Protein-Protein Interactions During Acyl Transfer In Polyketide Biosynthesis

by

Tonia Jane Buchholz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2010

Doctoral Committee:

Professor David H. Sherman, Chair
Professor Robert T. Kennedy
Professor Janet L. Smith
Associate Professor Anna K. Mapp

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. David H. Sherman, for the opportunity to play a role in better understanding a number of exciting natural product biosynthetic pathways. I have collaborated with Dr. Jeff Kittendorf, Dr. Sabine Grüschow, Dr. Nicole Lopanik, Noah Gardner, and Chris Rath from the Sherman laboratory on the projects presented here. I also mentored two undergraduate students, Frank Bartley, III and Paul Shotkin, who contributed their energy to this work.

Thanks to the laboratories of Dr. Janet L. Smith, Dr. Jason E. Gestwicki, and the High Throughput Protein Expression Core in the Life Sciences Institute. These labs have been willing to share ideas as well as resources, and in doing so have made conducting my dissertational research in the LSI very enjoyable. Specifically, I worked closely with Dr. Todd Geders, Dr. David Akey, Jaime Razelun, Srikanth Patury, Dr. Clay Brown, and Jim Delproposto.

I would also like to thank Dr. Kevin A. Reynolds and his laboratory members at Portland State University. The collaborative pikromycin teleconferences were quite memorable.

While conducting this research, I received funding from the Chemical Biology Doctoral Program, an Eli Lilly Graduate Fellowship, and a Rackham Travel Grant.

Finally, thank you to the all who have joined me in scientific discussions during my time in the Chemical Biology Doctoral Program.

PREFACE

This thesis contains five chapters covering my dissertational studies into the protein-protein interactions involved in polyketide biosynthesis. Chapter 1 is an introduction into bacterial polyketide biosynthesis and is an expansion of a published review entitled "Polyketides, Modular Polyketide Synthases" in *Wiley Encyclopedia of Chemical Biology*, **2008**, doi:10.1002/9780470048672.wecb459. Chapter 2 is focused on the structural and biochemical characterization of the modular polyketide synthase docking domains and is adapted from *ACS Chemical Biology*, **2009**, 4, 41-52 and *Chemistry & Biology*, **2007**, 14, 944-954. Chapter 3 describes our progress toward *in vitro* characterization of enzymes from the putative bryostatin biosynthetic pathway from the uncultured symbiont of *Bugula neritina*, Ca. *Endobugula neritina*. It covers key protein-protein interactions of BryR, the free-standing HMG-ACP synthase involved in β-branching, and it is adapted from a manuscript in preparation as well the review mentioned above. Chapter 4 highlights our initial characterization of the acyl carrier protein usage of the type III PKS, Germicidin synthase (Gcs) and is an extension of the work published in *Chembiochem*, **2007**, 8, 863-868. Finally, in Chapter 5, conclusions and future directions are presented.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Polyketide metabolites are produced by diverse bacterial taxa, including soil-dwelling bacteria, cyanobacteria, and bacterial symbionts living within insects or marine invertebrates, and are generated by decarboxylative condensations of simple coenzyme A (CoA) building blocks. At present, polyketide natural products find clinical utility as antibiotics, antiparasitics, antifungals, anticancer drugs, and immunosuppressants. These comprehensive pharmacological activities provide continued motivation to unravel polyketide biosynthetic mechanisms to enable the discovery of novel compounds for the benefit of human health. This dissertational research explores the molecular basis for guiding ACP-mediated protein-protein interactions in three diverse pathways (pikromycin, bryostatin and germicidin).

Accessing new members of the ketolide class of macrolide antibiotics remains an important goal given the increasing prevalence of drug-resistant pathogens. As a naturally occuring ketolide, the pikromycin could serve as a scaffold to build a diverse set of polyketides. The Sherman laboratory has spent over a decade investigating the catalytic mechanisms of pikromycin biosynthesis by the modular PKS-containing pathway found in *Streptomyces venezuelae*. Here, we explore the protein-protein interactions of the large, multifunctional polypeptides at the PikAIII/PikAIV interface. The combination of

structural characterization of the docking domains together with discrete docking domain affinity measurements supports a paradigm wherein the binding specificity that determines the linear arrangement of proteins in modular PKS systems is encoded in the small, terminal docking domains. Additionally, a model for the observed docking domain specificity across a matrix of interacting pairs from the pikromycin and erythromycin pathways is presented.

Secondly, we profiled the ACP binding and catalysis of BryR, the HMG-ACP synthase from an uncultured symbiont of *Bugula neritina,* a marine bryazoan. BryR functions to install β-branches in bryostatin, a PKC modulator with both anti-cancer and neuroprotective activities. Lastly, we have explored a unique type III PKS, Gcs, that is capable of using both acyl-CoA and acyl-ACP starter units in the catalysis of pyrones. A structural model for Gcs is reported. Together with the growing understanding of protein-protein interactions in PKSs, the knowledge and mechanistic understanding of the pikromycin, bryostatin, germicidin and other complex metabolic systems will provide additional opportunities to engineer chemical diverse polyketides using rational approaches.

Chapter 1

Protein-Protein Interactions in Natural Product Biosynthesis

1.1 Introduction

Polyketides constitute a large class of microbial and plant-derived secondary metabolites that displays a vast array of structural diversity. These organic molecules vary in molecular weight, functional group modification, and include linear, polycyclic, and macrocyclic structural forms. At present, polyketide natural products find clinical utility as antimicrobials, antiparasitic agents, anticancer drugs, and immunosuppressants (Figure 1-1) (1-5) Given these impressive and wide-ranging pharmacological activities, an ever-increasing demand is placed upon natural products research to uncover novel polyketide metabolites for the benefit of human health.

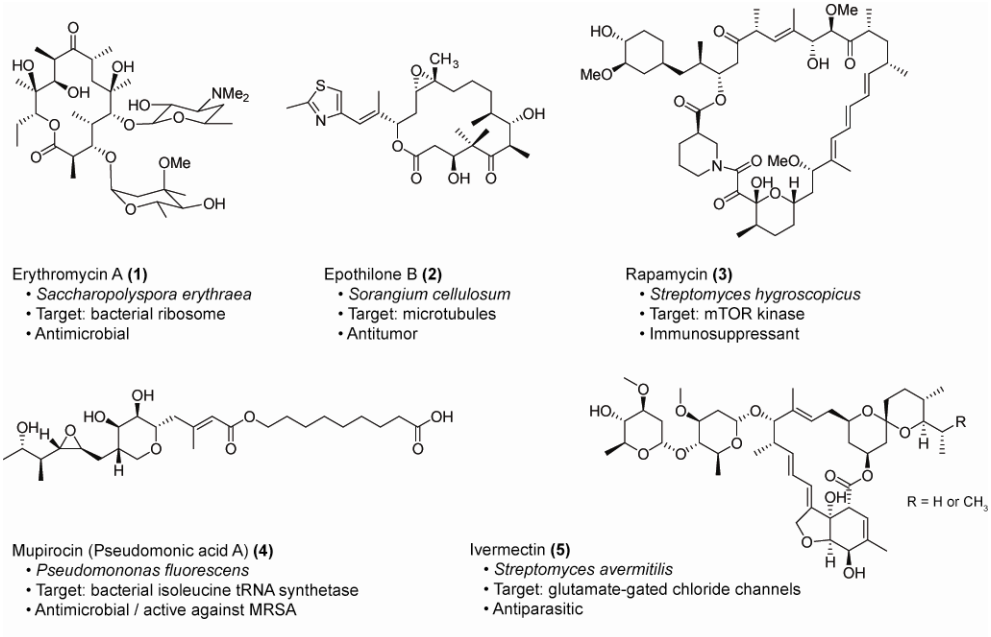


Figure 1-1. Clinically relevant polyketides, their producing organisms and activities

The quest for novel natural product-based compounds to be used as drug candidates and/or tools for chemical biology and biomedical research is moving forward on many fronts. Unbiased extraction and bioassay-guided isolation of small molecules from terrestrial or marine sources continue to reveal new compounds with interesting activities (*6*). However, it is often challenging to obtain the quantities (or sustainable sources) necessary to truly pursue many of these complex compounds as therapeutic agents (*7*). Efforts are also ongoing to devise effective synthetic methods that would give access to natural product-inspired compound libraries encompassing a large range of structural classes (carbocycles, oxa- and aza-heterocycles, and linear molecular organizations) (*8*). Although these scaffolds are often thought of as *prevalidated*, multistep synthesis of small libraries of these molecules still requires significant effort (though it is becoming increasingly feasible) (*8*). Analogs derived directly from natural products can often be accessed through semi-synthetic methods (*9-11*), or as an alternative, via genetic engineering of biosynthetic enzymes and pathways (*12, 13*).

Before researchers can successfully alter the basic activities of polyketide (PK), fatty acid (FA) and non-ribosomal peptide (NRP) megasynthetases and their accessory enzymes to produce novel natural products, the inherent structures, mechanisms and potential limitations of these proteins must first be understood. The past two decades have resulted in an explosion of structural and mechanistic information on biosynthetic proteins ranging from small, individual enzymatic domains all the way to full module-sized multifunctional polypeptides (*14-22*). Individual domains, or groups of domains, have also been studied extensively *in vitro* using biochemical methods to determine the chemical flexibility of individual catalytic activities (*23-28*). Though growing, a thorough

understanding of the protein-protein interactions used by nature to create these efficient pathways still lags behind the structural and mechanistic information (*29*). This dissertational research explores the molecular basis for guiding ACP-mediated protein-protein interactions in three diverse biosynthetic pathways (pikromycin, bryostatin and germicidin).

1.1.1 The Polyketide Synthase Superfamily

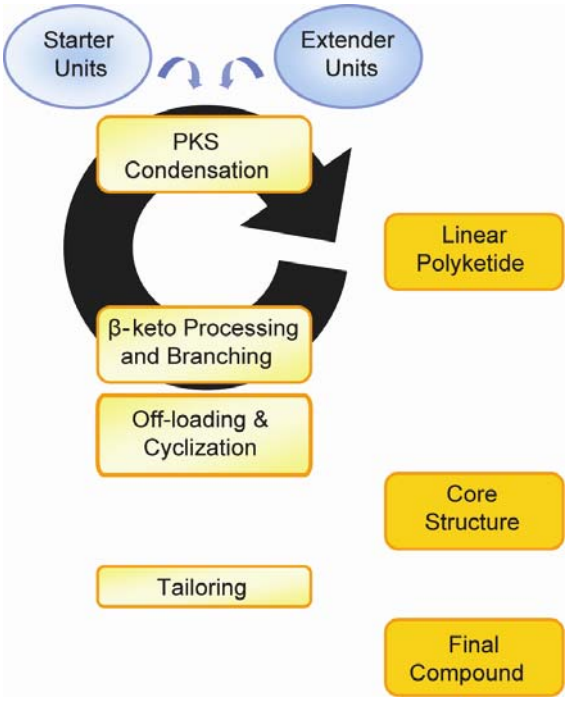


Figure 1-2. General strategy for polyketide biosynthesis.

The general strategy employed for the generation of polyketide natural products is elegantly simple, yet it allows for the introduction of chemical diversity at a variety of stages (Figure 1-2). A number of polyketide synthase (PKS) architectures have evolved to accomplish the task of polyketide biosynthesis; the organization of the enzymes in these systems allows one to group them into three major classes (1, 30, 31). Type I

biosynthetic systems have been identified that produce polyketide (PK) as well as fatty acid (FA) and non-ribosomal peptide (NRP) natural products (*15, 20, 32, 33*). Those biosynthetic systems classified as type I (or modular) are comprised of a number of large, multifunctional enzymes that generate their products via a stepwise, assembly-line type mechanism. The modular polyketide synthases (PKSs) are responsible for catalyzing the initiation, elongation, and processing steps that ultimately give rise to the characteristic macrolactone scaffold (Figures 1-2, 1-3, 1-4A-B). As seen for the pikromycin pathway, the genes encoding type I PKS proteins (*pikA*) are typically found clustered together near the genes encoding the tailoring enzymes (*des* and *pikC*) and resistance-conferring proteins (*pikR*) (Figure 1-3). In type I megasynthetases, modules can act either processively or iteratively to grow the linear chain intermediates.

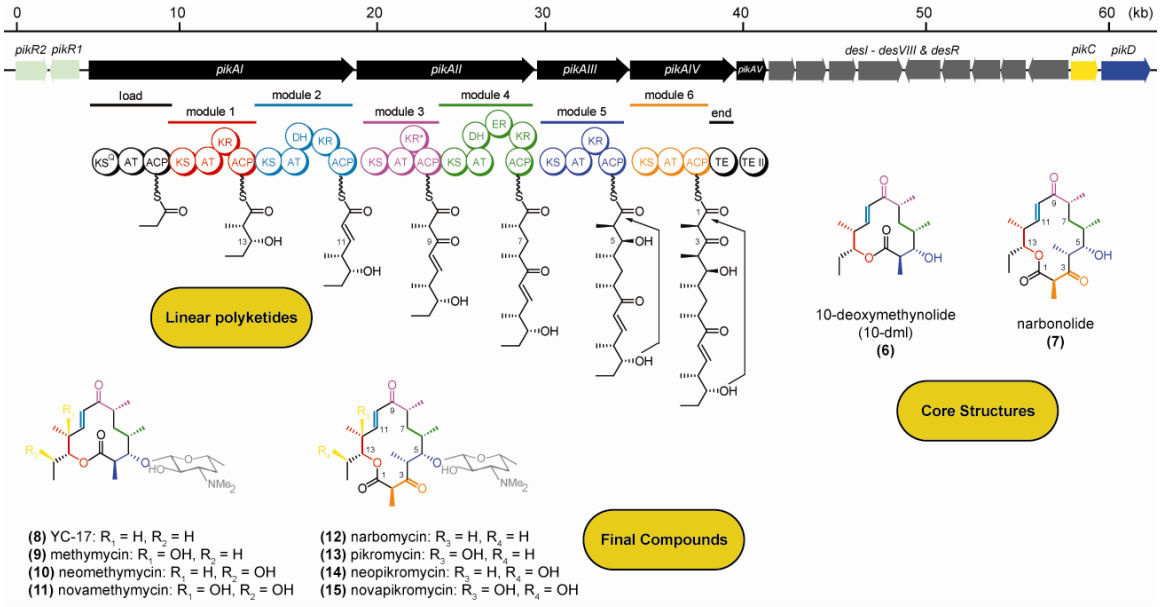


Figure 1-3. Pikromycin biosynthetic pathway. The *pikA* genes encode a type I PKS responsible for generating **1** and **2**. The *des* genes encode the necessary proteins for sugar (desosamine) biosynthesis and attachment to yield macrolides **8** and **12**. *pikC* encodes a cytochrome P450 monooxygenase responsible for hydroxylations of **8** amd **12** to generate the remaining suite of macrolide antibiotics shown above. The *pikR* locus encodes two resistance genes, and the *pikD* gene product is a transcriptional regulator (*34-36*).

Conversely, type II systems capable of synthesizing FAs and PKs contain a dissociable complex of discrete and usually monofunctional enzymes/domains (Figure 1-4C-D). Their functions are similar to those found in a single module from a type I system (*37, 38*). Type II systems function mostly in an iterative fashion to create FAs and PKs of varying lengths and often produce aromatic polyketide products. Finally, in PK synthesis, a third type exists. Type III PKSs are multifunctional, single domain enzymes often creating natural products solely from small molecule pools (39). These architecturally simple systems are able to initiate, extend and cyclize their products using a single active site. There is recent evidence that a small number of type III PKSs partner with other proteins to generate their polyketide products (*40-42*). In exploring protein-protein interactions during acyl transfer, my work has focused on the type I and type III PKSs.

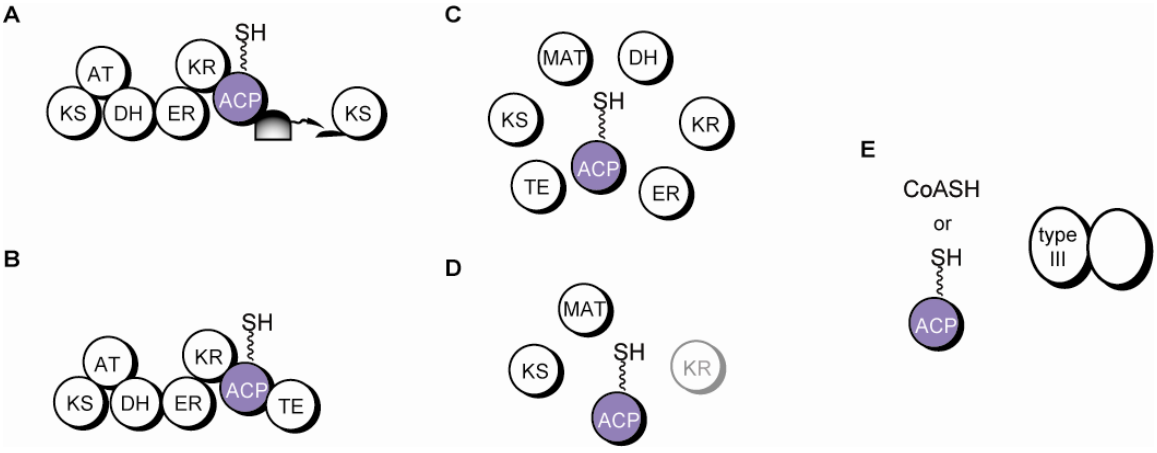


Figure 1-4.Organization of PKS and FAS enzyme systems. A) Type I PKS elongation module at the C-terminus of a polypeptide. The ACP must interact with all domains in its given module plus the downstream KS domain. B) Terminal module in a type I PKS system or a type I FAS system. C) Type II FAS system where the ACP must interact with each domain iteratively *in trans*. D) In a type II PKS system, the ACP interacts with the KS and MAT iteratively, but may only use the KR on selected cycles. E) Type III PKSs – ACP interactions are the exception to the rule, but examples are growing.

1.1.2 Type I PKSs

The sequential arrangement of modules within a type I PKS system effectively serves as a biosynthetic program, responsible for dictating the final size and structure of the polyketide core. Sets of domains grouped together to accomplish a single round of extension are referred to as modules (Figures 1-4, 1-5). Each module is typically used once to generate the linear polyketide. The number, arrangement, and type of modules within type I systems serve as a template to determine the core structure of the natural product (for specific examples, see Figure 1-3 and 2-1). Typically, initiation of polyketide biosynthesis begins by the acyltransferase (AT) catalyzed linkage of a coenzyme A (CoA) priming unit (e.g., methylmalonyl-CoA, malonyl-CoA, propionyl-CoA) to the acyl carrier protein (ACP) of the loading module. Once initiated, downstream elongation modules carry out repetitive extensions of the starter unit. In most type I PKS systems, each elongation module contains at minimum an AT domain, an ACP domain, and a ketosynthase (KS) domain. The AT domain is responsible for loading the appropriate CoA extender unit onto the ACP domain (i.e., malonyl-CoA, methylmalonyl-CoA, etc.). The KS domain then catalyzes a decarboxylative condensation of the extender unit with the growing polyketide chain obtained from the preceding module to generate an ACP-bound β-ketoacyl product.

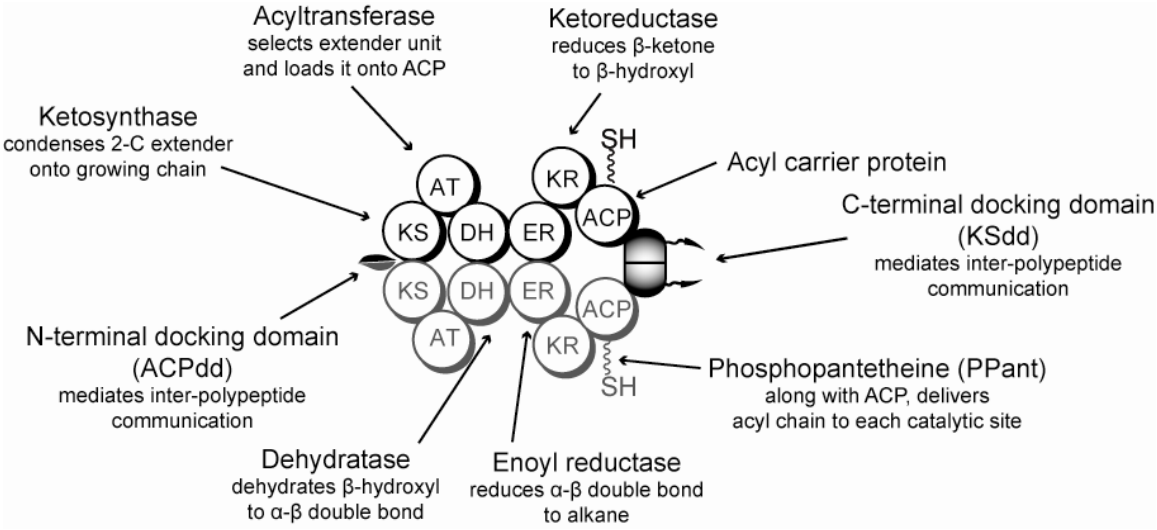


Figure 1-5. Components of a dimeric type I PKS elongation module. An internal monomodular polypeptide with both an N- and a C-terminal docking domain is depicted here.

