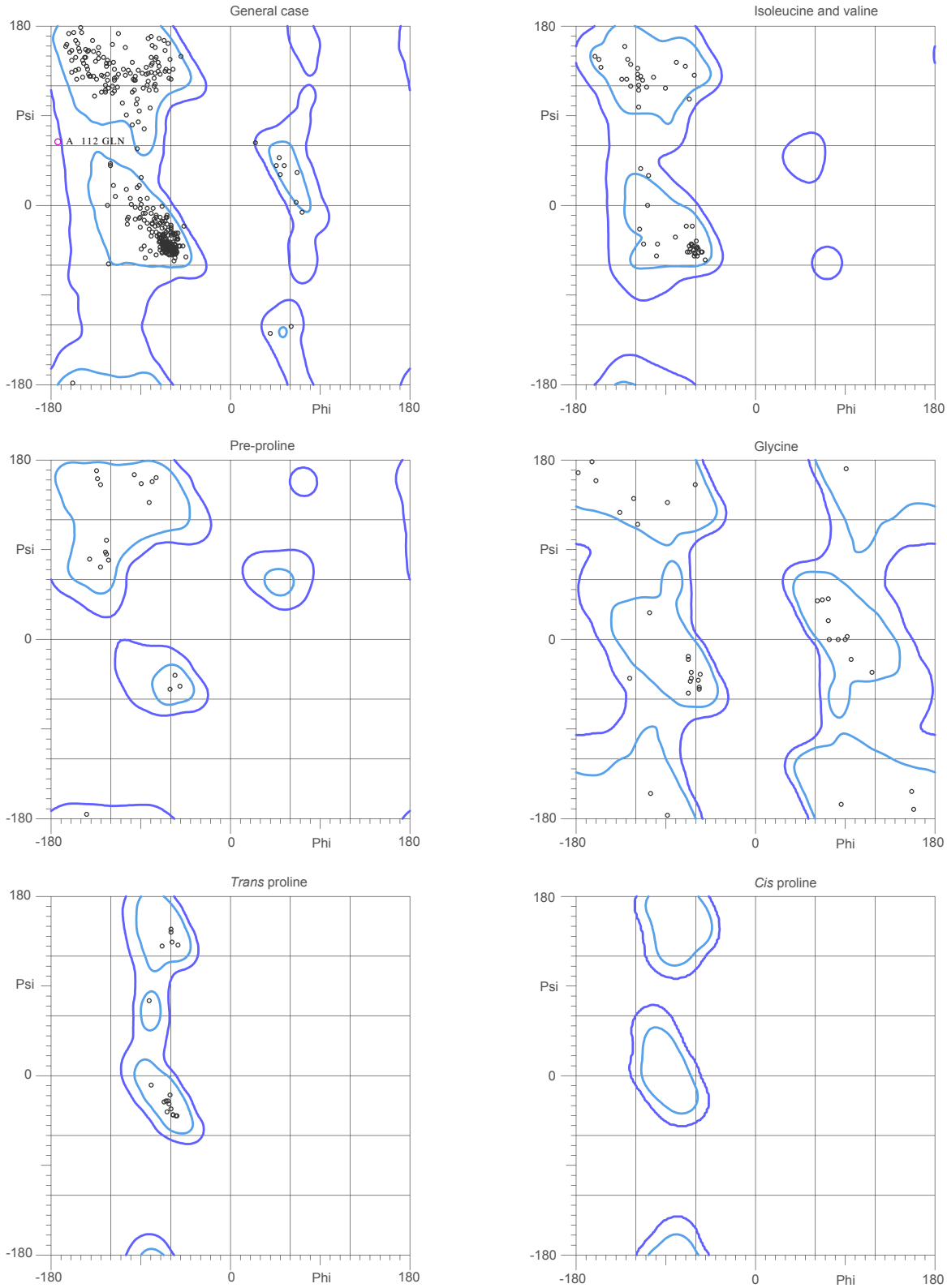


Figure 2.9 Ramachandran analysis of apo-ACP$_D$/HMGS complex structure.

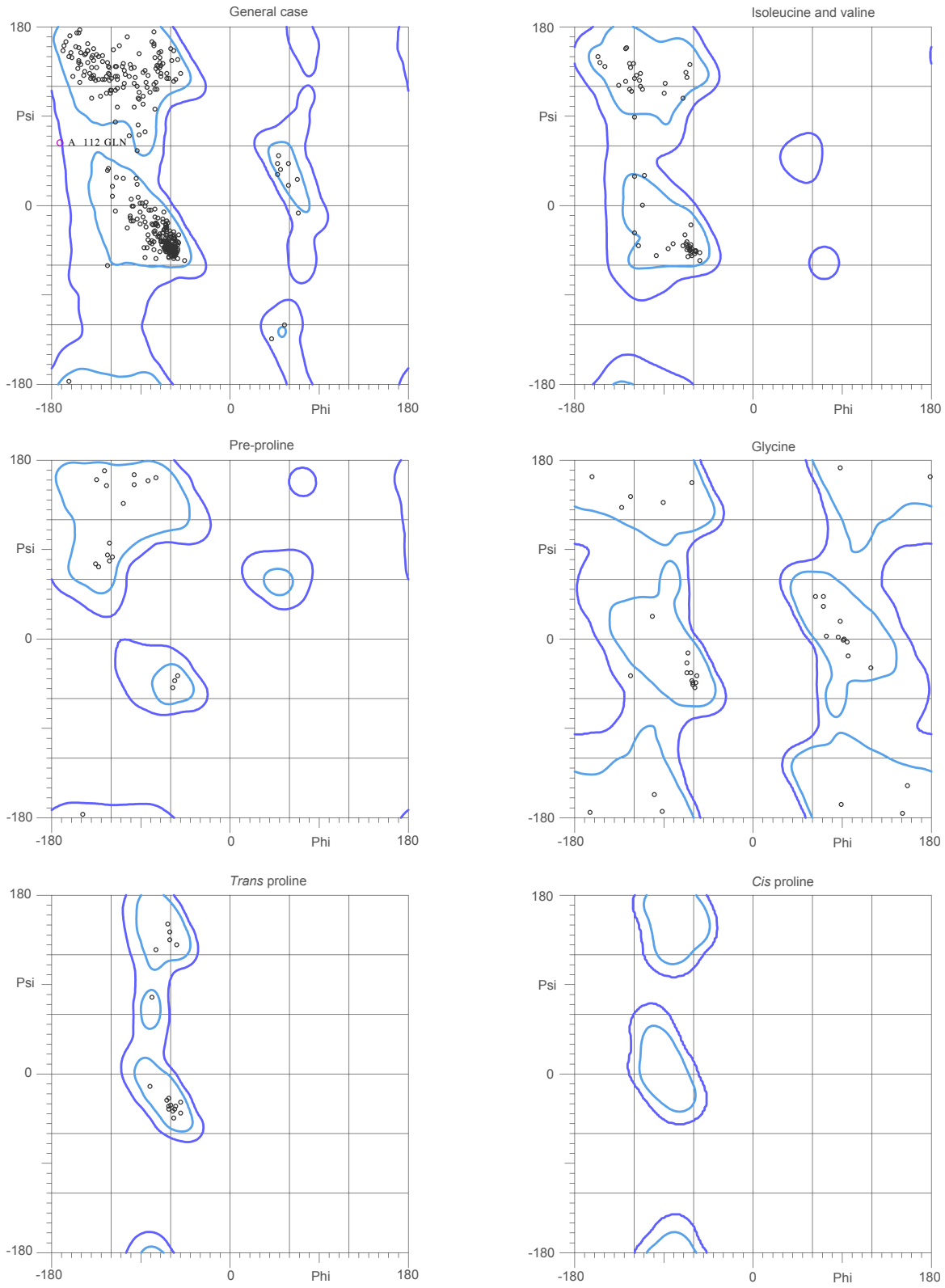


Figure 2.10 Ramachandran analysis of holo-ACP$_D$/HMGS complex structure.

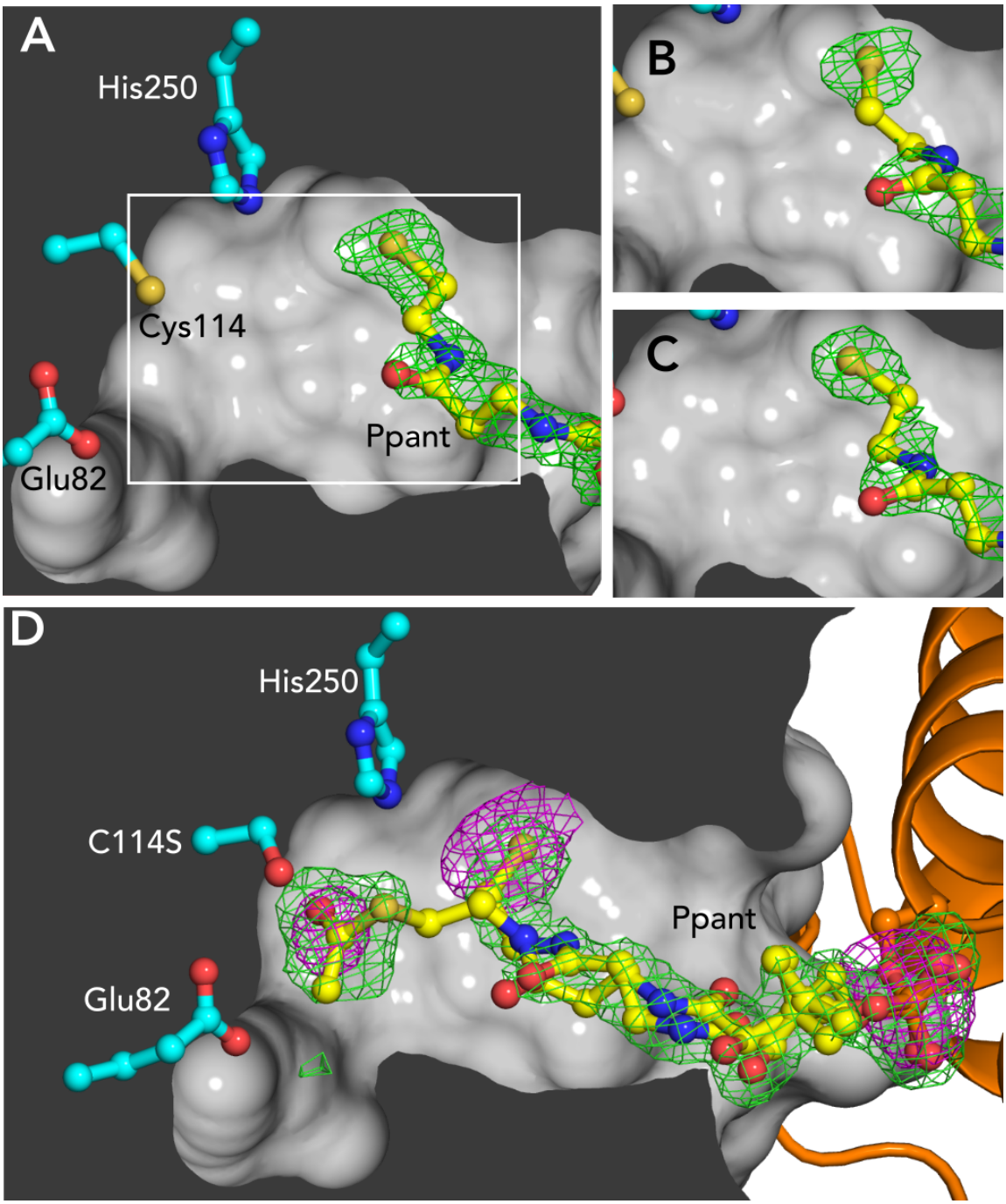


Figure 2.11 Acetylation-dependent position of Ppant. Panels show Fo-Fc omit density (3σ, Ser-Ppant omitted, green) for structures of ACP_D-HMGS in different biochemical states crystallized in identical conditions. **A.** Holo-ACP_D and $HMGS_{WT}$. White box indicates field of view for B and C. **B.** Acetyl-ACP_D and $HMGS_{WT}$, showing that the acetyl group has been lost. **C.** Holo-ACP_D and $HMGS_{C114S}$. **D.** Acetyl-ACP_D and $HMGS_{C114S}$. Anomalous difference density (3σ, magenta) indicates that S is present in both terminal densities for the Ppant and also shows the Ppant P atom. Atoms are colored as in Figure 2.8.

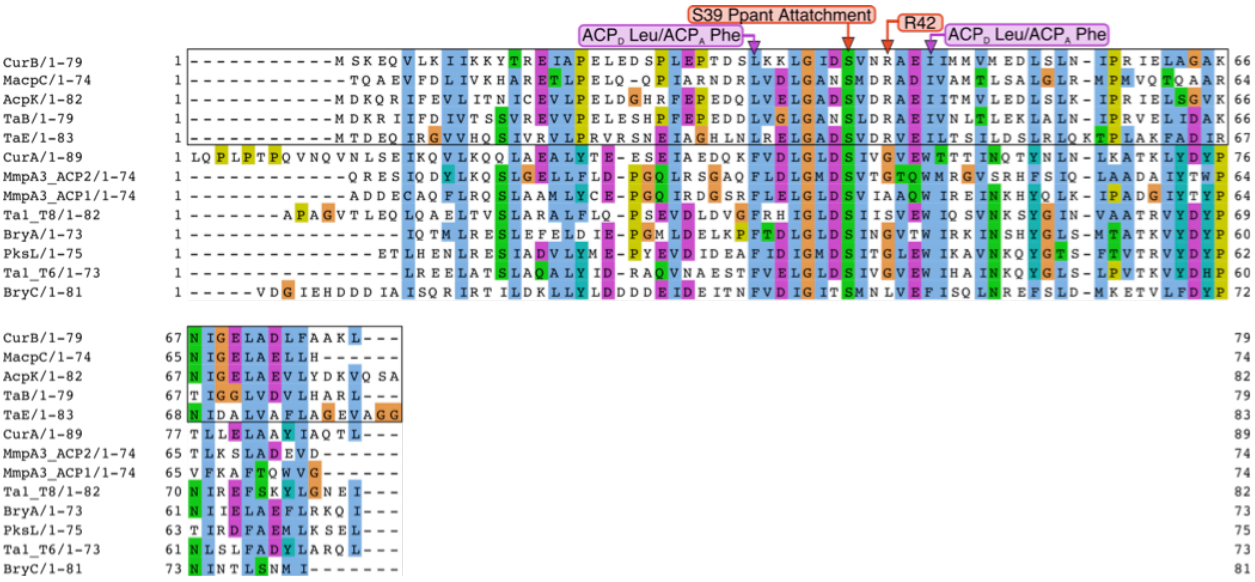


Figure 2.12 Alignment of ACP$_D$ and ACP$_A$ sequences. The alignment was generated using Clustal Omega[60] and analyzed in Jalview[61]. Boxed sequences are for ACP$_D$ and below for ACP$_A$. Important residues are numbered by the CurB ACP$_D$ sequence. The positions of hydrophobic side chains that are conserved aliphatics in ACP$_D$ and aromatic in ACP$_A$ are indicated by arrows below the alignment (see also Figure 2.14). Species of origin and accession codes for HMGS sequences are as follows: CurB and CurA from *Moorea producens 3L* (F4Y434, F4Y435); MacpC and MmpA3_ACPs from *Pseudomonas fluorescens* (Q8RL65, Q8RL76); AcpK and PksL from *Bacillus subtilis* (Q7PC63, Q05470); TaB, TaE, and Ta1 ACPs from *Myxococcus xanthus* (Q9XB07, Q9XB04, Q9Z5F4), BryA and BryC from *Candidatus Endobugula sertula* (A2CLL5, A2CLL2).

To trap an acetyl-ACP_D complex, we co-crystallized the inactive $HMGS_{C114S}$ with acetyl-ACP_D (Figure 2.13), resulting in multiple positions for the Ppant terminus (Figure 2.11), including the previously identified distal holo-Ppant position, again lacking density for an acetyl group. A second position was interpreted from a new strong density (15 σ) near Ser114 that could represent acetyl-Ppant (Figure 2.11). Two experiments validated the interpretation of the second "proximal" acetyl-Ppant position, as its density was discontinuous with the rest of the Ppant. We solved the structure of holo-ACP_D/$HMGS_{C114S}$, yielding density in the distal Ppant position and no density in the Ser114-proximal position, establishing that the new density was not due to the C114S substitution (Figure 2.11, Table 2.4). To distinguish whether the density near Ser114 was due to free acetate or the acetyl-Ppant terminus, we used anomalous scattering to identify atomic positions of S atoms. Data were recorded at an X-ray energy of 7.0 keV from acetyl-ACP_D/$HMGS_{C114S}$ co-crystals, yielding anomalous difference electron density for the Ppant S in both the distal site and at the site proximal to Ser114 (Figure 2.11, Table 2.4). A similar experiment with $HMGS_{C114S}$ crystals (no ACP_D) lacked anomalous difference electron density near Ser114. Thus, we conclude that during crystallization some of the acetyl-ACP_D hydrolyzed spontaneously, and the remaining acetyl-Ppant was adjacent to the nucleophilic side chain (C114S), defining a pre-acetyl-transfer position.

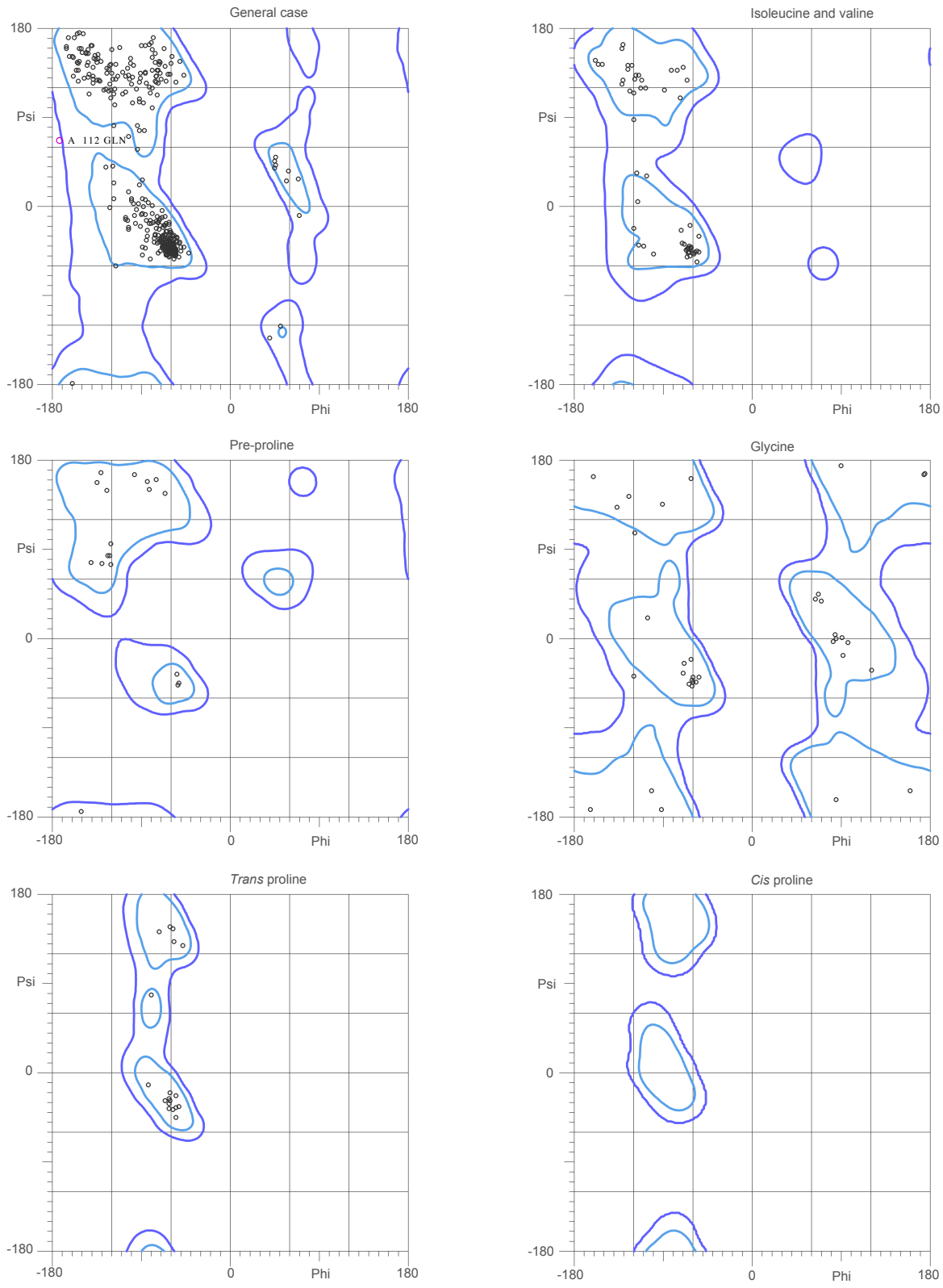


Figure 2.13 Ramachandran analysis of acetyl-ACP$_D$/HMGS complex structure.

ACP_D/HMGS interface. The interacting surfaces of ACP_D and HMGS are complementary in shape and charge (Figure 2.8, 2.14). The primary contact is between the N-terminal half of HMGS helix α8 (Figure 2.8) and an ACP_D cleft between the Ppant-Ser39 and helix III. On the ACP_D surface, conserved basic residues form a strongly electropositive region, which is separated from an electronegative region by a hydrophobic stripe (Figure 2.14). We compared the surface features of other ACPs in complexes with cognate enzymes[37,63–67] to ACP_D (Figure 2.14). Positive and negative surface regions are typical of PKS ACPs (Figure 2.14 B-E) whereas fatty acid synthase (FAS) ACPs are highly electronegative (Figure 2.14 F-I). Among PKS ACPs, ACP_D has two distinctive features: a strongly electropositive region and a cleft that is complementary to a conserved hydrophobic patch (Figure 2.8) (Leu217, Leu218) on the outer surface of HMGS helix α8. The analogous surface of the HMGCS helix is polar.

We evaluated several salt bridges in the ACP_D-HMGS interface by mutagenesis and, for each variant, measured HMGS activity and ACP_D affinity (Table 2.1). Each of the HMGS charged residues (Arg33, Asp214, Asp222, Glu225, Arg266) was substituted with alanine and an oppositely charged amino acid. Affinities were measured by fluorescence anisotropy with a fluorophore-conjugated ACP_D (Figure 2.15). The HMGS K_d was 1.1 μM for apo-ACP_D and 0.5 μM for holo-ACP_D, indicating significant protein-protein affinity and a twofold contribution from the Ppant cofactor. Asp222, Glu225, and Arg266 are involved in only the protein-protein interface and do not contact the Ppant (Figure 2.8). Correspondingly, substitutions to these residues had a greater impact on

the affinity of apo-ACP_D than holo-ACP_D. Substitutions to phosphate-interacting Arg33 and Asp214 resulted in equal affinities for holo and apo-ACP_D.

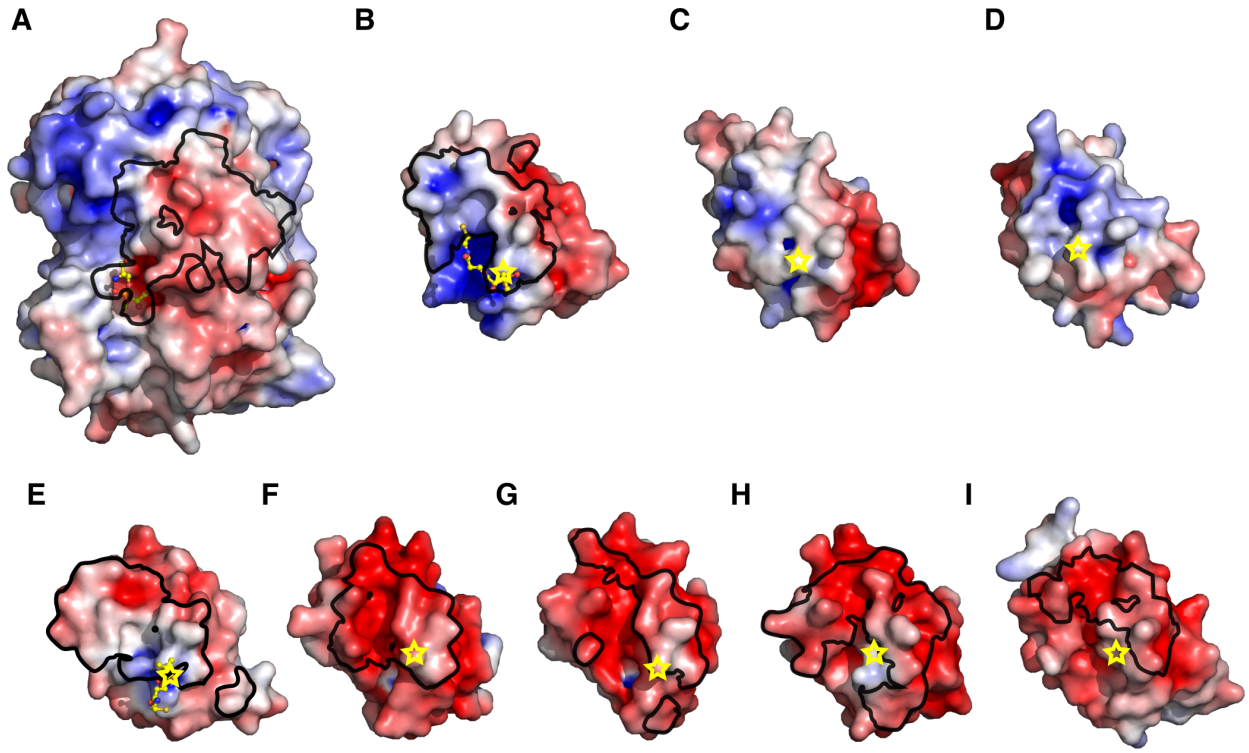


Figure 2.14 Electrostatic surface potentials and interacting surfaces for HMGS and selected ACPs. For ACPs from structures of enzyme complexes, the black outline delineates the molecular surface within 5 Å of any atom in the interacting enzyme and the yellow star denotes the site of Ppant attachment. **A.** CurD HMGS (complex with CurB ACP_D). **B.** CurB holo-ACP_D (complex with CurD HMGS). **C.** CurA ACP_A (2LIW [37], RMSD 2.2 Å). **D.** DEBS module 2 ACP (2JU2 [63], RMSD 2.7 Å). **E.** VinL ACP (5CZD, [68], complex with VinK Acyltransferase, RMSD 2.6 Å). **F.** *E. coli* AcpP (complex with *E. coli* FabA dehydratase, 4KEH [64], RMSD 2.1 Å). **G.** *E. coli* AcpP (complex with *E. coli* LpxD, 4IHG [65],RMSD 3.5 Å). **H.** *B. subtilis* ACP (complex with *B. subtilis* ACP synthase, 1F80 [67], RMSD = 2.2 Å). **I.** *R. communis* ACP (complex with *R. communis* ACP desaturase, 2XZ1 [66], RMSD 1.7). Molecular surfaces are colored by electrostatic potential (±5 kT/e, blue electropositive, red electronegative) [69]. RMSD values are from Cα superposition with CurB ACP_D.

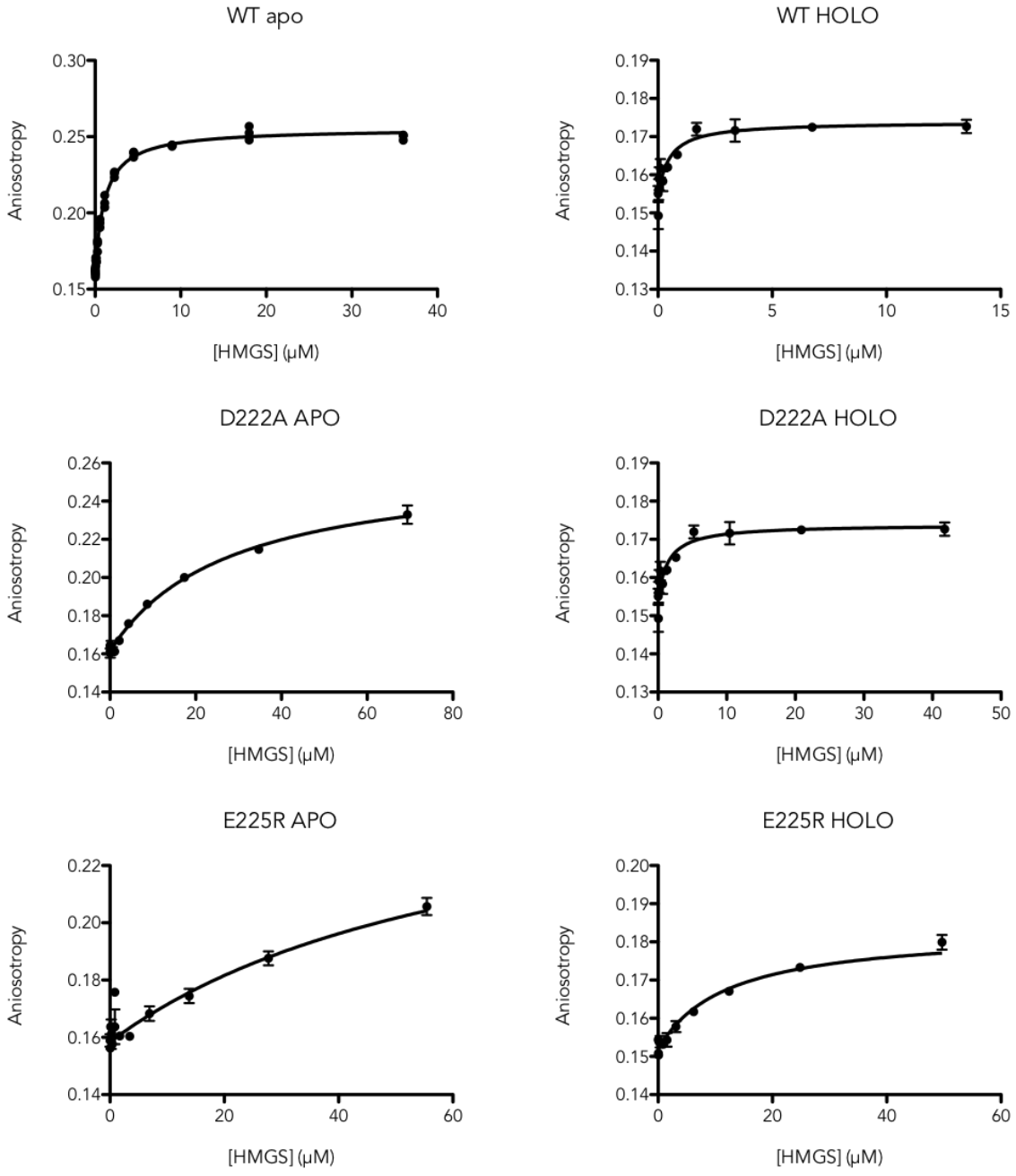


Figure 2.15 Affinity of ACP_D for wild type and variant HMGS. Fluorescence anisotropy of BODIPY-tagged ACP_D was recorded as a function of HMGS concentration. For each HMGS variant, binding curves are shown for apo-ACP_D (left) and holo-ACP_D (right). Data were recorded and analyzed with Graphpad Prism[70]. Data represent the average of 3 measurements.