In addition to the three core domains, each elongation module may contain up to three additional domains (ketoreductase (KR), dehydratase (DH), enoyl reductase (ER)) that are responsible for the reductive processing of the β-keto functionality prior to the next extension step (Figure 1-6). These reductive steps contribute to the overall structural diversity that is observed among polyketide natural products. The presence of a KR domain alone gives rise to a β-hydroxyl functionality, the presence of both a KR and a DH domain generates an alkene, while the combination of KR, DH, and ER results in complete reduction to the alkane. Finally, termination of polyketide biosynthesis is catalyzed by a thioesterase (TE) domain located at the carboxy terminus of the final elongation module. The activity of this domain results in the cleavage of the acyl chain from the adjacent ACP; typically intramolecular cyclization results in the formation and release of a macrolactone ring. Tailoring enzymes, such as hydroxylases and glycosyl transferases often serve to further modify the polyketide, to yield the final bioactive compound (Figures 1-2 and 1-3).

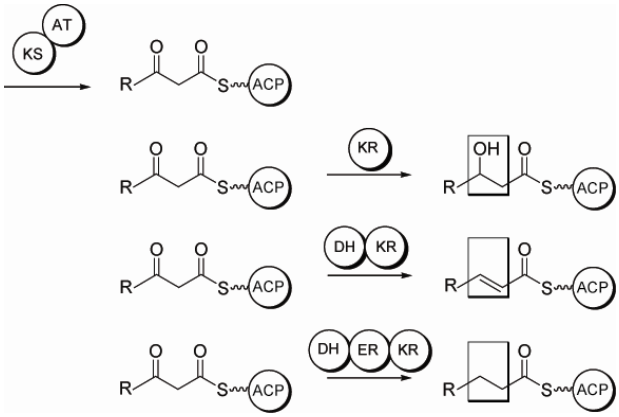


Figure 1-6. Reductive domains in type I systems. In type I PKSs, the presence or absence of processing domains determines the chemical outcome on the growing polyketide chain.

The modular organization of type I PKSs has made them attractive targets for rational bioengineering. Combinatorial biosynthetic efforts centered on prototypical modular PKSs have been the topic of many recent review articles (*12, 43, 44*). Several strategies are currently being pursued that attempt to leverage PKS systems for the generation of structurally diverse polyketides. For example, it has been demonstrated that alterations of individual catalytic domains (i.e., inactivation, substitution, addition, deletion) within a PKS module can result in predicted structural alterations of the final PKS product. Likewise, the addition, deletion, or exchange of intact modules can also impart structural variety into polyketide metabolites (*45*). Using these and other approaches, hundreds of novel polyketide structures have been generated, thereby establishing the potential of these applications (*12, 44*). However, often these PKS re-engineering efforts have resulted in low efficiency of product formation (*46-49*). This suggests that much remains to be learned regarding the molecular intricacies of these complex biosynthetic machines.

1.1.3 Type III PKSs

Type III PKSs differ from type I and type II PKSs in several key aspects. First, they function as homodimers with each subunit consisting of a single domain of ~40 kDa. Additionally, rather than utilizing substrates that are covalently linked to acyl-carrier proteins, they are thought to use mainly acyl-coenzyme A (acyl-CoA) thioesters as substrates (39). Notable exceptions to this starter unit preference have recently been reported and will be discussed in chapter 4 (*40-42*). Furthermore, the same active site that catalyzes the Claisen condensations is also responsible for loading of the starter unit and cyclizing the polyketide intermediate. Members of this family share a conserved active site triad (Cys-His-Asn), and their mechanism for condensation resembles that of the ketosynthase domain of type I PKSs (39, 50, 51). Through extension, condensation and cyclization, a wide range of aromatic polyketides are made by plant and bacterial type III PKSs (39). The diversity observed in type III PKS-derived natural products results from the selection of the acyl starter unit, control over the number of extension rounds, and the mechanism of cyclization (39). Moreover, the type III PKSs have demonstrated broad substrate specificity and have been shown to convert non-native acyl-CoA substrates to pyrones *in vitro*, making them accessible targets for bioengineering (40, 52, 53).

1.1.4 Simple, Activated Monomers are Used in Polyketide Biosynthesis

Despite the architectural diversity in arrangement of catalytic domains/enzymes across the variety of FAS and PKS types, individual reaction steps are essentially conserved across all kingdoms of life. Fatty acid chains and polyketides are built up through the condensation of a small set of activated carboxylic acid monomers. Details of the stepwise catalytic reactions in fatty acid and polyketide biosynthesis are well

understood (15). Three reaction types are used to accomplish a single round of elongation. First, activation of monomers occurs. Second, two monomers are condensed to form the extended chain; and third, processing of the newly extended chain occurs.

To create the ready-to-use building blocks, the carboxylate-containing substrates are converted to acyl thioesters by conjugation to the free thiol of the 4'-phosphopantethiene (PPant) prosthetic group. The PPant moiety can be part of either a carrier protein (CP) or a small molecule cofactor (Figure 1-7). Enzymatic thioester formation on coenzyme A (CoA) creates an acyl-CoA species. Alternatively, the carboxylic acid can be attached to the terminal thiol of the PPant that has been linked to the CP at its conserved serine residue in a post-translational modification step utilizing CoA as a substrate. While the chemical environment of the thioester is quite similar between the two species, there appear to be significant differences in the abilities of some metabolic enzymes to recognize the acyl chains when attached to CoA as opposed to ACP (40, 54). Additionally, there are instances where substituting one ACP for another leads to a change in activity (55, 56).

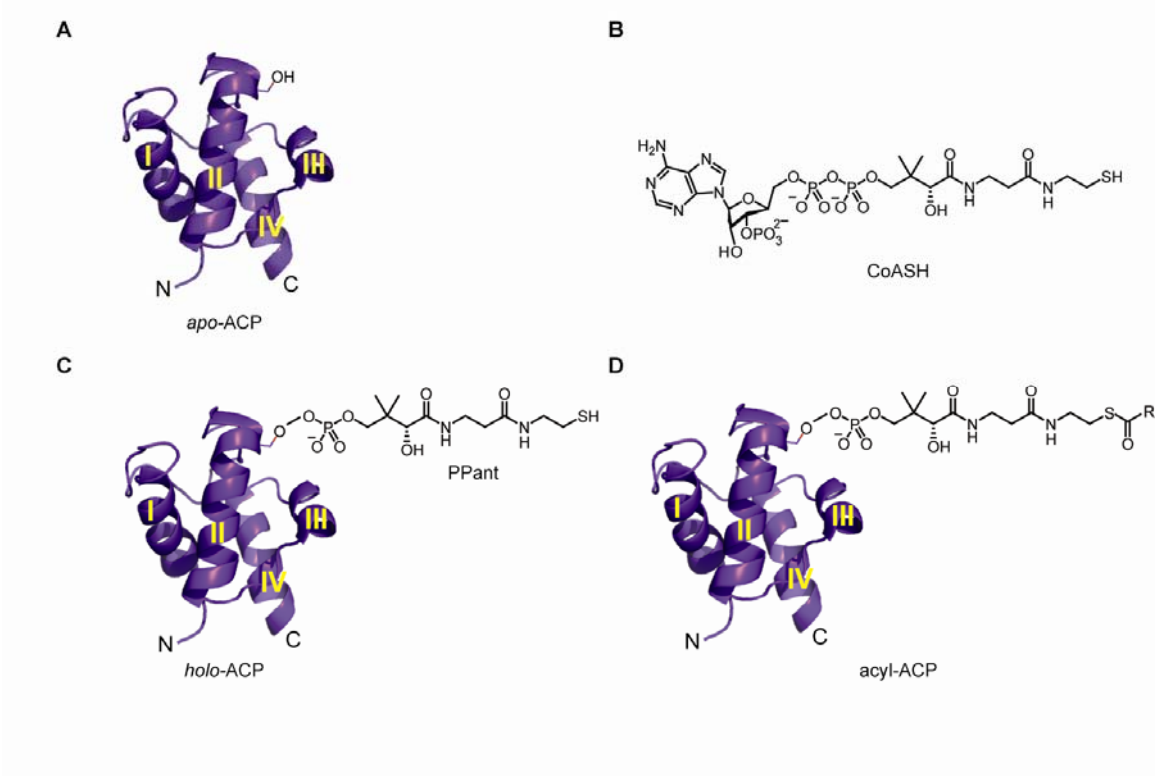


Figure 1-7. ACP Forms. As an example of the three-dimensional structure of a typical ACP, AcpM, is shown in its three states of modification (57). The N- and C- termini are labeled as well as the four helices. The transition from *apo* (A) to *holo* (C) requires the coenzyme A cofactor shown in (B).

Interest in studying CP structure and its possible modes of interaction with its attached acyl chain and/or other partner enzymes is well justified, and research in this field has been ongoing since the late 1980s (58, 59). Carrier proteins are widely used for delivering carboxylate-containing substrates and intermediates to a variety of enzymes and enzymatic domains in both primary and secondary metabolite production. It is likely that much of the efficiency and fidelity of complex polyketide biosynthesis can be attributed to the tethering of intermediates to the CPs during assembly. Additionally, via binding to the various catalytic domains or discrete enzymes in the pathway, CPs can help to coordinate reaction sequence (58) (Figure 1-6).

1.1.5 Acyl-, Aryl-, and Peptidyl-Carrier Proteins

The distinguishing structural features of the carrier protein family are their small size (8-10 kDa), their predominantly alpha helical secondary structure (as a rule, CPs contain three or four helices connected by short loops), the presence of a conserved 4'-phosphopantetheine attachment site (a serine residue in a loop near the start of helix II), their conformational flexibility, and their overall acidic nature (though most NRPS CPs have a more neutral net charge) (Figure 1-7) (58, 60). During fatty acid and polyketide synthesis, acyl carrier proteins (ACPs) are employed. Depending on the class of synthase (see section **1.1.1**), the ACPs are either integrated into the multidomain proteins (type I FAS and PKS) or they exist as free-standing proteins (bacterial type II FASs and type II PKSs) (Figure 1-4). Peptidyl carrier proteins (PCPs) are similarly found embedded within the related multimodular nonribosomal peptide synthetases (NRPSs) (PCPs are sometimes referred to as thiolation (T) domains). Lastly, aryl carrier proteins (ArCPs) are designed to shuttle aromatic compounds, such as salicylic and 3,4-dihydroxybenzoic acids for the generation of aryl-containing NRPs and siderophores (61-64).

CPs can exist in one of three forms; the native protein *apo* form, the phosphopantetheinylated carrier protein referred to as the *holo* form, and finally, the activated acyl- or aminoacyl-loaded species (Figure 1-7). The apo (inactive) to holo (activated) conversion is mediated by holo-acyl carrier protein synthases (ACPSs) or more generally, by phosphopantetheinyl transferases (PPTases) (65). PPTases transfer the PPant moiety from coenzyme A to the conserved serine (usually in the recognizable "DSL" motif) of apo-ACP to produce holo-ACP and 3',5'-ADP in a Mg^{2+}-dependent reaction. Typically, loading of an acyl group onto the holo-CP occurs through the action

of an acyltransferase (AT) domain in fatty acid and polyketide synthesis and amino acyl adenylation (A) domains in NRPSs (15, 66). The AT domain (or enzyme) can catalyze the acyl chain attachment to ACP via a thioester linkage to the terminal cysteamine thiol of the PPant prosthetic group. At this point, elongation and processing take place via interaction with remaining FAS or PKS domains/enzymes in the system as outlined above. For this reason, the PPant prosthetic group is sometimes referred to as a "tether" that transfers the growing acyl chains between the various enzymes or active sites in the biosynthetic system. Some promiscuous secondary metabolism phosphopantetheinyl transferases (PPTases) have been exploited to convert apo-ACPs directly to acyl-ACPs. The PPTases from surfactin (Sfp) and bleomycin (Svp) biosynthesis are able to generate acyl-ACP species *in vitro* by substituting an acyl-CoA for reduced CoA as the substrate (65, 67).

1.1.6 Carrier Protein - Partner Enzyme Interactions

A variety of binding strategies are employed to facilitate acyl transfer of polyketide and non-ribosomal peptide intermediates between the CP and its partner enzymes. Recognition of the CP can occur either *in cis* or *in trans*, and these binding events can be driven by protein-protein, protein-PPant, or protein-acyl (or aminoacyl) interactions. In modular FAS, PKS and NRPS systems, the majority of CP interactions occur between domains that are physically attached to one another (Figure 1-4). It is tempting to speculate that the interactions between the ACP and its type I PKS partner domains could therefore be more passive than active (that is allowed, but not favored). Within the type I PKS paradigm, the ACP must still react with noncovalent partners, including the initial pantetheinylation reaction via the PPTase, building block loading by

AT domains in the subset of *trans*-AT PKS pathways, transfer of the growing chain to a downstream module on a separate polypeptide (often mediated by docking domains), and interactions with a growing number of free-standing enzymes such as editing thioesterases (TE IIs) and cassettes of enzymes involved in β-branching. It is also possible that the ACP has specific interactions witheach domain. In type II FAS and PKS systems, the free-standing CP interfaces with all of its partners *in trans* (*37*). In these systems, specific recognition between CPs and the catalytic domains has been demonstrated, though the systems are often tolerant of CP substitution from a related type II system (56). Lastly, a small, but growing, number of type III PKSs have been identified that likely use CP-linked substrates (40, 42, 52, 68). Understanding the rationale behind each of the specific modes of association will help unlock the logic for multiprotein assembly line construction and enable efficient reengineering of, or *de novo* design of, novel biosynthetic pathways.

1.2 Prior Work

1.2.1 Pikromycin

The Sherman laboratory has been engaged in investigations of the biosynthesis of the pikromycin (**13**) and methymycin (**9**) macrolides by *Streptomyces venezuelae* for more than a decade (Figure 1-3) (*34, 36*). As a naturally-occurring ketolide in the macrolide antibiotic class of compounds (69), the potential of this scaffold for generation of clinically useful anti-infectives motivates the continued study of the molecular mechanisms of this pathway (70). Additionally, the intriguing observation that the *pik* gene cluster was responsible for the generation of an entire family of compounds (as

opposed to a single polyketide product) is suggestive that inherent flexibility may exist within the system (34, 71). Using *in vitro* biochemistry and structural biology, PikAIII and PikAIV have been studied extensively (14, 21, 26-28, 72). Both the chain elongation and chain termination mechanisms have been investigated, and a model has been proposed for the method of multiple macrolactone product formation (26). However, the role for the docking domains in the generation of the 12- and 14-membered lactones had not yet been studied. The dissertational research described in chapter 2 reports on our examinination into the docking domain-mediated protein-protein interactions in the pikromycin pathway at the PikAIII-PikAIV interface. In addition, potential cross-talk between docking domains elsewhere in the pikromycin pathway as well as in the related erythromycin pathway are reported.

1.2.2 Bryostatin

Another area of interest in the Sherman laboratory is the discovery of novel natural products and their production from marine sources (73). Exploring biosynthetic pathways found within marine invertebrates or bacteria can be particularly challenging when the organisms are reluctant to grow in the laboratory setting. Advances in cloning and expression of microbial secondary metabolism genes in heterologous hosts has increased our ability to gain a fundamental understanding of the mechanisms used to generate this pool of bioactive compounds (7, 73). The putative bryostatin gene cluster was proposed after sequencing bacteria-enriched *B. neritina* DNA libraries (74). Recently, we demonstrated *in vitro* biochemical activity for the first enzyme from the *bry* pathway (75). *bryP* encodes two acyltransferase domains and resides upstream of the genes encoding five multimodular type I PKS proteins. BryP was shown to selectively

load malonyl-CoA onto a variety of ACPs as well as a full module from the bryostatin pathway (75). These successes motivated the characterization of a second discrete enzyme from the *bry* cluster, BryR, an HMG-ACP synthase.

1.2.3 Germicidin

The potential to build novel biosynthetic pathways on microchips was explored by the Sherman and Dordick groups using tetrahydroxynaphthalene synthase (THNS/SCO1206), a type III PKS from *Streptomyces coelicolor* (53, 76). This technology appeared suited for the generatation of pyrone-containing small molecules. However, THNS prefers to use malonyl-CoA as a building block for both initiation and extension, which results in pyrone formation being secondary to flaviolin synthesis. In the search for other type III PKSs whose starter unit specificity was orthogonal to extender unit preference and would, therefore, not compete with malonyl-CoA, a remarkable finding was made. One of the *Streptomyces coelicolor* type III PKSs, SCO7221, now known as Germicidin synthase (Gcs) (*41*), has the ability to accept efficiently its starter unit as an acyl-ACP (*40*). Recently it has been proposed that a small number of other type III PKSs are able to accept type I FAS biosynthetic products still tethered to their ACPs as starter units (42, 52, 68) in the generation of aromatic polyketides. These reports, and direct demonstration of ACP usage by a type III PKS *in vitro* by the Sherman laboratory (*40*) challenged the prevailing paradigm that type III PKSs used exclusively acyl- or aryl-CoA starter units (39). To explore the molecular basis of this unusual protein selectivity, we initiated the structural characterization of Gcs.

1.3 Summary

The projects presented in this dissertational study are linked through a basic question in natural product biosynthesis. What strategies do ACPs use to pick up and deliver their tethered substrates during polyketide production, assuring efficient and accurate assembly of the compound? Three different modes of ACP-mediated acyl transfer have been investigated. First, the observation that the *PikA* pathway from *Streptomyces venezuelae* produces two macrolactones of different ring sizes (10-deoxymethynolide, a 12-membered heterocycle, and narbonolide, a 14-membered ring) prompted biochemical investigations into the structure and binding specificity of the docking domains at the PikAIII and PikAIV interface. Second, the search for the acetate substrate for BryR, the HMG-CoA synthase homolog implicated in β-branching in the bryostatin pathway, led to us to investigate the binding profile of a small subclass of discrete ACPs with this enzyme. Lastly, the discovery of a type III PKS, Germicidin Synthase (Gcs), capable of accepting both CoA- and ACP-linked starter units initiated structural and mutagenesis studies of this enzyme.

Chapter 2

Type I PKS Inter-Polypeptide Docking in the Pikromycin and Erythromycin Systems

2.1 Introduction

Type I modular PKSs, consisting mainly of large, multi-functional enzymes, use sequential decarboxylative condensation reactions to form carbon – carbon bonds between simple carboxylic acid extender units (2, 31) during the production of a diverse family of structurally-rich and often biologically-active natural products (e.g. antimicrobial, antifungal, antiviral, anticancer and immunosuppressant compounds) (2, 3). Recently, structural studies have provided important new insights relating to the architecture and mechanism of type I PKSs and the related fatty acid synthases (15, 17, 77, 78). Found in a variety of bacteria, modular PKSs direct biosynthesis via covalently-linked catalytic domains that are organized into linear modules where each module houses the requisite catalytic domains to perform a single elongation step in the building of the polyketide chain (Figure 2-1). Each elongation module receives the nascent chain from the previous module, extends the polyketide by two carbons, and (typically) modifies this portion before passing the intermediate to the downstream PKS protein (15, 79). The final chemical structure is determined by the number of modules in the pathway, their catalytic domain composition, and arrangement in the biochemical assembly line (Figure 2-1). Extensive research has identified signature amino acid sequences within the

catalytic domains that guide substrate specificity (15, 80). However, details about the protein-protein interactions that govern acyl transfer between modules have only recently been explored (29, 81, 82).

The modular nature of type I PKSs has led many to envision rational "mix and match" bioengineering for the generation of novel polyketide products. As such, metabolic engineering or combinatorial biosynthesis has emerged as one potential route to create novel polyketide agents (43, 46, 47, 83). Specific changes can be introduced to the final polyketide core in a controlled fashion by manipulating the genes that encode modular PKSs. Modifications at the level of the modules or the individual catalytic domains within a PKS module have been used to generate hundreds of novel polyketide structures, thereby establishing the potential of these applications (43, 46, 47, 83, 84). However, engineered PKS modules often fail to produce significant quantities of the desired product (84). Fundamental studies to establish the mechanistic basis for efficient molecular interactions between PKS multifunctional proteins will likely facilitate effective design and assembly of productive bioengineered pathways. The importance of this new information motivated the studies described in this chapter.

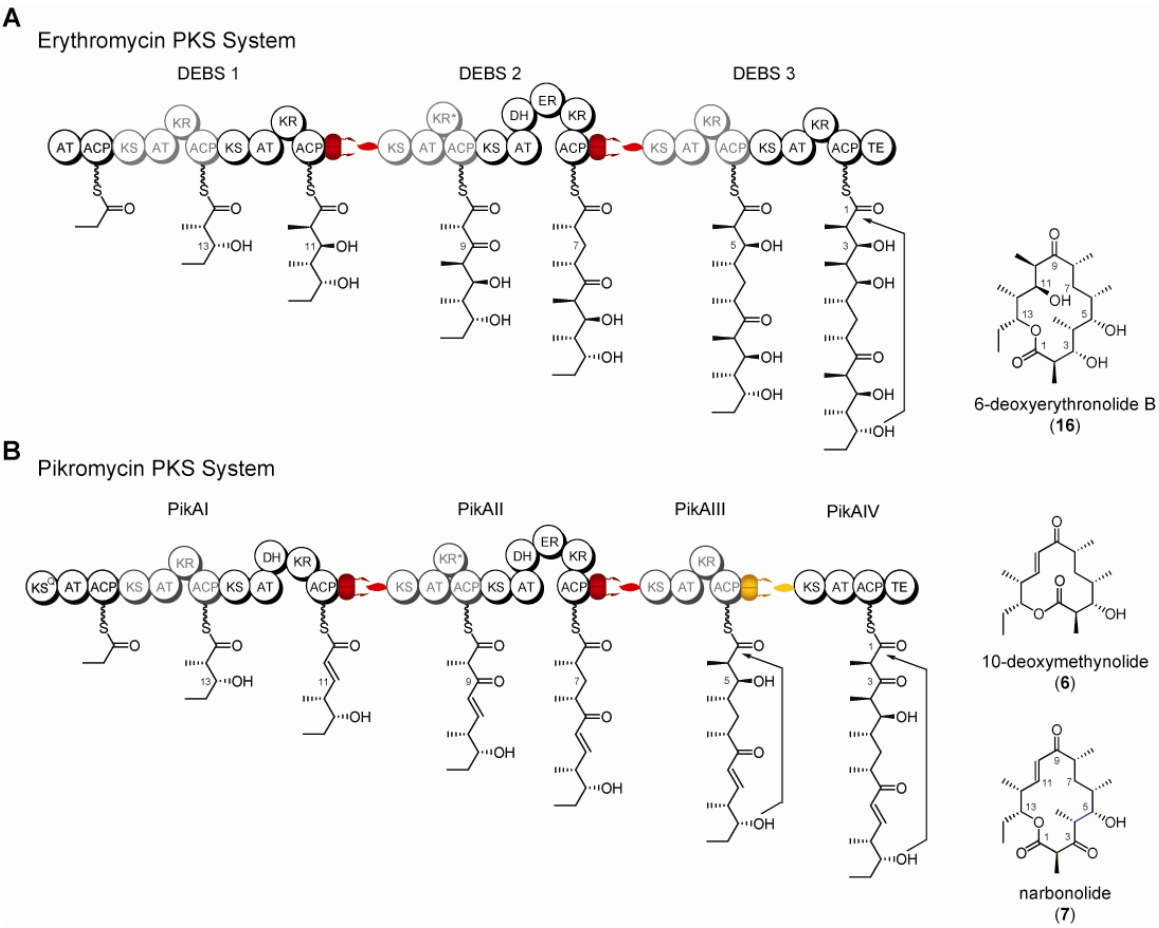


Figure 2-1. Type I polyketide synthases. Arrangement of the polyketide portions of the erythromycin and pikromycin biosynthetic pathways and their macrolactone products. Docking domains are colored by proposed subclass; H1-T1 are red, and H2-T2 is gold. Domain sizes are not drawn to scale.

The fidelity and efficiency of acyl transfer at the interfaces of the individual PKS proteins is thought to be governed by helical regions, termed docking domains (dd), located at the C-terminus of the upstream and N-terminus of the downstream polypeptide chains (Figure 2-1) (85). Two main strategies have been employed to study the specificity determinants for inter-polypeptide (e.g. module→module) communication. In the first strategy, modules (or excised domains) from the erythromycin PKS system were used to create a variety of *in vitro* intermodular transfer and elongation assays (Figure 2-2) (86-89). Typically, a variety of chimeric proteins were generated to investigate the effect of matched or mismatched docking domains in combination with a series of ACP/KS

pairings. Detection of triketide lactones resulting from the transfer and elongation of diketide intermediates established that complementary docking domain pairs are required for efficient transfer of polyketide intermediates between polypeptides (81, 90). In some cases, formation of the cognate ACP and KS pairs also appears to impart a catalytic advantage, although tolerance for mispairing at this junction is also evident (88, 89).

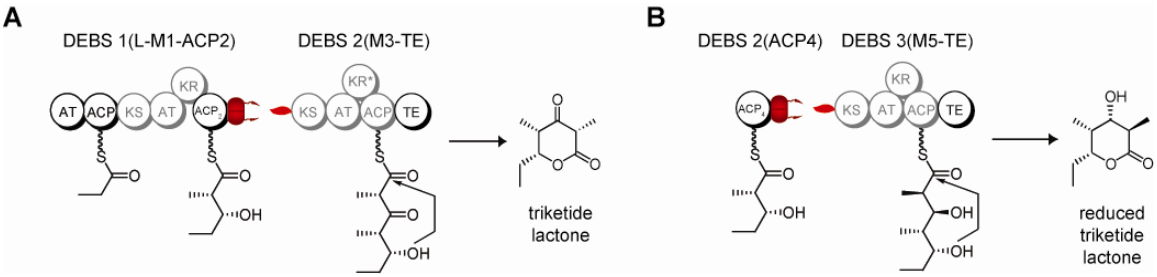


Figure 2-2. Intermodular transfer and elongation assays. Examples of assays featuring the erythromycin PKS system.

The second strategy for analysis of PKS module-module molecular recognition has been to structurally characterize the docking interface. A docking domain complex model for the DEBS 2/DEBS 3 interface (Figure 2-1A) was developed via protein NMR spectroscopy (85). The structure established that the docking domains are helical and revealed two roles for the C-terminal PKS docking domain (ACP-side docking domains, ACPdd). First, this region appears important for stabilizing the PKS homodimer. Second, ACPdd is poised to interact with the downstream KS polypeptide through its terminal helix. The N-terminal PKS docking domain (KS-side docking domain, KSdd) exhibits a coiled-coil motif that has been observed both in the solution structure of the fused DEBS 2/DEBS 3 construct (85) and subsequently in the X-ray crystal structure of the DEBS 3 KS-AT didomain (78). The KSdd dimer presents a small hydrophobic patch, sometimes

flanked by charged residues, as a narrow binding groove where the ACPdd terminal helix can bind.

Extension of the current DEBS 2/DEBS 3 structural model to the full range of docking domains across modular PKSs has not been possible due to low sequence similarity for a large subset of sequences. However, in a recent report based on the DEBS 2/DEBS 3 structural model and computational analysis of docking domain sequences from 42 characterized PKS systems, Thattai *et al.* proposed a new organization of PKS docking domains into distinct subclasses (91). Based on this classification system, the majority of docking domains (including the structurally characterized DEBS2/DEBS3 pair) fall into a single group termed H1-T1 (for head 1 and tail 1). Until this work, there was no structural information available for the proposed H2-T2 group of PKS docking domains.

2.2 Results and Discussion

To develop further our understanding of docking domain interactions in modular PKSs, and to expand fundamental information about docking domain protein structure, we pursued both biochemical and structural characterization of docking domains from two well-studied PKS biosynthetic systems. Here, we report an analysis of the binding affinities of discrete docking domain pairs excised from the erythromycin (DEBS) and pikromycin (Pik) PKSs, using surface plasmon resonance and fluorescence polarization methods. In addition, we report the first X-ray crystal structure of a member of the recently proposed H2-T2 class of PKS docking domains, derived from the interface between PikAIII (module 5) and PikAIV (module 6) proteins from the Pik PKS system (Figure 2-1B) (*36*). Combining structural characterization of the PikAIII/PikAIV

interface with discrete docking domain affinity measurements, we provide evidence in support of the prevailing model wherein the binding specificity that determines the linear arrangement of proteins in the biosynthetic assembly line is encoded in these small, terminal docking domains. Finally, we present a model for the observed docking domain specificity across a matrix of interacting pairs from the pikromycin and erythromycin pathways.

2.2.1 Binding Affinities of Discrete Docking Domains Determined by Surface Plasmon Resonance

To test the capacity of discrete docking domains to discriminate between possible partners within a single biosynthetic pathway and/or between related pathways, we produced peptides corresponding to each ACPdd and KSdd region of the pikromycin and erythromycin PKS pathways (Figure 2-1). Peptides were overexpressed in *E.coli* and purified using a His_6-affinity handle followed by removal of the His-tag via TEV protease cleavage where necessary. While each of the docking domain constructs resulted in stable, soluble protein, the yields of the PikAII KSdd and PikAIII KSdd were low.

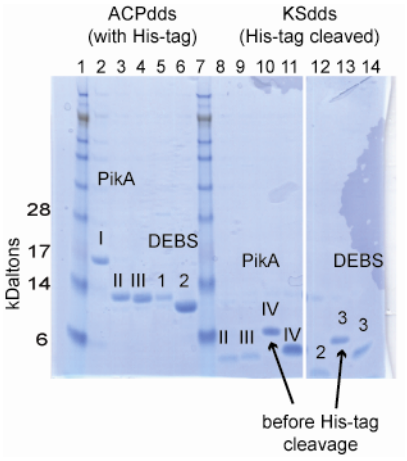


Figure 2-3. Purified discrete docking domains. Apparent molecular weight of the SeeBlue Plus2 molecular weight marker (Invitrogen) is shown for reference. Proteins were run on a NuPAGE 12% SDS-PAGE gel using MES buffer.

Hence, for these two peptides, chemical synthesis was employed to produce larger quantities. The ability of KSdds to bind native immobilized ACPdd partners was evaluated using surface plasmon resonance (SPR). Biosensors based on SPR technology have been used to measure binding interactions across a wide range of affinities between partners (including discrete docking domains from a related mixed PKS/NRPS megasynthase system) varying from small molecules to large protein complexes (92-94). In modular PKSs, individual docking domains are identifiable by considering sequences directly downstream from the ends of the C-terminal ACP domain or directly upstream from the conserved start sites of the N-terminal KS domain. Using multiple sequence alignments of a number of characterized type I PKS systems, we designed, overexpressed, and purified (and in two cases synthesized) a complete set of discrete ACPdds and KSdds from the erythromycin and pikromycin system (Figure 2-3). In these studies, we used a noncovalent method to immobilize the N-terminally His-tagged ACPdds to a nickel-loaded NTA sensor chip (Figure 2-4A) (95). The measured affinity (K_D) of His-tagged PikAIII ACPdd to the nickel-NTA surface was 4.0 ± 0.04 nM (Figure 2-5). This binding was sufficiently tight to enable measurement of the desired ACPdd-KSdd interactions when paired with tagless KSdds in solution.

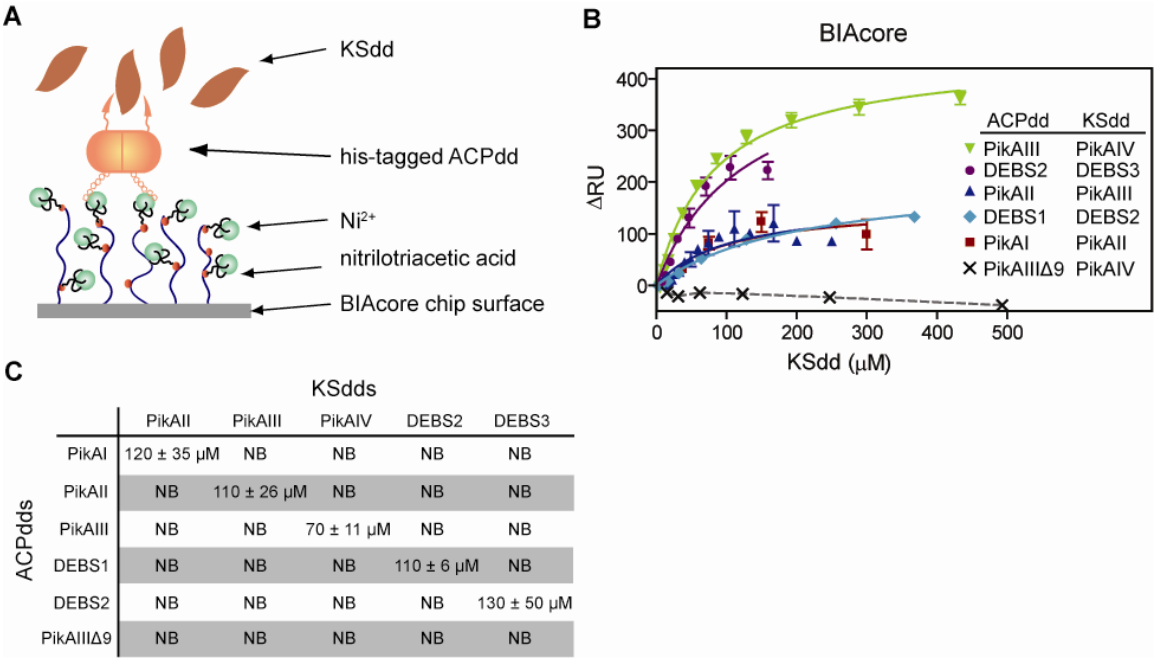


Figure 2-4. Binding of discrete docking domains determined by SPR. Binding analysis of discrete ACPdd's to matched and mismatched KSdd's measured by SPR. A) SPR assay design: after immobilizing the ACPdd's via their His-tags, varying concentrations of discrete KSdd's were injected over the ACPdd and control surfaces. B,C) K_Ds were calculated using a one-site binding model, $Y = B_{max}*X(K_D + X)$. Dose-response curves were performed in triplicate, and the error bars are SEM. NB = no binding.

After immobilization of ACPdd, equilibrium analysis of a variety of matched or mismatched docking domains pairs was performed using sequential injections of KSdd at varying concentrations. Using docking domains from the erythromycin and pikromycin PKSs, we measured K_D's for the matched docking domain pairs between 70 - 130 μM (Figure 2-4B,C). Additionally, we were able to calculate individual kinetic parameters for the PikAIII/PikAIV binding pair (k_{on} = 3000 ± 1800 $M^{-1}s^{-1}$, k_{off} = 0.21 ± 0.03 s^{-1}, K_D = 73 ± 43 μM) (Figure 2-6) that were in good agreement with the equilibrium analysis. As a negative control, a PikAIII ACPdd construct lacking the final nine amino acids was unable to bind to its partner KSdd (PiKAIV) or any other KSdds (Figure 2-4B, C). A similar C-terminal deletion of the PikAIII ACPdd was also shown to be incompetent for production of narbonolide in our *in vitro* PikAIII/PikAIV chemoenzymatic system (*26*).

Furthermore, studies with mismatched docking domains clearly demonstrate that the ability to discriminate between potential PKS protein partners is encoded within the docking domains themselves (Figure 2-7).

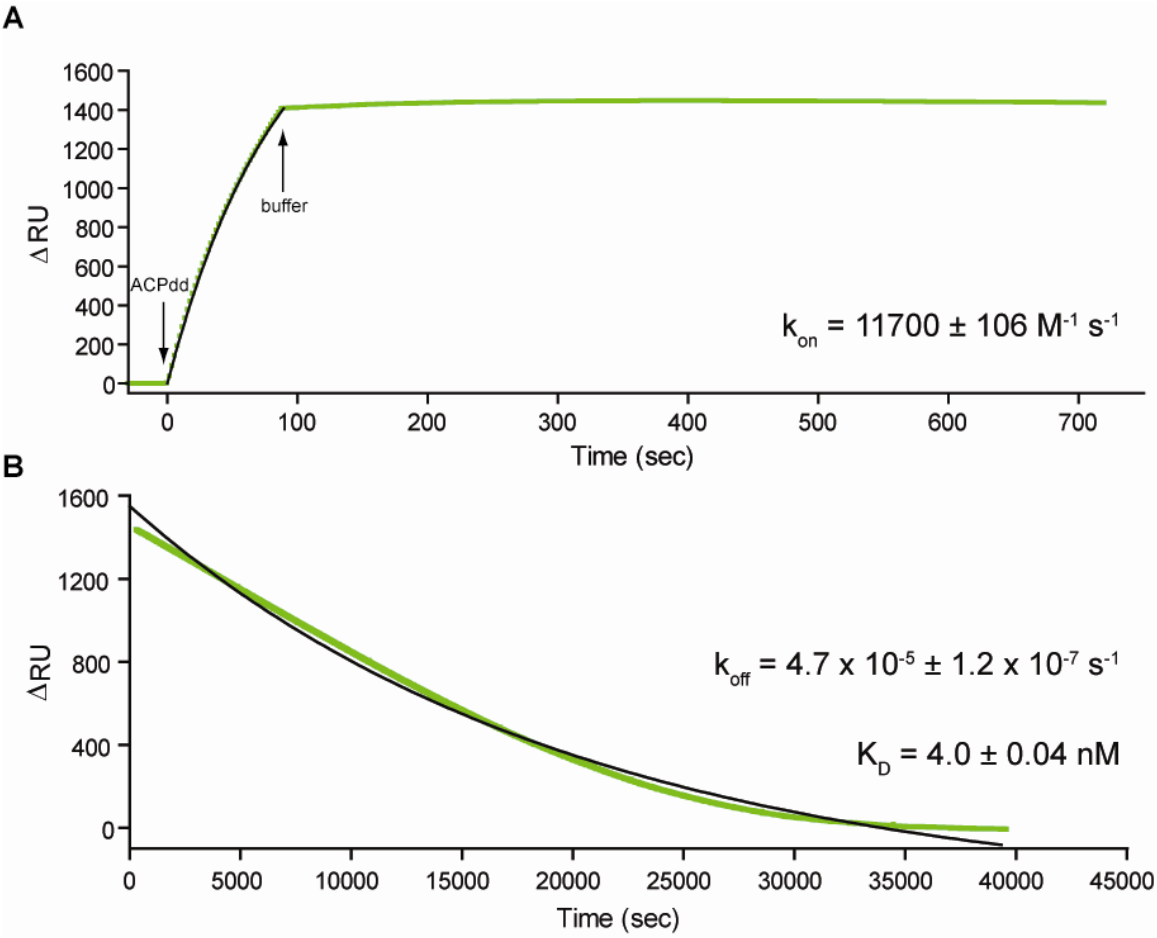


Figure 2-5. His-tagged ACPdd immobilization. Immobilization of His-tagged docking domains to the Ni-NTA BIAcore chip surface. A) Subtracted sensorgram (green) of 1 μM His-tagged PikAIII ACPdd binding to the nickel-NTA surface showing stable immobilization. The association phase was fit to a one-site association model (Y=Ymax*(1-exp(-K*X)). B) Overnight dissociation of the His-tagged PikAIII ACPdd. The dissociation phase was fit to a one phase exponential decay (Y=Span*exp(-K*X) + Plateau).