We evaluated several salt bridges in the ACP_D-HMGS interface by mutagenesis and, for each variant, measured HMGS activity and ACP_D affinity (Table 2.1). Each of the HMGS charged residues (Arg33, Asp214, Asp222, Glu225, Arg266) was substituted with alanine and an oppositely charged amino acid. Affinities were measured by fluorescence anisotropy with a fluorophore-conjugated ACP_D. The HMGS K_d was 1.1 μM for apo-ACP_D and 0.5 μM for holo-ACP_D, indicating significant protein-protein affinity and a twofold contribution from the Ppant cofactor. Asp222, Glu225, and Arg266 are involved in only the protein-protein interface and do not contact the Ppant (Figure 2.8). Correspondingly, substitutions to these residues had a greater impact on the affinity of apo-ACP_D than holo-ACP_D. Substitutions to phosphate-interacting Arg33 and Asp214 resulted in equal affinities for holo and apo-ACP_D.

The effect of the HMGS substitutions on activity did not show a clear pattern for Ppant-interacting and protein-protein contact residues. The Arg33 variants had little or no activity, suggesting that the Arg33 may orient Ppant in the active site. In contrast, substitutions to Asp214 did not significantly affect activity. To further test the importance of HMGS-Ppant interactions, we made alanine substitutions to Ser167 and to Pro166, which we hypothesized would increase helicity of Phe163 and prevent its carbonyl from interacting with Ppant. Despite the conservation of these residues, both variants had similar activity to wild-type HMGS, suggesting their interaction with Ppant may be unimportant for the acetylation step of the HMGS mechanism.

Substitutions to charged side chains in the protein-protein interface caused modest reductions in HMGS activity that were not correlated with changes in affinity,

indicating that HMGS-ACP_D affinity is not limiting in the assay conditions. At each position, the effect of the charge-reversal substitution was more deleterious to HMGS activity (3- to 10-fold) than was the Ala substitution (3-fold or less). The greatest effects occurred for charge-reversal substitutions at Asp222 (5-fold) and Glu225 (10-fold). Located on adjacent turns of helix α8 (Figure 2.8), Glu225 forms a salt bridge with ACP_D Arg59, and the Asp222 carboxylate "caps" ACP_D helix III, which is an atypical 3_{10} helix in an unusual position in the ACP (Figure 2.8). Thus, Asp222 and Glu225 may help orient ACP_D on HMGS, or may be anti-selective for ACP_A at this position.

Discussion

HMGS catalyzes the key reaction of polyketide β-branching, a critical process for chemical diversification in this important class of natural products. The β-branching HMGS of the curacin A biosynthetic pathway exhibits a remarkable selectivity for its donor (ACP_D) and acceptor (ACP_A) substrates (Table 2.1), like related enzymes[24,20,25]. This selectivity enables proper sequencing of substrates during catalysis and prevents aberrant β-branching by mis-association with other ACPs within the PKS pathway.

The ACP_D-HMGS structures capture the ACP_D Ppant in pre- and post-acetyl transfer positions. In the Cys-distal, post-acetyl-transfer position, the Ppant thiol was bound deep within a conserved thiol pocket that occludes the acetyl group, whereas the poorly ordered acetyl-Ppant was in the Cys-proximal position (captured in the Ser114 variant). This contrasts with structures of HMGCS-CoA complexes[53–55,57,71] where, in all cases supported by electron density, the Ppant was bound near the distal site with the thiol directed into the active site cavity regardless of its acylation state. Nevertheless,

the HMGS and HMGCS active sites are nearly identical, and thus we infer that HMGS and HMGCS have identical stereochemical outcome, generating only S-HMG products.

A major motivation for our study was to investigate how HMGS distinguishes the two ACP substrates. The structures of the ACP_D-HMGS complexes indicate that HMGS excludes ACP_A from the ACP_D site, consistent with the inability of ACP_A to act as an acetyl donor in the reaction (Table 2.1). ACP_D has a surface shape that is complementary to the HMGS surface and that differs from surfaces of ACP_A and other ACPs. A hydrophobic cleft, due to the unique position of ACP_D helix III, is matched with a hydrophobic ridge on the surface of HMGS helix α8 (amino acids 213-234) and is flanked by polar contacts, including Asp222 on helix α8 with backbone amides at the N-terminus of ACP_D helix III. ACP_A has no hydrophobic surface cleft because helix III is in a more typical position. None of the salt bridges between amino acids in ACP_D and HMGS was critical for binding or catalysis when tested by single-residue mutagenesis (Table 2.1), leading us to conclude that surface complementarity is the dominant factor in the HMGS-ACP_D interface.

ACPs from PKS pathways[37,45,63,68] have distinctive surface charge distributions compared to FAS ACPs[64–67], but in both systems the ACP helix II-III surface interacts with enzymes (Figure 2.14). The PKS ACPs have regions of positive surface potential near the point of contact with their cognate enzymes, whereas the analogous surface of the FAS ACPs is negatively charged. The distinctive pattern of negative/neutral/positive surface potential of β-branching ACPs (ACP_D, curacin A and mupirocin ACP_As) may contribute to HMGS selectivity against the ACPs within PKS modules where the surface

potential pattern differs (Figure 2.14). The unusual ACP_D helix III, accompanying surface cleft, and striking electrostatics specialize ACP_D for selective interaction with only the β-branching HMGS and KS_{DC} enzymes (Figure 2.2). In contrast, ACP_A shuttles substrates to seven enzymes in the curacin A pathway, including β-branching and module enzymes, necessitating a more promiscuous ACP-surface for enzyme interaction.

The ACP_D-HMGS interface provides clues about the selectivity of β-branching in the myxovirescin pathway, which has two ACP_D/HMGS pairs (TaB/TaC and TaE/TaF). TaB/TaC generates a methyl branch, while the unusual TaE/TaF pair installs an ethyl branch[72]. Each HMGS (TaC and TaF) is highly selective for its donor ACP (acetyl-TaB and propionyl-TaE, respectively) and does not react with the non-cognate ACP_D[20]. We generated homology models of TaB and TaE, based on the CurB ACP_D structure and of TaC and TaF based on CurD HMGS. Some ACP_D/HMGS interactions are recapitulated in both myxovirescin β-branching reactions, and the TaB/TaC pair retains most of the interactions of CurB/CurD. However, at other interacting positions, complementary sequence changes in TaE ACP_D and TaF HMGS would interfere with binding to the non-cognate partner. For example, TaE Lys57 pairs well with TaF Glu267, but not with TaC Lys267; and TaB Arg59 pairs well with TaC Glu225, but not with TaE Lys225.

Catalytic fidelity in a reaction with distinct donor and acceptor substrates is a common problem for HMGS and two homologs: the HMGCS of primary metabolism and the KS domain of modular PKS pathways. The enzymes employ different mechanisms of substrate selectivity, although they have analogous active site entrances for their acyl donor substrates. HMGCS has a single binding site for the acetyl-donor and

acetoacetyl-acceptor CoAs[54,71], but excludes the acetyl-CoA donor following acetyl transfer to the enzyme[31]. Like HMGS, the PKS KS domain has two ACP substrates, but uses separate active site entrances for the donor (upstream module) and acceptor (intra-module)[8]. The PKS KS restricts access of each ACP to the appropriate entrance by the module architecture and ACP tethering via fusion or interaction of docking domains[8,12,14,43,44]. In contrast, HMGS interacts with two ACPs *in trans*, and robust binding of ACP_D (K_d 0.5 μM for holo-ACP_D) is not acyl-group dependent. Nor is the ACP_D docking site on HMGS analogous to either of the ACP-KS docking sites observed in cryo-EM maps of a PKS module[8]. We find no evidence of a second active site entrance in the HMGS structure (for example, poorly ordered loops that could expose the active site, as in the KS[8,13,44,73]). We conclude that ACP_A and ACP_D insert acyl-Ppant through the same active site entrance, as do the donor and acceptor-CoAs of HMGCS, but that the ACPs interact with different regions of the HMGS surface.

The HMGS flexible loop, which does not exist in either HMGCS or KS, is a prime candidate for ACP_A interaction, as it is adjacent to the active site. A critical question is whether the ACPs engage HMGS simultaneously or sequentially. The affinity of HMGS for ACP_D is tenfold greater than a native docking domain pair (K_d 0.5 μM *vs*. 5-25 μM)[43,44]. HMGS has lower affinity for ACP_A than for ACP_D, based on the K_d of 150-200 μM for the bryostatin HMGS and its cognate ACP_A[25]. Cooperativity may enhance weak intrinsic ACP_A binding to HMGS because β-branching cassettes typically encode tandem ACP_As, for example the CurA tandem ACP_A tridomain of nearly identical sequences with an additional dimerization element at the C-terminus[49]. In the HMGS

dimer, the two active site entrances are separated by 40-50 Å, a distance that may be spanned by tandem or dimeric ACPs. However, the high activity of HMGS under assay conditions with equal concentrations of ACP_D and single-domain ACP_A implies that HMGS may either promote dissociation of ACP_D, following acetyl transfer to Cys114, or simultaneously engage acetoacetyl-ACP_A to prevent formation of a dead-end complex. Thus, interaction with acetoacetyl-ACP_A could trigger both catalytic and conformational events, including rearrangement of the flexible loop to disengage ACP_D or to widen the active site entrance to accommodate two Ppant cofactors. The possibilities are not resolved by the present structures.

In conclusion, the first structure of an HMG synthase involved in polyketide biosynthesis reveals features that distinguish HMGS from its primary metabolism homolog and allow it to interact selectively with its cognate ACPs. Analysis of the HMGS-ACP_D interface provides insight into HMGS selectivity in other pathways, including those with multiple β-branching functions. The HMGS structures with acetyl and holo-ACP_D provide a unique view of the molecular interactions between a PKS enzyme and its substrate, revealing the mechanism by which HMGS prevents its substrate from adopting a non-productive orientation in the active site. Finally, the unusual position of helix III in ACP_D is a new structural motif in acyl carrier proteins that can be selectively recognized by specialized enzymes such as HMGS.

Methods

Protein expression and purification

Our expression plasmid for the curacin A HMGS (pHMGScur) was generated by amplifying the CurD gene from the pLM54[52] cosmid (Table 2.5) and inserting it into pMCSG7[74] by ligation independent cloning (LIC). The *curB* gene encoding ACP_D was amplified from a pET28b construct[74] (Table 2.5) and inserted by LIC into pMCSG7 ($pACP_d$cur). pHMGScur was coexpressed with pGro7[75] (Takara) in *E. coli* BL21 (DE3) cells. $pACP_D$cur was coexpressed with pRARE-CDF[44] in *E. coli* BL21 (DE3) (for apo/acyl-loaded proteins) or Bap1[76] (for holo proteins) cells containing pRARE. *E. coli* BL21 (DE3) cells were transformed with pLG003[49] for expression of CurA ACP II, the second of the CurA tandem ACP_A tridomain. Transformed bacteria were grown with 100 mg/L ampicillin and 35 mg/L chloramphenicol for HMGS, 100 mg/L ampicillin and 50 mg/L spectinomycin for ACP_D, or 50 mg/L kanamycin for ACP_A at 37 °C in TB to an OD_{600} of 1. GroEL/ES chaperone expression was induced with 2 g/L arabinose for HMGS cultures at an OD_{600} of 0.3. Cells were cooled to 20 °C and target gene expression was induced with 200 μM IPTG. After 18 hr of expression, cells were harvested and stored at -20 °C. Cell pellets were resuspended in buffer A (300 mM NaCl, 50 mM tris pH 7.5, 20 mM imidazole pH 8.0, and 10% glycerol for ACPs; 50 mM $(NH_4)_2SO_4$, 50 mM HEPES pH 7.0, 20 mM imidazole pH 8.0, and 10% glycerol for HMGS). Cell suspensions were treated with 1 mg/mL lysozyme (Sigma), 2 mM $MgCl_2$, and 20 U/mL Pierce universal nuclease for 30 minutes at room temperature (RT), then cooled on ice. Cells were lysed by sonication and cleared by centrifugation. Each

protein was purified by immobilized metal affinity chromatography on a 5 mL His Trap column (GE Healthcare) with a 50 mL gradient from 20 mM (buffer A) to 400 mM imidazole (buffer B), followed by size exclusion chromatography (SEC). HMGS was purified on a Superdex S200 column in 50 mM $(NH_4)_2SO_4$, 20 mM HEPES pH 7.0, and 10% glycerol (HMGS buffer C), and ACPs on a Superdex S75 column in 150 mM NaCl, 20 mM Tris pH 7.5, and 10% glycerol (ACP buffer C). Proteins were concentrated to ~10 mg/mL, flash cooled in liquid N_2 and stored at -80 °C.

Site directed mutagenesis

Constructs encoding HMGS and ACP_D variants were generated by site directed mutagenesis (SDM) on pHMGScur and pACP_Dcur by PCR with overlapping primers (Table 2.5). SDM PCR components were 2 ng/µL of pHMGScur or pACP_Dcur, 200 nM sense and 200 nM antisense primer, 0.05 U/µL Pfu Turbo polymerase (Agilent), and 250 nM dNTPs, in 1X Pfu buffer (Agilent). After PCR, 0.4 U/µL DpnI (NEB) was added to each reaction and incubated for 2 hr at 37 °C. 2µL of this mixture was transformed into *E. coli* XL-1 Blue cells. Mutant expression constructs were purified by miniprep kit (Qiagen), and the mutations were verified by Sanger sequencing.

Table 2.5. Primers for cloning and mutagenesis of CurD HMGS and CurB ACP_D. Ligation independent cloning (LIC) overhangs are bolded.

Cloning Primers

Construct	Sense Primer	Antisense Primer
pHMGScur	**TACTTCCAATCCAATGCT**ATGCAACAAGT TGGC	**TTATCCACTTCCAATGCTA**TACCCACTCG TATTTTCGG
pACP_Dcur	**TACTTCCAATCCAATGCA**ATGAGCAAgGA ACAAGTAC	**TTATCCACTTCCAATGCTA**CAATTTTGCT GCA

Mutagenic Primers

Mutation	Sense Primer	Antisense Primer
HMGS: C114S	tttgaactcaagcaagctagctactc	cggttcctgagtagctagcttgcttga

	aggaaccg	gttcaaa
HMGS: AAA	gggctgatataccttgttgagctgca attgccgccccttctggagtcacaac cccac	gtggggttgtgactccagaaggggcgg caattgcagctcaacaaggtatatcag ccc
HMGS: P166A	ccagcaccactactggcttcagcaaa agaccaatc	gattggtcttttgctgaagccagtagt ggtgctgg
HMGS: S167A	ggctccagcaccactagcgggttcag caaaagac	gtcttttgctgaacccgctagtggtgc tggagcc
HMGS: R33A	catcattagattgtcaaaggcggaga tatccaactgacgc	gcgtcagttggatatctccgcctttga caatctaatgatg
HMGS: D214A	aagaaagcaaggataaggcggcatct cctgcttct	agaagcaggagatgccgccttatcctt gctttctt
HMGS: D222A	taggcattttcgcaacaggctaggta agaaagcaagg	ccttgctttcttacctagcctgttgcg aaaatgccta
HMGS: E225A	taatgtcggtaggcatttgcgcaaca gtctaggtaag	cttacctagactgttgcgcaaatgcct accgacatta
HMGS: R266A	ggtttagctcttttcaacctagccat catatttctatgagcgcc	ggcgctcatagaaatatgatggctagg ttgaaaagagctaaacc
HMGS: R33D	tcatcattagattgtcaaagtcggag atatccaactgacgcg	cgcgtcagttggatatctccgactttg acaatctaatgatga
HMGS: D214R	gtaagaaagcaaggataagcgggcat ctcctgcttctgaa	ttcagaagcaggagatgcccgcttatc cttgctttcttac
HMGS: D222R	ggcattttcgcaacagcgtaggtaag aaagcaaggataagtc	gacttatccttgctttcttacctacgc tgttgcgaaaatgcc
HMGS: E225R	ttgataatgtcggtaggcatttctgc aacagtctaggtaagaaagc	gctttcttacctagactgttgcagaaa tgcctaccgacattatcaa
HMGS: R266E	gcaggtttagctcttttcaacctctc catcatatttctatgagcgcctt	aaggcgctcatagaaatatgatggaga ggttgaaaagagctaaacctgc
ACP_D: M1C	atttttagtacttgttctttgctgca tgcattggattggaagtacaggttct cg	cgagaacctgtacttccaatccaatg catgcagcaaagaacaagtactaaaa at
ACP_D: R42A	catcatgataatttctgccgcattaa ctgaatcgatacctaatttttttaag ctatca	tgatagcttaaaaaaattaggtatcga ttcagttaatgcggcagaaattatcat gatg

Activity assay

ACPs in the apo state were loaded *in vitro* with acyl-Ppants by incubation with *Streptomyces verticillus* phosphopantetheinyl transferase[77] (SVP), in a 10:1 ratio of ACP:SVP, and acyl-CoA, in a 1:10 ratio of ACP:CoA, in a buffer of 10 mM $MgCl_2$, 50 mM Tris pH 7.5 for 3 hr at 30 °C. Acylated ACPs were purified from loading reaction components by SEC on a Superdex S75 column equilibrated with ACP buffer C. The

HMGS assays were performed at 25 °C in a buffer of 20 mM $(NH_4)_2SO_4$, 20 mM HEPES pH 7.0, and 2 mM *tris* (2-carboxyethyl)phosphine (TCEP) for 10 min, followed by quenching with 5% formic acid. The concentrations of assay components were 5 μM HMGS, 50 μM acetyl-ACP_D, and 50 μM acetoacetyl-ACP_A. ACP_A was separated from other reaction components by HPLC on a Jupiter C4 column with a gradient of 30% to 95% acetonitrile in 0.1% trifluoroacetic acid (TFA) over 15 min with a flow rate of 0.2 mL/min. The ACP_A peak fraction was stored at -20 °C until mass spectrometry analysis by Q-TOF (Agilent) with phosphopantetheine ejection[59]. Total ion counts for holo, acetyl, acetoacetyl, and HMG-loaded Ppant species were recorded (Agilent MassHunter software). Reaction progress was calculated by the percent of all ejected Ppant ions loaded with HMG and averaged for 3 replicates.

ACP_D Hydrolysis Assay

100 μM acetyl-ACP_D was incubated 18 hr at 20 °C in a 20 mM $(NH_4)_2SO_4$ + 1X MMT pH 6.5 (Qiagen) with 10 μM $HMGS_{WT}$, 10 μM $HMGS_{C114S}$, or alone. Samples were acidified with 5% formic acid and analyzed by HPLC and MS Ppant ejection as described above. An additional 100 μM acetyl-ACP_D was analyzed without incubation. Percent loading of the ACP_D with acetyl was determined using ion counts for acetyl-Ppant and holo-Ppant species.

Fluorescence anisotropy binding assay

Met1 of CurB ACP_D was mutagenized to cysteine. ACP_D M1C was reduced for 15 min at RT with 2 mM TCEP, then incubated with 10-fold excess BODIPY-FL 1-iodoacetamide (Invitrogen) at room temperature for 3 hr. BODIPY-FL-conjugated ACP_D

was purified from free BODIPY using a PD-10 column with ACP buffer C, flash frozen, and stored at -80 °C. HMGS was dialyzed overnight at 4 °C into 20 mM $(NH_4)_2SO_4$ and 20 mM HEPES pH 7.0. 10 serial 2X dilutions of the HMGS stocks were made, then 45 µL of each of these 11 HMGS samples was mixed with 5 µL of 0.1 nM ACP_D-1-BODIPY in an opaque 384 well plate. 45 µL buffer was mixed with 5 µL of 0.1 nM ACP_D-1-BODIPY for a control. The plate was incubated at RT for 30 min, then fluorescence anisotropy was measured with a λ_{EX} of 485 nm and λ_{EM} of 525 nm (Figure 2.15). The data were fit to $A = A_{free} + (\Delta A_{bound} \times [HMGS]/ (K_d+[HMGS]))$ (GraphPad Prism) to determine affinities. In cases where ΔA_{bound}, the change in anisotropy when ACP_D is maximally bound, could not be fit from the data, ΔA_{bound} was held at a constant value as determined from the average of trials where it was experimentally measured.

Crystallization

Substitutions to HMGS to make it more amenable to crystallization were predicted using the surface entropy reduction server[78]. Using these predictions, a triple alanine variant, $HMGS_{AAA}$, (K344A/Q345A/Q347A) was generated by SDM. $HMGS_{AAA}$ was crystallized by hanging-drop vapor diffusion from 1:1 and 2:1 mixtures of protein stock (10 mg/mL in HMGS buffer C) and well solution (20 mM $(NH_4)_2SO_4$, 1X MMT buffer pH 6.5 (Qiagen), 6% w/v PEG 8000) at 20 °C. For HMGS/ACP_D complex crystals, each protein was concentrated to 25 mg/mL then mixed in an HMGS:ACP_D ratio of 2:3. HMGS/ACP_D (apo) co-crystals were obtained with a well solution of 100 mM $(NH_4)_2SO_4$, 1X MMT buffer pH 6.5, 2% w/v PEG 8000. HMGS/ACP_D (holo) crystallized with a well solution of 120 mM $(NH_4)_2SO_4$, 1X MMT buffer pH 6.5, 10% w/v PEG 8000 at 20 °C.

Crystals were cryoprotected in well solution with 25% v/v glycerol prior to flash cooling in liquid N_2. Crystallization trials using refined conditions and $HMGS_{WT}$ (again cocrystallized with a 7:1 molar ratio of ACP_D:HMGS) yielded only microcrystals that were not suitable for harvesting or diffraction.

Structure solution and refinement

Data were collected at the Advanced Photon Source (APS) GM/CA 23ID-D and processed with XDS[79]. Diffraction limits were determined by $CC_{1/2}$ and I/σ_I statistics, and subsequent calculations were done with the ccp4 suite[80]. The HMGS structure was solved by molecular replacement in phaser[81] from an HMG-CoA synthase model[55] (1YSL) that was trimmed with chainsaw[82]. A 97%-complete initial model was auto-built using ARP/wARP[83], model refinement was done in refmac5[84] with TLS[85] parameters, and real space building with Coot[86]. Waters were manually added in coot to positive F_O-F_C difference density in positions consistent with hydrogen bonding to the protein. Waters were removed after refinement if they corresponded to negative difference density, had 2Fo-Fc density weaker than 0.6σ, or resulted in clashes. HMGS/ACP_D complex structure was solved by molecular replacement with the HMGS structure, followed by manual building of the ACP_D model in Coot and refinement in refmac5 with TLS parameters. Despite high B-factors for residues distal from HMGS that had no crystal contacts, we found continuous main chain density for residues 1-78 in all deposited structures, and density for side chains contacting HMGS was well resolved. Subsequent HMGS/ACP_D complex structures were solved by rigid body refinement with HMGS and ACP_D models. To validate the Ppant position in our model of the pre-acetyl

transfer complex, we collected data on holo-ACP_D/$HMGS_{C114S}$ crystals at 12 keV and Friedel-pair data at an X-ray energy of 7 keV on acetyl-ACP_D/HMGS (C114S) and HMGS (C114S) crystals. Data were processed with XDS, and phased by rigid body refinement in refmac5. Following model refinement and addition of waters, sulfur atomic positions were identified in an anomalous difference map The peak heights for the holo and acetyl thiol positions were 4.0 and 3.7, respectively, and peaks for cysteines and methionines ranged from 4.0 to 8.1. Refined models were validated and Ramachandran plots were generated by Molprobity[87], followed by a final round of building in coot and refinement in refmac5. Molecular figures were generated in PyMOL[88].

Homology modeling of myxovirescin HMGS and ACP_D

Homology models of the TaB and TaE ACP_Ds and of the TaC and TaF HMGSs were generated using Modeller[89] using sequence alignments by Clustal Omega[60] (Figure 2.6 2.7, 2.12) and our structure of the holo-ACP_D/HMGS complex. Models for TaC and TaF were aligned to CurD HMGS and models for the TaB and TaE ACP_Ds were aligned to holo-CurB ACP_D in PyMOL for analysis of the TaB/TaC and TaE/TaF complexes.

Chapter 3. Importance of conserved structural motifs to HMGS reactivity and ACP interaction

Introduction

Modular polyketide synthase (PKS) pathways are enzyme assembly lines that process acyl-CoA building blocks to create complex, bioactive small molecules[4–6]. Each module is a set of fused catalytic domains and an acyl carrier protein (ACP) domain, which tethers intermediates via a phosphopantetheine arm, shuttles them to each enzymatic domain in the module, and passes the intermediate to the next module[10]. Modules include enzymes to extend the polyketide and to make modifications after extension, allowing different functional groups to be incorporated into each building block of the natural product. Polyketides are an astoundingly diverse group of natural products and an important source of pharmaceuticals[50]. The architecture of PKS pathways has attracted interest as a target for engineering, as successfully reprogramming the assembly line could allow production of new pharmaceutical leads[2,3]. To this end, biochemically characterizing PKS enzymes that introduce unusual functional groups may lead to new enzymatic tools for engineering pathways or as reagents for chemoenzymatic synthesis.

Several modular PKS pathways include enzymes for alkylation of the nascent polyketide at a specific intermediate in the assembly line (β-branching)[18]. This invariably occurs after extension by a module with no modification enzymes, which leaves the β-

carbonyl generated by the ketosynthase extension enzyme intact. β-branches are often methyl groups, but also include many unique or rarely-seen substituents in polyketides (Figure 1.3). In β-branching, a hydroxymethylglutaryl synthase (HMGS) catalyzes the initial aldol addition of an acetyl nucleophile to the β-carbonyl intermediate generated by polyketide extension[24,25]. The HMGS reaction generates a chemical starting point that can be tailored to a diverse array of alkyl substituents by the other β-branching enzymes[18]. HMGS is selective for an ACP within the biosynthetic pathway (the acceptor ACP, ACP_A) that tethers the intermediate subject to branching, and a standalone ACP (the donor ACP, ACP_D) found with the β-branching enzymes that delivers the donor acyl group for branching. Each of these ACPs has distinguishing sequence motifs, and HMGS does not react with substrates delivered by CoA or non-branching ACPs[24,20,25,90]. Additionally, HMGS is weakly reactive when ACP_D delivers both substrates and nonreactive when both are delivered by ACP_A[90].

The mechanism by which HMGS catalytically differentiates between substrates tethered by each ACP is unknown. HMGS, like many PKS enzymes, was at some point poached from a primary metabolism pathway and adapted to polyketide biosynthesis. The primary metabolism homolog of HMGS is HMG-CoA synthase (HMGCS), which catalyzes the aldol addition of acetyl-CoA to acetoacetyl-CoA (Figure 2.2)[30]. The structure and biochemistry of HMGCS has been extensively characterized. HMGS is very similar to HMGCS but has a distinctive structural elements that may play a role in its reaction and ACP selectivity (Figure 3.1).

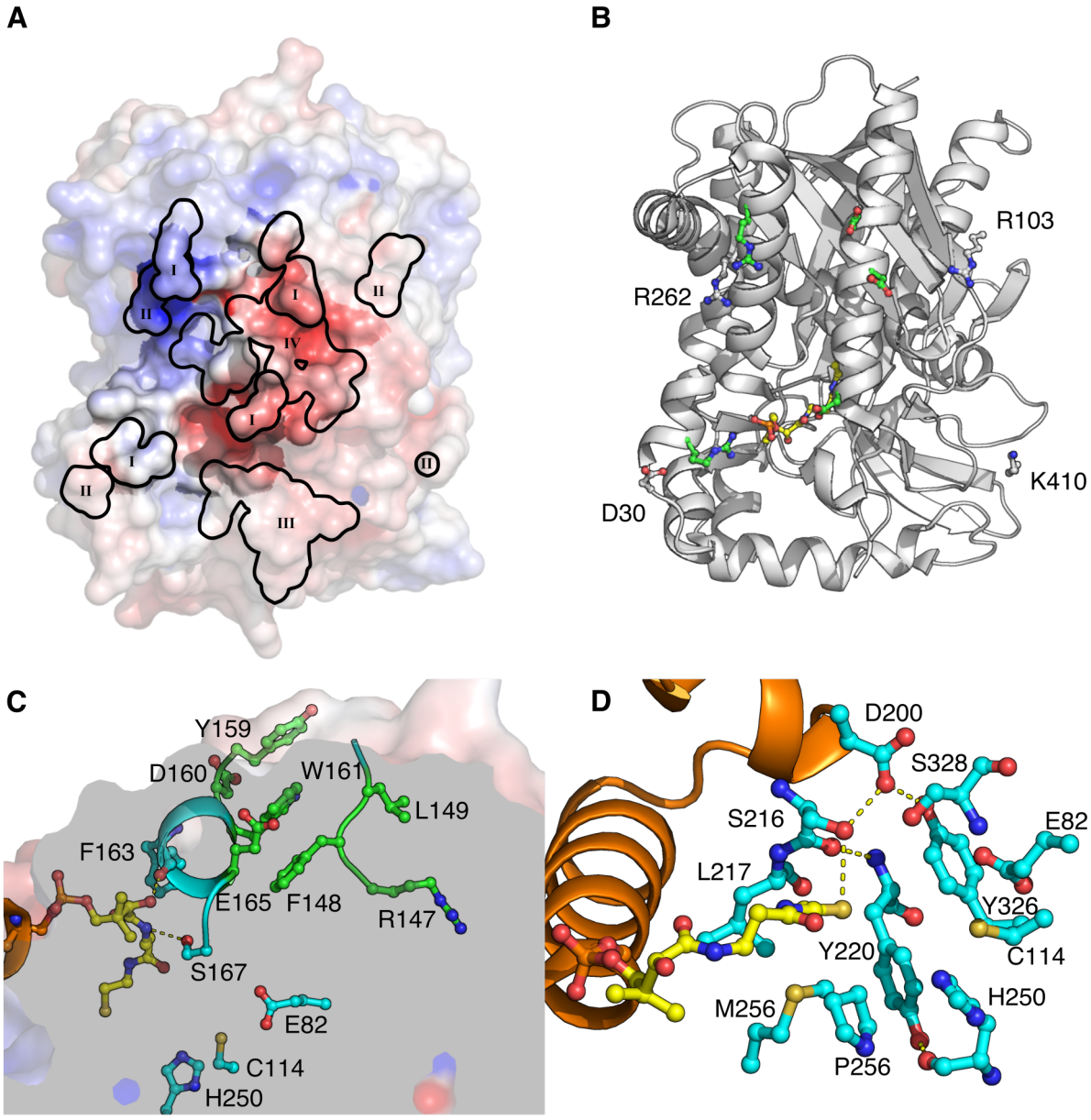


Figure 3.1 Structure of conserved structural elements of HMGS. A. Electrostatic surface of HMGS and conserved residues**.** The HMGS surface within 5 Å of ACP_D is opaque with the rest transparent. Outlined residues are as follows: I, charged residues previously studied for interaction with ACP_D[90]; II, previously unstudied conserved charged residues; III, conserved residues within the HMGS flexible loop; IV, residues of the thiol pocket. **B.** Surface charged residues (I, II)**.** Green residues correspond to region II and the Ppant is shown in yellow, as ordered in the structure of the HMGS-ACP_D complex (5KP7). Arg103, Arg382, and Lys410 extend from the partner subunit of the HMGS homodimer. **C.** HMGS flexible loop (III). Active site residues and those that interact with Ppant are colored in cyan, other conserved residues of the flexible loop are colored in green. Tyr159, Trp161, and F148 form a π-stack. Leu149 and Arg147 project

towards the interior of HMGS with Asp160 exposed on the surface. **D.** Residues of the thiol pocket (IV). Active site residues and residues of the thiol pocket are shown in cyan, the Ppant in yellow, and ACP_D in orange. The Ppant thiol makes a hydrogen bond to Ser216, which in turn interacts with Asp200 and Tyr326. Tyr220 packs against the thiol and makes a hydrogen bond to the carbonyl of H250, an essential catalytic residue. Leu217 appears to be important for both defining the thiol pocket and making a hydrophobic contact with ACP_D.