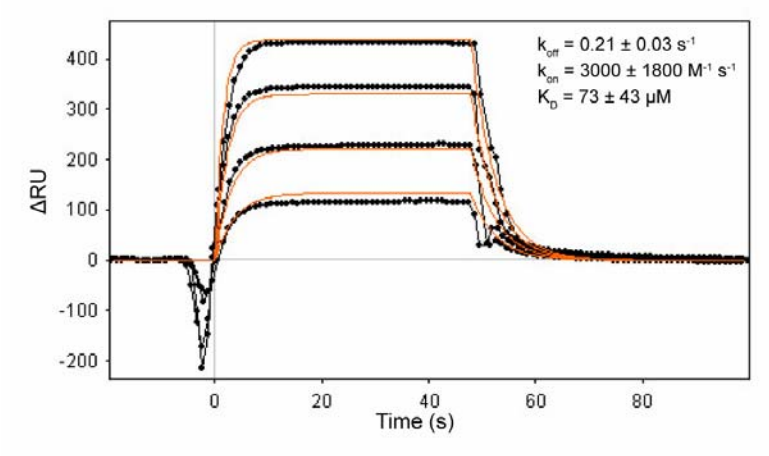


Figure 2-6. Kinetic fit of PikAIII ACPdd – PikAIV KSdd binding via SPR. Corrected sensorgram (in black) after subtraction of the control lane (nickel only). The association and dissociation phase data for four concentrations (150, 75, 37.5, 19 µM) were fit simultaneously using a global data analysis program, Scrubber2 (BioLogic Software) (shown in red) to determine the KD. Error for the individual association and dissociation values was calculated by fitting each concentration separately.

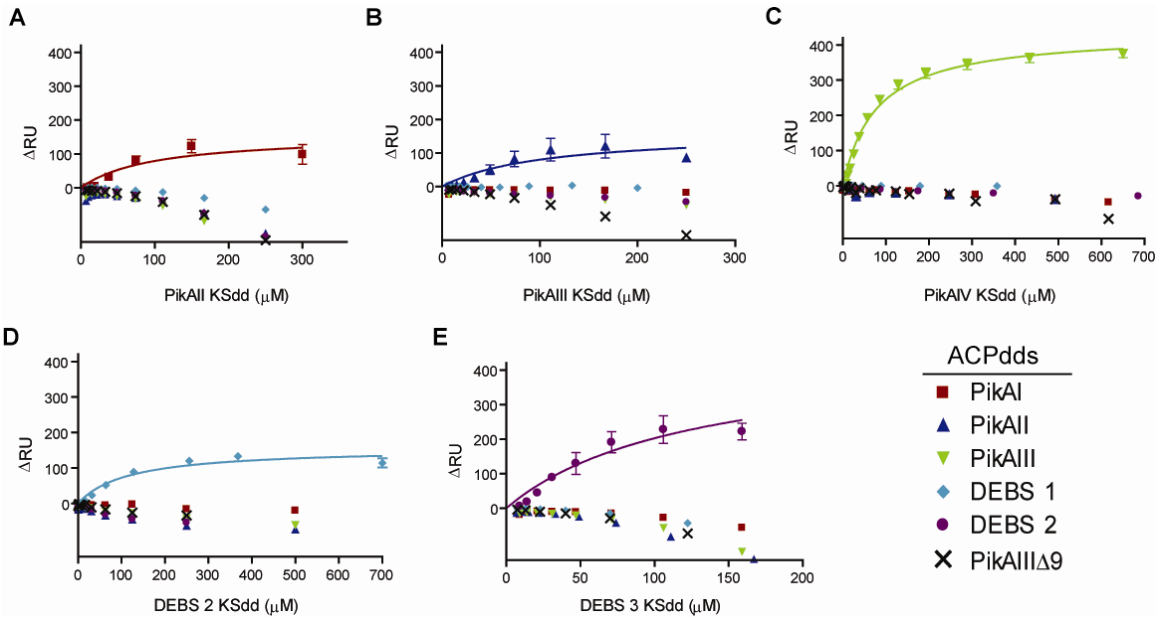


Figure 2-7. Matched and mismatched docking domain binding determined by SPR. The difference in response units between the ACPdd loaded surface and the nickel-only control surface in ΔRU is shown.

Ultimately, docking domains function not as discrete peptides, but as small appendages on much larger proteins (Figure 2-1). In addition to testing the complete library of discrete ACPdds and KSdds from the pikromycin and erythromycin PKS systems, we extended our analysis of the PikAIII/PikAIV docking interface to the

neighboring domains (Figure 2-8). Assigning affinity and specificity determinants to 1) the docking domains, 2) the neighboring catalytic domains, 3) the phosphopantethiene arm, and 4) the growing polyketide chain will begin to separate the importance of the correct protein-protein interaction from the questions of substrate specificity at the catalytic centers. Although binding of a larger KSdd-containing protein (PikAIV KSdd-KS-AT) to the His-tag immobilized PikAIII ACPdd via SPR was observed, we were unable to calculate affinity values due to the high background refractive index change exhibited. We thus sought an alternative method to address this question.

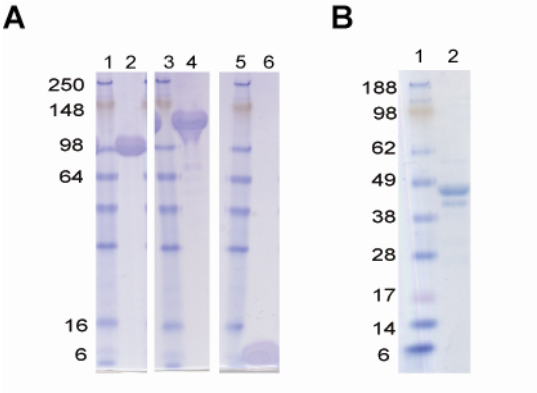


Figure 2-8. Protein purification of PikAIV proteins. Apparent molecular weight of the SeeBlue Plus2 molecular weight marker (Invitrogen) is shown for reference. A) PikAIV KSdd-KS-AT (lane 2), PikAIV full module (lane 4), and P3P4dock (lane 6) resolved on a 4-20% Tris-Glycine SDS-PAGE gel. B) PikAIV KS (TEV-cleaved) run on a NuPAGE 12% SDS-PAGE gel using MES buffer.

2.2.2 Binding Affinities of Discrete Docking Domains Determined by Fluorescence Polarization

To assess the effect of larger protein complexes on docking domain binding affinity, a fluorescence polarization (FP) assay was employed (96). We empirically determined the best fluorophore placement through the addition of a cysteine residue at each of the four possible termini (N-terminus and C-terminus of PikAIII ACPdd and PikAIV KSdd). Inclusion of a single cysteine residue enabled site-specific labeling with

iodoacetamide-BODIPY-FL. Titration of increasing concentrations of the unlabeled matched docking domain identified the C-terminus of PikAIII ACPdd (termed PikAIII ACPdd-FL) as the optimal fluorophore placement, as this tracer exhibited the largest change in FP upon protein binding. The binding affinity of the PikAIV KSdd for PikAIII ACPdd-FL measured using this method provided an independent confirmation of the discrete docking domain binding affinities generated using SPR (Figure 2-9). When the larger KSdd-containing PikAIV proteins were titrated against ACPdd-FL, a 2-10-fold increase in affinity was observed (Figure 2-9). However, a construct consisting of only the KS domain of PikAIV (without its docking domain) did not bind to PikAIII ACPdd-FL (data not shown). Most likely, the presence of downstream domains in these longer constructs stabilizes the productive binding conformation of the PikAIV KSdd. Furthermore, it is possible that additional protein-protein contacts exist between the upstream ACPdd and the downstream KS-AT region of the module, although these regions have yet to be identified (29, 87).

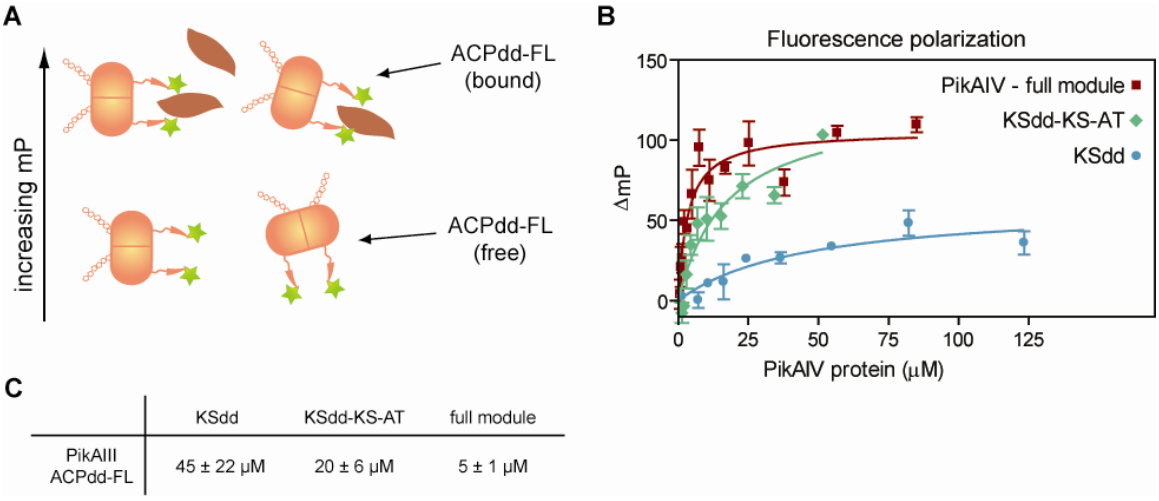


Figure 2-9. Binding of PikAIII ACPdd monitored by fluorescence polarization. A) FP assay design: 50 nM PikAIII ACPdd-Fl was mixed with varying concentrations of KSdd, KSdd-KS-AT, or full module PikAIV and allowed to equilibrate at RT before reading. B, C) K_Ds were calculated using a one-site binding model, $Y = B_{max}*X/(K_D + X)$. Dose response curves were performed in triplicate, and the error bars are SEM.

Our *in vitro* binding affinities for these canonical modular PKS docking domains are similar to those measured by SPR and ITC for the orthogonal discrete TubB/TubC docking elements (K_D ~50 μM), domains found in some mixed-PKS-NRPS synthetases, whose novel structure was reported recently (94). Additionally, the affinity of the PikAIV full module for PikAIII ACPdd as assessed by fluorescence polarization (5 ± 1 μM) is comparable to that estimated for the DEBS 1 / DEBS module 3+TE obtained by monitoring rates of tri- and tetraketide lactone synthesis (2.6 μM) *in vitro* (97). Thus, correct pairing of large multi-domain modules in both PKS and mixed PKS-NRPS biosynthetic assembly lines appears to result, at least in part, from specificity determinants with rather weak affinities. Despite these weak affinities, discrete docking domains from the related phoslactomycin (Plm) biosynthetic cluster have been used to separate the trimodular PikAI PKS (Figure 2-1B) into monomodular proteins in a *Streptomyces venezuelae* strain lacking *pikAI* (98). Remarkably, generation of the final macrolide products (methymycin and pikromycin) were within two-fold of the total yield compared with production when using native PikAI. The mechanism bacteria use to achieve such exquisite selectivity, albeit with modest protein-protein affinities, remains poorly understood. However, one clue might come from the analysis of the PksX megacomplex, a mixed PKS-NRPS responsible for producing bacillaene. In this system, the proteins of the biosynthetic machinery have been visualized via fluorescence microscopy to reside at a single organelle-like complex in the bacteria, perhaps suggesting that higher order multivalent interactions are available to further increase the affinities if needed (99).

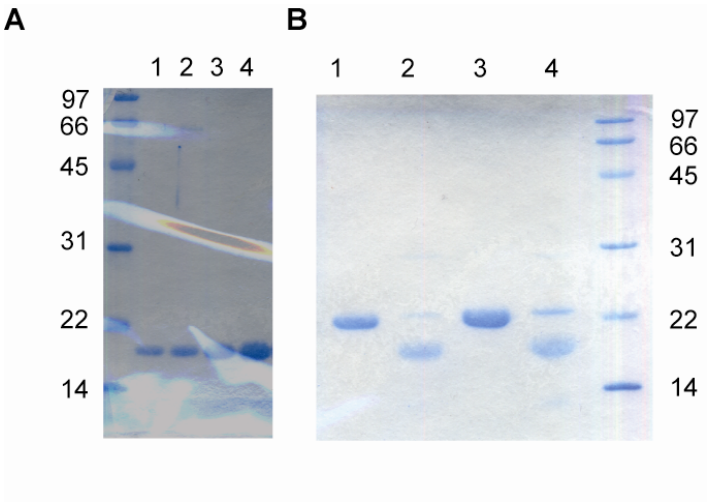


Figure 2-10. Protein purification of PikAIII-PikAIV docking domain fusions. For reference, molecular weights are listed along side each gel in kDa. A) dilutions of the direct PikAIII-PikAIV fusion. B) PikAIII-PikAIV fusion with one GGGS spacer before (lane 1) and after (lane 2) tag removal, and PikAIII-PikAIV fusion with two GGGS spacers before (lane 3) and after (lane 4) tag removal. All proteins were resolved on a 15% Tris-Glycine SDS-PAGE gel and stained with coomassie.

2.2.3 Structure of the PikAIII/PikAIV Docking Interface

We next explored the structural basis for the observed binding specificity between the discrete pikromycin and erythromycin docking domain pairs. Given the low sequence similarity between the structurally characterized H1-T1 class and the uncharacterized H2-T2 class of PKS docking domains, we targeted the recently proposed H2-T2 class for structure determination (91). A direct fusion strategy had been used successfully to generate a construct to solve the DEBS 2/DEBS 3 solution structure (85). To characterize the low-affinity PikAIII/PikAIV docking domain complex, we generated constructs where the C-terminus of PikAIII ACPdd was either directly fused to the N-terminus of PikAIV KSdd or separated by one or two Gly-Gly-Gly-Ser spacers. In the PikAIII/PikAIV system, this docking domain fusion strategy yielded proteins that were highly soluble, and purification yielded 25-75 mg protein/L of culture (Figure 2-10). Docking domain constructs derived from PikAIII/PikAIV containing all four predicted helices eluted as two oligomeric species on size exclusion chromatography, but these

proteins failed to form crystals. This is likely due to the existence of mobile linker regions, as were found in the DEBS 2/DEBS 3 docking complex (85). We then targeted a smaller construct focused only on the putative inter-polypeptide docking helices (amino acids 1534-1562 of PikAIII ACPdd fused to amino acids 1-37 of the KSdd of PikAIV, together termed P3P4dock) (Figure 2-11). The P3P4dock crystal structure was solved by single wavelength anomalous diffraction using selenomethionyl protein. The 1.75 Å crystal structure of P3P4dock includes residues 1544-1562 of PikAIII ACPdd and 1-37 of PikAIV KSdd, whereas residues 1534-1543 were disordered and lacked elctron density.

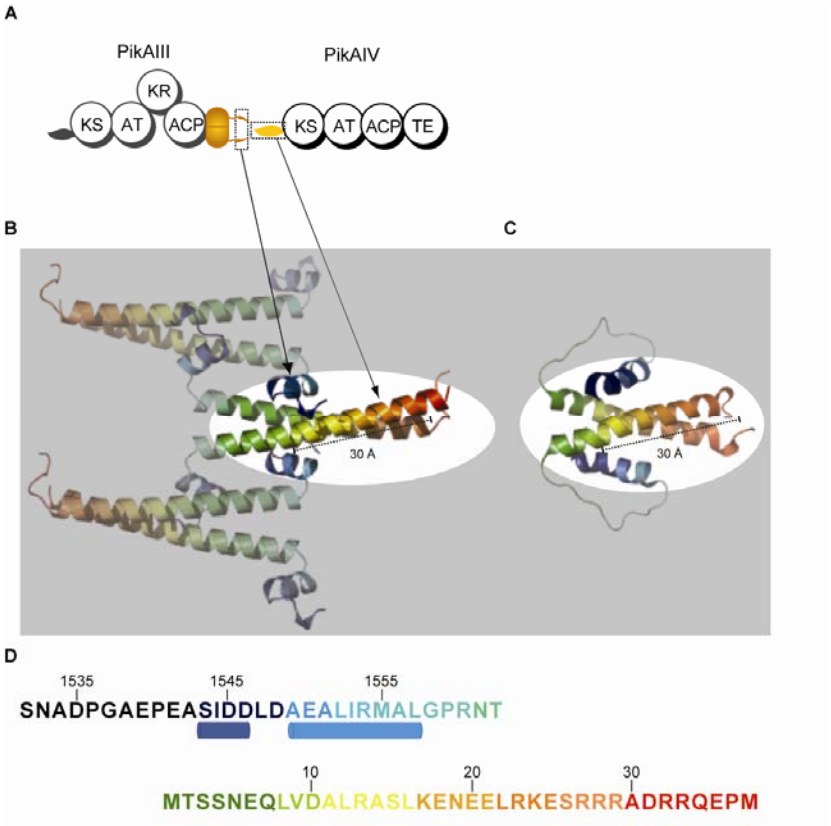


Figure 2-11. PKS docking interface structures. A, B) Packing of the PikAIII ACPdd/PikAIV KSdd crystal structure (PDB id 3F5H). Three P3P4dock dimers are shown. The docking interaction formed by neighboring dimers is highlighted. C) In the NMR structure of the DEBS 2 ACPdd-DEBS3 KSdd (PDB id 1PZR), the docking interaction is intramolecular. In panels B and C, polypeptide chains are colored blue to red from the N-terminus to the C-terminus of the construct. D) P3P4dock sequence; top line is residual purification tag (SN) followed by PikAIII (residues 1534-1562), bottom line is PikAIV (residues 1-37). ACPdd helices are indicated below the sequence.

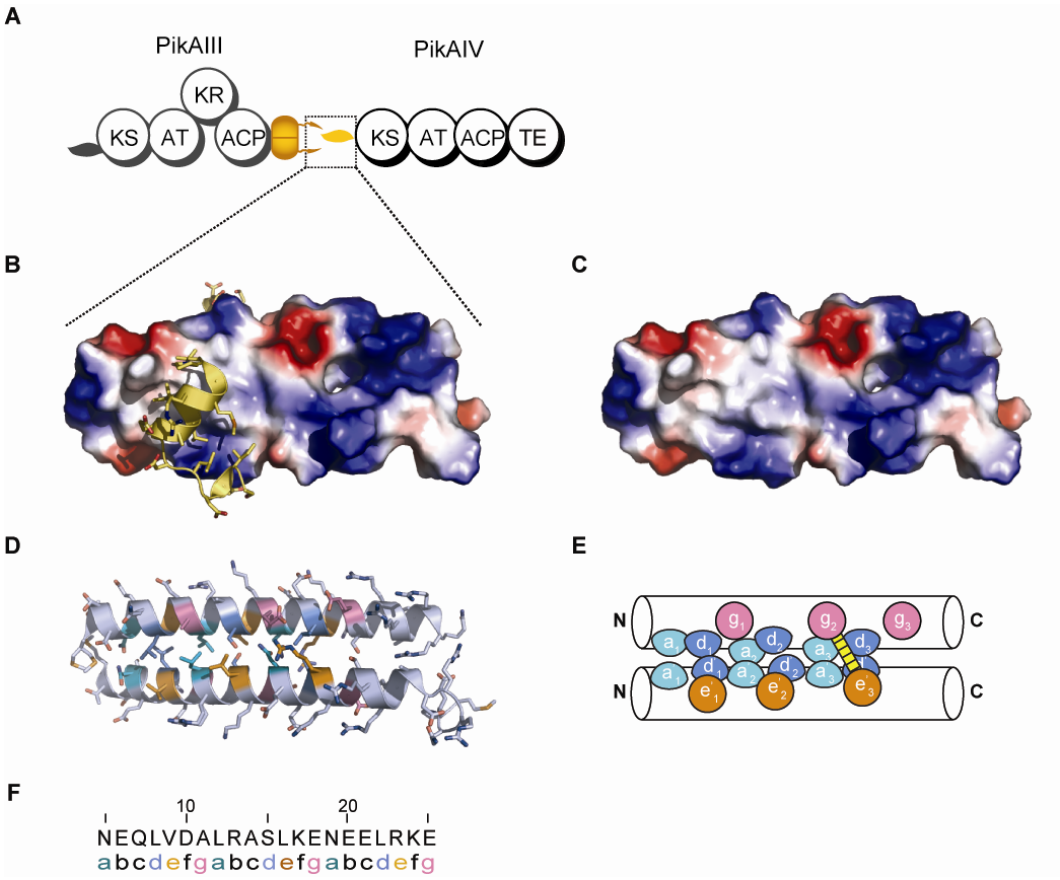


Figure 2-12. PikAIV KSdd coiled-coil structure. A) PikAIII and PikAIV multidomain proteins. The docking domains are colored gold, and the region used to construct the P3P4dock fusion is boxed. B, C) The electrostatic potential map of the coiled-coil generated in PyMOL; in yellow, the short helix of the PikAIII ACPdd (residues 1543-1559) is bound to the small hydrophobic patch of the PikAIV KSdd. Blue represents positive charge and red negative charge. D, E) The PikAIV KSdd parallel coiled-coil. The "a" and "d" residues make up the core of the homodimeric interface and are colored in cyan and light blue. The "e" and "g" residues are colored orange and pink, respectively. An inter-coil hydrogen bond is found between residue g2 and e'3. F) The coiled-coil heptad repeat of the PikAIV KSdd.

The P3P4dock protein structure consists of a short helix bound to a parallel coiled-coil (Figure 2-11B, C and 2-12) (85). The relevant docking interface is made up of a coiled-coil of a single homodimer flanked by two individual ACPdd helices from neighboring protein molecules in the crystal lattice (Figure 2-11B and 2-12). The coiled-coil packing exhibits the familiar heptad repeat architecture with the "a" and "d" amino acids forming the core of the coiled-coil, and the "e" and "g" positions providing the majority of the residues for contacting the upstream PiKAIII ACPdd helix (Figure 2-12 D-F). The dominant interaction of the PikAIII ACPdd helix occurs in a hydrophobic

patch on the PikAIV KSdd coiled-coil (Figure 2-12B, C). The interacting hydrophobic surfaces display exquisite shape complementarity (Figure 2-13). Additional inter-domain interactions are found where residues 1544-1547 of PikAIII ACPdd fold back to interact further downstream on the KSdd dimer (Figure 2-13A, B). This positioning of residues 1544-1547 in PikAIII ACPdd is mediated by a charge-charge interaction between Asp1545 of PikAIII and Lys17 of PikAIV, as well as hydrogen bonds between main chain carbonyls from Ile1544 and Leu1547 of PikAIII and Arg13 of PikAIV (Figure 2-13A). No other charge-charge interactions are seen at the PikAIII/PikAIV docking interface. These electrostatic interactions and remarkable shape complementarity represent a potential selectivity filter (Figure 2-13D).

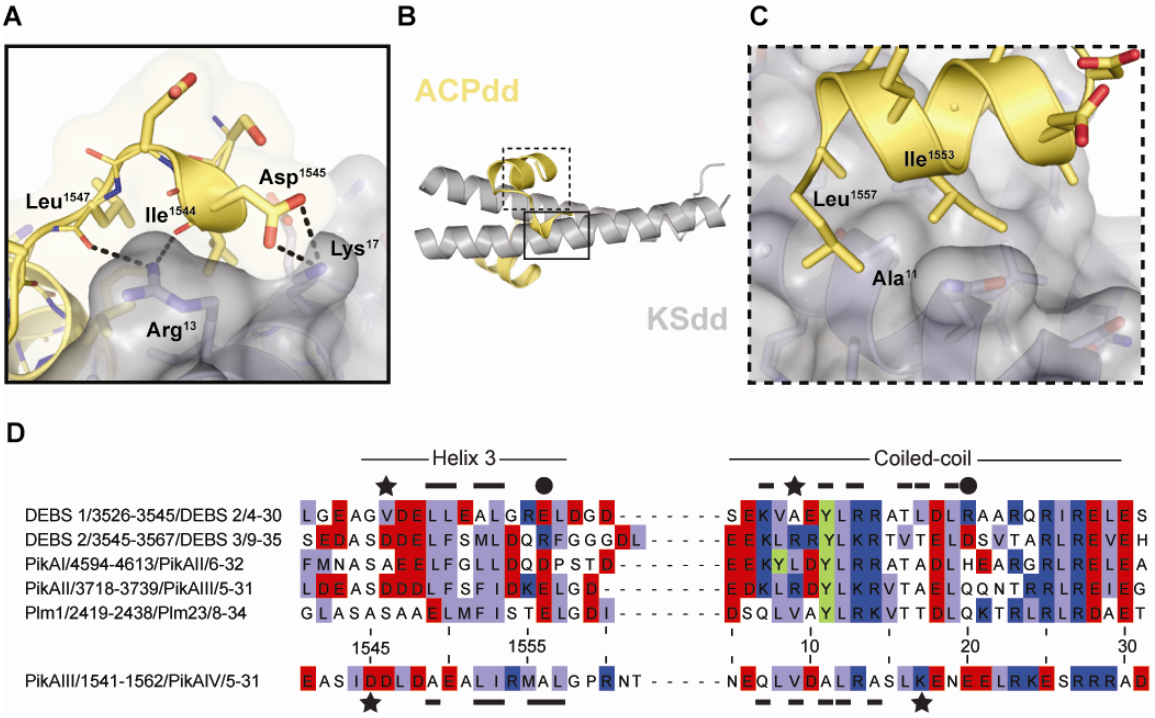


Figure 2-13. Shape complementarity in PKS docking domains. A-C) Docking interface of PikAIII-PikIAV; KSdd of PikAIV is colored gray, and the terminal helix of PikAIII ACPdd is light yellow. D) Multiple sequence alignment of the docking domains tested (or highlighted) in this chapter generated in JalView. Residue numbering is shown for the PikAIII/PikAIV pairing. Basic residues are colored dark blue, acidic residues are red, hydrophobic amino acids are light blue, and tyrosine is highlighted in green. Regions of interaction for each subclass are denoted by matching symbols.

The PikAIII/PikAIV docking domain structure revealed an overall architecture similar to that of the DEBS 2/DEBS 3 docking domain model obtained via NMR spectroscopy (Figure 2-11B, C) (85). In both the PikAIII/PikAIV and DEBS 2/DEBS 3 docking domain structures, the ACPdd helix binds to the KSdd coiled-coil approximately 30 Å (Figure 2-11B, C) from the downstream KS catalytic domain (not present in either structure). However, many details of the structures differ. The most apparent structural difference between the PikAIII/PikAIV and DEBS 2/DEBS 3 docking domains is the length of the terminal ACPdd helix (Figures 2-11, 2-12, and 2-13). The 9-residue PikAIII ACPdd helix (residues 1549-1557) is considerably shorter than its 15-residue DEBS 2 ACPdd counterpart. Although both the PikAIII/PikAIV and DEBS 2/DEBS 3 docking domain interfaces display well-defined shape complementarity between matched pairs, the mis-matched pairs appear highly non-complementary. Polar interactions also differ between the two interfaces (Figure 2-13D).

The classification of PKS docking domains proposed by Thattai *et al.* is generally consistent with the two experimental structures of paired docking domains in which DEBS 2/DEBS 3 is type H1-T1 and PikAIII/PikAIV is type H2-T2 (Figure 2-13D). For example, the residues analogous to PikAIII Asp1545 and PikAIV Lys17 of the H2-T2 subclass are most frequently an Asp/Lys pair. In contrast, small or hydrophobic residues occupy those positions in the H1-T1 subclass of PKS docking domains (Figure 2-13D and Appendix A). In addition, the key residues involved in the hydrophobic interface are shifted between the proposed H1-T1 and H2-T2 subclasses of PKS docking domains (see bars above the sequence alignments in Figure 2-13D). Furthermore, residue 11 is an alanine in PikAIV and all other H2-T2 KSdds, whereas in over 90% of H1-T1 KSdds the

analogous residue is a tyrosine (91). (Figure 2-13 and Appendix A). On the ACPdd helix, the large, hydrophobic residues Ile1553 and Leu1557 are across the interface from Ala11 of PikAIV KSdd (Figure 2-13C). Due to the size of these residues on the ACPdd helix, accommodating the tyrosine side chain of an H1-T1 KSdd at the position analogous to Ala11 of PikAIV KSdd appears unfavorable and likely represents another selectivity filter between the H1-T1 and H2-T2 subclasses. Altogether, these data lend support to the hypothesis that H1-T1 and H2-T2 are structurally distinct subclasses of PKS docking domains.

Specificity within the H1-T1 class appears to be driven by three distinct interaction zones. The hydrophobic core of the protein-protein interface is symbolized (Figure 2-13D) by solid bars above both the ACPdd and the KSdd in the multiple sequence alignment. On either side of the hydrophobic core are positions of potential charge-charge interaction (Figure 2-13D, stars & circles). Mismatching at any of the three zones appears to be sufficient to inhibit non-cognate docking domains from binding productively to one another *in vitro*. For example, at the "star" position, the DEBS 2/DEBS 3 pair and the PikAII/PikAIII pair both contain the same attractive charge-charge pair (Asp/Arg). However, at the "circle" position, a mismatched PikAII/DEBS 3 pairing would bring two negatively charged residues (Glu/Asp) in close proximity. A report of productive association and transfer between PiKAII and DEBS 3 *in vivo* suggests that within the H1-T1 class, inter-polypeptide interactions beyond the docking domains may also come into play (83), although further structural work is needed to identify additional contact regions.

2.3 Summary

Applying techniques from both analytical biochemistry and structural biology to the pikromycin biosynthetic pathway, we have increased our understanding of the key steps involved in intermodular acyl group transfer. Specifically, we have demonstrated the first direct binding analysis of the discrete PKS docking domains via two complementary methods, surface plasmon resonance and fluorescence polarization. We combined our binding analysis with a structural protein model generated by x-ray crystallography to elucidate the molecular details of the PikAIII-PikAIV docking domain interface. Collectively, with the predictive tools already in place for PKS catalytic domains, the enhanced diagnostic power should enable more accurate assignment of individual pathway metabolic products.

2.4 Experimental Methods

2.4.1 Design of Expression Constructs

Plasmids for the expression of the discrete docking domain fragments, PikAIV KSdd-KS-AT, and the full module of PikAIV were generated by amplification using PCR with LIC overhangs and inserted into the vector pMCSG7 (100). DEBS 1, DEBS 2, and DEBS 3 docking domains were amplified from cosmid pDHS9746. PikAI ACPdd and KSdd were amplified from plasmid pDHS0030. PikAII ACPdd and KSdd were amplified from plasmid pDHS0805. PikAIII ACPdd, PikAIII ACPdd-C-FL and KSdd were amplified from plasmid pDHS8011. PikAIV KSdd, KSdd-KS-AT, and the full module were amplified from plasmid pDHS0137. All primers are listed 5' to 3' in Table 2-1. Sequences in all capital letters represent the LIC overhangs necessary for insertion

into the pMCSG7 vector. The engineered cysteine codon for the P3ACPdd-C-Cys is indicated in bold. The overlaps used to generate the PikAIII-PikAIV fusion are shown in italics. All PCR fragments were inserted in the vector pMCSG7 via ligation independent cloning. Similarly, a construct lacking the N-terminal docking domain, termed PikAIV KS, was amplified from plasmid pDHS0137 and inserted into pMocr (101). The C-terminus of the PikAIV discrete ketosynthase construct terminates at a position near that of a recently reported soluble DEBS module 3 KS (*25*).

A plasmid encoding the full PikAIII ACPdd fused to the PikAIV KSdd (pDHS9672) was generated via sequential PCR amplification of 1) individual dd's PikAIII and PikAIV containing appropriate overlapping DNA at the ends using plasmid DNA for PikAIII (pDHS8011) and PikAIV (pDHS0137) and 2) the fused construct from PCR amplification of the combined fragments using outside primers. The plasmid pDHS9570 (encoding P3P4dock) was generated by PCR amplification of a fragment of pDHS9672 followed by insertion into the vector pMCSG7. All DNA sequences were confirmed by sequencing.

Table 2-1. Primer list for PikA and DEBS protein expression.

Primer name	Primer sequence	Plasmid(s)
P1ACPddLICF	TACTTCCAATCCAATGCC ctg cct cgc ggc gcc tcg gac ca	pDHS9683
P1ACPddLICR	TTATCCACTTCCAATGCTA gtc cgt gct ggg gtc ctg gt	
P2ACPddLICF	TACTTCCAATCCAATGCC gcg ggc ggg tcc tgg gcg gaa	pDHS9687
P2ACPddLICR	TTATCCACTTCCAATGCTA gaa gtc gga gtc gcc cag ctc ctt	
P3ACPddLICF	TACTTCCAATCCAATGCC ctc cac gag gcg tac ctc gca c	pDHS9560 & pDHS9672
P3ACPddLICR	TTATCCACTTCCAATGCTA ggt gag gcg cag gac ggt gt	pDHS9560
P3ACPddCysLICR	TTATCCACTTCCAATGCTA **aca** ggt gag gcg cag gac ggt gt	pDHS9681
P3ACPddOverR	*act gtt cgt tgg aac tcg tca t*gg tgt tac ggg ggc cga gag c	pDHS9672
P4KsddOverF	tct cgg ccc ccg taa cac c*at gac gag ttc caa cga aca g*	
P4KSLICF	TACTTCCAATCCAATGCC cag gag ccc atg gca atc gt	pDHS9703
P4postATLICR	TTATCCACTTCCAATGCTA ctc gcg ccc gga agc ggt gtg	pDHS9696
P4CtermLICR	TTATCCACTTCCAATGCTA ctt gcc cgc ccc ctc gat gc	
P4KSddLICF	TACTTCCAATCCAATGCC atg acg agt tcc aac gaa cag ttg	pDHS9561, pDHS9736, pDHS9734
P4KSddLICR	TTATCCACTTCCAATGCTA cat ggg ctc ctg ccg acg gt	pDHS9561 & pDHS9672
P4KSLICR	TTATCCACTTCCAATGCTA cgg ggc ctc ctc cag gac aac gt	pDHS9703
P3P4dockLICF	TACTTCCAATCCAATGCC gac cct ggt gcg gag ccg gag g	pDHS9570
D1ACPddLICF	TACTTCCAATCCAATGCC ggc acc gag gtc cga ggg gag	pDHS9691
D1ACPddLICR	TTATCCACTTCCAATGCTA atc gcc gtc gag ctc cc	
D2KSddLICF	TACTTCCAATCCAATGCC gtg act gac agc gag aag gtg gc	pDHS9693
D2KSddLICR	TTATCCACTTCCAATGCTA gga ttc cag ctc acg gat g	
D2ACPddLICF	TACTTCCAATCCAATGCC ttc gcg gcc agt ccg gcg gtg gac a	pDHS9695
D2ACPddLICR	TTATCCACTTCCAATGCTA cag gtc ctc tcc ccc c	
D3KSddLICF	TACTTCCAATCCAATGCC atg agc ggt gac aac ggc atg a	pDHS9696
D3KSddLICR	TTATCCACTTCCAATGCTA acc ggc ccg gtg ctc gac tt	

2.4.2 Expression and Purification of Docking Domain Proteins

Plasmids encoding TEV protease-cleavable N-terminal His_6-fusion proteins were transformed into *E. coli* BL21(DE3) and grown at 37 °C in TB medium to an OD_{600} of

~1.0 in 2L flasks. The cultures were cooled to 18 °C, and isopropyl β-D-thiogalactopyranoside was added to a final concentration of 0.2 mM and grown 12-16 h with shaking. The cells were harvested by centrifugation and frozen at either -20 or -80 °C. Selenomethionyl protein was produced in a similar fashion using selenomethionine minimal medium (102). Cell pellets were thawed to 4 °C and resuspended in 5X volume of lysis buffer (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 1 mM $MgCl_2$, and ~100 mg CelLytic Express (Sigma-Aldrich)) before lysis via sonication. For discrete KSdds, Complete Protease Inhibitor Cocktail tablets (Roche) were added to the lysis buffer. Centrifugation at 25,000*x*g for 30 min provided clarified lysates. Proteins were purified using Ni-Sepharose affinity chromatography on an Akta FPLC. Briefly, after filtration of the supernatant through 0.45 μm membrane, the solution was loaded onto a 5-mL HisTrap nickel-nitrilotriacetic acid column. The column was washed with 10 column volumes of buffer A (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole) and eluted with a linear gradient of buffer B (20 mM HEPES, pH 7.8, 300 mM NaCl, 400 mM imidazole). His-tag removal was achieved by TEV protease incubation overnight at 4 °C in HEPES buffered saline (20 mM HEPES, pH 7.8, 150 mM NaCl, HBS) or buffer A containing 1 mM TCEP. His-tagged peptides and TEV protease were removed by repassaging the solution over the HisTrap column. Flow-through fractions were pooled, concentrated, and loaded onto a HiLoad 16/60 Superdex 75 (GE Healthcare) column equilibrated with HBS. Fractions were combined, concentrated, frozen, and stored at -80 °C. Because many of the small peptides lack amino acids with appreciable absorbance at 280 nm, protein concentration was determined using the bicinchoninic acid (BCA) method using BSA as a standard. Protein yields varied from 1 – 75 mg/L of cell culture.

PikAII and PikAIII KSdd were chemically synthesized by Genscript corporation. Proteins were further purified by size exclusion chromatography on the HiLoad 16/60 Superdex 75 as above to remove residual HPLC purification contaminants before using the peptides in binding assays. To assure that no undesired cleavage products were formed during TEV protease incubation, the PikAIV, DEBS2 and DEBS3 KSdds were subjected to high resolution mass spectrometry (data not shown). For each peptide, the observed molecular weight was consistent with cleavage exclusively at the predicted TEV protease site (Figure 2-4).

2.4.3 Expression and Purification of PikAIV KS, PikAIV KSdd-KS-AT and PikAIV Full Module

Proteins were expressed as described for the docking domains above except that the PikAIV full module construct was grown in BAP1 *E.coli* cells (103) to provide post-translational modification of its ACP domain. Proteins were purified as above using the following buffers. For cell lysis, lysis buffer with reductant (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 0.5 mM TCEP 1 mM $MgCl_2$, and ~100 mg CelLytic Express) was used. During FPLC purification, wash buffer was buffer C (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 0.5 mM TCEP) and the elution buffer used was buffer D (20 mM HEPES, pH 7.8, 300 mM NaCl, 400 mM imidazole, 10% glycerol, 0.5 mM TCEP). Size exclusion chromatography was performed on a HiLoad 16/60 Superdex 200 (GE Healthcare) column equilibrated with storage buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP). Protein concentrations were determined using absorbance at 280 nm and calculated extinction

coefficients (PikAIV KS, 1 A_{280} = 1.0 mg/ml; PikAIV KSdd-KS-AT, 1 A_{280} = 0.91 mg/ml; PikAIV full module, 1 A_{280} = 0.94 mg/ml) (Figure 2-8).

2.4.4 Surface Plasmon Resonance Assays

Sensor chips (NTA) and HBS-P buffer were purchased from GE Healthcare Life Sciences. SPR experiments were performed on a BIAcore 3000 instrument. Running buffer for SPR was HBS-P+E (10 mM Hepes, pH 7.4, 0.15 M NaCl, 0.005% surfactant P20, 50 μM EDTA). The surface was prepared for immobilization of ACPdd by activating with 12 μL of 500 μM $NiCl_2$ in HBS-P. Both the loading concentration and contact time were empirically determined for each ACPdd so that the maximum amount of protein was immobilized on the chip and that this protein was stably bound for the course of the experiment. ACPdd concentrations used for loading varied depending on the protein between 50 nM – 1 μM. Typically, 700 – 1500 RU of ACPdd was bound to the Ni-NTA sensor chip for each experiment. To measure binding to ACPdd by SPR, solutions of KSdd in HBS-P+E were injected over the prepared surface as well as a nickel only flow cell at a flow rate of 10 μL/min. After multiple injections (8 – 10 concentrations), the surface was regenerated using 30 μL of 175 mM EDTA in HBS-P, pH 8.3. Maximum testable concentrations for the KSdds were limited by either the solubility of the peptide or its level of nonspecific binding to the nickel-only control lane. Kinetic data analysis was carried out using Scrubber2 (BioLogic Software) and BIAevaluation (GE Healthcare Life Sciences). Nonlinear curve fitting of the equilibrium binding response was carried out using GraphPad Prism software.

2.4.5 Fluorescence Polarization Assays

Labeled ACPdds were generated by reaction of BODIPY® FL C_1-IA (N-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-yl)methyl)iodoacetamide) (Invitrogen) with cysteine-containing ACPdds. Briefly, 4 µL of 100 mM TCEP in water and 40 µL of 10 mM BODIPY® FL C_1-IA in DMSO was added to 360 µL of 500 µM ACPdd in HBS. Reactions were protected from light and proceeded for 2 hours at room tempetature. Unreacted BODIPY® FL C_1-IA was removed from the labeled protein by passing the mixture over a preequilibrated Zeba spin desalting column (Pierce) and dialyzing into HBS. FP assays were performed at 20 µL total volume in a low volume black opaque polystyrene plate (Matrix Technologies). Proteins (50 nM PikAIII ACPdd-FL tracer and varying concentrations of unlabeled KSdd-containing PikAIV constructs) were allowed to incubate together for 10 minutes at room temperature in HBS-P (10 mM Hepes, pH 7.4, 0.15 M NaCl, 0.005% surfactant P20). Fluorescence polarization measurements were made at high sensitivity setting on a SpectraMax M5 (Molecular Devices) using 485 nm excitation, 538 nm emission, and 530 nm cutoff filter. The G factor was determined experimentally by setting a standard of 50 nM fluorescein in 0.1 N NaOH to 20 mP. Nonlinear curve fitting of the equilibrium binding response was carried out using GraphPad Prism software. Control experiments using up to 1 mg/ml BSA confirmed that the polarization increase upon incubation of PikAIII ACPdd-FL with unlabeled PikAIV KSdd was due to a specific protein-protein interaction (data not shown).