The most prominent of these structural elements is an insertion in HMGS relative to HMGCS that corresponds to a flexible loop near the active site entrance. The loop itself has several conserved positions (Appendix 3.1), including an "EGGEAL" motif (EGGEAI in CurD, residues 152-157), which is disordered in the HMGS structure. HMGCS sequences from bacteria include part of the motif but not the full HMGS flexible loop, and HMGS that use a propionyl donor group lack the motif entirely. The GGE in structures of *S. aureus* (1TVZ) and *E. faecalis* (1X9E) HMGCS is ordered and corresponds to Ala164 and Phe163 in HMGS, implying a different role for this motif in HMGCS and HMGS. The C-terminal end of the flexible loop orders upon holo-ACP_D binding and forms a helix that makes two hydrogen bonds to the Ppant (Figure 3.1C). This helix appears to be further stabilized by a π-stacking interaction between Phe148, Trp161, and Tyr159. Tyr159 and Trp161 are part of a conserved 159-EDWSF/Y-163 motif (Y159 in CurD is usually E159). Though Asp160 in this motif is nearly invariant among HMGS (excluding those that use propionyl), it hangs out into space in the HMGS structure and has no apparent importance.

HMGS and HMGCS both contain a hydrophobic pocket in the active site. In HMGS, the Ppant thiol of holo-ACP_D localizes to this pocket and makes a hydrogen bond to Ser216. Positioning of the thiol in the thiol pocket may be incompatible with

acetyl transfer due to an 8 Å distance between the holo-Ppant thiol and the catalytic cysteine. In contrast, acetyl-Ppant adopts a different conformation that positions the thiol near to the cysteine[90].Residues pack tightly around the Ppant thiol and could occlude acetyl-Ppant from the deepest end of the pocket, preventing the substrate from adopting a non-reactive conformation in the active site. In structures of HMGCS with bound CoA, the Ppant does not penetrate as deeply into the active site and the thiol angles toward the HMGCS cysteine in a catalytic position.

Finally, the electrostatic landscape surrounding the active site is different in HMGS and HMGCS (Figure 3.2). HMGCS has regions of positive charge that interact with the 5' diphosphate and 3' phosphate of CoA. The surface surrounding the HMGS active site has positive and negative areas that are separated by a hydrophobic stripe. These regions complement the surface charges of ACP_D and include conserved acidic and basic residues that make salt bridges to the ACP. These regions may exclude other ACPs from docking to HMGS, as ACP_D has an atypical electrostatic surface.

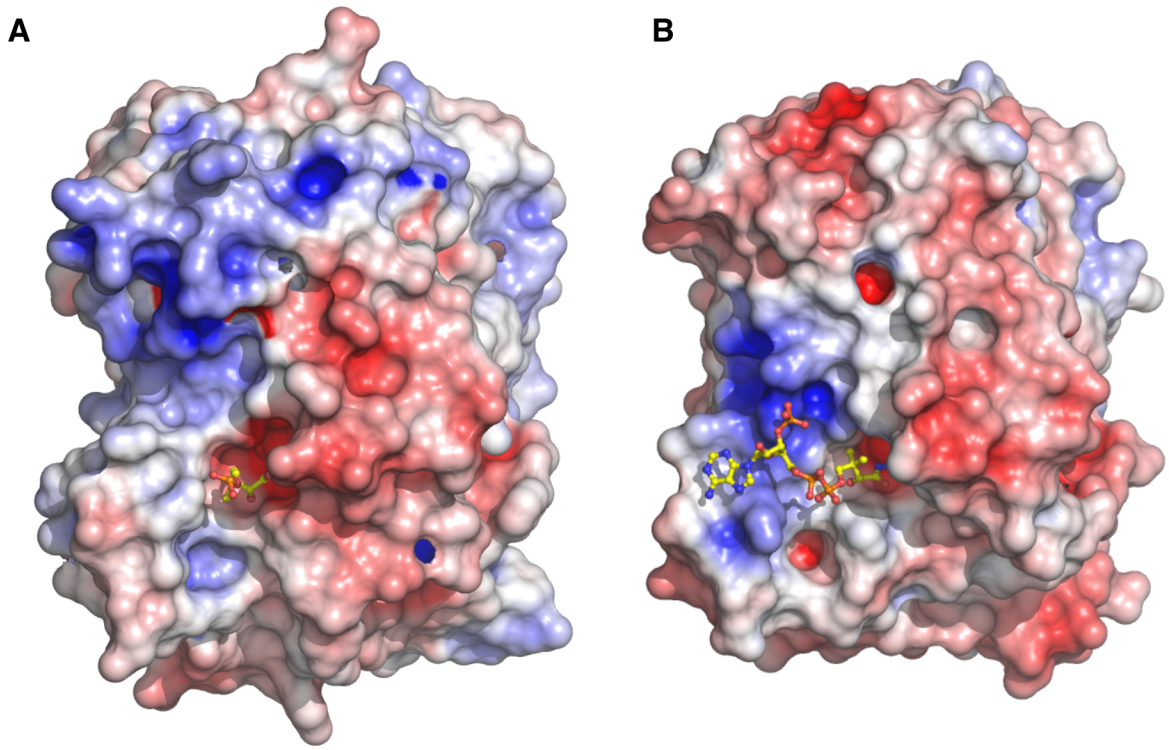


Figure 3.2 Electrostatic comparison of HMGS and HMGCS. A. Curacin A HMGS (5KP7)[90]**. B.** *E.faecalis* HMGCS (1YSL)[55]**.**

In this chapter we evaluate the importance of these conserved structural elements to HMGS reactivity, with a secondary goal of determining how HMGS distinguishes between ACP_A and ACP_D linked substrates. We observed a critical dependence of HMGS activity on residues within the thiol pocket region, several of which also affected ACP_D affinity. Apart from a few residues, reductions in ACP_A affinity were modest. The results provide a groundwork for studying ACP interactions with the curacin A HMG, more definitively resolve questions about how ACPs interact with HMGS, and identify new residues that are critical to HMGS catalysis.

Results

Dependence of ACP interactions on surface electrostatics. Based on the previous observation that ACP_D binding to HMGS was strongly influenced by charge-charge interactions[90], we identified conserved charged residues on the surface of HMGS that were not conserved in HMGCS and did not appear to interact with ACP_D (Figure 3.1B, 3.2). Most HMGS variants behaved well through purification and were assayed for activity, holo-ACP_D affinity, and holo-ACP_A affinity (Table 3.1).

Table 3.1 Surface electrostatic substitutions

	HMGS activity[1] (% conversion)	ACP_D K_d[2] (μM)	ACP_A K_d[2] (μM)
WT	78.3 ±0.9	0.2 ± 0.1	8 ± 2
D30A	50 ± 1	16 ± 11	5 ± 2
R103A	47 ± 2	7 ± 3	4 ± 2
K410A	71.9 ± 0.9	1 ± 0.5	8 ± 2
R262A	27 ± 8	0.8 ± 0.3	33 ± 17
R33A	N.D.[3,4]	8 ± 2[5]	25 ± 13
R33D	2 ± 3[3]	25 ± 3[5]	N.B.[6]
D214A	86 ± 8[3]	9 ± 2[5]	10 ± 2
D214R[7]	45 ± 4[3]	11 ± 3[5]	--
E225A[8]	55 ± 3[3]	--	--
E225R	8 ± 1[3]	12 ± 1[5]	11 ± 3
D222A	27 ± 1[3]	1.5 ± 0.6[5]	15 ± 7
R266A	26 ± 1[3]	17 ± 5[5]	19 ± 13

[1]Conversion of equimolar acetyl-ACP_D and acetoacetyl-ACP_A to HMG-ACP_A in 7 min at 25 °C, averaged for 3 reactions.
[2]Affinity was measured by fluorescence polarization using Bodipy-FL conjugated ACPs.
[3]Previously published data[90]. Assay conditions were slightly different. WT HMGS converted 82.8 ± 0.3 % of substrate to HMG-ACP_A.
[4] No product detected in activity assay.
[5]Previously published data[90]. Buffer for affinity assay was 20 μM $(NH_4)_2SO_4$, 20 μM HEPES pH 7.0. Measured HMGS-ACP_D affinity was 0.5 ± 0.2 μM.
[6]No binding detected by fluorescence anisotropy. See Figure 3.5 for binding data.
[7]$HMGS_{D214R}$ was unstable in low salt buffer for the fluorescence anisotropy experiment and reliable binding data for ACP_A interaction could not be obtained.
[8]$HMGS_{E225A}$ was unstable during protein production/purification and only low amounts of stable protein could be obtained for activity assays.

D30A, R103A, K410A, and R262A substitutions did not significantly impact HMGS activity or ACP_A affinity. $HMGS_{D30A}$ and $HMGS_{R103A}$ had 80-fold and 35-fold decreases in affinity for ACP_D, although they do not directly contact ACP_D. As previously observed[90], decreases in ACP affinity are not correlated with loss of activity. $HMGS_{R33A}$ and $HMGS_{R33D}$ are similarly inactive, yet R33D has a larger impact on ACP_A and ACP_D affinity, with the ACP_A affinity too weak to be measured. $HMGS_{R262A}$ had 4-fold reduced activity, signifying it may be important to the ACP_A interface.

We also investigated the possibility of competition between ACP_D and ACP_A binding using our FP assay (Figure 3.3). We mixed 2 -16 μM unlabeled holo-ACP_D with 8.8 μM HMGS and 20 nM fluorescent ACP_A. At concentrations of unlabeled ACP_D greater than ~8 μM, where most HMGS would be bound by unlabeled ACP_D, we observed reduction in fluorescence anisotropy of ACP_A, as would be predicted if ACP_D and ACP_A binding to HMGS were mutually exclusive. As a positive control, we repeated the experiment with unlabeled holo-ACP_A (5-80 μM) and observed a similar reduction in ACP_A fluorescence anisotropy. Thus the ACP_A binding site on HMGS may overlap with that of ACP_D, at least in part, but ACP_A does not show the same strong dependence on charge-charge interactions we observe for ACP_D.

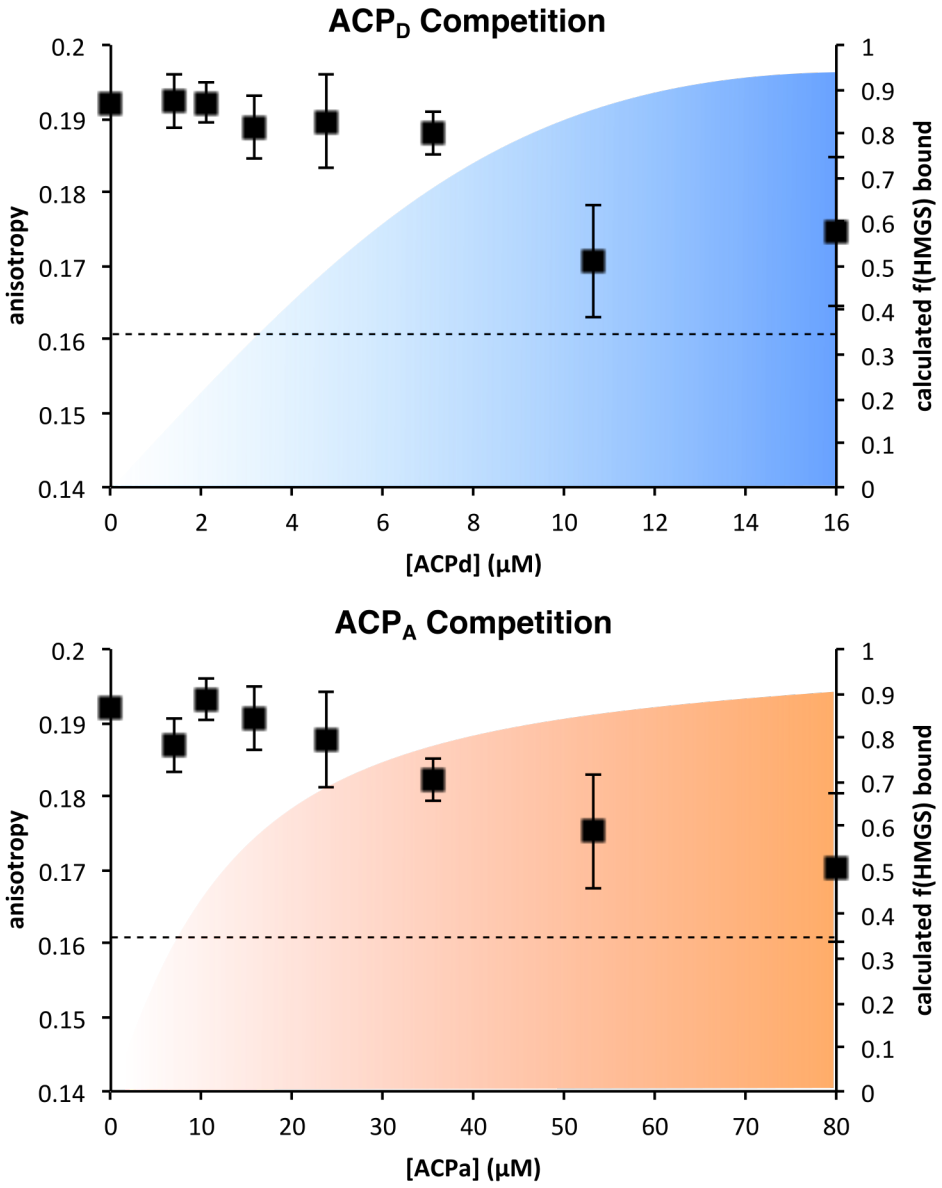


Figure 3.3 ACP competitive binding. Fluorescence anisotropy data for fluorescent ACP_A binding to HMGS is shown. Various concentrations of unlabeled ACP_A or ACP_D were mixed with a constant 8.8 μM HMGS and 20 nm fluorescent ACP_A. Dotted lines signify fluorescence anisotropy for ACP_A in the absence of HMGS (0.16 ± 0.01). At each concentration, the fraction of HMGS bound by the unlabeled ACP was calculated, based on the measued ACP_A K_d of 8 μM, and is shown as a faded background.

Importance of the active site thiol pocket to HMGS activity. We observed an extensive hydrogen-bonding network in the HMGS active site, which included residues previously identified as important to HMGS and HMGCS catalysis, as well as residues that interact with the ACP_D Ppant thiol. Our hypothesis for the role of this pocket during donor-acyl transfer was that it sterically occluded acetyl-Ppant before transthioesterification, preventing the substrate from adopting a non-catalytic pose. Lacking a structure of the ACP_A-HMGS complex, we additionally sought to determine whether these conserved residues were important for the aldol addition step. We made alanine substitutions to thiol pocket residues (Figure 3.1D), as well as a P252G substitution and phenylalanine substitutions of Tyr220 and Ty4326. We assayed these HMGS variants for activity and ACP affinity using the methods described above (Table 3.2).

Table 3.2 Thiol pocket substitutions.

	HMGS activity[1] (% conversion)	ACP_D K_d[2] (μM)	ACP_A K_d[2] (μM)
WT	78.3 ± 0.9	0.2 ± 0.1	8 ± 2
D200A	< 5[3]	--	9 ± 4
S216A	< 5	1.1 ± 0.5	10 ± 2
L217A	< 5	2.3 ± 0.7	19 ± 6
Y220A	21 ± 2	22 ± 17	16 ± 6
Y220F	58 ± 2	6 ± 2	49 ± 14
P252G	< 5	17 ± 13	--
M256A	23.9 ± 0.8	2 ± 1	5 ± 1
Y326A	< 5	100 ± 40	12 ± 5
Y326F	84 ± 4	17 ± 5	11 ± 2
S328A	21 ± 1	0.9 ± 0.3	5.6 ± 0.8

[1]Conversion of equimolar acetyl-ACP_D and acetoacetyl-ACP_A to HMG-ACP_A in 7 min at 25 °C, averaged for 3 reactions.
[2]Affinity was measured by fluorescence polarization using Bodipy-FL conjugated ACPs.
[3]Approximate detection limit was 1 μM acyl-ACP_A, or less than 5% conversion of substrate to product.

Several residues in the thiol pocket were essential to HMGS activity. $HMGS_{Y220F}$ and $HMGS_{Y326F}$ had activity similar to wild-type, with only a ~25% reduction in activity from Y220F. We did observe a ~85 fold reduction in the affinity of $HMGS_{Y326F}$ for ACP_D, presumably due to the loss of the hydrogen-bond of Tyr326 to Asp200. Due to the complete loss of HMGS activity from some other thiol pocket substitutions, we tested acetyl transfer from ACP_D to these HMGS variants to determine whether donor transfer and/or aldol addition was impaired (Figure 3.4).

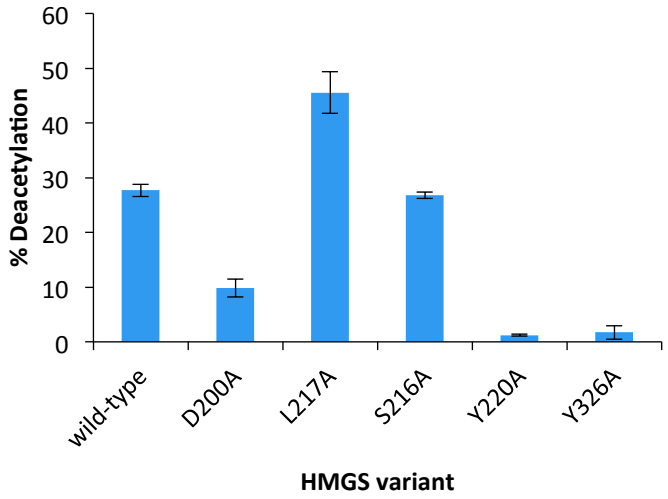


Figure 3.4 Deacetylation of ACP_D by HMGS. Mixtures of 20 µM acetyl-ACP_D and 50 µM HMGS were incubated for 30 sec at 25 °C, then acidified with formic acid and analyzed by LC-MS with Ppant ejection[59]. The percent of Ppant ions loaded with acetyl was averaged for three replicate incubations with each HMGS variant, then subtracted from a no-enzyme control to calculate percent deacetylation.

We observed no activity from $HMGS_{Y326A}$ and a 4-fold reduction in activity from $HMGS_{Y220A}$. Tyr220 appears more directly involved in interaction with the ACP_D-Ppant thiol and makes a hydrogen bond to the His250 backbone carbonyl, while Tyr326 sits more deeply in the active site, makes a hydrogen bond to Asp200, and packs against Ser328. Substitution of Tyr220 with alanine or phenylalanine both disrupted interaction

with ACP_A, although the alanine substitution was more deleterious to HMGS activity. $HMGS_{S328A}$ also displayed a 4-fold decrease in HMGS activity, presumably due to the role of the Ser328 amide in stabilizing the donor enolate intermediate.

Leu217 appears to make a hydrophobic contact and to play a role in packing of the thiol pocket around the Ppant. We observed a ~4-fold decrease in ACP_D affinity, a 2-fold decrease in ACP_A affinity, no $HMGS_{L217A}$ activity, and increased donor-transfer, supporting a role in either active site organization or Ppant positioning, but not in ACP interaction. $HMGS_{S216A}$ similarly had no activity, unimpaired donor-transfer, and wild-type affinities for ACP_D and ACP_A. Ser216 makes a hydrogen bond to both the ACP_D-Ppant thiol and the Asp200 carboxylate. Asp200 also makes a hydrogen bond to Tyr326 and $HMGS_{D200A}$ was inactive. $HMGS_{D200A}$ shows a ~3-fold decrease in donor-transfer, though deficient donor-transfer does not explain lack of overall activity, as $HMGS_{Y220A}$ displayed negligible donor-transfer but a more robust conversion of acetoacetyl-ACP_A to HMG-ACP_A.

The final two tested substitutions, P252G and M256A were both deleterious to HMGS activity. Pro252, in addition to interacting with the ACP_D Ppant is near the catalytic His250. Whereas $HMGS_{M256A}$ had a 4-fold reduction in activity and a modest decrease in ACP_D affinity, $HMGS_{P252A}$ was inactive, had a 50-100 fold decrease in ACP_D affinity, and the ACP_A interaction was too weak to fit the fluorescence anisotropy data.

Role of the HMGS flexible loop region. The most obvious structural difference between HMGS and HMGCS is a flexible loop that includes an insertion in HMGS relative to HMGCS[90]. We were interested in determining whether this region may

become ordered upon ACP_A binding and play a role in ACP selectivity, so we designed substitutions (Table 3.3) to disrupt the "EGGE" motif (residues 152-155), the π-stack that orders upon ACP_D interaction, and several other conserved residues. We previously observed reductions in activity and ACP_D affinity from alanine substitutions to charged residues that interact with the ACP[90]. We generated E152A and E155A variants of HMGS under the assumption that the substitutions would be similarly deleterious to HMGS activity and ACP_A affinity if the glutamates interact with ACP_A. We made substitutions to Gly153 and Gly154 to cause steric clashes or to eliminate flexibility conferred to this loop. F148A, W161A, and Y159A substitutions were made to eliminate contributions to the π-stack comprising these residues. We identified five other positions in the flexible loop region that are conserved in sequences for HMGS but not for HMGCS (Appendix 1). We substituted two hydrophobic residues with aspartate (L149D, I157D), a conserved aspartate with alanine (D160A), and two other conserved charged residues with the corresponding HMGCS residue (R147L, E165R).

Table 3.3 Flexible loop substitutions.

	HMGS activity[1] (% conversion)	ACP_D K_d[2] (μM)	ACP_A K_d[2] (μM)
WT	78.3 ±0.9	0.2 ± 0.1	8 ± 2
EGGE motif			
E152A	68 ± 2	9 ± 9	14 ± 7
G153F	76 ± 1	3 ± 2	4 ± 2
G154P	57 ± 2	16 ± 5	16 ± 4
G154F	47.5 ± 0.7	0.7 ± 0.3	8 ± 2
E155A	59 ± 1	13 ± 10	6 ± 2
π - stack			
F148A	92.4 ± 0.6	5 ± 2	5 ± 2
W161A	75 ± 3	1.1 ± 0.5	13 ± 1
Y159A	25 ± 4	0.5 ± 0.2	8 ± 2
Other conserved residues			
R147L	50 ± 3	1.6 ± 0.8	6 ± 4
L149D	91 ± 1	13 ± 8	1 ± 0.7
I157D	58 ± 3	2 ± 1	7 ± 2
D160A	72 ± 2	2.2 ± 0.9	4 ± 1
E165R	20 ± 2	3.5 ± 0.9	5 ± 1

[1]Conversion of equimolar acetyl-ACP_D and acetoacetyl-ACP_A to HMG-ACP_A in 7 min at 25 °C, averaged for 3 reactions.
[2]Affinity was measured by fluorescence polarization using Bodipy-FL conjugated ACPs.

The activity and ACP affinities of these variants indicate that in general the flexible loop is not essential to HMGS. The E152A, G154P, E155A substitutions resulted in 40-80 fold decreases in ACP_D affinity but did not affect ACP_A affinity. $HMGS_{G154F}$ did not affect ACP affinity but had a ~40% reduction in activity, greater than any other substitution to the EGGE residues. While $HMGS_{F148A}$ and $HMGS_{W161A}$ had activity indistinguishable from wild type, $HMGS_{Y159A}$ had 4-fold diminished activity. None of the π-stack substitutions affected ACP affinities, with the exception of the F148A substitution, which caused a 25-fold reduction in affinity for ACP_D. Of the other

conserved positions in the flexible loop that we studied, only Glu165 seemed to be important to HMGS activity, with a 4-fold reduction activity observed for $HMGS_{E165R}$. Several of these variants had moderately reduced affinity for ACP_D, with the only significant reduction (65-fold) coming from the L149D substitution. The data for HMGS activity and ACP affinity reiterate that affinity is not well correlated with activity.

Discussion

One of our objectives for disrupting these conserved structural motifs in HMGS was to study ACP selectivity in the absence of a structure for the ACP_A/HMGS complex. Affinity reductions of $HMGS_{R33A}$ and $HMGS_{R33D}$ for ACP_A suggest that Arg33 contacts the acceptor-Ppant phosphate. The only other charged residue exposed near the ACP_D-interacting surface on HMGS was Arg262. Substitutions to other charged residues that were surface exposed (Table 3.1) did not cause reductions to ACP_A affinity similar to those observed for ACP_D (on the order of 10-100 fold). Competition between the two ACPs for HMGS interaction may indicate that the ACP binding sites are overlapping or that two Ppants cannot occupy the active site at once. Reduction in ACP_A affinity from Arg262 and Arg266 substitutions suggest that that overlap may be in the basic patch of the HMGS surface, which ACP_D binding partially occludes. In general, the ACP_A-HMGS interaction appears to depend less strongly on electrostatic interactions as compared to the ACP_D-HMGS interaction, which is consistent with the more subtle surface charge distribution of ACP_A (Figure 2.9). Where ACP_D has sharply contrasting areas of positive

and negative charge, positive charge on the helix III face of ACP_A is more attenuated and the protein surface appears to be more hydrophobic.

HMGS is selective between ACP_D and ACP_A delivered substrates, which could be explained by different enzyme-ACP interactions that result in different poses of the donor and acceptor-Ppants in the active site. Looking solely at the intermediates present in the active site before donor transfer and before aldol addition, a possible explanation becomes apparent. Cys114 attacks the acetyl thioester carbon directly during donor transfer, but, during aldol addition, the α-carbon of the enolate intermediate attacks the β-carbon of the acetoacetyl acceptor (Figure 2.1). Thus, the ACP_D-Ppant must, by necessity, reach farther into the active site than the ACP_A-Ppant. In our previous tests of ACP selectivity by the curacin A HMGS[90], we observed weak activity when both the donor and acceptor substrates were presented by ACP_D and no activity when both were on ACP_A. If the above reasoning stands, this could be because the ACP_A-Ppant may not reach far enough into the active site to deliver an acetyl to Cys114. The ACP_D-Ppant, however, may be able to adopt a conformation in which the aldol addition can occur, explaining the weak activity observed.

How the ACP_A and ACP_D interactions differ is not clarified by mutagenesis of the conserved HMGS flexible loop region (Appendix 3.1, residues 151-158) does not appear essential to HMGS activity or ACP_A affinity, which does not help to clarify how ACP_A interaction differs from ACP_D interaction. The π-stack formed in the HMGS-flexible loop is not important to HMGS activity. Only $HMGS_{Y159A}$ had a significant (4-fold) reduction to activity, and no substitution seemed to affect ACP_D or ACP_A affinity, which

may be due to interaction with the ACP_A Ppant that is not disrupted by loss of Trp161 or Asp160.

Curiously, residues that were disordered in holo-ACP_D/HMGS complex structures appeared to influence ACP_D affinity. Both Glu152 and Glu155 were disordered, yet alanine variants at these positions had 20-30 fold reduced affinity for ACP_D. $HMGS_{D30A}$ and $HMGS_{R103A}$ displayed similar reductions in ACP_D affinity, even though these residues also appear not make direct electrostatic interactions with ACP_D. Loss of affinity from these substitutions may come from disruption of other residues that are more directly responsible from ACP_D interaction, or from a defect in the overall electrostatic landscape of the surface surrounding the active site entrance. $HMGS_{G154P}$ also exhibited reduced affinity for ACP_D, again presumably to disruption of other residues that interact with ACP_D. This may also explain the loss of activity in $HMGS_{E165R}$. Glu165 is ordinarily involved in a salt bridge with Arg27, to which the substituted Arg165 would be repulsive, which could indirectly have an effect on the catalytically essential Arg33.

Of the studied variants of HMGS, those that target the “thiol pocket” region had the greatest effect on HMGS activity and ACP affinities. We observed that each of the two tyrosine residues (Tyr220 and Tyr326) was important. Phenylalanine substitutions of both of these residues indicate that that the hydrogen bonds from Tyr220 to His250 and from Tyr326 to Asp200 are not essential. Complete elimination of the side chains by alanine substitution in both cases resulted in nearly complete loss of donor transfer. Neither protein was unstable, so instead these deficiencies in donor transfer support our

previously assigned role to the thiol pocket. The Ppant is very well ordered in holo-ACP_D/HMGS complex structures and distal from the catalytic thiol, so these residues may help to occlude the acetyl-Ppant before transfer. Variants of Tyr220 and Tyr326 also exhibited reductions in ACP_D and ACP_A affinity. This could be due to mispositioning of other residues due to loss of the bulky side chain, which could also occur for catalytically relevant residues that are near the tyrosines, including Ser328, Cys114, and His 250.

The Asp200 and Ser216 pair was also critical to activity. Ser216 is hydrogen-bonded to the holo Ppant thiol, but $HMGS_{S216A}$ had unimpaired donor transfer. Loss of activity should then be attributed to the aldol addition step. In structures of human HMGCS in complex with HMG-CoA[71] (2WYA), Ser258, which is analogous to the HMGS Ser216, makes a hydrogen bond to the cysteamine amide of Ppant and does not interact with Asp200 (Asp263). In HMGS, an analogous interaction between Ser216 and the ACP_A Ppant may explain loss of activity from the S216A substitution. Asp200, meanwhile, makes several hydrogen-bonds that could explain loss of activity in $HMGS_{D200A}$. The carboxylate interacts with Ser216, Tyr326, and the backbone amide of Asp212, while the Asp200 backbone carbonyl hydrogen-bonds to Gly329. In this way, Asp200 may serve as a "keystone" that locks the positioning of several other loops in the HMGS active site.

Finally, the losses to HMGS activity from M256A and P252G substitutions may result from mispositioning of His250. Pro252 is two residues preceding His250 and sits 4.0 Å away from the sidechain. Met256 was putatively identified to interact with ACP_A by

docking ACP_A structures to a homology model of the mupirocin HMGS[45], and was supported by loss of pathway throughput from substitutions to its proposed partner residue in ACP_A. In the curacin A HMGS, we do not observe a loss of ACP_A affinity from an M256A substitution, but we do see a loss in HMGS activity. Because Met256 packs between Pro252 and Tyr220, the Met256 substitution may disrupt activity by indirectly disrupting positioning of His250. Meanwhile, the apparent non-effect of M256A on the ACP_A interaction may be rationalized if either the ACP_A tyrosine does not interact with Met256, as proposed, or it does but the loss of Met256 is compensated for by other hydrophobic residues in the vicinity, including Leu217 and Phe253.