2.4.6 Crystallization, Data Collection and Structure Determination

Initial screening with P3P4dock produced small crystals of needle morphology under a variety of conditions containing high concentrations of organic solvents such as dioxane and 2-methyl-2,4-pentanediol (MPD) (Figure 2-14). The best-diffracting native crystals grew in 4-8 weeks at 4 °C using hanging-drop vapor diffusion techniques. Similarly, selenomethionyl P3P4dock crystals grew in 1-2 weeks at 4 °C after microseeding with native crystals. For crystal growth, an equal volume of protein solution (2.5 - 5 mg/ml) in HBS (20 mM HEPES, pH 7.8, 150 mM NaCl) was mixed with mother liquor containing 55% MPD, 150 – 200 mM sodium acetate, pH 5.0. The crystals were harvested in loops and frozen in liquid N_2.

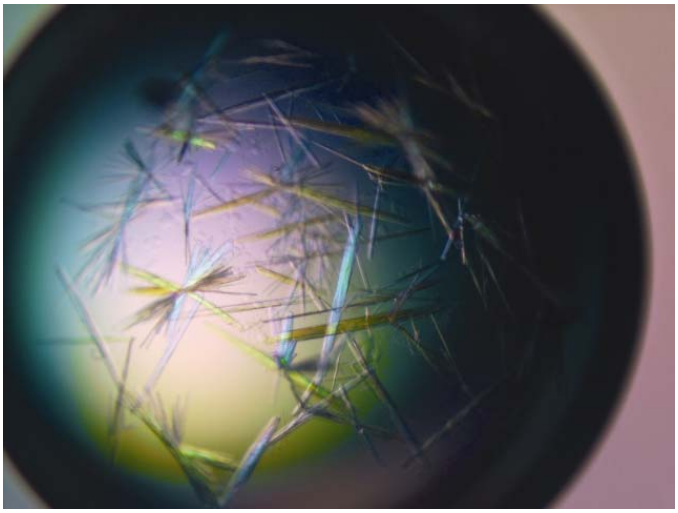


Figure 2-14. P3P4dock crystals. 2.5 mg/ml selenomethionyl P3P4dock protein in 20 mM HEPES, pH 7.8, 150 mM NaCl. Crystals seeded with a 1:1000 dilution from a 5 mg/ml native protein well. Crystals are shown from a well containing 55% MPD and 150 mM sodium acetate.

Diffraction data were collected at 100 K on GM/CA-CAT beamlines 23ID-B and 23ID-D at the Advanced Photon Source in the Argonne National Laboratory (Argonne, IL). The data were processed using the HKL2000 suite (104). Initial phasing by the single-wavelength anomalous diffraction (SAD) method was performed using data collected at the wavelength with strongest anomalous signal from a single

selenomethionyl-labeled protein crystal (Table 2-2). To minimize radiation damage, the dataset was assembled from 45° wedges of data collected from multiple points along a single crystal using a 10-micron X-ray beam (105). The PHENIX software package (106) located five of the six selenium atoms and approximately two-thirds of the structure was automatically built from the 3.0 Å SAD-phased map. Two molecules were present in the asymmetric unit (Vm = 2.40, 49% solvent). Modeling was completed manually using COOT (107). The model was refined against the 1.75 Å native dataset using REFMAC5 of the CCP4 suite (108-110). (Table 2-3). The atomic coordinates of the PikAIII / PikAIV complex have been made publicly available through the Protein Data Bank (www.rcsb.org/pdb) with the PDB ID 3F5H.

Table 2-2. P3P4dock diffraction data

Parameter	Native	SeMet
Space Group	$C222_1$	$C222_1$
Dimensions (Å) *a,b,c*	59.0, 117.9, 41.8	59.7, 118.5, 41.9
X-ray source	23ID-D	23ID-D
Wavelength λ (Å)	0.97934	0.97940
d_{min}(Å) [a]	1.75 (1.81-1.75)	2.80 (2.90-2.80)
Unique observations	15,084	3,917
R_{merge} (%) [a,b]	6.9 (50.7)	12.3 (29.7)
$\langle I/\sigma \rangle$ [a]	15.6 (2.1)	11.5 (4.3)
Completeness (%) [a]	99.2 (98.0)	100 (100)
Avg. redundancy [a]	3.6 (3.1)	5.1 (5.1)

[a] Values in parenthesis are for outer shell
[b] $R_{merge} = \Sigma|I_i - \langle I \rangle|/\Sigma I_i$, where I_i is the intensity of the *i*th observation and $\langle I \rangle$ is the mean intensity

Table 2-3. P3P4dock refinement statistics

	P3P4dock
Date range	50-1.75
R/R_{free} [a,b]	0.201/0.250
RMSD bond length (Å)	0.011
RMSD bond angle (°)	1.216
Avg. Protein B-factor (Å^2)	24.4
Avg. Solvent B-factor (Å^2)	39.5
Wilson B (Å^2)	20.3
Ramachandran plot[c]	
Favored	100
Allowed	0.0
Disallowed	0.0
Protein atoms	919
Water molecule	151
Other atoms	1

[a] $R = \Sigma|F_o-|F_c||/\Sigma|F_o|$ where F_o is the observed structure factor and F_c is the calculated structure factor used in the refinement
[b] $R_{free} = \Sigma|F_o-|F_c||/\Sigma|F_o|$ where F_o is the observed structure factor and F_c is the calculated structure factor from 5% of reflections not used in the refinement
[c] From output of MOLProbity

2.4.7 Sequence and Structure Analysis.

Multiple sequence alignments were performed using the ClustalX method within JalView software (111). Structural figures were generated with PyMOL (DeLano Scientific).

Notes:

This work has been published as "Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains," Buchholz, T. J., Geders, T. W., Bartley 3rd, F. E., Reynolds, K. A., Smith, J. L. and Sherman, D. H. ***ACS Chem. Biol.*** 2009, 4 (1) 41-52.

A*uthor contributions:*
Tonia Buchholz, Janet Smith, Kevin Reynolds and David Sherman designed the experiments;
Tonia Buchholz conducted the protein expression, purification, binding experiments and protein crystallization;

Frank Bartley, III and Todd Geders set and harvested crystals;
Todd Geders solved the x-ray crystal structure;
Tonia Buchholz, Todd Geders, David Sherman and Janet Smith analyzed the x-ray crystal structure.

Christopher M. Rath provided mass spectrometry analysis of KSdds. BAP1 *E. coli* cells were a generous gift from C. Khosla (Stanford University).

GM/CA CAT has been funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract DE-AC02-06CH11357.

Chapter 3

BryR, an HMG-ACP Synthase with Specificity for HMG-CoA Synthase Cassette ACPs

3.1 Introduction

The wide distribution of PKSs in the microbial world and the extreme chemical diversity of their products result from a varied use of the well known catalytic domains described above (section 1.1) for the canonical type I PKS systems. Taking a theoretical view of polyketide diversity, the Hatzimanikatis and Broadbelt groups have suggested that even if the starter and extender units are fixed, over 100,000 linear heptaketide structures are possible using only the 5 common reductive outcomes at the β-carbon position (ketone, (*R*- or *S*-) alcohol, *trans*-double bond, or alkane) (112). Recently it has become apparent that even this does not represent the upper limit for polyketide diversification. In order to create chemical functionalities beyond those mentioned above, nature has modified some enzymes from sources other than fatty acid synthesis (the mevalonate pathway in primary metabolism is one example) not typically thought of as type I PKS domains. Presented here is one way PKS containing systems have modified these domains for the catalysis of some unique chemistries observed in their natural products.

3.1.1 Bryostatins

The bryostatins are antifeedant polyketide natural products putatively produced by a bacterial symbiont of the marine bryozoan *Bugula neritina* (74). They are highly potent protein kinase C (PKC) modulators (113), and, as such, bryostatin 1 (Figure 3-1, **18**) has been tested in numerous clinical trials as a potential anticancer agent (114). Separately, the neuroprotective activity of PKC activators has recently been demonstrated in preclinical studies where bryostatin 1 was able to rescue memory loss after postischemic stroke (115). Additional studies suggest that bryostatin 1 (and a synthetic analog) may be able to reduce the levels of Aβ, a toxic peptide implicated in Alzheimer's disease (116, 117). However, like many marine-derived natural products, fulfilling the promise of these initial studies may be hindered by the low abundance of bryostatins available from either natural sources or chemical synthesis (118). The intriguing biological activities and lingering supply questions motivate our continued study of the bryostatin biosynthetic pathway (Figure 3-2). Increasing our knowledge of the molecular mechanisms employed may help open the door to new methods of bryostatin production as well as the generation of related bryostatin analogs.

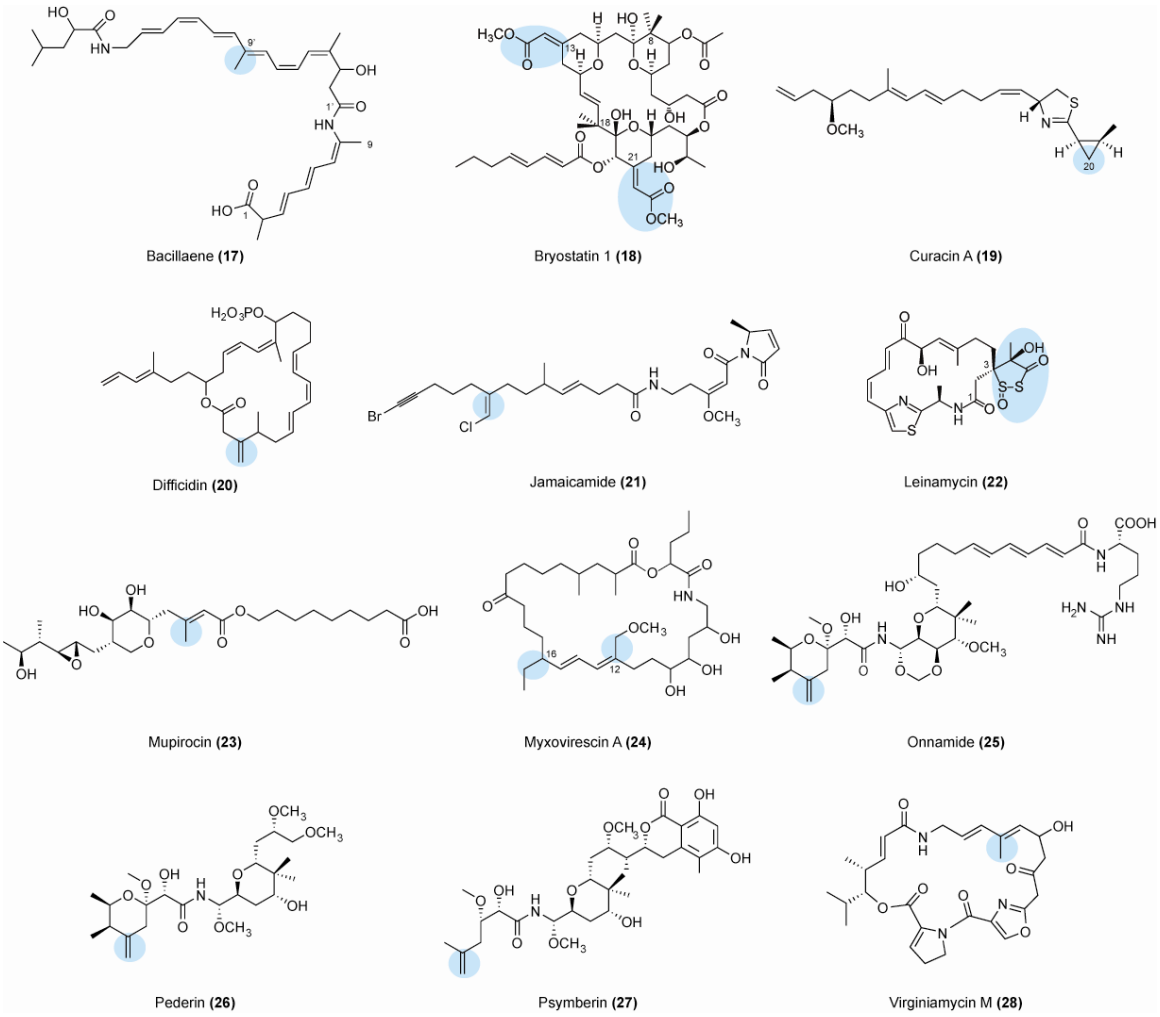


Figure 3-1. Products from HMGS-containing biosynthetic pathways. Positions of β-branching are shaded blue.

3.1.2 Methylation at the α- and β-Carbons

As described in section 1.2.1 for type I polyketide biosynthesis, the presence or absence of a methyl group on the α-carbon position of the growing polyketide chain is most often governed by the selection of the extender unit (malonyl-CoA versus methylmalonyl-CoA). However, in PKS systems that use *trans* acyltransferases (AT-less type I PKSs) (31, 119), the module by module control over extender unit selection is

sometimes not possible. In most of these cases, malonyl-CoA is used as the extender unit, and a methyl group can be added to selected positions through the action of embedded methyl transferase (MT) domain or discrete MT enzyme. For example, the C6 methyl group of leinamycin **22** is thought to be installed by the MT embedded in LnmJ (120), and the gem dimethyl groups on C8 and C18 of bryostatin **18**, most likely are the consequence of the MT domains in BryB and BryC (74).

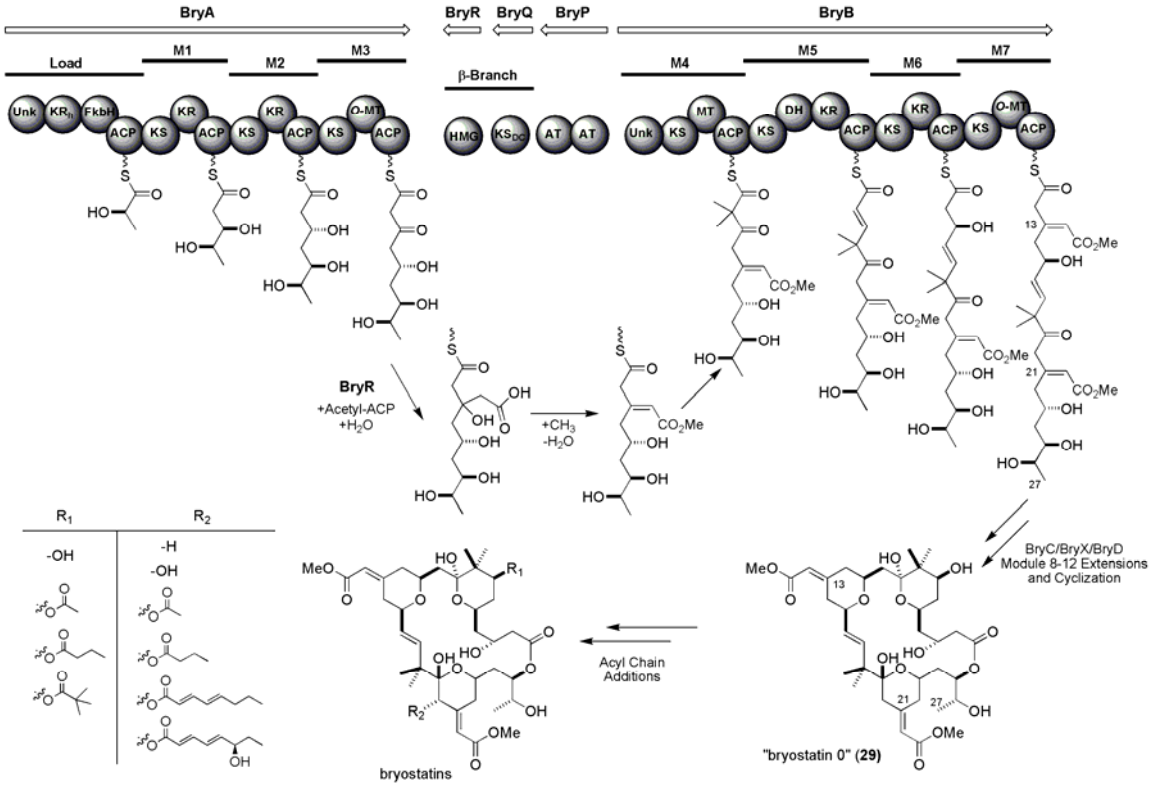


Figure 3-2. *bry* gene cluster. Portions of the pathway utilized in β-branching are highlighted in this depiction of the bryostatin biosynthetic pathway. BryC, BryX and BryD are not shown. ACP, acyl carrier protein; AT, acyltransferase; DH, dehydratase; FkbH, homolog to FkbH (121); HMGS, HMG-CoA synthase homolog; KR, ketoreductase; KS, ketosynthase; KS_{DC}, decarboxylative ketosynthase; MT, methyltransferase; Unk, domain with unknown function.

In contrast to the α-carbon methylations, the incorporation of methyl or methylene groups (or functional groups derived from such groups) at the β-position represents the assimilation of a full suite of enzymes into the typical PKS machinery. A

subset of type I modular PKSs (and hybrid NRPS/PKS megasynthases) recently have been identified that contain multiple enzymes acting in *trans* during the traditional linear assembly-line process. These systems provide a unique method of expanding the repertoire of the traditional reductive domains (KR, DH, ER). Significant genetic and biochemical evidence has been accumulated to demonstrate that β position alkyl side-chains are typically introduced through an HMG-CoA synthase (HMGS) cassette of enzymes/domains performing chemistry similar to that seen in mevalonate biosynthesis. These enzymes work in conjunction with the PKS machinery to create unique functionalities seen at the branch points that include the pendant methyl groups of bacillaene **17** (122-124)**,** mupirocin **23** (125), and virginiamycin M **28** (126), the methoxymethyl and ethyl groups of myxovirescin A **24** (127-129), the exomethylene groups of difficidin **20** (124), onnamide **25** (130), pederin **26** (130-132), and psymberin **27** (133), the cyclopropyl ring of curacin A **19** (54, 134, 135), the vinyl chloride of jamaicamide **21** (136), the unique 1,3-dioxo-1,2-dithiolane moiety of leinamycin **22** (120, 137), and the exocyclic olefins in bryostatin **18** (74) (Figure 3-2).

3.1.3 HMG-CoA Synthase Cassettes

In primary metabolism, HMG-CoA synthase (HMGS) is responsible for the condensation of C2 of acetyl-CoA onto the β-ketone of acetoacetyl-CoA to form 3-hydroxyl-3-methylglutaryl-CoA and free CoASH (138) (Figure 3-3A). A number of secondary metabolite pathways have been identified over the past 5 years that perform an analogous reaction, although they appear to use ACP-tethered acyl groups as opposed to acyl-CoA substrates. After generation of the HMG-ACP analog on the growing polyketide chain, the product is usually dehydrated and decarboxylated to yield the

branched intermediate. Found in 12 pathways to date (54, 74, 120, 122-134, 136, 137) (Table 3-1 and Appendix B), included in the cassette are a discrete donor ACP (ACP_D), a decarboxylative KS (cysteine to serine active site variant, KS_{DC}), an HMGS homolog, one or two enoyl CoA hydratase-like (ECH) enzymes/domains, and an acceptor ACP site (ACP_A) (Figure 3-3B). Often, the pathways contain tandem ACP_As at the site of modification, though the purpose of these ACP repeats is not clear. Additionally, acyltransferase activity is needed, though this may come from a variety of sources.