Reductions to activity and ACP affinity as a result of the described substitutions collectively indicate that the most important factor in HMGS reactivity is Ppant positioning. Substitttions to residues that are near Ppant in the active site had the greatest effects on activity. Several of these substitutions abolished HMGS activity while having minimal effect on donor transfer, while $HMGS_{D200A}$ and $HMGS_{Y220A}$ had reduced donor transfer but only a 4-fold loss in activity. We conclude from this data that the aldol addition step is much more easily disrupted than donor transfer. Intuitively this makes sense, since the aldol addition requires stabilization of an enolate nucleophile by Glu82 and Ser328 and precise positioning of the enolate to attack the acetoacetyl elecrophile. Ironically, ACP_D is capable of presenting an acceptor substrate that is suitably positioned for this more difficult step, while the ACP_A is incapable of transferring acetyl to Cys114.

As we have several times stated, HMGS variants that affect activity do not necessarily affect ACP affinities, and vice versa, but based on the observed ACP selectivity by HMGS changes in the conformation of ACP binding translate to changes in Ppant positioning. Several variants (R262A, R266A, R33A, R33D) that diminish ACP_A binding and also affect activity may do so by indirectly affecting Ppant positioning by affecting ACP_A binding. Moreover ACP_A is recognized by each of the other β-branching enzymes, and, based on what is currently known about enzyme ACP interactions[47], likely does so with approximately the same surface by which it interacts with HMGS. By further characterization of the ACP_A/HMGS complex one may glean information about the distinguishing features of the ACP that β-branching enzymes recognize.

Methods

Protein expression and purification

Plasmids encoding the curacin A HMGS[90], ACP_A[49], and ACP_D[90], were expressed *E. coli* BL21 (DE3). Cultures were grown in terrific broth (TB) at 37 °C to an OD_{600} of 1, cooled 1 hour to 20 °C, then induced with 200 μM IPTG. pHMGScur[90] was coexpressed with pGro7[75], which was induced with 2 g/L arabinose at an OD_{600} of 0.3. Cell pellets were lysed by resuspending in 4 mL/g buffer, incubating for 30 minutes at room temperature with 1 mg/mL lysozyme, 2 mM $MgCl_2$, and 0.1 mg/mL DNase, followed by cooling on ice and sonication. Lysis buffer components were as follows: 50 mM ammonium sulfate, 50 mM HEPES pH 7.0, 20 mM imidazole, and 10% v/v glycerol for HMGS; 300 mM NaCl, 50 mM Tris pH 7.5, 20 mM imidazole, and 10% v/v glycerol for ACP_A; and 300 mM NaCl, 20 mM imidazole, and 50 mM PBS pH 7.5 for ACP_D. Proteins

were purified by immobilized metal affinity chromatography on a 5 mL His Trap column (GE Healthcare) by a 50 mL gradient of 20 to 400 mM imidazole, followed by size exclusion chromatography (SEC). HMGS was purified by SEC on a Superdex S200 column (GE Healthcare) with a buffer of 50 mM ammonium sulfate, 20 mM HEPES pH 7.0, and 10% v/v glycerol (Buffer H). ACPs were purified by SEC on a Superdex S75 column (GE Healthcare) with a buffer of 150 mM NaCl, 50 mM Tris pH 7.5, and 10% v/v glycerol (Buffer A). Purified proteins were flash cooled in liquid N_2 and stored at -80 °C.

Site directed mutagenesis

pHMGScur mutants were generated by PCR with overlapping primers (Table 3.4). 0.2 ng/µL of pHMGScur, 150 nM sense and 150 nM antisense primer, 0.02 U/µL KOD Hot Start polymerase (Novagen), and 0.2 mM dNTPs (Novagen), in 1X KOD buffer (Novagen), 2mM $MgSO_4$, and 10% v/v DMSO. The program was 18 cycles of 20 sec melting at 95 °C, 10 sec annealing at 60 °C, and 3 min extension at 70 °C. After PCR, 0.2 U/µL DpnI (NEB) was added to each reaction mixture and incubated for 3 hr at 37 °C. *E. coli* XL1-Blue cells were transformed with 2 µL digested PCR mixtures. Mutant expression constructs were purified by miniprep kit (Qiagen), and the mutations were verified by Sanger sequencing.

Table 3.4 Mutagenic primers

Substitution	Sense Primer	Antisense Primer
	Surface Electrostatics	
D30A	caagcgcgtcagttggctatctcccgctttgac	gtcaaagcgggagatagccaactgacgcgcttg
R103A	caagagtatttgggcttaagtgctaactgccgcatgtttgaac	gttcaaacatgcggcagttagcacttaagccaaatactcttg
R262A	ggtatggttaaaggcgctcatgcaaatatga	ttcaacctacgcatcatatttgcatgagcg

	tgcgtaggttgaa	cctttaaccatacc
R382A	tgcagttgcctttggcactgcaaatgtcacc ttggactat	atagtccaaggtgacatttgcagtgccaaa ggcaactgca
K410A	agggtcgattagtcttaaaagcaatcgcgga attccaccgaaaatac	gtattttcggtggaattccgcgattgcttt taagactaatcgaccct
	Flexible Loop	
R147L	ctcggcaaccagaaacagagagatatcggtg gc	gccaccgatatctctctgtttctggttgcc gag
F148A	tccctcggcaaccagagcccgagagatatcg gtg	caccgatatctctcgggctctggttgccga ggga
L149D	tcgccaccgatatctctcggtttgatgttgc cgagggag	ctccctcggcaacatcaaaccgagagatat cggtggcga
E152A	ggtttctggttgccgcgggaggagaagcaat	attgcttctcctcccgcggcaaccagaaac c
G153P	taattgattgcttctcctggctcggcaacca gaaaccg	cggtttctggttgccgagccaggagaagca atcaatta
G153F	atcgtaattgattgcttctccgaactcggca accagaaaccgaga	tctcggtttctggttgccgagttcggagaa gcaatcaattacgat
G154P	cgtaattgattgcttctggtccctcggcaac cagaaacc	ggtttctggttgccgagggaccagaagcaa tcaattacg
G154F	ccaatcgtaattgattgcttcgaatccctcg gcaaccagaaaccg	cggtttctggttgccgagggattcgaagca atcaattacgattgg
E155A	gttgccgagggaggagcagcaatcaattacg att	aatcgtaattgattgctgctcctccctcgg caac
I157D	tgccgagggaggagaagcagacaattacgat tggtctttt	aaaagaccaatcgtaattgtctgcttctcc tccctcggca
Y159A	tgccgagggaggagaagcaatcaatgccgat tggtcttttg	caaaagaccaatcggcattgattgcttctc ctccctcggca
D160A	ggagaagcaatcaattacgcttggtcttttg ctgaaccc	gggttcagcaaaagaccaagcgtaattgat tgcttctcc
W161A	tactgggttcagcaaaagacgcatcgtaatt gattgcttctcc	ggagaagcaatcaattacgatgcgtctttt gctgaacccagta
E165R	caccactactgggtctagcaaaagaccaatc gtaattgattgc	gcaatcaattacgattggtcttttgctaga cccagtagtggtg
	Thiol Pocket	
D200A	tatggttatgaagtcatggctacctgtagac ctaatccc	gggattaggtctacaggtagccatgacttc ataaccata
S216A	gaagcaggagatgccgacttagccttgcttt ctta	taagaaagcaaggctaagtcggcatctcct gcttc

L217A	aggagatgccgacttatccgcgctttcttac ctagactgt	acagtctaggtaagaaagcgcggataagtc ggcatctcct
Y220A	agatgccgacttatccttgctttctgcccta gactgttgc	gcaacagtctagggcagaaagcaaggataa gtcggcatct
Y220F	cgacttatccttgctttctttcctagactgt tgcgaaa	tttcgcaacagtctaggaaagaaagcaagg ataagtcg
P252G	gttttgattatttgagttttcatactggttt tggtggtatggttaaaggcg	cgcctttaaccataccaccaaaaccagtat gaaaactcaaataatcaaaac
M256A	agttttcatactccttttggtggtgcggtta aaggcgctcatag	ctatgagcgcctttaaccgcaccaccaaaa ggagtatgaaaact
Y326A	cagaattgggatgttttccgctggctctggt tgttgttcc	ggaacaacaaccagagccagcggaaaacat cccaattctg
Y326F	tcgcagaattgggatgttttcctttggctct ggttg	caaccagagccaaaggaaaacatcccaatt ctgcga

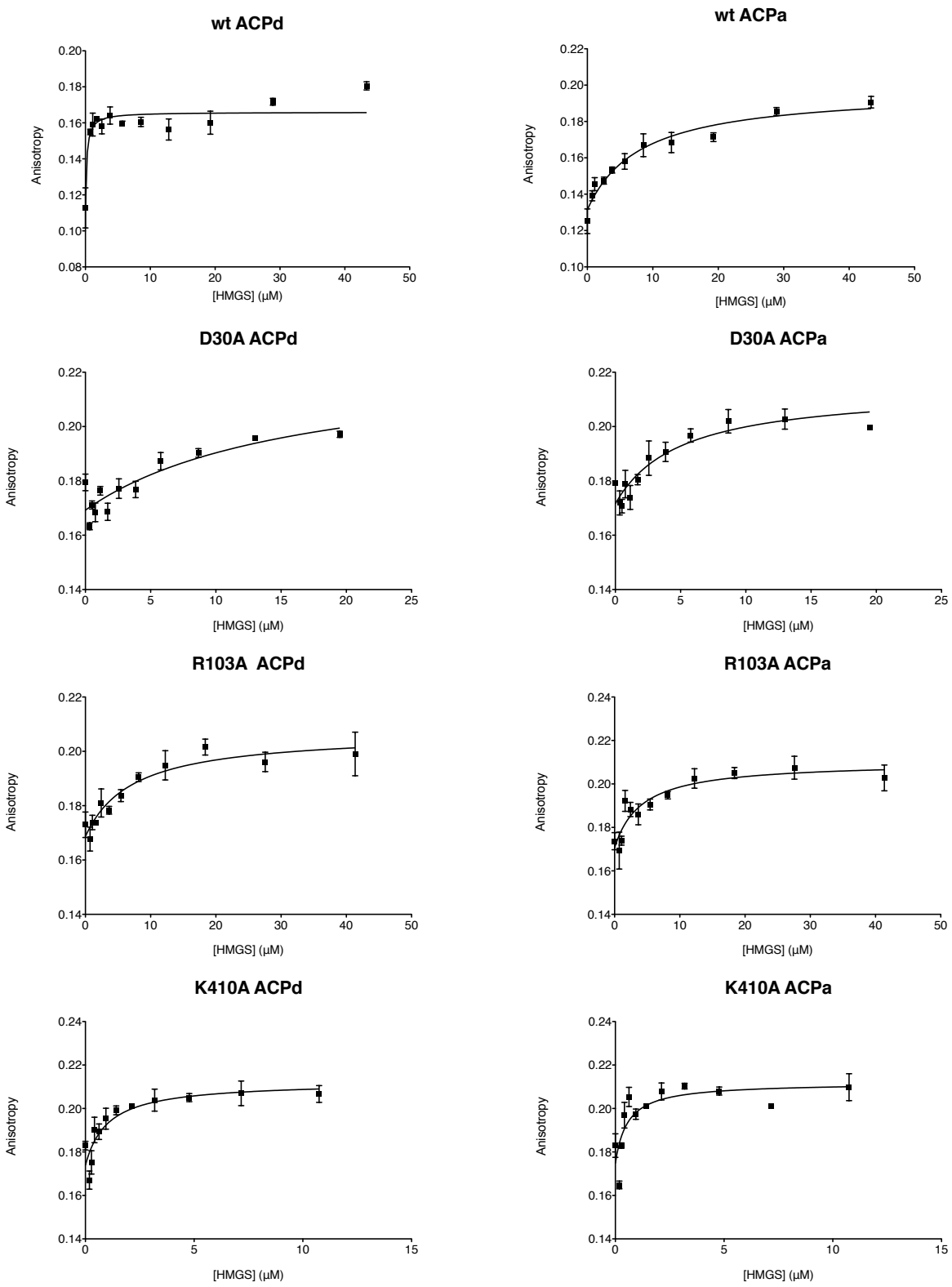
Activity assay

ACPs purified in the apo state were loaded *in vitro* using *B. subtilis* phosphopantetheinyl transferase (Sfp). ~200 µM purified ACP was incubated with 10 mM $MgCl_2$, 50 mM Tris pH 7.5, 0.2 mg/mL Sfp, 1 mM acyl-CoA, and 5% v/v glycerol ~18 h at 4 °C, followed by SEC. 2 µM HMGS was reacted with 20 µM acetyl-ACP_D and 20 µM acetoacetyl-ACP_A in 20 mM ammonium sulfate, 20 mM HEPES pH 7.0, and 2 mM TCEP for 7 minutes at 25 °C. Reaction mixtures were quenched by addition of 5% v/v formic acid and analyzed by HPLC-mass spectrometry by QTOF (Agilent) with Ppant ejection. Reaction progress is reported as the HMG percent of the HMG- and acetoacetyl-Ppant ions combined and averaged for three replicates.

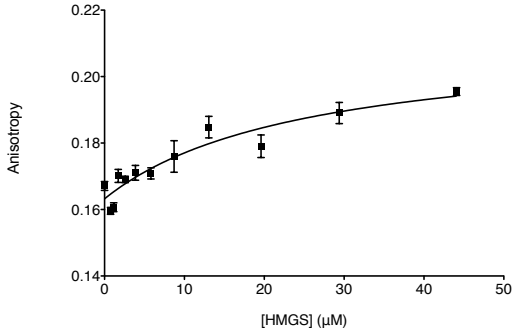
FP affinity assay

HMGS affinity for ACP_D and ACP_A was measured by fluorescence anisotropy using a BODIPY-FL conjugated M1C variant of the curacin ACP_D, as described previously[90], or an E2020C variant of CurA ACPII (pLG003[49]). HMGS was buffer exchanged into 5 mM NaCl and 10 mM HEPES pH 7.0 (FP Buffer) then incubated at various concentrations with 20 nM fluorescent ACP_D or ACP_A at room temperature for 30 minutes. Fluorescence anisotropy was measured by a λ_{EX} of 485 nm and λ_{EM} of 538 nm (Figure 2.15). The data were fit to $A = A_{free} + (\Delta A_{bound} \times [HMGS]/ (K_d+[HMGS]))$ (GraphPad Prism). ACP competition was assayed by fluorescence anisotropy using fluorescent ACP_A. Unlabeled holo-ACP_A, holo-ACP_D, and HMGS were buffer exchanged into FP buffer. 20 nM fluorescent ACP_A was incubated for 30 min with 8.8 μM HMGS and varying concentrations of unlabeled ACP_D or ACP_A.

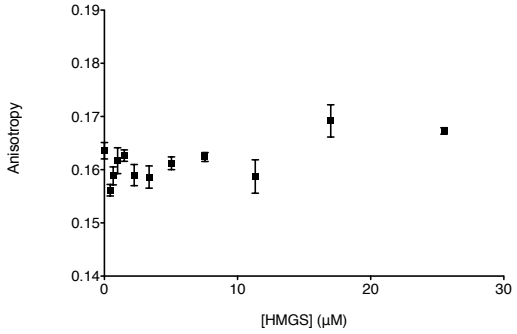
Figure 3.5 Fluorescence anisotropy for ACP_D and ACP_A binding experiments.



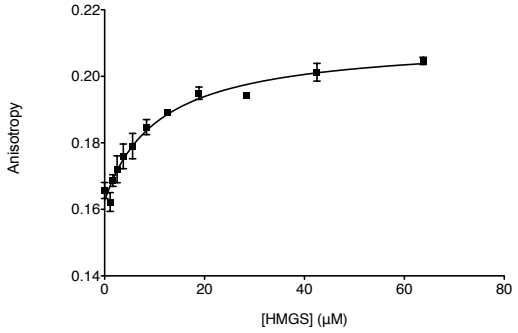
R33A ACPa



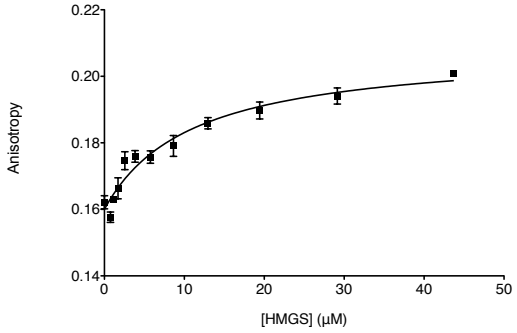
R33D ACPa



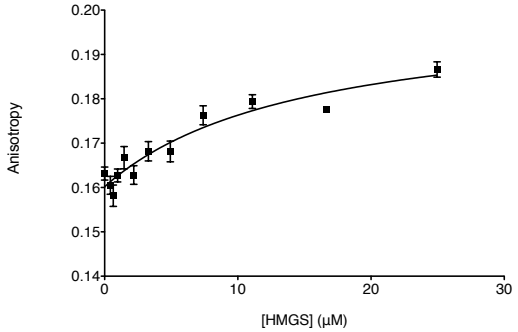
D214A ACPa



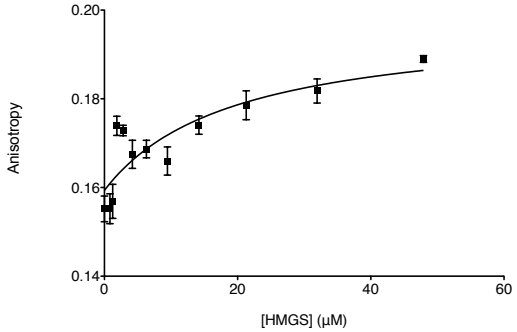
E225R ACPa



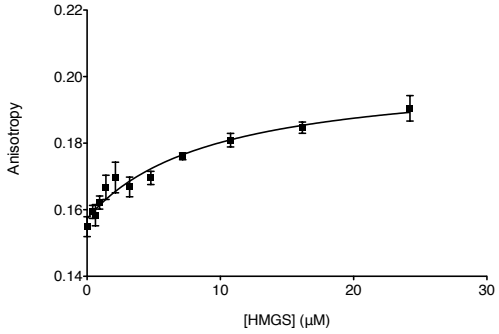
D222A ACPa



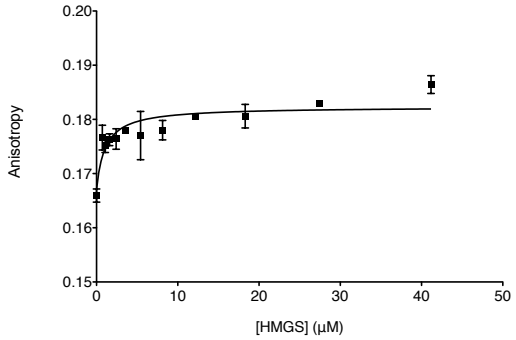
R266A ACPa



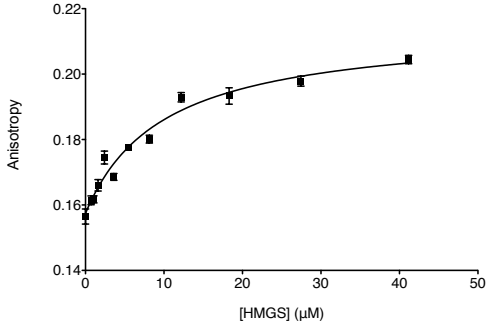
D200A ACPa



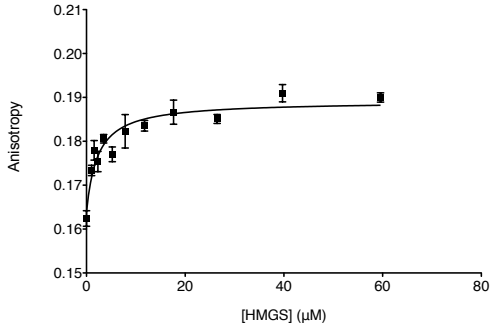
S216A ACPd



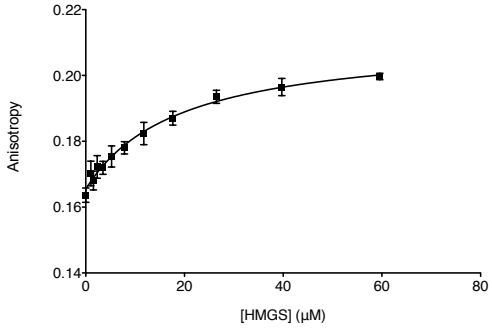
S216A ACPa



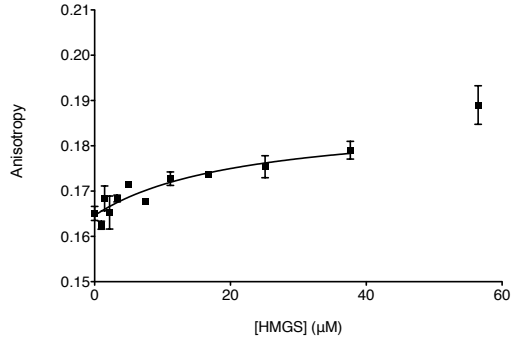
L217A ACPd



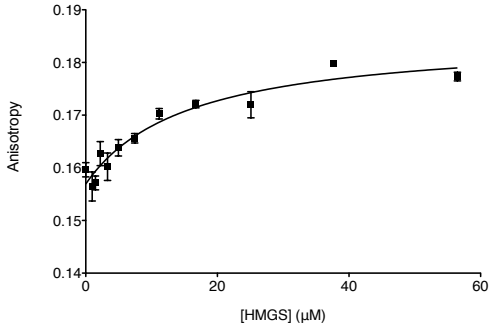
L217A ACPa



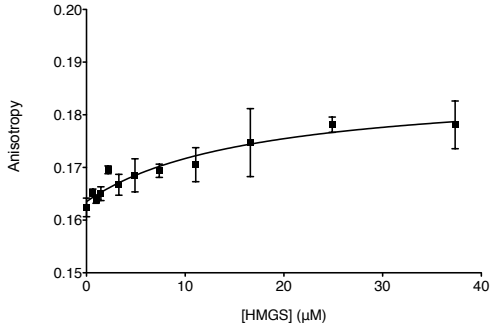
Y220A ACPd



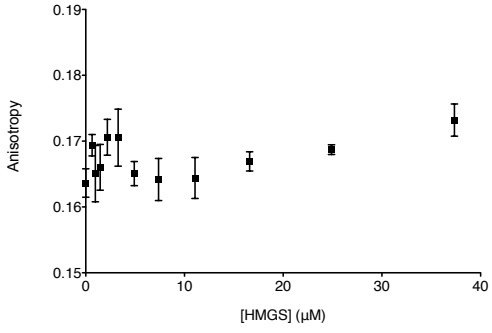
Y220A ACPa



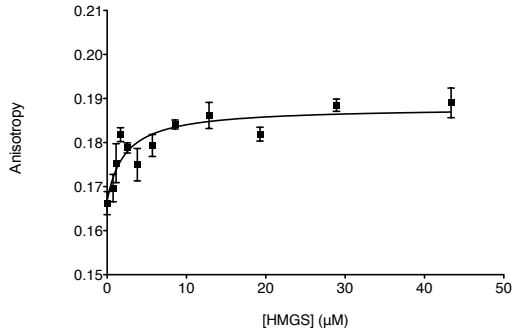
P252G ACPd



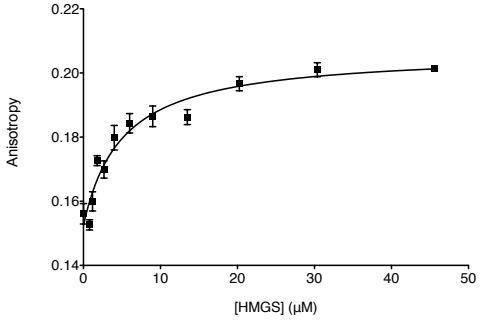
P252G ACPa



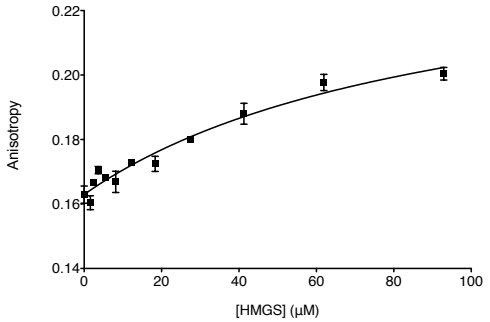
M256A ACPd



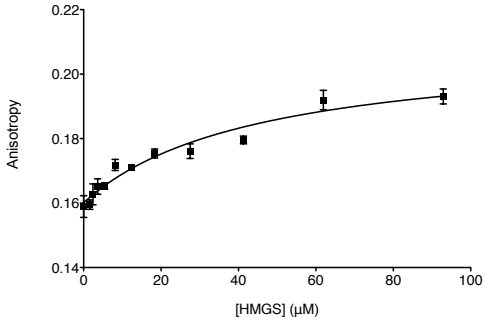
M256A ACPa



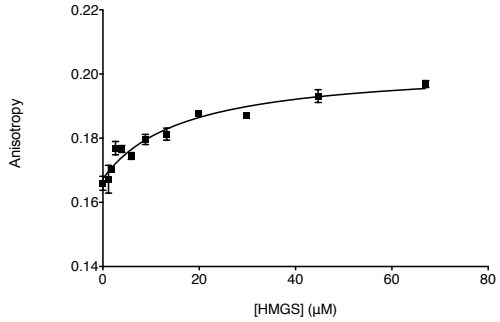
Y326A ACPd



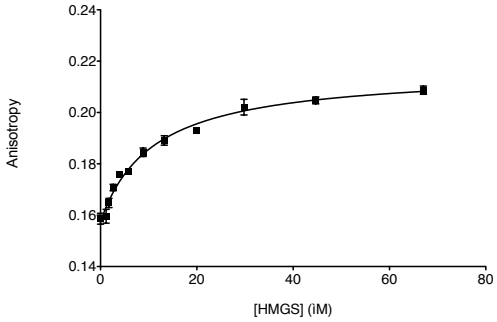
Y326A ACPa



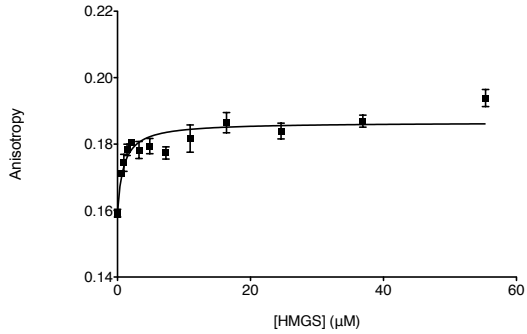
Y326F ACPd



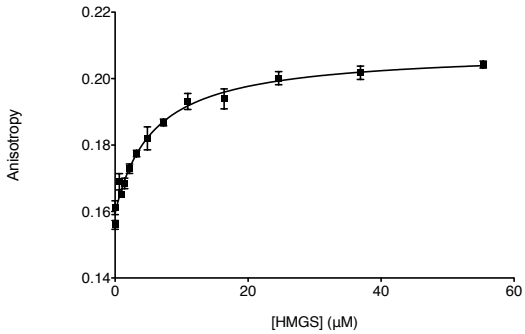
Y326F ACPa



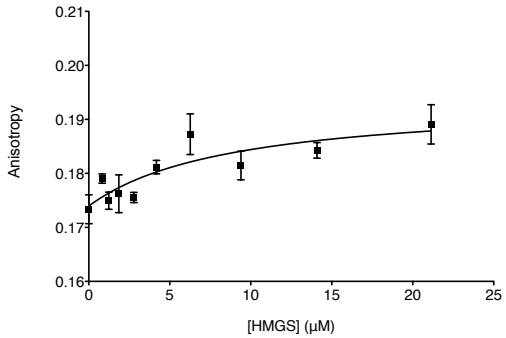
S328A ACPd



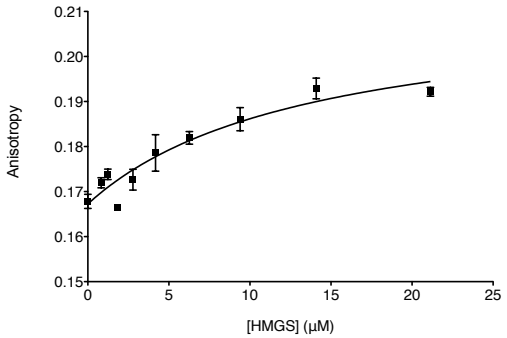
S328A ACPa



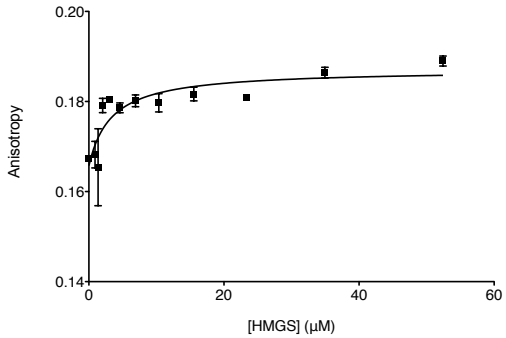
E152A ACPd



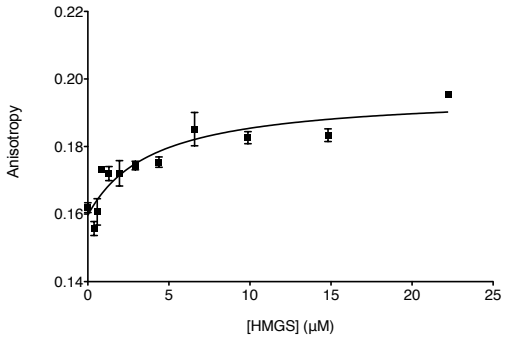
E152A ACPa



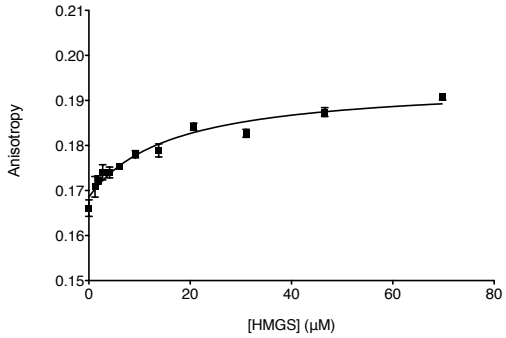
G153F ACPd



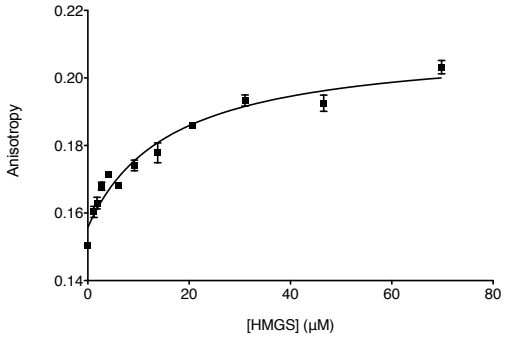
G153F ACPa



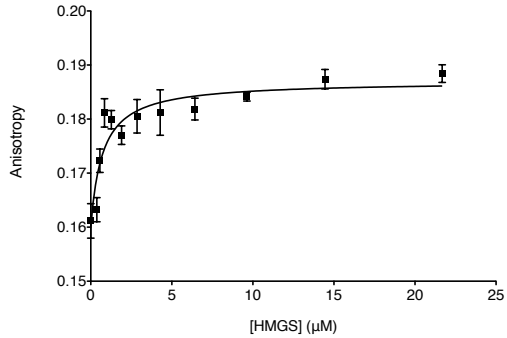
G154P ACPd



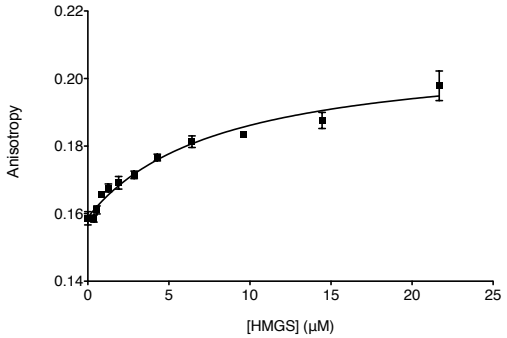
G154P ACPa



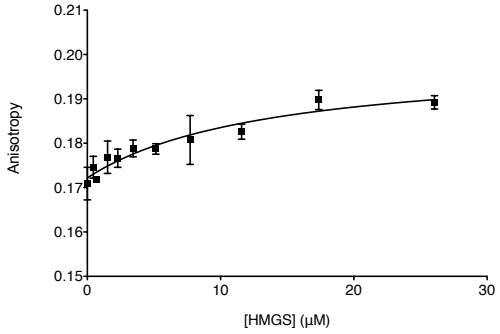
G154F ACPd



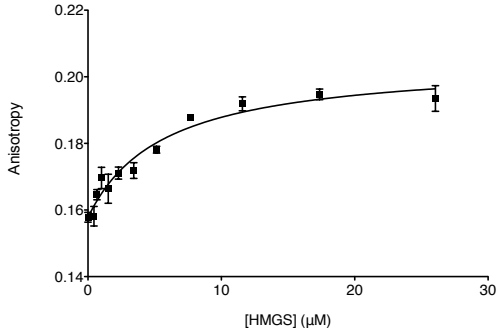
G154F ACPa



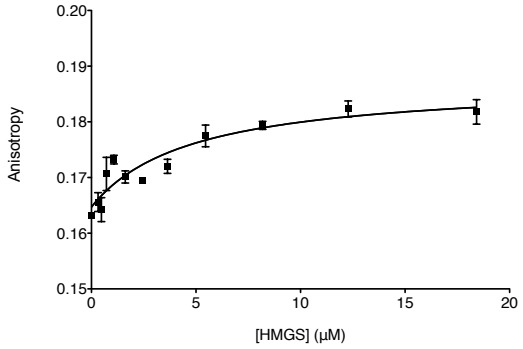
E155A ACPd



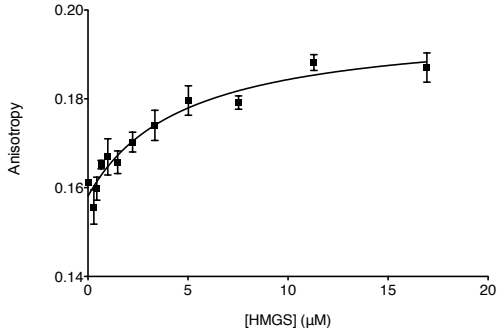
E155A ACPa



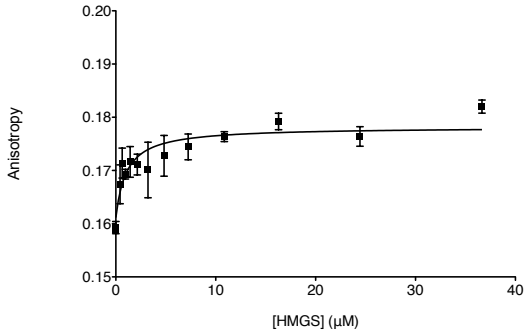
F148A ACPd



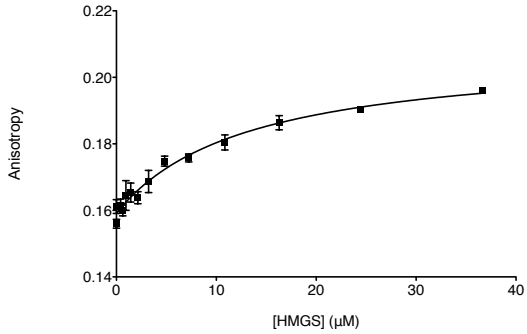
F148A ACPa



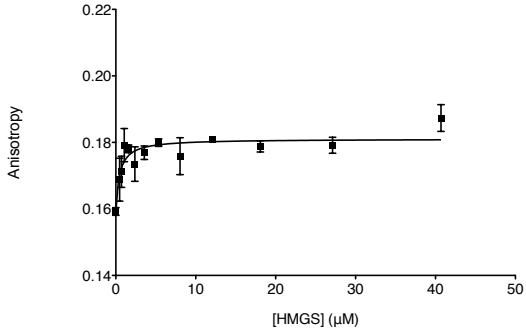
W161A ACPd



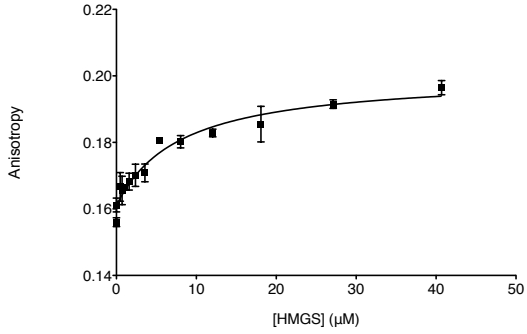
W161A ACPa



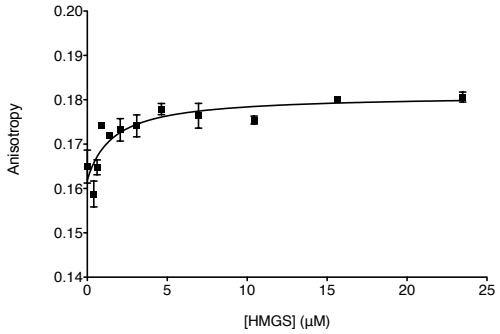
Y159A ACPd



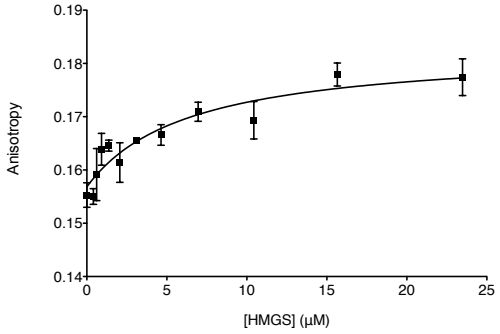
Y159A ACPa



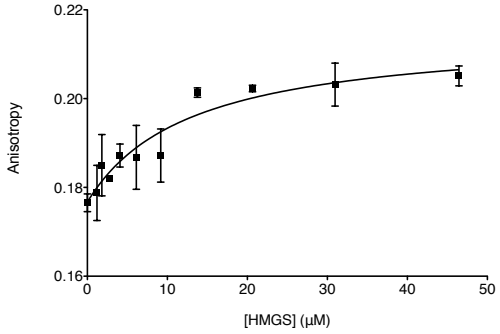
R147L ACPd



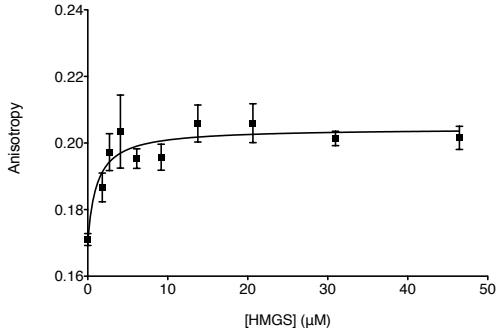
R147L ACPa



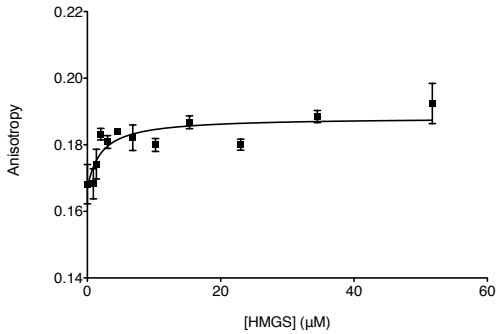
L149D ACPd



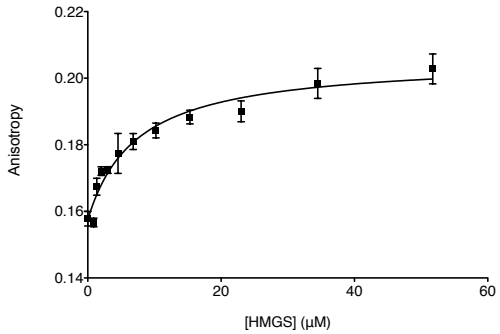
L149A ACPa



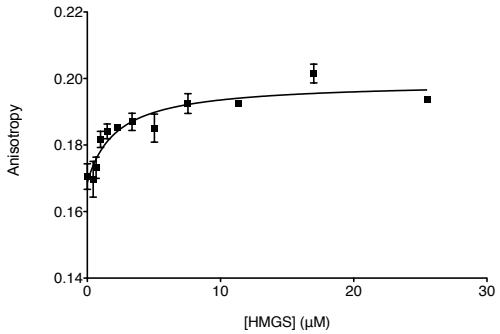
I157D ACPd



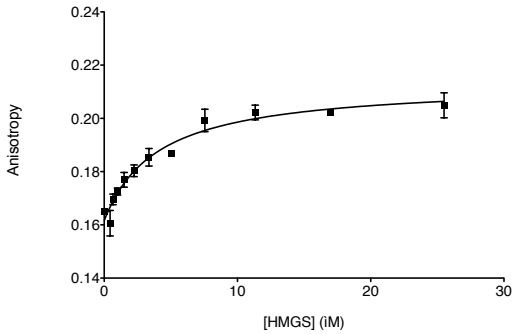
I157D ACPa

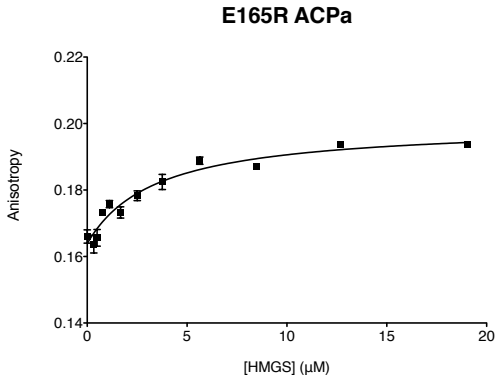


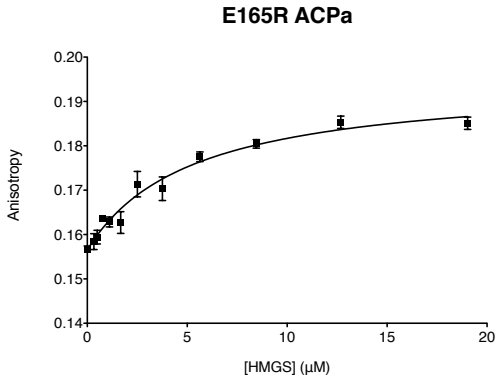
D160_CurB

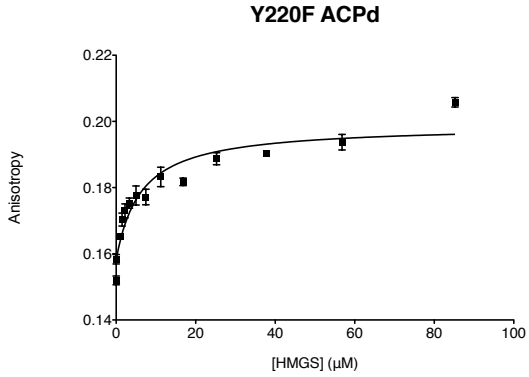


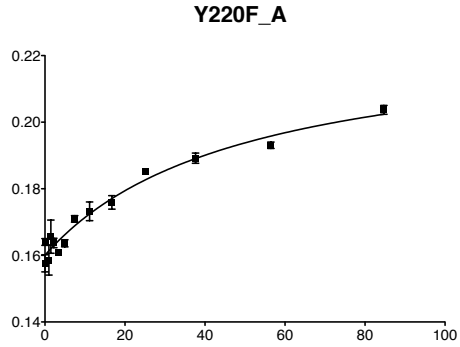
D160A ACPa




E165R ACPa
Anisotropy
0.22
0.20
0.18
0.16
0.14
0
5
10
15
20
[HMGS] (μM)

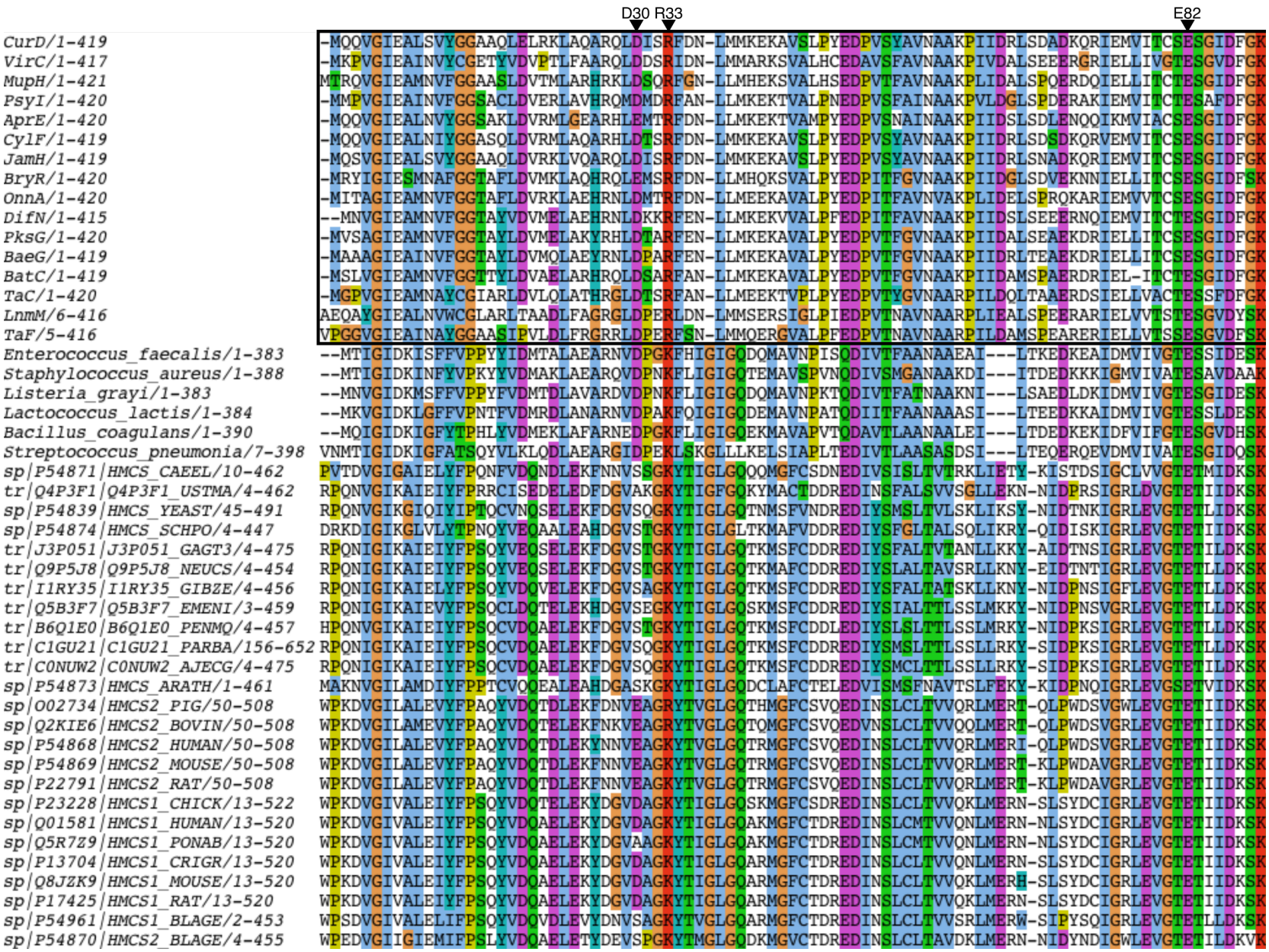

E165R ACPa
Anisotropy
0.20
0.19
0.18
0.17
0.16
0.15
0.14
0
5
10
15
20
[HMGS] (μM)

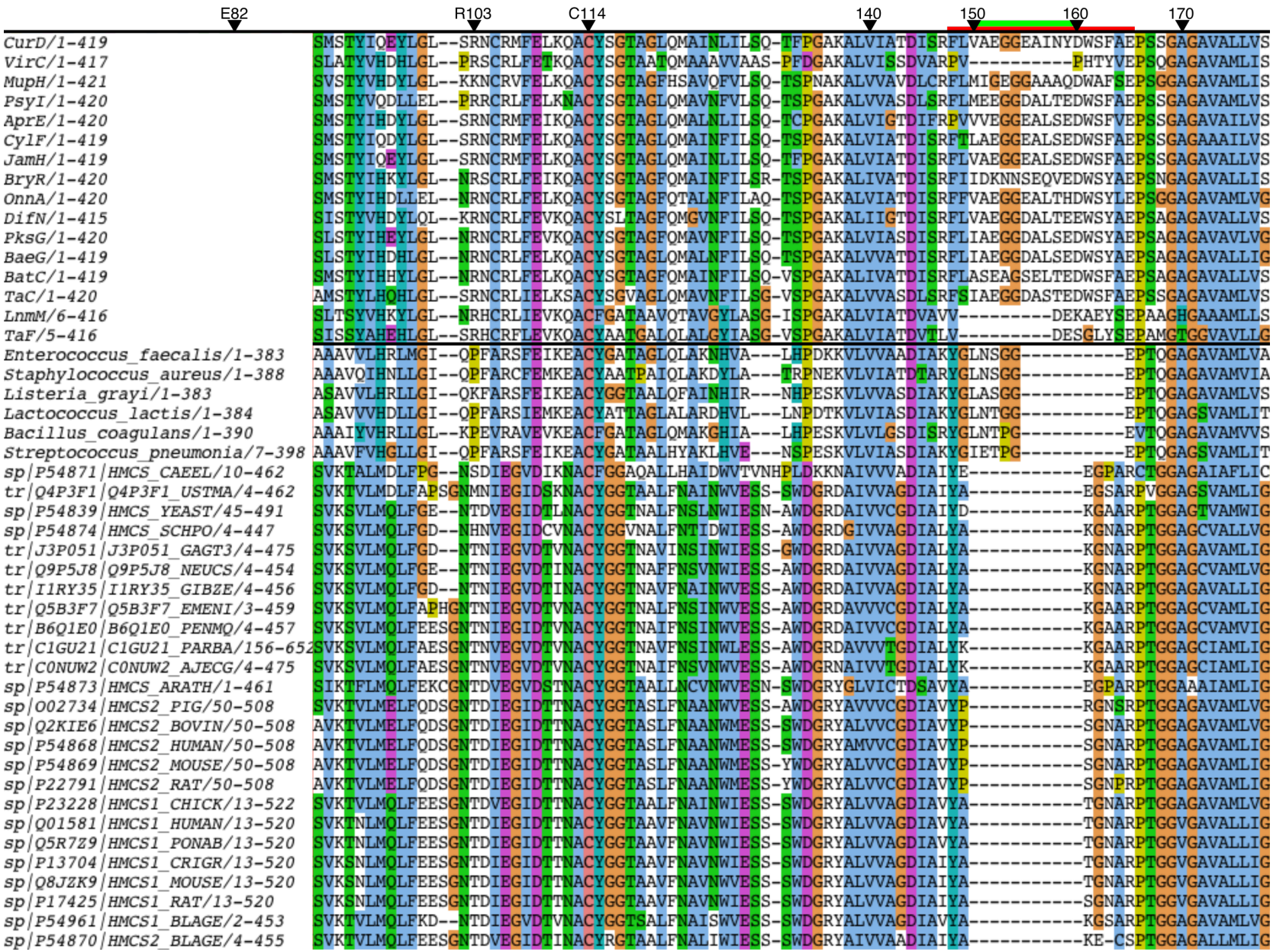

Y220F ACPd
Anisotropy
0.22
0.20
0.18
0.16
0.14
0
20
40
60
80
100
[HMGS] (μM)

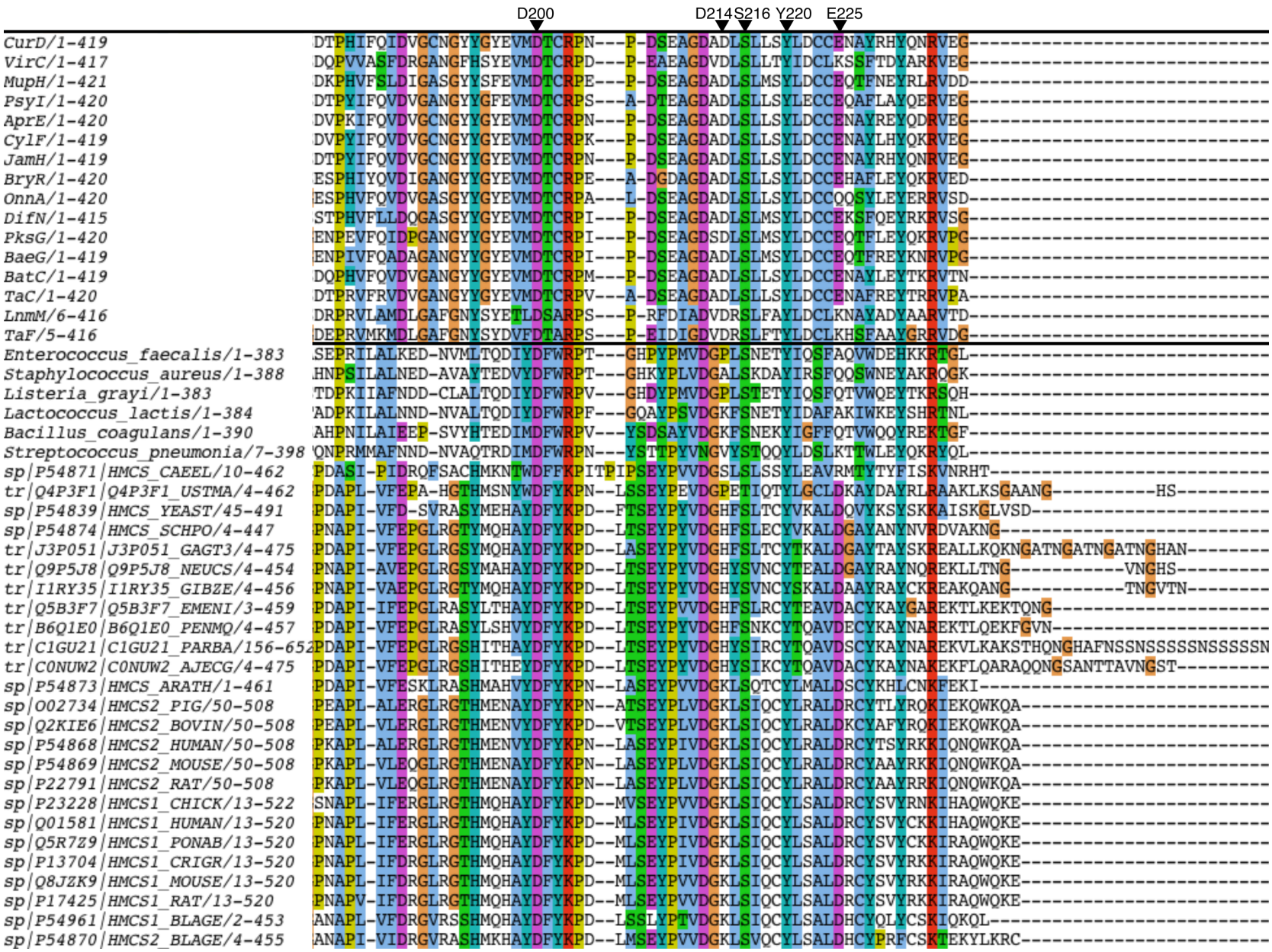

Y220F_A
0.22
0.20
0.18
0.16
0.14
0
20
40
60
80
100

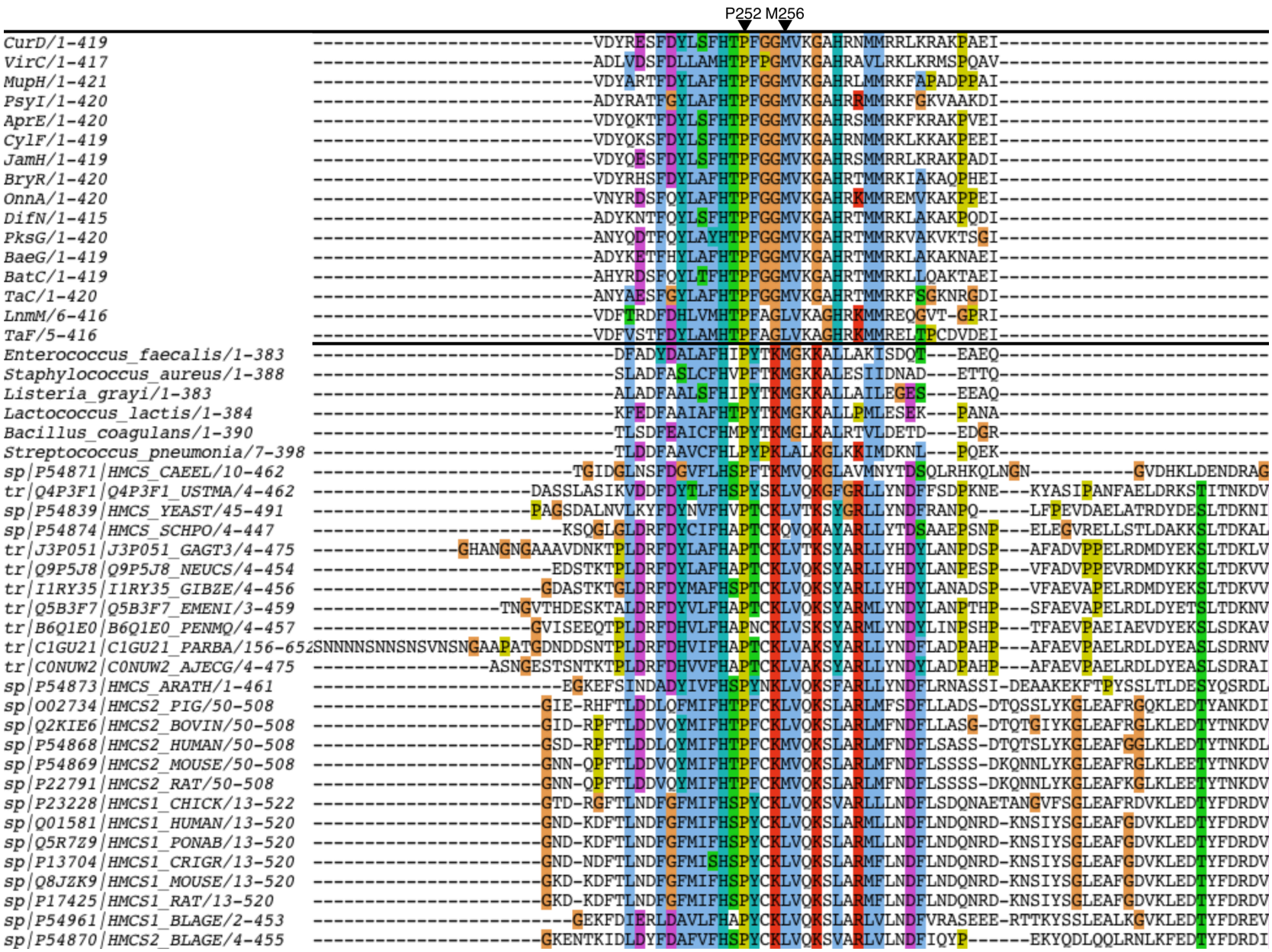
Appendix

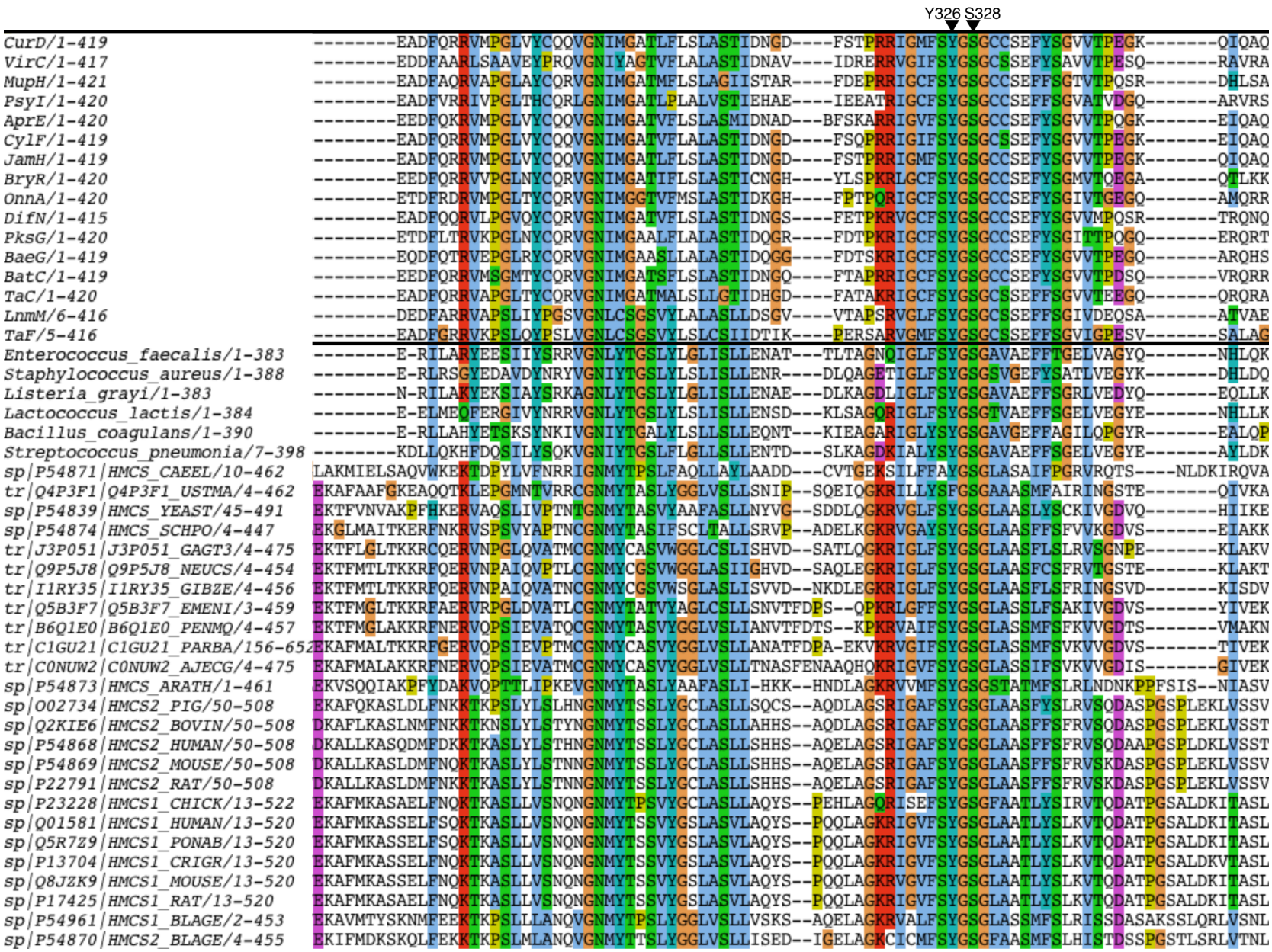
Conservation between HMGS and HMGCS. Sequences are numbered according to CurD. Boxed sequences are HMGS from PKS, unboxed are primary metabolism HMGCS. Red bar indicates disordered residues in HMGS (5KP5) and the green bar indicates residues that are disordered in the holo-ACP_D-HMGS complex structure (5KP7). Accession codes and species for HMGS sequences are as follows: CurD F4Y432, *M.producens 3L*; VirC A4PHM8, *S.virginiae*; MupH Q8RL63, *P.fluorescens*; PsyI ADA82596, uncultured bacterium psy1; AprE WP_075900452, *Moorea Bouillonii*; CylF ARU81120, Cylindrospermum licheniforme; JamF Q6E7K2; *M.producens*; BryR A2CLL9, *Candidatus E.sertula*; OnnA Q5MP08, symbiont of *T.swinhoei*; DifN Q1RS50, *B.amyloliquefaciens*; PksG P40830, *B.subtilis*; BaeG E1UUM7, *B.amyloliquefaciens*; BatC D4NZD8, *P.fluorescens*; TaC Q1D5E5; *M.xanthus*; TaF Q1D5E8, *M.xanthus*; LnmK Q8GGP1, *S.atroolivaceus*.