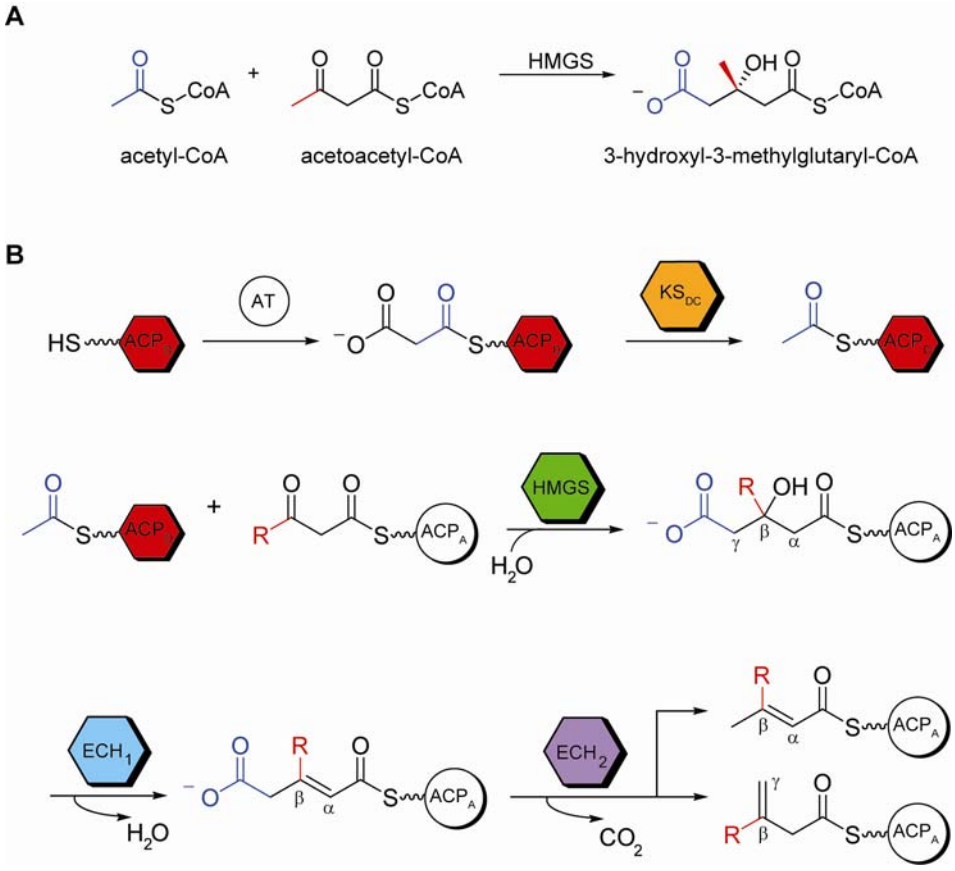


Figure 3-3. HMG generation in primary and secondary metabolism A) the mevalonate pathway and B) the full HMGS cassettes found in PKS and mixed biosynthetic pathways incorporating β branches.

Three HMGS-containing cassettes (those in the curacin A, bacillaene and myxovirescin pathways) have been biochemically validated in the past three years and will serve as the basis to analyze the individual components in this complex (54, 123,

135, 139-141). The mechanistic and structural details for HMG-CoA synthase in primary metabolism have been elucidated for both bacterial and eukaryotic HMGSs (142-146). While polyketide HMGSs share only 20-30% sequence identity with their primary metabolism homologs (in both prokaryotes and eukaryotes), multiple sequence alignment reveals that the key catalytic residues (Glu/Cys/His) are conserved. As shown in Figure 3-3, the first step in the formation of the HMG-intermediate is the generation of acetyl-ACP_D. This is typically accomplished through the loading of malonyl-CoA via either an embedded or discrete AT (123, 139, 147). The decarboxylative KS then converts the malonyl-ACP into acetyl-ACP, after which the tethered acetyl group is condensed onto the β-ketone of the polyketide intermediate. However, a unique enzyme with dual acyltransferase/decarboxylase activity was recently characterized from the leinamycin pathway, LnmK (137). LnmK is responsible for generation of a propionyl-ACP_D. A homolog of this enzyme (TaD) has been found in the pathway of myxovirescin, an ethyl-branch-bearing compound (129). Finally, formation of the HMG-analog is completed upon addition of water.

Table 3-1. Known HMGS cassettes and their producing organisms

Natural Product	Producing Organism	Discrete ACP	KS (Cys→Ser)	HMGS	ECH_1 (dehydration)	ECH_2 (decarboxylation)
Bacillaene	*Bacillus subtilis* 168 / *B. amyloiquefaciens*	AcpK/BaeF	PksF	PksG/BaeG	PksH/BaeH	PksI/BaeI
Bryostatin	*Candidatus* Endobugula sertula		BryQ	BryR		
Curacin	*Lyngbya majuscula*	CurB	CurC	CurD	CurE	CurF N-terminal domain
Difficidin	*Bacillus amyloiquefaciens*	DifC		DifN		DifO
Jamaicamide	*Lyngbya majuscula* JHB	JamF	JamG	JamH	JamI	JamJ N-terminal domain
Leinamycin	*Streptomyces atroolivaceus*	LnmL		LnmM	LnmF	
Mupirocin	*Pseudomonas fluorescens*	Macp14	MupG	MupH	MupJ	MupK
Myxovirescin A (antibiotic TA)	*Myxococcus xanthus* DK1622	TaB & TaE	TaK	TaC & TaF	TaX	TaY
Onnamide	Symbiont bacterium of *Theonella swinhoei*			OnnA		OnnB (embedded)
Pederin	Symbiont bacterium of *Paederus fuscipes*	PedN	PedM	PedP	PedL	PedI (embedded)
Psymberin	Symbiont bacterium of *Psammocinia* aff. *bulbosa*	PsyL	PsyM	PsyI	PsyA (embedded)	PsyA (embedded)
Virginiamycin M	*Streptomyces virginiae*		VirB	VirC	VirD	VirE

Processing of the HMG-intermediate can vary considerably, but typically proceeds via dehydration and decarboxylation catalyzed by two enoyl-CoA reductase domains (Figure 3-3B). Based on sequence similarity, the members of the crotonase fold family seen in these HMGS cassettes can be subdivided into two groups, termed ECH_1 and ECH_2 (54, 141). The successive dehydration and decarboxylation steps are catalyzed by the ECH_1 and ECH_2 enzymes/domains, respectively. Evidence for the specific function of the curacin ECH_1 and ECH_2 enzyme pair from the curacin pathway has been demonstrated using a coupled enzyme assay, ESI-FT-ICR MS, and x-ray crystallography (54, 141). Using purified ECH_1 (CurE) and ECH_2 (the N-terminal domain of CurF) overexpressed in *E. coli*, (*S*)-HMG-ACP was converted first to 3-methylglutaconyl-ACP then to 3-methylcrotonyl-ACP, the proposed intermediate for subsequent formation of the cyclopropyl ring. Further *in vitro* evidence for the function of these enzymes has been

generated using proteins from the PksX pathway of *Bacillus subtilis* (123) and the myxovirescin pathway from *Myxococcus xanthus* (139). Before the identification of bacillaene as the product of the PksX pathway, Calderone *et al.* reported the function of a number of its discrete enzymes (123, 140). Using radioactive biochemical assays together with mass spectrometry, they were able to assign functional roles to AcpK, PksC, the tandem ACPs in PksL, PksF, PksG, PksH and PksI. Using the model acceptor ACP, acetoacetyl-ACP, and malonyl-CoA in combination with the above proteins, a Δ^2-isoprenyl-*S*-carrier protein was generated (123). More recently, a similar *in vitro* investigation was carried out using the homologous enzymes from the myxovirescin pathway (139). The HMGS cassette logic proposed above held fast for the myxovirescin pathway. Though the generation of the propionyl- or methylmalonyl-*S*-ACP could not be demonstrated, by analogy to the leinamycin pathway, TaD is the likely missing piece (137).

In vivo evidence for the function of these HMGS cassettes has recently come from the Müller lab (127, 128). Both of the HMGS homologs in the myxovirescin A pathways (TaC and TaF) were individually deleted and the impact on the products of the engineered *Myxococcus xanthus* strains were analyzed. Though masses of predicted products were not detectable in the *ΔtaC* strain, disruption of TaF led to the production of a myxovirescin analog with a methyl group in place of the ethyl group at C16 of **24**, presumably via compensation by TaC.

Analysis of the placement of the known HMGS cassettes identified to date into their biosynthetic clusters reveals a variety of possible architectures (Appendix B). For example, the ECH_2 can exist as a discrete enzyme downstream of the ECH_1 (mupirocin

and others), as an N-terminal domain of a large PKS (curacin and jamaicamide), and as an embedded domain (pederin and onnamide). While most of the clusters published to date are mixed PKS/NRPS systems with *in trans* ATs and tandem ACPs at the site of HMGS modification, exceptions exist for each of these examples (difficidin is PKS only, curacin and jamaicamide contain embedded ATs, and bryostatin and myxovirescin do not contain tandem ACPs at the site of HMGS modification).

Partial HMGS cassettes have been identified in the onnamide, difficidin, psymberin, leinamycin (120, 124, 130, 133, 137) and bryostatin (missing $ACP_{D,}$ ECH_1 & ECH_2) (74) biosynthetic pathways (Table 3-1). Lack of complete gene cluster sequencing or annotation is one possible explanation for the presence of a partial HMGS cassette. This is likely the case for the onnamide, virginiamycin, difficidin and bryostatin systems where either firm pathway boundaries have yet to be determined for contiguous pathways or the pathway is possibly dispersed over multiple chromosomal loci. In other instances (leinamycin, bryostatin), product formation is unlikely to require enzymatic transformations performed by the ECH homologs (dehydration and decarboxylation). Alternately, functions typically performed by the cassette members may be carried out by other domains/enzymes within the pathway. For example, the leinamycin pathway does not need a KS_{DC} enzyme, as LnmK fulfills this as one of its roles as an acyltransferase/decarboxylase to generate the acyl donor propionyl-LnmL (137). Similarly, the β-methoxylacylidene moieties found in the bryostatins are hypothesized to be the result of a β-γ dehydration (whereas the dehydration performed by ECH_1 enzymes typically occurs at the α-β positions) (147). Possible candidate domains for these transformations are the N-terminal domains of unknown function found on BryB and

BryC, immediately downstream of both HMGS modification sites in the bryostatin pathway (74).

3.1.4 Protein-Protein Specificity in HMG-CoA Synthase Cassettes

As HMGS enzyme cassettes have only been identified and functionally characterized recently, some of the mechanistic details as well as many of the key protein-protein interactions needed to orchestrate communication among the polypeptide components remain unclear. Details on how the individual proteins are brought to the correct place in the pathway to perform their functions are still unknown for the majority of the pathways. In the case of the PksX/bacillaene pathway, some intriguing microscopy performed on *B. subtilis* suggests that the bacillaene proteins are clustered into a huge mega-enzyme factory inside the bacterial cell (99). Whether this organization extends (or is limited) to the other members of HMGS cassette containing pathways remains to be seen.

3.2 Results and Discussion

The type I PKS biosynthetic gene cluster presumed responsible for the synthesis of the bryostatins, *bry*, has been identified and sequenced from two sibling species of "*Ca.* Endobugula sertula/*B. neritina*" (74). The shallow-water North Carolina (NC) sibling species appears to be located within a contiguous DNA fragment approximately 77 kb in length, whereas the deep-water California (CA) species is split between two or more locations on the chromosome. Apart from the transposition of the HMGS cassette and AT enzymes, the two sequences exhibit >99.5% sequence identity at the DNA level.

For all of the following studies, we have used BryR from the presumed NC species of "*Ca.* Endobugula sertula/*B. neritina*".

3.2.1 BryR -ACP Binding Determined by Surface Plasmon Resonance

A key step for selectivity in the HMGS cassette appears to be the HMGS reaction itself. Biochemical studies of PksG, the HMGS homolog of the bacillaene pathway, revealed that the enzyme only accepts the acetyl group when presented on AcpK, its cognate ACP_D (123). In addition, gene deletion studies of the myxovirescin HMGS cassette enzymes indicate that the two HMGS homologs present (TaC/TaF) utilize separate ACP_Ds (TaB/TaE) (128). The ability of PksG to accept a model substrate, ACP_A-bound acetoacetyl (Acac), is consistent with the importance of protein-protein interactions for HMGS-ACP_D specificity (123).

By analogy to the previously characterized secondary metabolite HMGS homologs PksG and TaC (123, 139), the discrete HMGS in the bryostatin pathway, BryR, is likely to be involved in the β-branching at C13 and C21 of the bryostatins (Figures 3-1 and 3-2). To date, no discrete ACP_D for the HMGS cassette of the bryostatin pathway has been identified in either the CA or NC *bry* cluster sequences. It has been suggested that BryR, like its primary metabolism counterparts, may be able to use acetyl-CoA as the acyl donor in its reaction (147). However, the presence of a KS-type (BryQ) decarboxylase, whose putative role is to generate acetyl-ACP from malonyl-ACP, makes this an unlikely scenario. In the absence of a Bry ACP_D, we sought to identify surrogate acetyl donors for substrate loading of BryR.

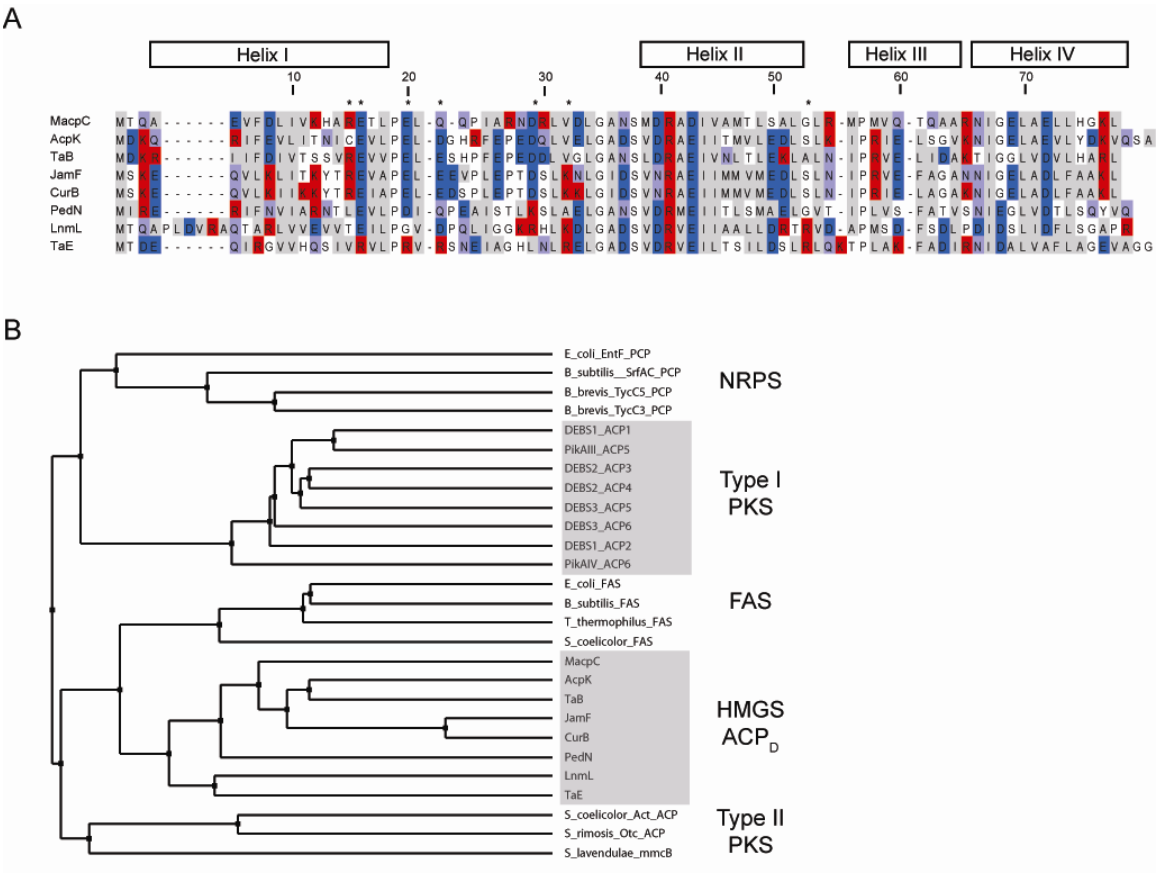


Figure 3-4. The HMGS cassette ACP_D subclass of acyl carrier proteins. Both the A) multiple sequence alignment (Clustal) and B) phylogenetic tree (average distance BIOSUM2) were generated using Jalview software (148). Numbering is based on MacpC. Helix designations are predicted from alignment with DEBS1_ACP2 structure (149). Basic residues are colored in red, acidic in blue, and hydrophobic in grey.

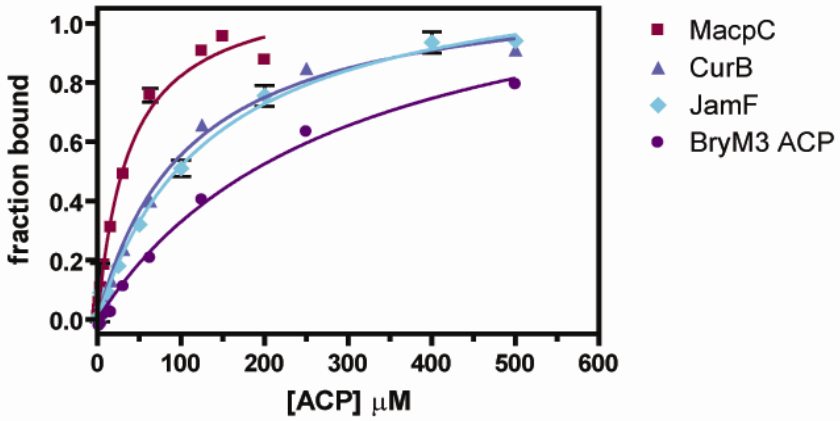
Several types of ACPs were surveyed (discrete ACP_Ds from HMGS-cassettes, type II PKSs and bacterial FASs, and excised ACPs from type I PKSs) in search of suitable ACP_D partners for BryR (Figure 3-4). The direct binding of BryR to a variety of potential acetyl-ACP donors (Ac-ACP_Ds) and a model acetoacetyl acceptor substrate, Acac-BryM3 ACP, an embedded ACP excised from the BryA tetramodule at one of the predicted HMGS modification sites (Figure 3-2) was evaluated by surface plasmon resonance (SPR) (Figure 3-5). The unmodified (apo-), phosphopantetheine (PPant)-containing (holo-) and Ac- or Acac-loaded forms of the ACPs were accessed by overexpression and purification in *E. coli* followed by loading with either Sfp or Svp

(flexible phosphopantetheinyl transferases (PPTases) from surfactin and bleomycin biosynthesis) (65, 67). Modified ACPs were separated from unreacted CoAs before testing.

After BryR immobilization to a BIAcore CM5 SPR ship using standard amine coupling chemistry (Figure 3-6), equilibrium binding analysis was performed using sequential injections of apo-, holo-, Ac-, or Acac-ACPs at varying concentrations (Figure 3-5 and 3-7). Wild-type (WT) and an active site BryR mutant (C114A) behaved similarly in our binding studies. BryR was able to bind to ACPs from the curacin **19** (CurB), jamaicamide **21** (JamF), and mupirocin **23** (MacpC) HMGS cassettes as well as to the excised native acceptor (BryM3 ACP). Affinities (K_Ds) were in the middle to high micromolar range for ACP_Ds (40 – 110 μM) and the ACP_A (180 μM). A discrete ACP upstream of the *bry* gene cluster was found adjacent to genes that encode proteins resembling components of fatty acid synthases (FAS). Though other FAS ACPs have not been reported as part of HMGS cassettes, no other endogenous ACP_D candidates were evident in or near the *bry* cluster. Referred to here as Bry FAS ACP, no significant binding was observed between it and BryR at up to 500 μM. A related FAS ACP from the well-studied *Streptomyces coelicolor*, (SCO2389, *Sc* FAS ACP) (150), also failed to interact with BryR at 650 μM (Figure 3-5B). The affinity of BryR for the ACPs seems to be mediated mainly by protein-protein contacts (as opposed to protein-acyl chain or protein-PPant contacts). No enhancement of affinity was observed between apo- and Ac-ACP_D or between apo-, holo-, Ac-, Acac-, or HMG-ACP_A (Figure 3-5B). These data also suggest that specificity for a protein-bound acyl group is a distinguishing feature between

HMGS homologs found in PKS or mixed PKS/non-ribosomal peptide synthase (NRPS) biosynthetic pathways and those of primary metabolism.