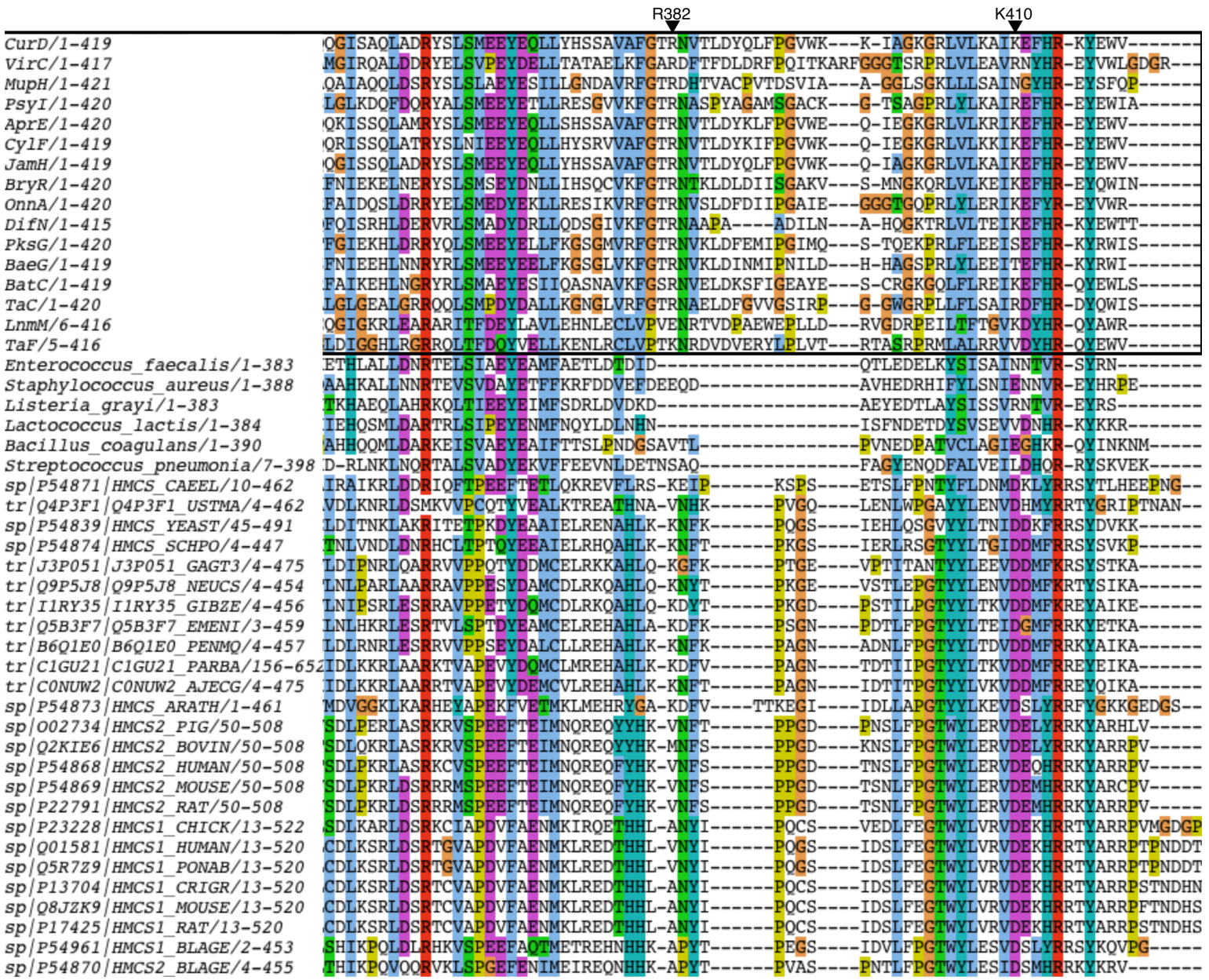

D30 R33
E82
CurD/1-419
VirC/1-417
MupH/1-421
PsyI/1-420
AprE/1-420
CylF/1-419
JamH/1-419
BryR/1-420
OnnA/1-420
DifN/1-415
PksG/1-420
BaeG/1-419
BatC/1-419
TaC/1-420
LnmM/6-416
TaF/5-416
Enterococcus_faecalis/1-383
Staphylococcus_aureus/1-388
Listeria_grayi/1-383
Lactococcus_lactis/1-384
Bacillus_coagulans/1-390
Streptococcus_pneumonia/7-398
sp|P54871|HMCS_CAEEL/10-462
tr|Q4P3F1|Q4P3F1_USTMA/4-462
sp|P54839|HMCS_YEAST/45-491
sp|P54874|HMCS_SCHPO/4-447
tr|J3P051|J3P051_GAGT3/4-475
tr|Q9P5J8|Q9P5J8_NEUCS/4-454
tr|I1RY35|I1RY35_GIBZE/4-456
tr|Q5B3F7|Q5B3F7_EMENI/3-459
tr|B6Q1E0|B6Q1E0_PENMQ/4-457
tr|C1GU21|C1GU21_PARBA/156-652
tr|C0NUW2|C0NUW2_AJECG/4-475
sp|P54873|HMCS_ARATH/1-461
sp|O02734|HMCS2_PIG/50-508
sp|Q2KIE6|HMCS2_BOVIN/50-508
sp|P54868|HMCS2_HUMAN/50-508
sp|P54869|HMCS2_MOUSE/50-508
sp|P22791|HMCS2_RAT/50-508
sp|P23228|HMCS1_CHICK/13-522
sp|Q01581|HMCS1_HUMAN/13-520
sp|Q5R7Z9|HMCS1_PONAB/13-520
sp|P13704|HMCS1_CRIGR/13-520
sp|Q8JZK9|HMCS1_MOUSE/13-520
sp|P17425|HMCS1_RAT/13-520
sp|P54961|HMCS1_BLAGE/2-453
sp|P54870|HMCS2_BLAGE/4-455


E82
R103
C114
140
150
160
170
CurD/1-419
VirC/1-417
MupH/1-421
PsyI/1-420
AprE/1-420
CylF/1-419
JamH/1-419
BryR/1-420
OnnA/1-420
DifN/1-415
PksG/1-420
BaeG/1-419
BatC/1-419
TaC/1-420
LnmM/6-416
TaF/5-416
Enterococcus_faecalis/1-383
Staphylococcus_aureus/1-388
Listeria_grayi/1-383
Lactococcus_lactis/1-384
Bacillus_coagulans/1-390
Streptococcus_pneumonia/7-398
sp|P54871|HMCS_CAEEL/10-462
tr|Q4P3F1|Q4P3F1_USTMA/4-462
sp|P54839|HMCS_YEAST/45-491
sp|P54874|HMCS_SCHPO/4-447
tr|J3P051|J3P051_GAGT3/4-475
tr|Q9P5J8|Q9P5J8_NEUCS/4-454
tr|I1RY35|I1RY35_GIBZE/4-456
tr|Q5B3F7|Q5B3F7_EMENI/3-459
tr|B6Q1E0|B6Q1E0_PENMQ/4-457
tr|C1GU21|C1GU21_PARBA/156-652
tr|C0NUW2|C0NUW2_AJECG/4-475
sp|P54873|HMCS_ARATH/1-461
sp|O02734|HMCS2_PIG/50-508
sp|Q2KIE6|HMCS2_BOVIN/50-508
sp|P54868|HMCS2_HUMAN/50-508
sp|P54869|HMCS2_MOUSE/50-508
sp|P22791|HMCS2_RAT/50-508
sp|P23228|HMCS1_CHICK/13-522
sp|Q01581|HMCS1_HUMAN/13-520
sp|Q5R7Z9|HMCS1_PONAB/13-520
sp|P13704|HMCS1_CRIGR/13-520
sp|Q8JZK9|HMCS1_MOUSE/13-520
sp|P17425|HMCS1_RAT/13-520
sp|P54961|HMCS1_BLAGE/2-453
sp|P54870|HMCS2_BLAGE/4-455


D200
D214
S216
Y220
E225
CurD/1-419
VirC/1-417
MupH/1-421
PsyI/1-420
AprE/1-420
CylF/1-419
JamH/1-419
BryR/1-420
OnnA/1-420
DifN/1-415
PksG/1-420
BaeG/1-419
BatC/1-419
TaC/1-420
LnmM/6-416
TaF/5-416
Enterococcus_faecalis/1-383
Staphylococcus_aureus/1-388
Listeria_grayi/1-383
Lactococcus_lactis/1-384
Bacillus_coagulans/1-390
Streptococcus_pneumonia/7-398
sp|P54871|HMCS_CAEEL/10-462
tr|Q4P3F1|Q4P3F1_USTMA/4-462
sp|P54839|HMCS_YEAST/45-491
sp|P54874|HMCS_SCHPO/4-447
tr|J3P051|J3P051_GAGT3/4-475
tr|Q9P5J8|Q9P5J8_NEUCS/4-454
tr|I1RY35|I1RY35_GIBZE/4-456
tr|Q5B3F7|Q5B3F7_EMENI/3-459
tr|B6Q1E0|B6Q1E0_PENMQ/4-457
tr|C1GU21|C1GU21_PARBA/156-652
tr|C0NUW2|C0NUW2_AJECG/4-475
sp|P54873|HMCS_ARATH/1-461
sp|O02734|HMCS2_PIG/50-508
sp|Q2KIE6|HMCS2_BOVIN/50-508
sp|P54868|HMCS2_HUMAN/50-508
sp|P54869|HMCS2_MOUSE/50-508
sp|P22791|HMCS2_RAT/50-508
sp|P23228|HMCS1_CHICK/13-522
sp|Q01581|HMCS1_HUMAN/13-520
sp|Q5R7Z9|HMCS1_PONAB/13-520
sp|P13704|HMCS1_CRIGR/13-520
sp|Q8JZK9|HMCS1_MOUSE/13-520
sp|P17425|HMCS1_RAT/13-520
sp|P54961|HMCS1_BLAGE/2-453
sp|P54870|HMCS2_BLAGE/4-455


P252 M256
CurD/1-419
VirC/1-417
MupH/1-421
PsyI/1-420
AprE/1-420
CylF/1-419
JamH/1-419
BryR/1-420
OnnA/1-420
DifN/1-415
PksG/1-420
BaeG/1-419
BatC/1-419
TaC/1-420
LnmM/6-416
TaF/5-416
Enterococcus_faecalis/1-383
Staphylococcus_aureus/1-388
Listeria_grayi/1-383
Lactococcus_lactis/1-384
Bacillus_coagulans/1-390
Streptococcus_pneumonia/7-398
sp|P54871|HMCS_CAEEL/10-462
tr|Q4P3F1|Q4P3F1_USTMA/4-462
sp|P54839|HMCS_YEAST/45-491
sp|P54874|HMCS_SCHPO/4-447
tr|J3P051|J3P051_GAGT3/4-475
tr|Q9P5J8|Q9P5J8_NEUCS/4-454
tr|I1RY35|I1RY35_GIBZE/4-456
tr|Q5B3F7|Q5B3F7_EMENI/3-459
tr|B6Q1E0|B6Q1E0_PENMQ/4-457
tr|C1GU21|C1GU21_PARBA/156-652
tr|C0NUW2|C0NUW2_AJECG/4-475
sp|P54873|HMCS_ARATH/1-461
sp|O02734|HMCS2_PIG/50-508
sp|Q2KIE6|HMCS2_BOVIN/50-508
sp|P54868|HMCS2_HUMAN/50-508
sp|P54869|HMCS2_MOUSE/50-508
sp|P22791|HMCS2_RAT/50-508
sp|P23228|HMCS1_CHICK/13-522
sp|Q01581|HMCS1_HUMAN/13-520
sp|Q5R7Z9|HMCS1_PONAB/13-520
sp|P13704|HMCS1_CRIGR/13-520
sp|Q8JZK9|HMCS1_MOUSE/13-520
sp|P17425|HMCS1_RAT/13-520
sp|P54961|HMCS1_BLAGE/2-453
sp|P54870|HMCS2_BLAGE/4-455


Y326 S328
CurD/1-419
VirC/1-417
MupH/1-421
PsyI/1-420
AprE/1-420
CylF/1-419
JamH/1-419
BryR/1-420
OnnA/1-420
DifN/1-415
PksG/1-420
BaeG/1-419
BatC/1-419
TaC/1-420
LnmM/6-416
TaF/5-416
Enterococcus_faecalis/1-383
Staphylococcus_aureus/1-388
Listeria_grayi/1-383
Lactococcus_lactis/1-384
Bacillus_coagulans/1-390
Streptococcus_pneumonia/7-398
sp|P54871|HMCS_CAEEL/10-462
tr|Q4P3F1|Q4P3F1_USTMA/4-462
sp|P54839|HMCS_YEAST/45-491
sp|P54874|HMCS_SCHPO/4-447
tr|J3P051|J3P051_GAGT3/4-475
tr|Q9P5J8|Q9P5J8_NEUCS/4-454
tr|I1RY35|I1RY35_GIBZE/4-456
tr|Q5B3F7|Q5B3F7_EMENI/3-459
tr|B6Q1E0|B6Q1E0_PENMQ/4-457
tr|C1GU21|C1GU21_PARBA/156-652
tr|C0NUW2|C0NUW2_AJECG/4-475
sp|P54873|HMCS_ARATH/1-461
sp|O02734|HMCS2_PIG/50-508
sp|Q2KIE6|HMCS2_BOVIN/50-508
sp|P54868|HMCS2_HUMAN/50-508
sp|P54869|HMCS2_MOUSE/50-508
sp|P22791|HMCS2_RAT/50-508
sp|P23228|HMCS1_CHICK/13-522
sp|Q01581|HMCS1_HUMAN/13-520
sp|Q5R7Z9|HMCS1_PONAB/13-520
sp|P13704|HMCS1_CRIGR/13-520
sp|Q8JZK9|HMCS1_MOUSE/13-520
sp|P17425|HMCS1_RAT/13-520
sp|P54961|HMCS1_BLAGE/2-453
sp|P54870|HMCS2_BLAGE/4-455


R382
K410
CurD/1-419
VirC/1-417
MupH/1-421
PsyI/1-420
AprE/1-420
CylF/1-419
JamH/1-419
BryR/1-420
OnnA/1-420
DifN/1-415
PksG/1-420
BaeG/1-419
BatC/1-419
TaC/1-420
LnmM/6-416
TaF/5-416
Enterococcus_faecalis/1-383
Staphylococcus_aureus/1-388
Listeria_grayi/1-383
Lactococcus_lactis/1-384
Bacillus_coagulans/1-390
Streptococcus_pneumonia/7-398
sp|P54871|HMCS_CAEEL/10-462
tr|Q4P3F1|Q4P3F1_USTMA/4-462
sp|P54839|HMCS_YEAST/45-491
sp|P54874|HMCS_SCHPO/4-447
tr|J3P051|J3P051_GAGT3/4-475
tr|Q9P5J8|Q9P5J8_NEUCS/4-454
tr|I1RY35|I1RY35_GIBZE/4-456
tr|Q5B3F7|Q5B3F7_EMENI/3-459
tr|B6Q1E0|B6Q1E0_PENMQ/4-457
tr|C1GU21|C1GU21_PARBA/156-652
tr|C0NUW2|C0NUW2_AJECG/4-475
sp|P54873|HMCS_ARATH/1-461
sp|O02734|HMCS2_PIG/50-508
sp|Q2KIE6|HMCS2_BOVIN/50-508
sp|P54868|HMCS2_HUMAN/50-508
sp|P54869|HMCS2_MOUSE/50-508
sp|P22791|HMCS2_RAT/50-508
sp|P23228|HMCS1_CHICK/13-522
sp|Q01581|HMCS1_HUMAN/13-520
sp|Q5R7Z9|HMCS1_PONAB/13-520
sp|P13704|HMCS1_CRIGR/13-520
sp|Q8JZK9|HMCS1_MOUSE/13-520
sp|P17425|HMCS1_RAT/13-520
sp|P54961|HMCS1_BLAGE/2-453
sp|P54870|HMCS2_BLAGE/4-455

Chapter 4. HMGS selectivity for non-natural donor substrates.

Introduction

Modular polyketide synthases (PKS) are versatile enzyme assembly lines that synthesize chemically diverse and pharmaceutically interesting natural products[50,4,5], including erythromycin, rapamycin, and bryostatin. Modular PKS biosynthetic pathways assemble these molecules by appending an acyl building block to the nascent product, making chemical modifications, then transferring the intermediate to the next module in the pathway[5,6]. The most common modifications before transfer involve reduction and/or dehydration of the β-carbonyl generated by polyketide extension (Figure 1.2). Each intermediate is tethered via a thioester linkage to the phosphopantetheine arm of an acyl carrier protein (ACP) domain within each module. Many PKS pathways encode a set of standalone enzymes that introduce an alkyl substitutent at the β-position (β-branching[18]) instead of reducing the β-carbonyl. While most modification enzymes act *in cis* on a fused ACP-tethered substrate, β-branching enzymes act *in trans*, and recognize a specialized ACP tethered to a specific module in the pathway.

The standard enzymes encoded by PKS β-branching cassettes are an HMG-synthase (HMGS), which installs a carboxyalkyl branch at the β-carbon, an enoyl-CoA hydratase-like enzyme (ECH_1), which dehydrates the resulting β-hydroxyl, and a second ECH_2 that decarboxylates the carboxyalkyl branch[18]. The branch carbon originates from

an acetyl group that is delivered to HMGS by a standalone donor ACP (ACP_D). ACP_D is encoded in the β-branching cassette along with an unusual ketosynthase-decarboxylase (KS_{DC}), which generates acetyl-ACP_D by decarboxylating malonyl-ACP_D. Chemical diversification of polyketide β-branches is usually achieved by variation in the later steps by either extra enzymes to process the carboxyalkyl branch or versions of the standard enzymes with non-canonical activities (Figure 1.3). The HMGS reaction, by which an aldol addition of an acetyl donor to an acetoacetyl-like acceptor generates the initial carboxyalkyl branch (Figure 2.1), is usually non-variable.

There are two characterized examples (leinamycin and myxovirescin) in which HMGS uses a propionyl donor, in both cases resulting in an unusual branch (Figure 4.1)[20,21]. ACP_D delivers the propionyl group to a specialized HMGS, which produces a methylated initial branch. Instead of a KS_{DC}, the leinamycin biosynthetic pathway includes a specialized acyl-transferase/decarboxylase (AT/DC, LnmK) that directs formation of the propionyl-ACP_D substrate for HMGS (LnmM)[29]. Myxovirescin biosynthesis involves two β-branching events, which install ethyl and methoxymethyl branches[20]. The former initiates in an analogous manner to the leinamycin branch, using a propionyl donor and a specialized ACP_D/HMGS (TaE/TaF) pair, while the latter is generated using an acetyl donor and a standard ACP_D/HMGS (TaB/TaC) pair. Both initial branches are dehydrated and decarboxylated by the same ECH_1/ECH_2 enzymes, but the acetyl-derived branch is subsequently hydroxylated and methylated to form the final, methoxymethyl branch.

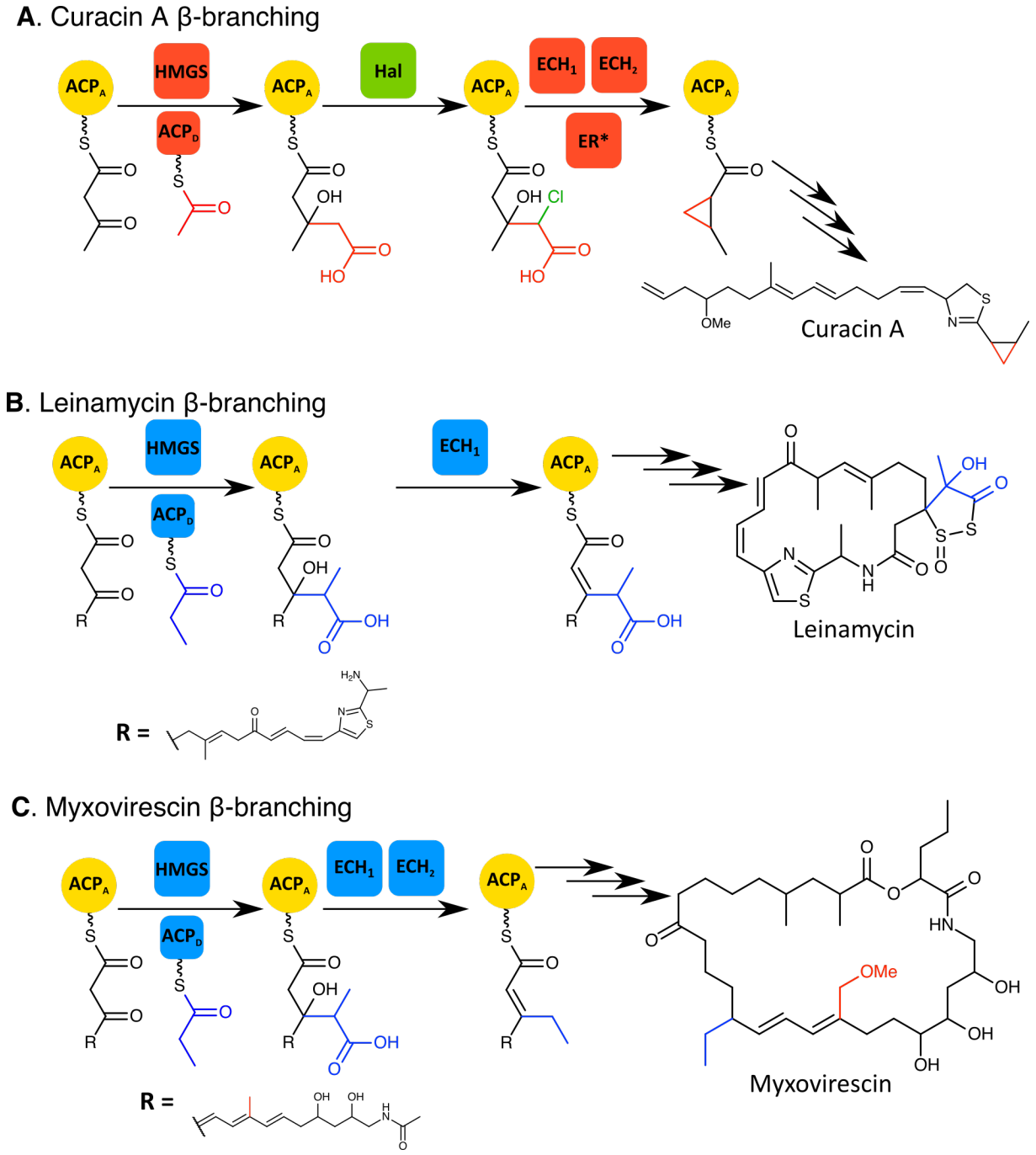


Figure 4.1 Polyketide β-branching with non-canonical outcomes. β-branches are shown in color, with red for those generated from an acetyl donor and blue for a propionyl donor. β-branching enzymes and ACP_D are colored accordingly. **A.** Curacin A branching. The curacin A HMGS recognizes an acetyl-ACP_D and is more typical of PKS HMGSs in sequence. The branch-processing enzymes include a halogenase and an ER* cyclopropanase, which are non-canonical. **B**. Leinamycin branching. HMGS recognizes a propionyl-loaded ACP_D, creating a methylated initial branch. An ECH_1 dehydrates the β-position. The final fate of the branch is as part of a 5-membered heterocycle in the final product. **C.** Myxovirescin branching. Myxovirescin biosynthesis

involves two β-branching events. In the second, an HMGS similar to that of LnmM installs a propionyl branch, which becomes an ethyl substituent in the final natural product.

Compared to other HMGS sequences (Figure 3.2), LnmM and TaF have several substitutions to ordinarily conserved residues. LnmM and TaF lack the flexible loop that is characteristic of PKS HMG-ACP synthases[90]. In TaF these substitutions may be involved in distinguishing between its cognate ACP_D/ACP_A substrates (TaE/Ta1-M8) and those of TaC (TaB, Ta1-M6), as discussed in chapter 2. In examining which residues adapt HMGS to use a propionyl donor, we focused on three active site residues, Ala164, Ser167, and Met298, which are conserved among HMGS sequences but vary in LnmM and TaF. Ser167, which makes a hydrogen bond to Ppant, is an alanine in both LnmM and TaF, while Ala164 on the same helix is serine. These may point to different interactions with Ppant to allow reaction with a propionyl donor. Met298 is on a loop adjacent to the catalytic cysteine in the HMGS active site and is a cysteine in LnmM and TaF. The presence of a smaller side chain in this position could make room for the extra methyl group on a propionyl-Cys intermediate.

Engineered PKS pathways have low throughput, and one reason is that enzymes of the engineered pathway repeatedly encounter non-cognate substrates[2,3]. Characterizing the substrate range of these enzymes may guide rational engineering efforts. In particular, we are interested in how structural features in PKS enzyme active sites influence substrate recognition and confer non-canonical activity. We examined the effect of A164S, S167A, and M298C substitutions to the curacin HMGS on overall reactivity and acyl transfer from ACP_D to HMGS with acetyl, propionyl, and butyryl

substrates, along with all combinations of those substitutions. We solved the structure of $HMGS_{C114Q}$, designed as an isosteric mimic of the acetyl-HMGS intermediate, to further investigate how these residues may interact with a propionyl-loaded HMGS.

Results

Structure of acetyl-HMGS intermediate. Each catalytic residue in the HMGS active site occurs in a loop connecting two secondary structure elements in the thiolase fold. The same is true for Ser328, the backbone amide of which forms the HMGS oxyanion hole, for the flexible loop region, and for Met298. Met298 sits in the active site close to the catalytic cysteine (Cys114) and glutamate (Glu82) (Figure 4.2). TaF and LmnM have a cysteine in the Met298 position, leading us to hypothesize that the smaller cysteine side chain may open up room for the extra methyl of a propionyl donor. We generated a C114Q variant of HMGS as an isosteric mimic of the acetyl-HMGS intermediate to further investigate the structural role of Met298. $HMGS_{C114Q}$ was crystallized both with and with out bound holo-ACP_D (Table 4.1, Figure 4.3-4). Active site residues were identically positioned in the wild-type and $HMGS_{C114Q}$ structures both with and without bound holo-ACP_D. The lone structural difference as compared to wild-type HMGS structures was in the enzyme-only structure, where residues 28-33 had discontinuous density. The same region in $HMGS_{WT}$ has poor density, but in the case of $HMGS_{C114Q}$ these residues could not be modeled confidently. Despite an apparent increase in flexibility of Arg33, Ppant bound in the same position as in the holo-ACP_D-$HMGS_{WT}$ complex. As expected, the Gln114 side chain carbonyl sits in the oxyanion

hole, making a hydrogen bond to the Ser328 backbone amide. The Met298 side chain packs closely with both Gln114 and the catalytic Glu82 side chain (Figure 4.2).

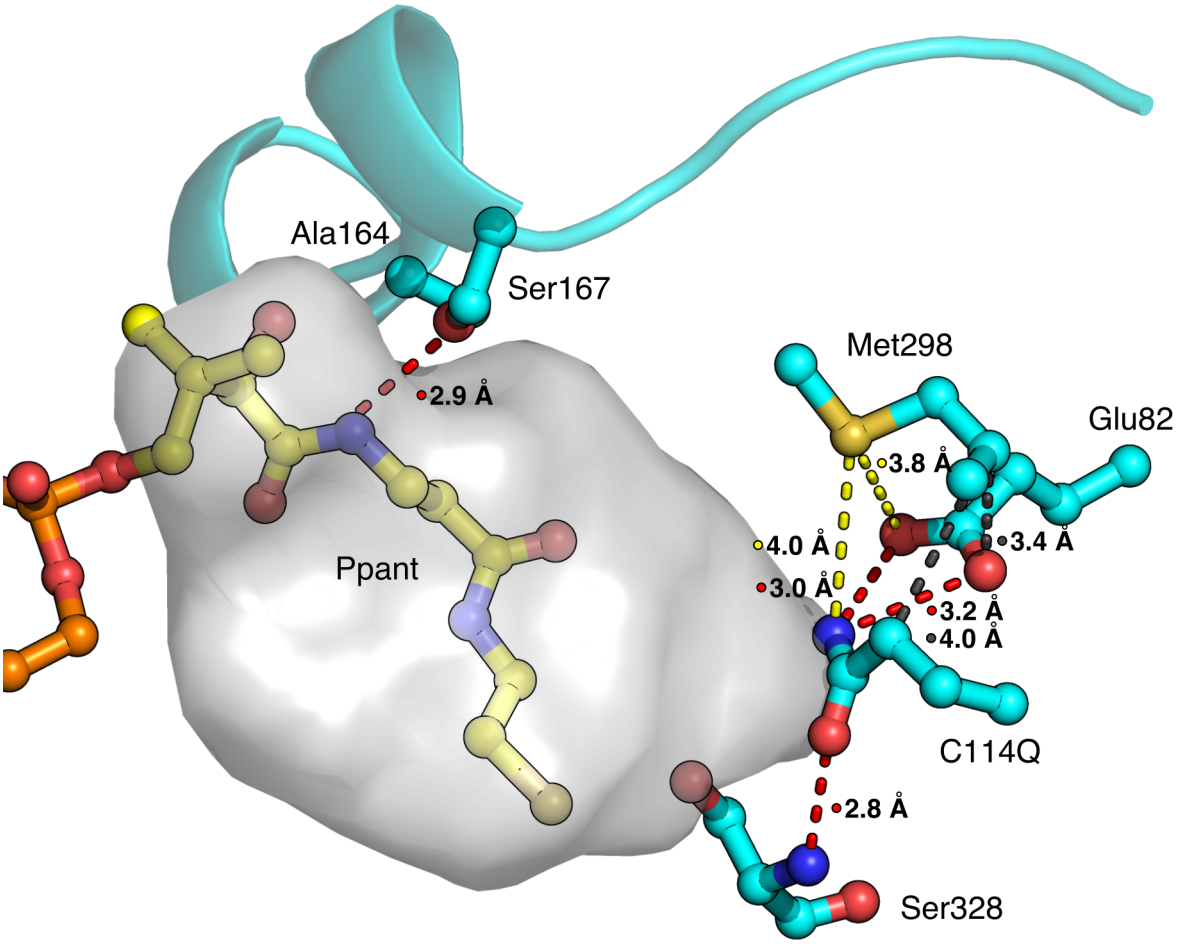


Figure 4.2 Role of Met298 in the acetyl-HMGS intermediate. HMGS residues are shown in cyan, Ppant in yellow, and the HMGS active site cavity in gray. C114Q isosterically mimics an acetyl-Cys114 intermediate, and the side chain makes an expected hydrogen bond to the Ser328 backbone amide. The Gln114 side chain amide points to the Glu82 carboxylate in a position consistent with deprotonation of acetyl-Cys114. Met298 packs closely against both catalytic components in the active site.

Table 4.1 Data collection and refinement statistics

	$HMGS_{C114Q}$	Holo-ACP_D/$HMGS_{C114Q}$
Wavelength (Å)	1.033	1.033
Space Group	$P3_121$	$P3_121$
a,b (Å)	100.32	101.29
c (Å)	105.89	104.91
Data range (Å)	45.33-2.00 (2.05-2.00)	45.61-2.35 (2.43-2.35)
Unique Reflections	42082 (3048)	26385 (2551)
Completeness (%)	100 (100)	100 (100)
Multiplicity	19.5 (20.0)	9.8 (10.2)
Wilson B-factor	30.5	36.8
R_{merge}	0.078 (0.816)	0.190 (1.473)
Average I/σ_i	25.1 (4.5)	9.4 (1.4)
CC 1/2	1.00 (0.90)	1.00 (0.52)
R_{work}	0.157	0.181
R_{free}	0.176	0.219
Protein atoms	3081	3845
Ppant atoms	--	21
Water/Ion atoms	155	209
RMSD bond length (Å)	0.010	0.013
RMSD bond angle (°)	1.35	1.63
Average B-factors		
HMGS	40.0	41.1
ACP_D		73.6
Ppant	--	67.5
Water	50.5	49.0
Ramachandran		
Favored (%)	97.4	97.1
Generously Allowed (%)	2.6	2.7
Outlier (%)	0	0.2

Figure 4.3 Ramachandran analysis of $HMGS_{C114Q}$ (enzyme only).

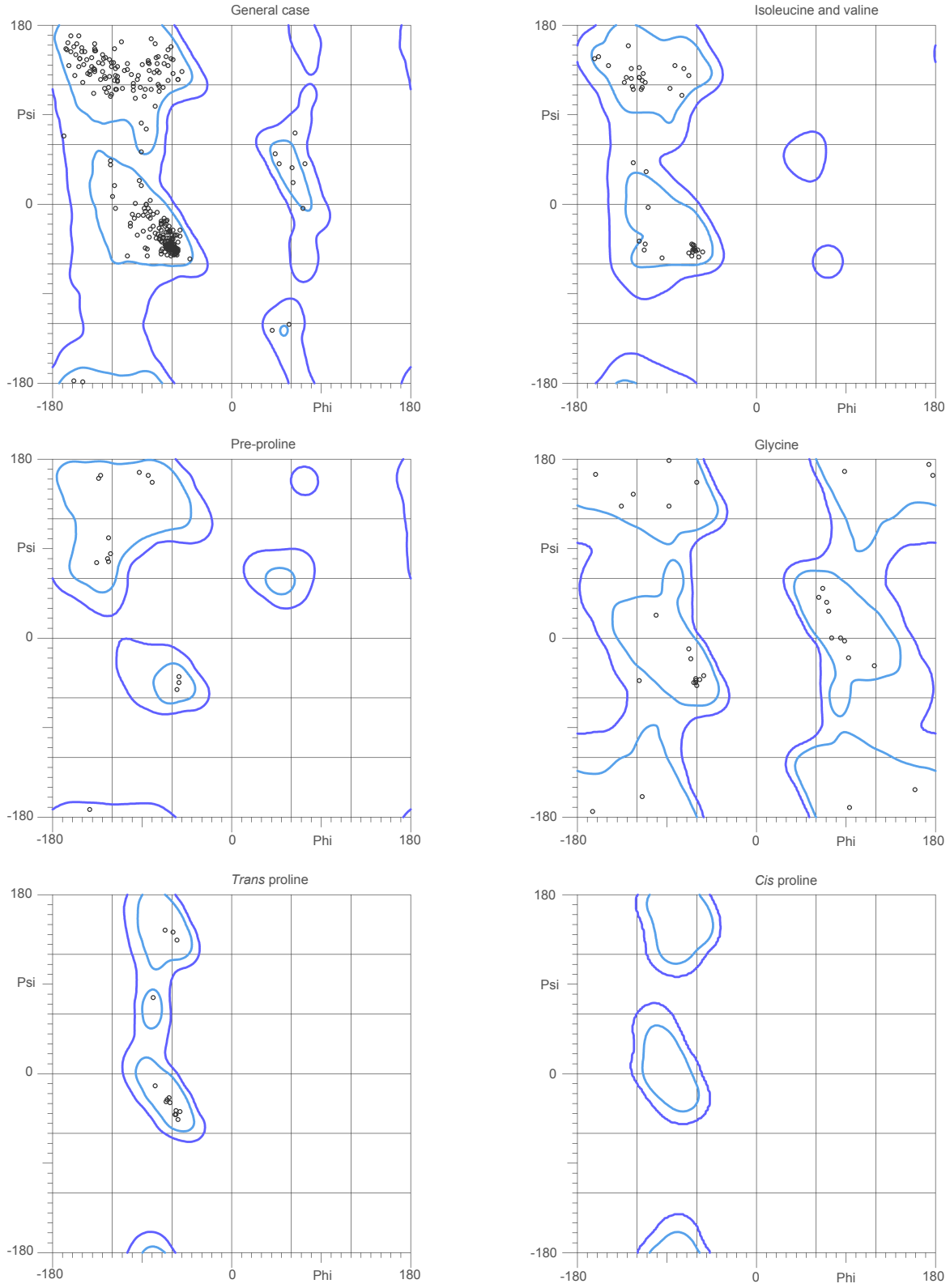
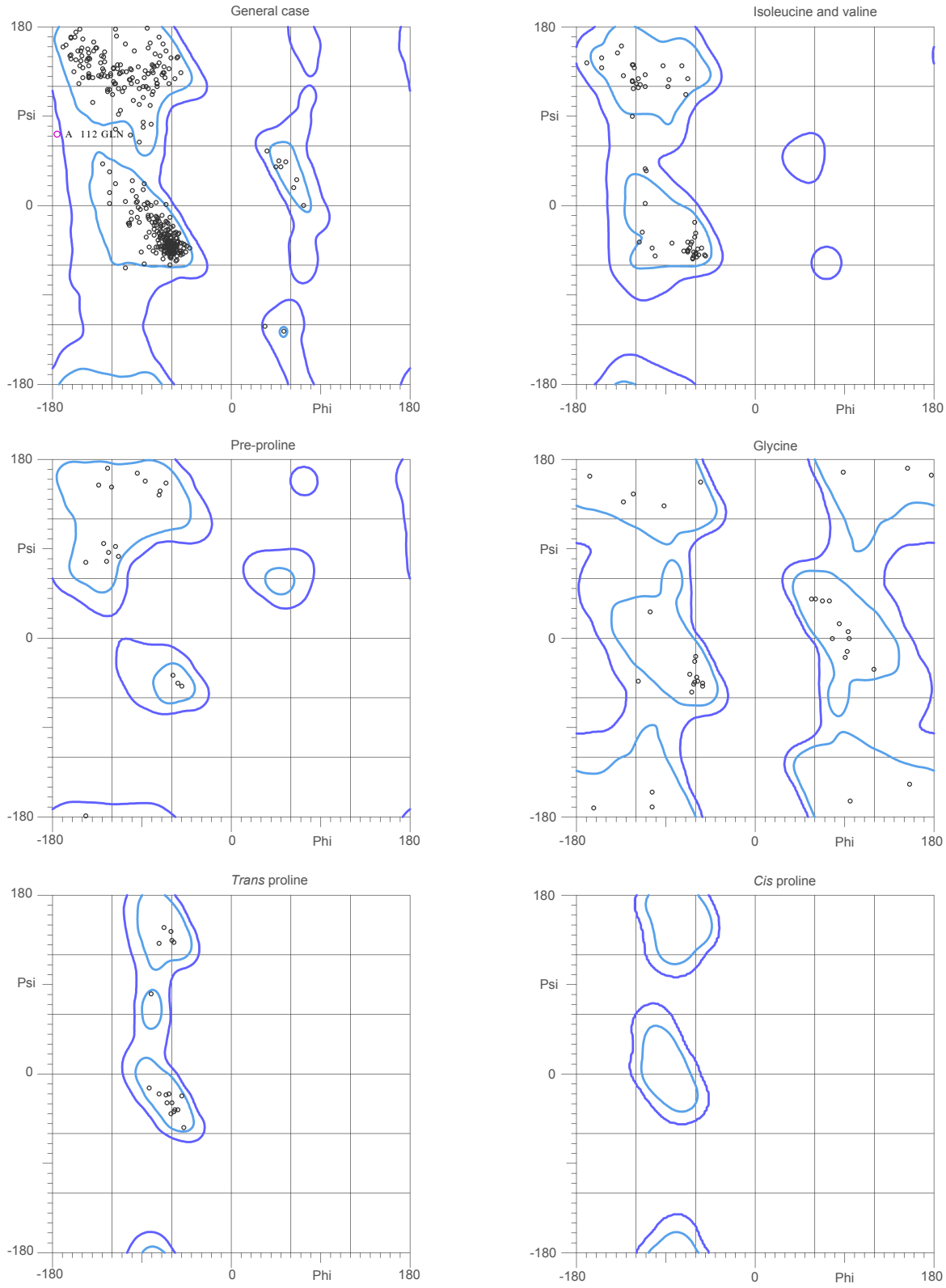


Figure 4.4 Ramachandran analysis of Holo-ACP_D/$HMGS_{C114Q}$ complex.



HMGS reaction with propionyl and butyryl substrates. CurB ACP_D was purified in the apo state and loaded *in vitro* with acyl-Ppant from acetyl, propionyl, or butyryl-CoA. Equimolar acyl-ACP_D and acetoacetyl-ACP_A[90] (20 μM) were mixed with 2 μM variants of CurD HMGS for a 30 minute reaction at 25 °C. Reactions were quenched by addition of formic acid, and mixtures were analyzed by LC-MS for detection of substrate (acetoacetyl-Ppant) and product ions[59] (Figure 4.5). We tested wild-type curacin A HMGS, as well as variants at the Ala164, Ser167, and Met298 positions designed based on LnmM and TaF sequences.

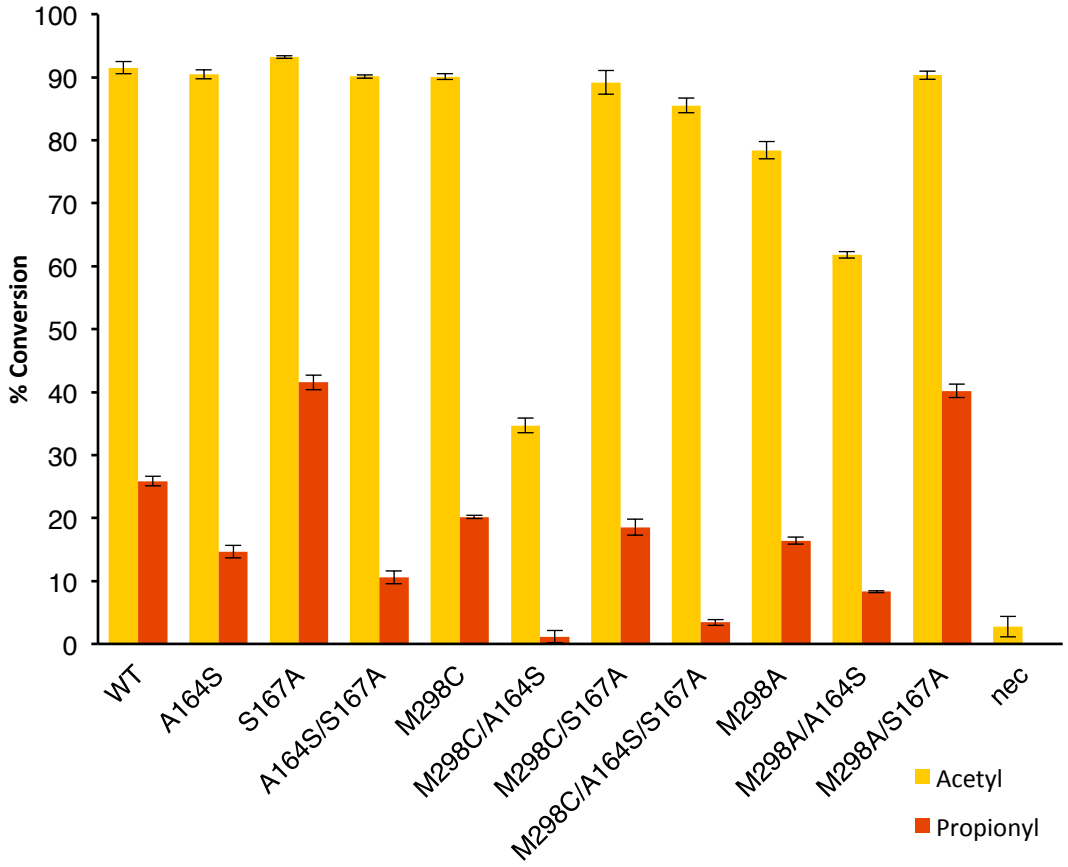


Figure 4.5 HMGS activity with a propionyl donor. Percent conversion is shown of acetoacetyl-ACP_A to HMG-ACP_A or to 3-hydroxy-3,4-dimethylglutaryl-ACP_A (HDMG-ACP_A) by reaction with acetyl or propionyl-ACP_D, respectively, and variants of HMGS.

HMGS showed robust conversion of propionyl-ACP_D and acetoacetyl-ACP_A to HDMG-ACP_A in a 30 min reaction, though it was nearly four-fold less active than with its cognate acetyl-ACP_D substrate. No reaction was observed with butyryl-ACP_D for any of the variants tested. $HMGS_{A164S}$, $HMGS_{S167A}$, $HMGS_{M298C}$, and $HMGS_{M298A}$ had reactivity with an acetyl donor indistinguishable from wild-type HMGS. In all cases, A164S reduced activity with a propionyl donor. A164S, when combined with M298 substitutions, was the only substitution that had substantially reduced activity with acetyl-ACP_D. $HMGS_{A164S/M298C}$ exhibited ~2-fold reduced production of HMG-ACP_A (acetyl donor) and negligible production of HDMG-ACP_A (propionyl donor). $HMGS_{A164S/M298A}$ displayed smaller reductions in activity than the $HMGS_{A164S/M298C}$. $HMGS_{S167A}$ converted 42 ± 1% of propionyl-ACP_D and acetoacetyl-ACP_A to product, an increase over 26 ± 1% from wild type.

The M298C and M298A substitutions also had differing effects on HMGS reactivity when combined with the S167A substitution. Production of HDMG-ACP_A from reaction with $HMGS_{S167A/M298C}$ was indistinguishable from wild-type. $HMGS_{S167A/M298A}$, however, had a similar gain to activity as was observed from $HMGS_{S167A}$. Finally, $HMGS_{A164S/S167A/M298C}$ displayed wild-type production of HMG-ACP_A and negligible production of HDMG-ACP_A. In this triple variant, the S167A substitution rescued activity with acetyl-ACP_D but did not significantly affect activity with propionyl-ACP_D.

Loss of HMGS activity with a propionyl donor could be due to inefficient transfer of the acyl group from ACP_D to HMGS and/or slower aldol addition of the propionyl nucleophile to acetoacetyl-ACP_A. To determine whether the acyl-transfer reaction was

impaired, we incubated 20 μM acetyl, propionyl, and butyryl-ACP_D with 50 μM wild type HMGS (Table 4.2) for 30 sec at 25 °C. The results demonstrate that acyl transfer is unimpaired with propionyl and butyryl donors.

Table 4.2 Donor acyl transfer

	Acetyl-ACP_D	Propionyl-ACP_D	Butyryl-ACP_D
Acyl transfer[1] (%)	29 ± 1	59 ± 2	60.6 ± 0.4

[1]Acyl transfer is calculated as percent of ppant ions loaded with acetyl, propionyl, or butyryl averaged over three incubations with $HMGS_{WT}$, subtracted from a no-enzyme control.

We hypothesized that the observed increase in deacylation of propionyl and butyryl-ACP_D could be due to overestimation of concentrations of the two protein stocks, and not due to increased transfer. The curacin A ACP_D (CurB) has no tryptophan residues and does not respond linearly to Bradford assays using BSA standards, necessitating estimation of the concentration based on a Bradford assay using the curacin A ACP_A (CurA ACPII) as a standard. We measured relative concentrations of each acyl-ACP_D by comparing peak areas for the ACP_D and ACP_A intact protein masses. Using deconvoluted mass spectra from activity assays with $HMGS_{WT}$, we calculated the relative total peak height for ACP_D relative to ACP_A. We determined a relative peak height of 4.15 ± 0.06 for acetyl-ACP_D, 1.88 ± 0.04 for propionyl-ACP_D, and 1.80 ± 0.05 for butyryl-ACP_D. Thus, we conclude that the reduced deacylation of acetyl-ACP_D was due to a greater abundance in the assay. We repeated the assay using a corrected concentration of ACP_D (using the calculations above) and observed 71.8 ± 0.6% acetyl transfer to HMGS. We also note that in all assays donor acyl-Ppant substrate was observed. Decreased reactivity with propionyl and butyryl substrates is a

result of impaired catalysis, and is not due to complete consumption of the donor substrate as a result of overestimating its concentration.

Discussion

Structures of $HMGS_{C114Q}$, the acetyl-HMGS intermediate mimic, and sequences of LmnM and TaF both point to the M298C substitution as possibly alleviating steric clashes with the larger propionyl-HMGS intermediate in LnmM and TaF. However, the M298C and M298A substitutions do not improve production of HDMG and may even diminish it, contradicting such an obvious explanation for the role of Met298. Because Met298 packs against both the acetyl-Cys intermediate and Glu82, the M298C and M298A substitutions may impair enolization in the curacin A HMGS. $HMGS_{S167A/M298C}$ and $HMGS_{S167A/M298A}$ show conflicting effects on HMGS activity with a propionyl donor. The M298C thiol may be positioned such that Glu82 is no longer in an ideal orientation to deprotonate the propionyl donor, whereas the M298A side chain would not disrupt Glu82 positioning (Figure 4.2). Alternatively, M298C may be able to commandeer substrate by nonproductive transfer of acetyl or propionyl from the ACP_D Ppant to Cys298. The $HMGS_{C114Q}$ structure shows that the acetyl-Cys114 carbonyl would be perfectly positioned in the oxyanion hole to stabilize an enolate nucleophile. An acyl-Cys298 adduct, presumably, would not be positioned for deprotonation by Glu82, for stabilization by the oxyanion hole, and for attack on the β-carbonyl of the acetoacetyl acceptor.

The A164S substitution impairs HMGS activity. Hypothetically, Ser164 could be positioned such that it could make a hydrogen bond to Ppant, either from ACP_A or

ACP_D. Met298 substitutions paired with the A164S substitution create the most significant reductions to HMGS activity with acetyl and propionyl donors, supporting the conclusion that these substitutions both interfere with the normal HMGS mechanism. In general, Ppant positioning appears to be critical for HMGS activity. We previously observed that Ppant positioning in the ACP_D/HMGS complex is dependent on acylation state[90] (Chapter 2), and that some active site substitutions can eliminate activity without seriously disrupting ACP affinity (Chapter 3). We presume that the curacin A HMGS catalytic residues are not ideally positioned for reaction with propionyl, and the HMGS reaction could be additionally sensitive to improper Ppant positioning as a result of a hydrogen bond from A164S. Conversely, the S167A substitution could eliminate a hydrogen bond to the acetoacetyl-Ppant and increase its flexibility, allowing it to adopt a more favorable position for reaction with a propionyl donor.

Although the A164S, S167A, and M298C substitutions are present in both LnmM and TaF, they are not sufficient to confer selectivity for a propionyl donor to the curacin A HMGS, which normally uses acetyl. TaF has differences from canonical HMGS sequences that are responsible for selecting the TaE ACP_D and not TaB, the other ACP_D in the myxovirescin biosynthetic cluster. Thus, sequences of acetyl-donating and propionyl-donating HMGSs contain differences responsible for ACP selectivity and differences that are catalytically relevant. It remains untested whether propionyl-donating HMGSs show a catalytic preference for propionyl over acetyl. It is also unknown if such a catalytic preference would be due to differences in ACP interaction.

HMGS discriminates between ACP_D and ACP_A linked substrates, so ACPs may interact with a propionyl-donating HMGS in a way that is optimized for the larger donor acyl.

Of the various strategies for engineering PKS pathways, modulating the activity of individual domains in a PKS pathway by mutagenesis has been relatively successful in proof of principle[2]. For several classes of PKS enzymes in which some members exhibit altered activity, however, structural motifs that modulate the altered activity have been identified. Recent examples from our lab include the cyclopropanase activity of the CurF ER[36] and control over regioselectivity of methyltransfer by the MycF methyltransferase[91]. The current results of HMGS mutagenesis reveal at least one position, Ser167, that can be substituted to enhance reactivity with a non-cognate substrate. Meanwhile the activity of other active site variants suggest that subtle perturbations of several substitutions may confer propionyl-selectivity to the enzyme, and there is no "switch" comprising only a few residues. Finally, a definitive role for the conserved flexible loop region of HMGS remains unknown, and a more complete understanding of the functional relevance of that region may provide further guidance for engineering HMGS.

Methods

Protein expression and purification

E. coli BL21 (DE3) were transformed with plasmids for the curacin A HMGS[90], ACP_A[49], and ACP_D[90]. Transformed *E. coli* were grown in terrific broth (TB) at 37 °C to an OD_{600} of 1, cooled to 20 °C, then gene expression was induced with 200 μM IPTG.

pHMGScur[90] was coexpressed with pGro7[75], which was induced at an OD_{600} of ~0.3 with 2 g/L arabinose. Cell pellets from pHMGScur expression were resuspended in 4 mL/g of buffer containing 50 mM $(NH_4)_2SO_4$, 50 mM HEPES pH 7.0, 20 mM imidazole, and 10% v/v glycerol. Pellets for ACP_A were resuspended in 300 mM NaCl, 50 mM Tris pH 7.5, 20 mM imidazole, and 10% v/v glycerol, and those for ACP_D were resuspended in 300 mM NaCl, 20 mM imidazole, and 50 mM PBS pH 7.5 for ACP_D. 1 mg/mL lysozyme, 2 mM $MgCl_2$, and 0.1 mg/mL Dnase were added to each suspension, followed by incubation for 30 min at room temperature, 30 min on ice, then sonication. Each protein was purified by immobilized metal affinity chromatography with a 5 mL His Trap column (GE Healthcare) using a 50 mL gradient of 20 to 400 mM imidazole (in addition to lysis buffer components), then size exclusion chromatography (SEC). SEC for HMGS was a Superdex S200 column (GE Healthcare) with a buffer of 50 mM $(NH_4)_2SO_4$, 20 mM HEPES pH 7.0, and 10% v/v glycerol. ACPs were purified on a Superdex S75 (GE Healthcare) using 150 mM NaCl, 50 mM Tris pH 7.5, and 10% v/v glycerol. Proteins were concentrated then flash cooled in liquid N_2 and stored at -80 °C.

Site directed mutagenesis

pHMGScur[90] mutants were made by overlapping-primer PCR. The reaction components were 1X KOD buffer (Novagen), 2mM $MgSO_4$, 10% v/v DMSO, 0.2 ng/µL of pHMGScur, 150 nM of each primer (Table 4.3), 0.2 mM dNTPs (Novagen), 0.02 U/µL KOD Hot Start polymerase (Novagen). Each reaction ran for 18 cycles of 20s melting at 95 °C, 10s annealing at 60 °C, and 3 min extension at 70 °C. 0.2 U/µL of DpnI (NEB) was added to each PCR mixture, then incubated for 3 hr at 37 °C. 2 µL of each digested

PCR mixture was transformed into *E. coli* XL-1 Blue cells. The mutant constructs were purified by miniprep (Qiagen), and checked for the target mutation by Sanger sequencing. Combinations of M298 mutations the A164S, S167A, or A164S/S167A mutations were made by PCR with the M298C or M298A primers and the corresponding A164S, S167A, or A164S/S167A mutant construct.

Table 4.3 Mutagenic primers

Subsitution	Sense Primer	Antisense Primer
A164S	caatcaattacgattggtctttagtgaacccagtagtggtgctg	cagcaccactactgggttcactaaaagaccaatcgtaattgattg
S167A	ggctccagcaccactagcgggttcagcaaaagac	gtcttttgctgaacccgctagtggtgctggagcc
M298C	ttagtctactgccaacaggtaggcaatatttgcggagcaactttgtttttatc	gataaaaacaaagttgctccgcaaatattgcctacctgttggcagtagactaa
M298A	gataaaaacaaagttgctccgcaaatattgcctacctgttggcagtagactaa	acaaagttgctcccgcaatattgcctacctgttggcagtag
A164S/S167A	agaagcaatcaattacgattggtctttagtgaacccgctagtggtgctggagc	gctccagcaccactagcgggttcactaaaagaccaatcgtaattgattgcttct
C114Q	atgtttgaactcaagcaagctcagtactcaggaaccgctggctta	taagccagcggttcctgagtactgagcttgcttgagttcaaacat

Protein Crystallization

HMGS readily crystallizes as a triple alanine variant (K344A/Q345A/Q347A, AAA)[78,90]. The three alanine mutations were made to pHMGScur (C114Q) as previously described[90]. $HMGS_{C114Q/AAA}$ was crystallized as previously described[90] by hanging-drop vapor-diffusion at 20 °C using a well solution of 6% v/v PEG 8k, 30 mM $(NH_4)_2SO_4$, and 1X MMT pH 6.5 (Qiagen). $HMGS_{C114Q/AAA}$ was similarly cocrystallized with a 7:1 molar

ratio of holo-ACP_D to HMGS in a well solution of 6% v/v PEG 8k, 60 mM $(NH_4)_2SO_4$, and 1X MMT pH 6.5 (Qiagen).

Structure solution and refinement

Diffraction data were collected at the Advanced Photon Source (APS) GM/CA 23ID-D and processed with XDS[79]. Diffraction limits were assessed using $CC_{1/2}$ and I/σ_I statistics. Data were phased by rigid body refinement in refmac5[80,84] using HMGS and ACP_D models (5KP5, 5KP7)[90]. Subsequent rounds of model-building were carried out in Coot[86] followed by refinement in refmac with TLS[85] parameters. Waters were modeled in coot based on F_O-F_C difference density and hydrogen bonding. Molprobity was used to validate final models and generate Ramachandran plots[87]. Molecular figures were generated in PyMOL[88].

HMGS activity assay

ACPs were loaded *in vitro* from acyl-CoAs using *B. subtilis* phosphopantetheinyl transferase (Sfp)[92]. ~200 µM purified ACP was incubated with 10 mM $MgCl_2$, 50 mM Tris pH 7.5, 0.2 mg/mL Sfp, 1 mM acyl-CoA, and 5% v/v glycerol ~18 h at 4 °C, followed by SEC. 2 µM HMGS was mixed with 20 µM acetyl, propionyl, or butyryl-ACP_D and 20 µM acetoacetyl-ACP_A in 20 mM ammonium sulfate, 20 mM HEPES pH 7.0, and 2 mM TCEP and incubated for 30 minutes at 25 °C. Reactions were quenched with 5% v/v formic acid. Product formation was detected by HPLC-mass spectrometry by QTOF (Agilent) with Ppant ejection[59]. Reaction progress is reported as the product percent of the product and acetoacetyl-Ppant ions over three replicates.

ACP_D acyl-transfer assay

20 μM Acetyl, propionyl, and butyryl ACP_D were mixed in equimolar quantity with HMGS in 20 mM ammonium sulfate, 20 mM HEPES pH 7.0, and 2 mM TCEP. Reaction mixtures were incubated for 30 sec at 25 °C, then reactions were quenched with 5% v/v formic acid. Reaction mixtures were analyzed by HPLC-mass spectrometry by QTOF (Agilent) with Ppant ejection[59]. The percent of Ppant ions loaded with acetyl was averaged for three replicates, then the percent of acetyl-Ppant ions from a no-enzyme control was subtracted to calculate percent acyl-transfer.

Chapter 5. Conclusions and Future Directions

Summary and Conclusions

Engineering natural product biosynthetic pathways has enormous potential for pharmaceutical production and discovery, but faces many challenges[2,3]. Chief among these is an incomplete understanding of the rules, dictated by protein structure, that govern substrate selectivity and interdomain interactions. Thus, structure determination for proteins and complexes involved in natural products biosynthesis complements research concerning pathway engineering. In modular polyketide synthase (PKS) pathways, alkyl substitution by β-branching could be a potent and versatile tool for creating variants of existing natural products or for creating natural product libraries. β-branching enzymes recognize a specialized acyl-carrier protein (ACP_A) in the pathway[24,20,25,90] and are almost always standalone proteins[18]. Introducing a non-natural β-branch into a PKS pathway then involves fewer complications with incompatible protein-protein interfaces.

In this thesis, we present the detailed structural and biochemical characterization of the hydroxymethylglutaryl synthase (HMGS) of curacin A biosynthesis (CurD). Our central goal at the outset of the project was to determine how differences in protein-protein interactions between HMGS and each of the two acyl-carrier proteins (ACP) it recognizes lead to selectivity. Chapter 2 described the purification, structure, and ACP_D-

bound complex of HMGS, laying the groundwork for later chapters. We furthermore identified several important interactions between HMGS and ACP_D and determined their effects on HMGS activity and ACP_D affinity. We found that Arg33 makes a critical interaction with the Ppant phosphate. Asp222 caps a 3_{10} helix III on ACP_D, with D222A and D222R substitutions causing 4-5 fold losses in HMGS activity. Overall we observed that electrostatic contacts are important to the ACP_D-HMGS interaction, that the ACP_D surface has striking shape complementarity with HMGS, and that decreases in ACP_D affinity and HMGS activity are not well correlated. Finally, we found that the ACP_D Ppant is flexible in the active site before acetyl transfer but can position the acetyl-thiol proximal to the catalytic Cys114, and is rigidly positioned into a Cys114-distal "thiol pocket" region in the active site after acetyl-transfer.

Chapters 3 and 4 follow up our description of the HMGS structure and examine the role of conserved structural motifs in HMGS and the donor acyl selectivity of HMGS. Most substitutions that affected ACP_D interaction did not cause corresponding effects on ACP_A affinity. However, we determined that ACP_A and ACP_D bind competitively to HMGS, and identified Arg33 and Arg262 as residues that interact with ACP_A. Based on ACP affinity and HMGS activity changes from several substitutions to HMGS we conclude that differences in ACP interaction likely create different Ppant positioning for each ACP. We also determined that several residues in the thiol pocket identified in chapter 2 were essential to HMGS activity during the aldol addition step.

We did not observe that residues of the flexible loop region were important to HMGS activity or ACP interaction. The A164S substitution is deleterious to HMGS

activity. Although the A164S and M298C substitutions were ineffective in our goal of engineering the CurD HMGS to efficiently react with a propionyl donor, they highlighted the importance of Ppant positioning. Finally, we speculate that many differences of LnmM or TaF from CurD may collectively reprogram the enzyme, but further structural characterization is needed.

There are still outstanding questions regarding the biochemistry β-branching. It is not understood how ACP_D is loaded in the case of *cis*-AT PKS pathways. While ACP selectivity in HMGS is well documented[24,20,25,90], the specific features of ACP_A that are recognized by HMGS and the other β-branching enzymes are unknown. Finally, β-branching requires a pause in the PKS assembly line whereby transfer of the polyketide intermediate to the next module should delay to allow the branching reactions to take place. There are no published studies of this phenomenon, which could result from a unique architecture of the branch-accepting module, the often observed sequence of tandem ACP_As, or the differential recognition of branched vs. unbranched intermediates by the downstream KS. In fact, recognition of β-branches, which can be relatively large substituents, by downstream enzymes in general is an unexplored area of PKS biochemistry and could be crucial to engineering pathways by addition of β-branches.

Future Directions

ACP_A Interaction. We did not observe strong decreases to holo-ACP_A affinities as a result of our HMGS substitutions, but a next step may be to measure apo-ACP_A affinities. The effects of substitutions to residues that interact with ACP_D were much greater for apo-ACP_D than for holo-ACP_D, so apo-ACP_A affinities may paint a much

clearer picture of the protein-protein interface. The main challenge is that the apo-ACP_A affinities for HMGS variants are expected to be lower than that of holo-ACP_A, especially for variants that affect affinity, necessitating a much greater investment of material for each experiment.

Extensive crystallization screening of CurD HMGS with excess ACP_A, excess ACP_D, and crosslinked with ACP_A via bismaleimidoethane has only ever yielded one crystal form and has never produced a structure with bound ACP_A. It seems unlikely at this point that attempts to cocrystallize curacin A HMGS with ACP_A will be successful. A future effort may be to dialyze all salt from an HMGS sample with excess holo-ACP_A and to screen for conditions with low salt concentrations. ACP_A affinity data could not be obtained in high salt buffer and we have observed CurD HMGS crystals in a very wide variety of conditions when ACP_D was present. If the CurD HMGS crystal form is compatible with an ACP_A/HMGS complex, the interaction appears to be salt dependent and this could be the best chance to obtain crystals. It appears, however, that CurD HMGS crystallization is incompatible with an ACP_A/HMGS complex. The next step for pursuing this crystal structure would then be to explore ACP_A/HMGS pairs from other pathways. While HMGS proteins are often insoluble under standard expression conditions in *E. coli*, coexpression of HMGS plasmids with pGro7 is broadly applicable for producing soluble HMGS. Finally, crystallization of the ACP_A/HMGS complex could be circumvented entirely by using NMR to observe chemical shift perturbations of HMGS as a result of ACP_A binding. This approach could be challenging due to the size of the HMGS dimer, which is 98.8 kDa, but might yield enough data to confidently dock

an ACP_A/HMGS complex, and may additionally provide information about the flexible loop region.

HMGS reaction with a propionyl-donor. The first follow up to our activity assays of HMGS active site variants should be to crystallize them to more conclusively determine their effects on active site architecture. In chapter 4, we speculated that the A164S substitution may create a new hydrogen bond to Ppant and that Met298 substitutions may disrupt the positioning of Glu82 and the acetyl-Cys114 intermediate, each of which could be confirmed by crystal structures.

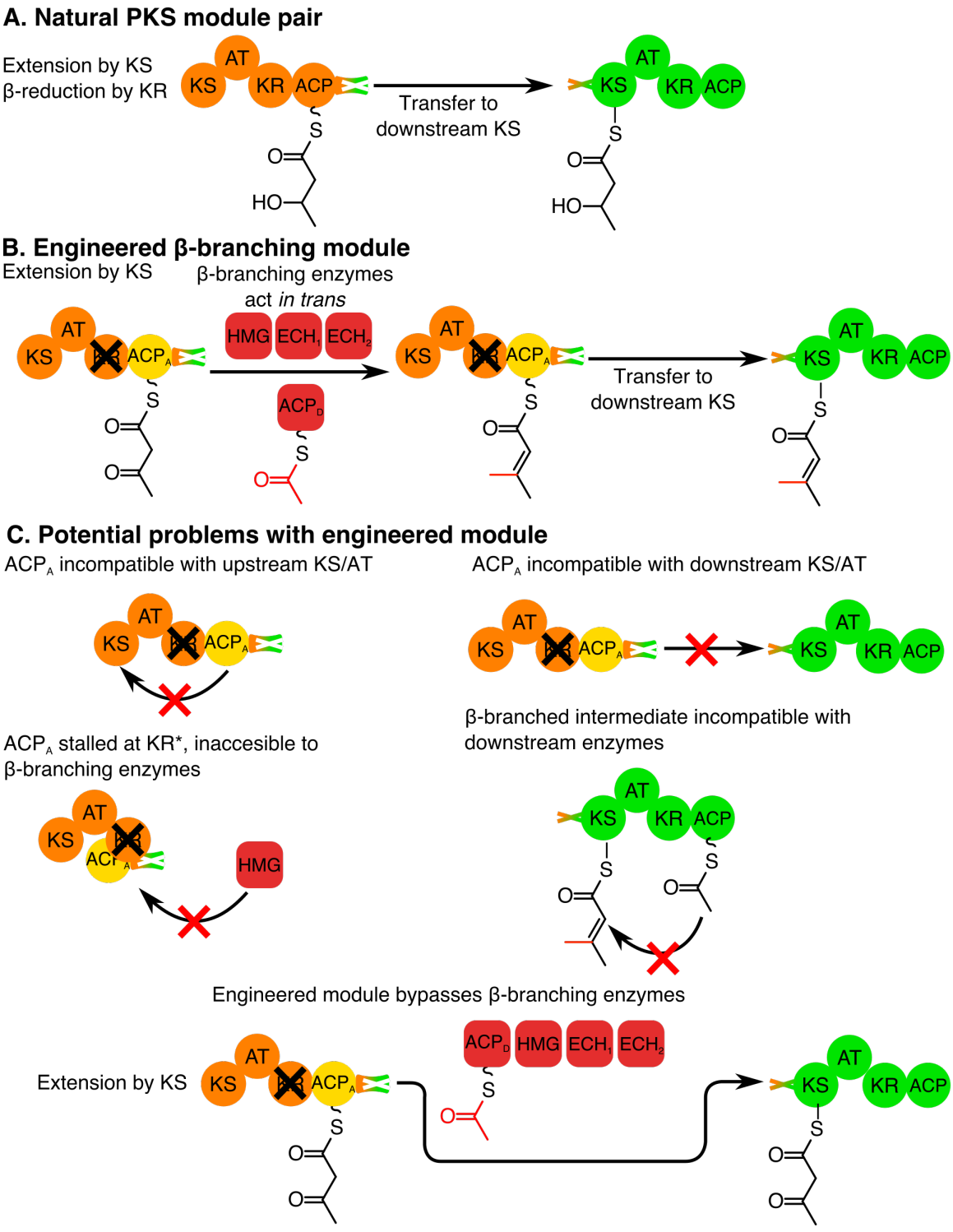
A second natural follow up to our assays would be to completely replace the flexible loop region of the CurD HMGS with a sequence from LnmM or TaF. Engineering this variant is straightforward by Gibson assembly, and the resulting protein may crystallize by the standard protocol for CurD HMGS. ACP selectivity has already been demonstrated for the TaF HMGS, and differences in ACP_D and ACP_A interaction lead to ACP selectivity in each characterized HMGS. Thus, differences in ACP interaction with an HMGS that donates propionyl may lead to greater efficiency with the propionyl donor, and could result from differences in the flexible loop region.

The CurD HMGS has diminished activity with a propionyl donor, but the observed level of conversion is still fairly robust by the scale of other PKS enzymes, for which it is not uncommon to use assay times on the order of hours to days. For this reason, the HMGS reaction with a propionyl donor could be used as a platform to test the substrate range of other β-branching enzymes. *In vitro* assays for the curacin A and jamaicamide branching enzymes are well established. If it can be demonstrated that other β-

branching enzymes still process the methylated branch generated by HMGS from a propionyl donor, it becomes feasible to use the previously identified propionyl-CoA acyltransferase/decarboxylase (AT/DC) enzyme to engineer pathways that produce methylated β-branches.

PKS engineering by β-branching. Structures have now been solved of almost all of the standard β-branching proteins[34,37,45,90], along with several proteins involved in the generation of nonstandard branches[29,36,38]. Only the ketosynthase/decarboxylase (KS_{DC}), which is presumed to decarboxylate malonyl-ACP_D to generate acetyl-ACP_D, still eludes structural characterization. Nevertheless, the next important step in the β-branching story is to attempt to engineer a pathway by introducing a branch, as well as to determine how β-branching interlocks with the other modules of PKS pathways. The ultimate goal would be to engineer a chimeric PKS module that can accept a β-branch and transfer it to the next module (Figure 5.1).

Figure 5.1 Engineering PKS by addition of β-branches. A. An unaltered PKS module (orange) extends the polyketide intermediate, modifies the β-position, then transfers the intermediate to the KS of a downstream module (green). Docking domains are shown appended to the orange ACP and the green KS. **B.** A module could be engineered to accept a β-branch by generating a chimera in which the ACP is replaced by a branch-acceptor ACP_A (yellow). Modification domains (in this case KR) must be knocked out by mutagenesis (black X). β-branching enzymes acting *in trans* substitute a methyl (red) for the β-carbonyl and the resulting branched product is transferred to the downstream module. **C.** The chimeric module may fail due to a number of different points due to incompatible ACP-enzyme interfaces or substrate-enzyme pairs.


A. Natural PKS module pair
Extension by KS
β-reduction by KR
AT
KS
KR
ACP
S
O
HO
Transfer to
downstream KS
B. Engineered β-branching module
Extension by KS
β-branching enzymes
act in trans
HMG
ECH1
ECH2
ACPD
ACPA
Transfer to
downstream KS
C. Potential problems with engineered module
ACPA incompatible with upstream KS/AT
ACPA incompatible with downstream KS/AT
ACPA stalled at KR*, inaccesible to
β-branching enzymes
β-branched intermediate incompatible with
downstream enzymes
Engineered module bypasses β-branching enzymes
Extension by KS

The most straightforward and minimally invasive approach would be to replace the ACP of a module with established *in vitro* activity with an ACP_A or a tandem ACP_A sequence, and then to knock out any modification domains by mutagenesis (Figure 5.1B). This approach avoids disruptions to the modular architecture that could occur if modification domains were excised, but still generates several potential bottlenecks (Figure 5.1C). Luckily, each of these potential pitfalls can in theory be singled out and evaluated *in vitro*. First, KS and AT activity in the chimeric module can be compared to the wild-type module to determine the compatibility of the ACP_A to react with these domains. Substrate dependent localization of the ACP to specific domains has been previously observed[14] and could stall the ACP_A at the knocked out KR. This possibility can be tested by loading the chimeric module with acetoacetyl, testing for HMGS reactivity, and then comparing to activity with the ACP_A-docking domain (ACP_A-dd) alone.

Bottlenecks may also occur due to incompatibility of the ACP_A or the branched substrate with the downstream KS. Efficient docking can be validated by fluorescence anisotropy, and our lab now has established protocols for this assay using fluorescently labeled docking domains[44] or ACP_A[90]. The ACP_A-dd can also be loaded with 3-methylcrotonyl-CoA, a commercially available analog of the “canonical” β-branched polyketide intermediate, and used to test both transfer of the branched intermediate to the downstream KS and validity of that intermediate as a substrate for the downstream enzymes.

Once these complications are controlled for, a one-pot reaction of the chimeric module containing an ACP_A (loaded with acetoacetyl) and docking domains for a downstream KS, the downstream KS, and the β-branching enzymes can be used to test the feasibility of producing an unnatural β-branch. This assay will test for potential bypass of the β-branching enzymes that could occur (Figure 5.1). One could then expand this scheme to test branching of longer polyketide intermediates, and, much further down the line, proceed to genetic engineering of the unnatural β-branch.

Outstanding questions about initiation of β-branching. Separate from ACP selectivity, the acceptor-acyl selectivity of HMGS has not been tested. Several PKS pathways include an HMGS that acts upon more than one intermediate, with some installing up to five branches[22,23]. This necessitates that the HMGS active site be able to accommodate long acyl chains without compromising positioning of the β-carbonyl for aldol addition. Selectivity may be tested by loading ACP_A with extended β-keto acyl groups and assaying for activity against a panel of HMGS enzymes. A second approach would be to pursue structures of HMGS enzymes that act more than once in a biosynthetic pathway.

A second open area for investigation is ACP_D loading with malonyl in *cis*-AT pathways. While embedded ATs are proposed as potential loading enzymes this hypothesis is untested. Due to the characterization of the propionyl-specific acyltransferase/decarboxylase (AT/DC) enzyme of leinamycin biosynthesis, the KS_{DC} enzyme encoded in most β-branching cassettes warrants examination. The proposed decarboxylation activity of this enzyme has been established, but there are no published

tests for acyl-transferase activity. KS_{DC} is sometimes referred to as a "KS_Q", but this is somewhat misleading (for example reference 46). The "Q" refers to a substitution in the active site of loading KS enzymes of the catalytic cysteine for a glutamine. KS_{DC} enzymes from β-branching cassettes actually have a cysteine to serine substitution and furthermore are highly distinct from other KS sequences. We presume that some of these differences confer ACP_D selectivity to the enzyme, but their catalytic implications should be further investigated.

Bibliography

1. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629–661 (2016).
2. Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33,** 203–230 (2015).
3. Hertweck, C. Decoding and reprogramming complex polyketide assembly lines: Prospects for synthetic biology. *Trends Biochem. Sci.* **40,** 189–199 (2015).
4. Fischbach, M. a & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106,** 3468–96 (2006).
5. Keatinge-Clay, A. T. The uncommon enzymology of *cis* -acyltransferase assembly lines. (2016). doi:10.1021/acs.chemrev.6b00683
6. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33,** 231–316 (2016).
7. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science (80-.).* **321,** 1315–1322 (2008).
8. Dutta, S., Whicher, J. R., Hansen, D. a, Hale, W. a, Chemler, J. a, Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **510,** 512–7 (2014).
9. Herbst, D. A., Jakob, R. P., Zähringer, F. & Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **531,** 533–537 (2016).
10. Keatinge-Clay, A. T. The structures of type I polyketide synthases †. *Nat. Prod. Rep.* **29,** 1050–1073 (2012).
11. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. 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 (2006).
12. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10,** 723–731 (2003).
13. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14,** 931–943 (2007).
14. Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L. & Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510,** 560–564 (2014).
15. Wong, F. T. & Khosla, C. Combinatorial biosynthesis of polyketides--a

perspective. *Curr. Opin. Chem. Biol.* **16,** 117–23 (2012).
16. Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. a & Santi, D. V. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23,** 1171–6 (2005).
17. Hagen, A., Poust, S., Rond, T. de, Fortman, J. L., Katz, L., Petzold, C. J. & Keasling, J. D. Engineering a polyketide synthase for *in vitro* production of adipic acid. *ACS Synth. Biol.* **5,** 21–27 (2016).
18. Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* **25,** 845–853 (2008).
19. Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V, Gerwick, L., Håkansson, K., Wipf, P., Smith, J. L., Gerwick, W. H. & Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **459,** 731–735 (2009).
20. Calderone, C. T., Iwig, D. F., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T. Incorporation of nonmethyl branches by isoprenoid-like logic: Multiple β-alkylation events in the biosynthesis of myxovirescin A1. *Chem. Biol.* **14,** 835–846 (2007).
21. Huang, Y., Huang, S. X., Ju, J., Tang, G., Liu, T. & Shen, B. Characterization of the LnmKLM genes unveiling key intermediates for β-alkylation in leinamycin biosynthesis. *Org. Lett.* **13,** 498–501 (2011).
22. Mattheus, W., Gao, L.-J., Herdewijn, P., Landuyt, B., Verhaegen, J., Masschelein, J., Volckaert, G. & Lavigne, R. Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. *Chem. Biol.* **17,** 149–59 (2010).
23. Bertin, M. J., Vulpanovici, A., Monroe, E. A., Korobeynikov, A., Sherman, D. H., Gerwick, L. & Gerwick, W. H. The phormidolide biosynthetic gene cluster: a *trans*-AT PKS pathway encoding a toxic macrocyclic polyketide. *ChemBioChem* **17,** 164–173 (2016).
24. Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T. & Dorrestein, P. C. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A.* **103,** 8977–8982 (2006).
25. Buchholz, T. J., Rath, C. M., Lopanik, N. B., Gardner, N. P., Håkansson, K. & Sherman, D. H. Polyketide β-branching in bryostatin biosynthesis: Identification of surrogate acetyl-ACP donors for BryR, an HMG-ACP synthase. *Chem. Biol.* **17,** 1092–1100 (2010).
26. Grindberg, R. V, Ishoey, T., Brinza, D., Esquenazi, E., Coates, R. C., Liu, W., Gerwick, L., Dorrestein, P. C., Pevzner, P., Lasken, R. & Gerwick, W. H. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One* **6,** e18565 (2011).
27. Edwards, D. J., Marquez, B. L., Nogle, L. M., Mcphail, K., Goeger, D. E., Roberts, M. A. & Gerwick, W. H. Structure and biosynthesis of the jamaicamides , new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *lyngbya majuscula*. **11,** 817–833 (2004).

28. Blokhin, A. V, Yoo, H.-D., Geralds, R. S., Nagle, D. G., Gerwick, W. H. & Hamel, E. Characterization of the interaction of the marine cyanobacterial natural product curacin A with the colchicine site of tubulin and initial structure-activity studies with analogs. *Mol. Pharmacol.* **48,** 523–531 (1995).
29. Liu, T., Huang, Y. & Shen, B. Bifunctional acyltransferase/decarboxylase LnmK as the missing link for β-alkylation in polyketide biosynthesis. *J. Am. Chem. Soc.* **131,** 6900–1 (2009).
30. Miziorko, H. M. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* **505,** 131–143 (2011).
31. Middleton, B. The kinetic mechanism of 3-hydroxy-3-methylglutaryl-coenzyme A synthase from baker's yeast. *Biochem. J.* **126,** 35–47 (1972).
32. Gu, L., Jia, J., Liu, H., H??kansson, K., Gerwick, W. H. & Sherman, D. H. Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: Functional identification of a mechanistically diverse enzyme pair. *J. Am. Chem. Soc.* **128,** 9014–9015 (2006).
33. Sudek, S., Lopanik, N. B., Waggoner, L. E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D. H. & Haygood, M. G. Identification of the putative bryostatin polyketide synthase gene cluster from 'Candidatus Endobugula sertula', the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod.* **70,** 67–74 (2007).
34. Geders, T. W., Gu, L., Mowers, J. C., Liu, H., Gerwick, W. H., Håkansson, K., Sherman, D. H. & Smith, J. L. Crystal structure of the ECH2 catalytic domain of CurF from *Lyngbya majuscula*: Insights into a decarboxylase involved in polyketide chain β-branching. *J. Biol. Chem.* **282,** 35954–35963 (2007).
35. Hamed, R. B., Batchelar, E. T., Clifton, I. J. & Schofield, C. J. Mechanisms and structures of crotonase superfamily enzymes - How nature controls enolate and oxyanion reactivity. *Cell. Mol. Life Sci.* **65,** 2507–2527 (2008).
36. Khare, D., Hale, W. A., Tripathi, A., Gu, L., Sherman, D. H., Gerwick, W. H., Håkansson, K. & Smith, J. L. Structural basis for cyclopropanation by a unique enoyl-acyl carrier protein reductase. *Structure* **23,** 2213–2223 (2015).
37. Busche, A., Gottstein, D., Hein, C., Ripin, N., Pader, I., Tufar, P., Eisman, E. B., Gu, L., Walsh, C. T., Sherman, D. H., Löhr, F., Güntert, P. & Dötsch, V. Characterization of molecular interactions between ACP and halogenase domains in the curacin A polyketide synthase. *ACS Chem. Biol.* **7,** 378–386 (2012).
38. Khare, D., Wang, B., Gu, L., Razelun, J., Sherman, D. H., Gerwick, W. H., Håkansson, K. & Smith, J. L. Conformational switch triggered by α-ketoglutarate in a halogenase of curacin A biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14099–104 (2010).
39. Simunovic, V. & Müller, R. Mutational analysis of the myxovirescin biosynthetic gene cluster reveals novel insights into the functional elaboration of polyketide backbones. *ChemBioChem* **8,** 1273–1280 (2007).
40. Liu, X., Biswas, S., Berg, M. G., Antapli, C. M., Xie, F., Wang, Q., Tang, M. C., Tang, G. L., Zhang, L., Dreyfuss, G. & Cheng, Y. Q. Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative

agents from *Burkholderia thailandensis* MSMB43. *J. Nat. Prod.* **76,** 685–693 (2013).
41. Eustáquio, A. S., Janso, J. E., Ratnayake, A. S., O'Donnell, C. J. & Koehn, F. E. Spliceostatin hemiketal biosynthesis in *Burkholderia* spp. is catalyzed by an iron/α-ketoglutarate-dependent dioxygenase. *Proc. Natl. Acad. Sci.* **111,** E3376–E3385 (2014).
42. Kwan, J. C. & Schmidt, E. W. Bacterial endosymbiosis in a chordate host: long-term co-evolution and conservation of secondary metabolism. *PLoS One* **8,** 1–22 (2013).
43. Buchholz, T. J., Geders, T. W., Bartley, F. E., Reynolds, K. A., Smith, J. L. & Sherman, D. H. Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4,** 41–52 (2009).
44. Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H. & Smith, J. L. Cyanobacterial polyketide synthase docking domains: A tool for engineering natural product biosynthesis. *Chem. Biol.* **20,** 1340–1351 (2013).
45. Haines, A. S., Dong, X., Song, Z., Farmer, R., Williams, C., Hothersall, J., Płoskoń, E., Wattana-amorn, P., Stephens, E. R., Yamada, E., Gurney, R., Takebayashi, Y., Masschelein, J., Cox, R. J., Lavigne, R., Willis, C. L., Simpson, T. J., Crosby, J., Winn, P. J., *et al.* A conserved motif flags acyl carrier proteins for β-branching in polyketide synthesis. *Nat. Chem. Biol.* **9,** 685–92 (2013).
46. Weissman, K. J. & Müller, R. Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* **9,** 826–848 (2008).
47. Crosby, J. & Crump, M. P. The structural role of the carrier protein – active controller or passive carrier. *Nat. Prod. Rep.* **29,** 1111 (2012).
48. Rahman, A. S., Hothersall, J., Crosby, J., Simpson, T. J. & Thomas, C. M. Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series. *J. Biol. Chem.* **280,** 6399–6408 (2005).
49. Gu, L., Eisman, E. B., Dutta, S., Franzmann, T. M., Walter, S., Gerwick, W. H., Skiniotis, G. & Sherman, D. H. Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect. *Angew. Chemie - Int. Ed.* **50,** 2795–2798 (2011).
50. Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* **1830,** 3670–3695 (2013).
51. Verdier-Pinard, P., Lai, J. Y., Yoo, H. D., Yu, J., Marquez, B., Nagle, D. G., Nambu, M., White, J. D., Falck, J. R., Gerwick, W. H., Day, B. W. & Hamel, E. Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. *Mol. Pharmacol.* **53,** 62–76 (1998).
52. Chang, Z., Sitachitta, N., Rossi, J., Roberts, M. A., Flatt, P. M., Sherman, D. & Gerwick, W. H. Biosynthetic pathway investigations and gene cluster of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. *Gene* **67,** 1356–1367 (2004).

53. Theisen, M. J., Misra, I., Saadat, D., Campobasso, N., Miziorko, H. M. & Harrison, D. H. T. 3-hydroxy-3-methylglutaryl-CoA synthase intermediate complex observed in 'real-time'. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16442–16447 (2004).
54. Campobasso, N., Patel, M., Wilding, I. E., Kallender, H., Rosenberg, M. & Gwynn, M. N. *Staphylococcus aureus* 3-hydroxy-3-methylglutaryl-CoA synthase: Crystal structure and mechanism. *J. Biol. Chem.* **279,** 44883–44888 (2004).
55. Steussy, C. N., Vartia, A. A., Burgner, J. W., Sutherlin, A., Rodwell, V. W. & Stauffacher, C. V. X-ray crystal structures of HMG-CoA synthase from *Enterococcus faecalis* and a complex with its second substrate/inhibitor acetoacetyl-CoA. *Biochemistry* **44,** 14256–14267 (2005).
56. Skaff, D. A., Ramyar, K. X., McWhorter, W. J., Barta, M. L., Geisbrecht, B. V. & Miziorko, H. M. Biochemical and structural basis for inhibition of *Enterococcus faecalis* hydroxymethylglutaryl-CoA synthase, mvaS, by hymeglusin. *Biochemistry* **51,** 4713–4722 (2012).
57. Pojer, F., Ferrer, J.-L., Richard, S. B., Nagegowda, D. A., Chye, M.-L., Bach, T. J. & Noel, J. P. Structural basis for the design of potent and species-specific inhibitors of 3-hydroxy-3-methylglutaryl CoA synthases. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11491–11496 (2006).
58. Shafqat, N., Turnbull, A., Zschocke, J., Oppermann, U. & Yue, W. W. Crystal structures of uman HMG-CoA synthase isoforms provide insights nto inherited ketogenesis disorders and inhibitor design. *J.Mol.Biol.* **398,** 497–506 (2010).
59. Dorrestein, P. C., Bumpus, S. B., Calderone, C. T., Garneau-Tsodikova, S., Aron, Z. D., Straight, P. D., Kolter, R., Walsh, C. T. & Kelleher, N. L. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* **45,** 12756–12766 (2006).
60. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
61. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
62. Miziorko, H. M. & Lane, M. D. 3-Hydroxy-3-methylglutaryl-CoA synthase. Participation of acetyl-S-enzyme and enzyme-S-hydroxymethylglutaryl-SCoA intermediates in the reaction. *J. Biol. Chem.* **252,** 1414–1420 (1977).
63. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain – domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16,** 2093–2107 (2007).
64. Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S. C. & Burkart, M. D. Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505,** 427–431 (2014).

65. Masoudi, A., Raetz, C. R. H., Zhou, P. & Pemble, C. W. Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505,** 422–426 (2014).
66. Guy, J. E., Whittle, E., Moche, M., Lengqvist, J., Lindqvist, Y. & Shanklin, J. Remote control of regioselectivity in acyl-acyl carrier protein-desaturases. *Proc. Natl. Acad. Sci.* **108,** 16594–16599 (2011).
67. Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J. & Somers, W. S. Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8,** 883–895 (2000).
68. Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F. & Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 1–20 (2016). doi:10.1073/pnas.1520042113
69. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10037–10041 (2001).
70. Graphpad Prism. Available at: www.graphpad.com.
71. Shafqat, N., Turnbull, A., Zschocke, J., Oppermann, U. & Yue, W. W. Crystal structures of human HMG-CoA synthase isoforms provide insights into inherited ketogenesis disorders and inhibitor design. *J. Mol. Biol.* **398,** 497–506 (2010).
72. Simunovic, V., Zapp, J., Rachid, S., Krug, D., Meiser, P. & Müller, R. Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. *ChemBioChem* **7,** 1206–1220 (2006).
73. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11124–9 (2006).
74. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R. & Donnelly, M. I. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* **25,** 8–15 (2002).
75. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* **64,** 1694–1699 (1998).
76. Pfeifer, B. a, Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli. Science* **291,** 1790–1792 (2001).
77. Sánchez, C., Du, L., Edwards, D. J., Toney, M. D. & Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem. Biol.* **8,** 725–738 (2001).
78. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: A Web server for the design of crystallizable protein

variants. *Protein Sci.* **16,** 1569–1576 (2007).
79. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).
80. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. a., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).
81. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
82. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41,** 641–643 (2008).
83. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3,** 1171–1179 (2008).
84. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D* **53,** 240–255 (1997).
85. Painter, J. & Merritt, E. a. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 439–450 (2006).
86. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).
87. Chen, V. B., Arendall III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* **66,** 12–21 (2010).
88. Schrödinger, L. The PyMOL Molecular Graphics System, Version~1.3r1. (2010).
89. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Protein Sci.* **Chapter 2,** 2.9.1-2.9.31 (2007).
90. Maloney, F. P., Gerwick, L., Gerwick, W. H., Sherman, D. H. & Smith, J. L. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proc. Natl. Acad. Sci.* **113,** 10316–10321 (2016).
91. Bernard, S. M., Akey, D. L., Tripathi, A., Park, S. R., Konwerski, J. R., Anzai, Y., Li, S., Kato, F., Sherman, D. H. & Smith, J. L. Structural basis of substrate specificity and regiochemistry in the MycF/TylF family of sugar O-methyltransferases. *ACS Chem. Biol.* **10,** 1340–1351 (2015).
92. Nakano, M. M., Corbell, N., Besson, J. & Zuber, P. Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol. Gen. Genet. MGG* **232,** 313–321 (1992).