A



B

		apo-	holo-	Ac-	Acac-	HMG-
ACP_D	MacpC	40 ± 5 μM		32 ± 9 μM		
	CurB	108 ± 7 μM		90 ± 8 μM		
	JamF	100 ± 20 μM		107 ± 7 μM		
ACP_A	BryM3 ACP	177 ± 7 μM	180 ± 24 μM	200 ± 12 μM	150 ± 51 μM	170 ± 13 μM
	Bry FAS ACP	No Binding				
	Sc FAS ACP	No Binding				

Figure 3-5. Binding of apo-ACPs to immobilized BryR, monitored by SPR. A) Each data point is the average of triplicate measurements; error bars are standard deviation. The data were fit to a one-site binding model ($Y = B_{max}*X/(K_D + X)$). Y = fraction bound, B_{max} = maximal response, X = ACP concentration. B) Dissociation constants (K_Ds) table.

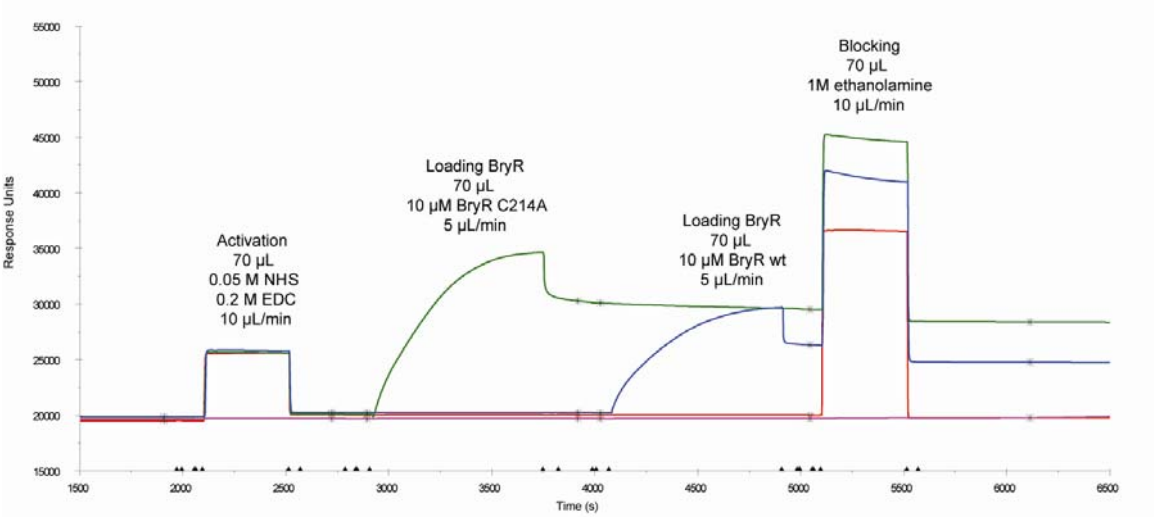


Figure 3-6. BryR immobilization sensorgram. Raw sensorgram data from BIACORE 3000 Control software. FC1 is shown in red, FC2 in blue, FC3 in green and FC4 in pink.

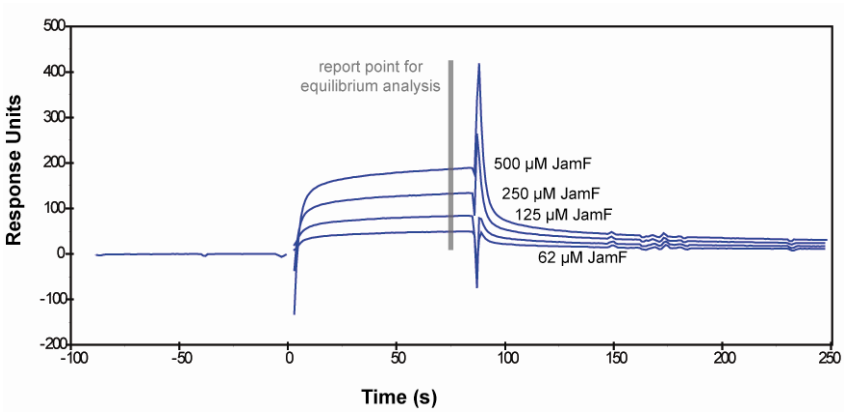


Figure 3-7. Subtracted BIAcore data for JamF:BryR binding. Data analyzed with BIAevaluation software. Report point was set at 75 seconds after the ACP injection.

3.2.2 Probing the HMGS - ACP_D Interface

To date, no structures have been reported for members of the secondary metabolism class of HMGS enzymes or the small set of discrete donor ACPs (Figure 3-4 and Appendix B). Sequence analysis of the discrete ACPs found in related HMGS cassettes suggested potential key regions flanking the PPant modification site that may present a unique ACP binding interface. Specifically, we focused on those positions that

differed between TaB and TaE, two ACP_D's found in a single biosynthetic pathway with orthogonal HMGS partners (128). Based on homology to ACPs of known structure, the chosen mutations are predicted to lie in the loop between helix 1 and helix 2 (R15A, E16R, E20A, Q22A, Q22E, Q22R, D29A, V32R) or the end of helix 2 (G53R) (18, 58). (Figure 3-4). Mutations were engineered into the mupirocin ACP_D, MacpC, as this protein exhibited the highest affinity for BryR in our testing set. Of the nine over-expressed mutant proteins, we were only able to obtain full equilibrium binding curves for three (E20A, Q22E, Q22R) of them due to limited solubility in the SPR assay conditions. Affinities of BryR for the the MacpC proteins were within 2-fold of WT MacpC (Figure 3-8). We were unable to draw significant conclusions on the residues likely to be involved in the BryR:MacpC interface based on such a limited data set.

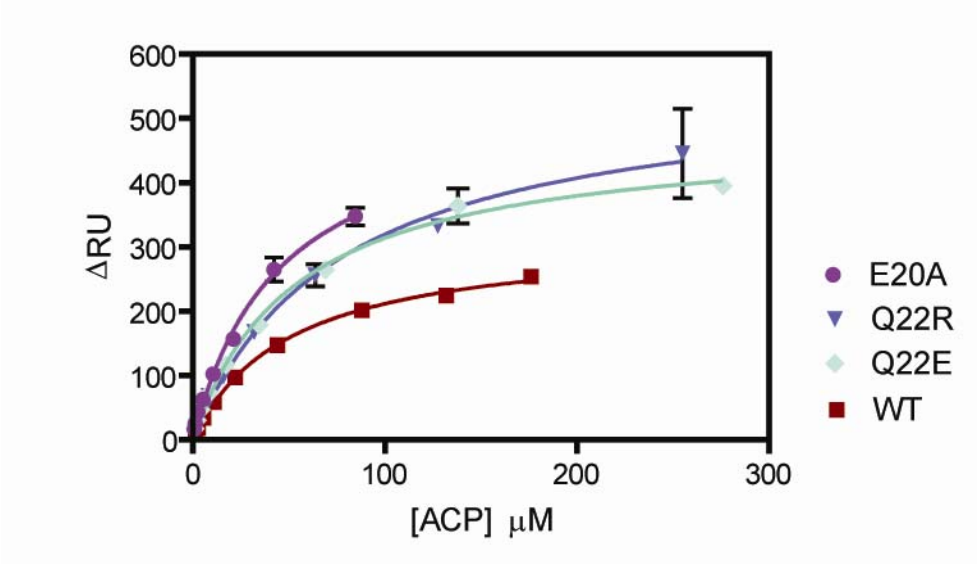


Figure 3-8. Binding of testable MacpC mutant proteins to immobilized BryR, monitored by SPR. Each data point is the average of triplicate measurements; error bars are standard deviation. The data were fit to a one-site binding model ($Y = B_{max}*X/(K_D + X)$). Y = fraction bound, B_{max} = maximal response, X = ACP concentration. Binding constants are reported in Table 1. K_Ds - E20A = 45 ± 6 μM / Q22R = 76 ± 14 μM / Q22E = 53 ± 5 μM.

Attempts to produce crystallographic quality polyketide HMGS homolog protein were made for both BryR, and CurD, the discrete HMGS of the curacin pathway. CurD

was targeted as our second HMGS to pursue based on the previous success solving protein structures for other domains from the curacin pathway such as the ECH_2 and $GNAT_L$ (141, 151). Both BryR and CurD were cloned into multiple vectors (pMCSG7, pMCSG9 and pMOCR) (100, 152) and their solubility was tested before and after tag removal via TEV protease. None of the over-expression constructs resulted in HMGS proteins with increased solubility.

3.2.3 BryR Enzymatic Activity

To assess the ability of BryR to catalyze HMG formation using the surrogate acetyl donors (Ac-ACP_D + Acac-ACP_A → HMG-ACP_A), the enzymatic activity of BryR when paired with different Ac-ACP_D and AcAc-ACP_A substrates was monitored. By analogy to primary metabolism HMGSs, the first step in the BryR enzyme mechanism should be acetylation of the active site cysteine in BryR (143, 144, 153). Subsequently, the C2 of acetate group reacts with the β-keto group of the acetoacetyl-ACP_A substrate to form HMG (or a related molecule during biosynthesis) (Figure 3-3). These steps were followed by both radio-SDS PAGE (Figure 3-9) and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) (Figure 3-10, 3-11, 3-12 and Table 3-2). Substrate transfer from acetyl-ACP (FTICR-MS) or [1-^{14}C]-acetyl-ACP (radio-SDS PAGE) to BryR was confirmed by both methods only when a member of the discrete HMGS-cassette ACP donor group (Figure 3-4B) was paired with wild-type BryR.

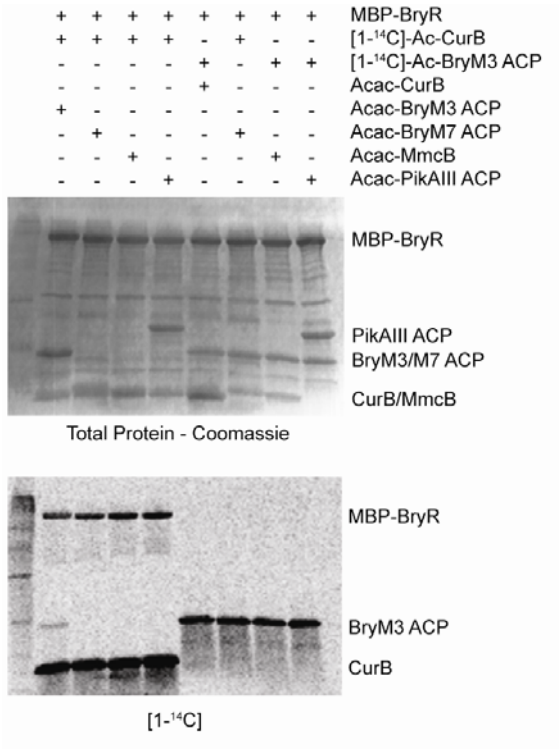


Figure 3-9. BryR-catalyzed formation of HMG-BryM3 ACP. Acetyl transfer from [1-^{14}C]-Ac-CurB to Acac-BryM3 ACP is shown in lane 1. All other lanes represent control reactions.

Generation of the Ac-BryR intermediate during the first half of the reaction can be visualized in the phosphorimage only when [1-^{14}C]-Ac-CurB donates the acetyl group (Figure 3-9). To verify that the BryR reaction does in fact proceed through the same enzyme intermediate as those observed in primary metabolism, we monitored the acetylation state of Cys114. BryR (10 μM) was reacted with Ac-MacpC (50 μM)in the absence of BryM3 ACP_A. After the sample was proteolyzed with trypsin, peptides were separated by HPLC, and using LC-FTICR MS and iontrap LC-MS/MS the BryR active site peptide was identified and acetylation of Cys114 was confirmed (Table 3-4). Additionally, a mutation at this location (C114A) was enzymatically inactive (data not shown). This is the first direct demonstration of the Ac-Cys species in an HMGS homolog in polyketide biosynthesis. By FTICR-MS, we were also able to observe the

loss of the Ac-ACP_D species and its conversion to holo-ACP_D in the presence of wild-type BryR (Figures 3-10, 3-11, 3-12).

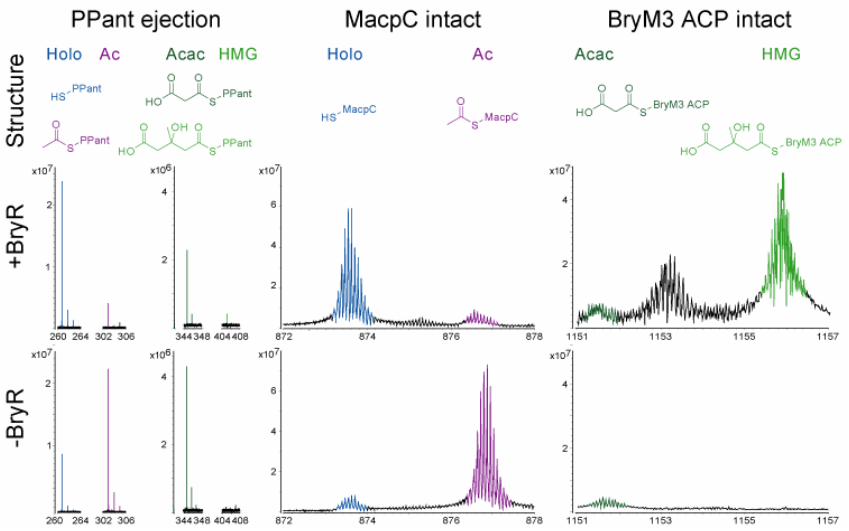


Figure 3-10. BryR catalyzed generation of HMG-BryM3 ACP from Ac-MacpC as monitored by FTICR-MS. Data are presented as m/z versus intensity. PPant ejection assay data from the entire charge state distribution are presented. As well, intact donor and acceptor ACP data are also illustrated.

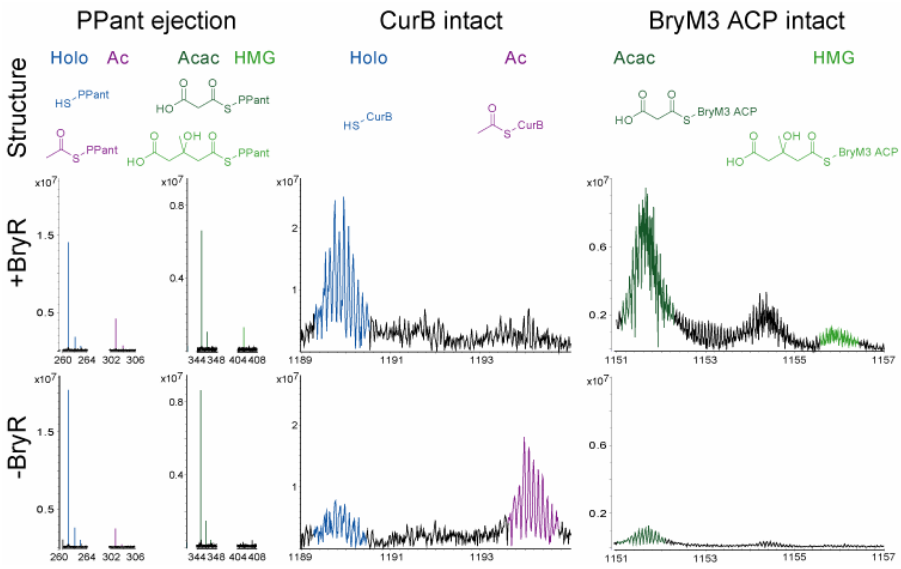


Figure 3-11. BryR catalyzed generation of HMG-BryM3 ACP from Ac-CurB as monitored by FTICR-MS. Data are presented as m/z versus intensity. PPant ejection assay data from the entire charge state distribution are presented. As well, intact donor and acceptor ACP data are also illustrated.

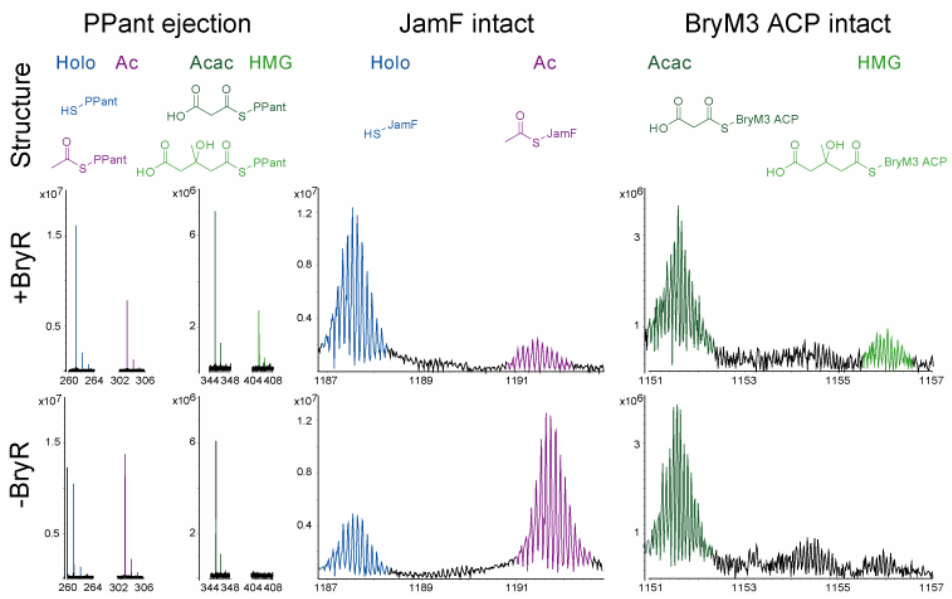


Figure 3-12. BryR catalyzed generation of HMG-BryM3 ACP from Ac-JamF as monitored by FTICR-MS. Data are presented as m/z versus intensity. PPant ejection assay data from the entire charge state distribution are presented. As well, intact donor and acceptor ACP data are also illustrated.

Table 3-2. Negative control reactions of ACP_D/ACP_A's with BryR HMGS monitored by FTICR-MS.

Donor	Acceptor	PPant ejection ion				Donor ACP		Acceptor ACP	
		Holo-	Ac-	Acac-	HMG-	Holo-	Ac-	Acac-	HMG-
Ac-Bry FAS ACP	Acac-Bry M3	=[1]	=	=	nd[2]	nd	nd	=	nd
Ac-CoA	Acac-Bry M3	=	=	=	nd	=	=	=	nd
none	Acac-Bry M3	=	nd	=	nd	=	=	=	nd
Ac-CurB ACP	Acac-*Sc* FAS ACP	=	nd	=	nd	=	=	=	nd
Ac-BryM3 ACP	Acac-MacpC	=	=	=	nd	=	=	=	nd
Ac-MacpC ACP	Acac-MacpC	=	=	=	nd	=	=	=	nd

[1] no change upon incubation with BryR
[2] none detected

As evidence of BryR's ability to catayze the complete reaction (Ac-ACP_D + Acac-ACP_A → HMG-ACP_A), we observed a third radioactive band, consistent with modification of BryM3 ACP (Figure 3-9). To identify the chemical modification on BryM3 ACP, the reaction mixtures were monitored by top-down FTICR-MS. The mass

shift of +60 Da on the intact BryM3 ACP between the +/- BryR samples is consistent with conversion of the Acac-BryM3 ACP_A substrate to HMG-BryM3 ACP_A (Figures 3-10, 3-11, 3-12). An MS/MS experiment was performed using the PPant ejection assay (140, 154). This confirmed that the mass shift between +/- BryR samples was due to a modification of the PPant prosthetic group. The pattern of BryR activity, the highest activity using MacpC, lesser activity with CurB and JamF, and no activity with other Ac-ACP_Ds, observed in our enzymatic activity assays was consistent with the protein-protein binding activity reported above (Figure 3-5B and Table 3-2). Finally, while it serves as the donor substrate for HMGSs from primary metabolism, acetyl-CoA was unable to load the BryR active site Cys (Table 3-2).

3.3 Summary

In summary, we have confirmed the enzymatic activity of BryR (condensation of acetyl-ACP_D with acetoacetyl-ACP_A to form HMG-ACP_A) using two complementary methods, radio-SDS PAGE and FTICR-MS. The activity of BryR was dependent on pairing of the native Acac-BryM3 acceptor ACP with an appropriate surrogate Ac-ACP_D from a related HMGS cassette (CurB, JamF, or MacpC). In addition, the ability of BryR to discriminate between various ACPs was assessed using an SPR-based protein-protein binding assay. BryR bound selectively to ACPs obtained from a series of HMGS cassettes (MacpC, CurB, JamF, and BryM3 ACP). Despite testing a small set of MacpC residues to identify key BryR/ACP_D contact points, no single amino acid residues were identified to control binding ability. Finally, this work, as well as other recent studies (75, 133) demonstrates further that natural product biosynthetic genes isolated from

uncultured endosymbiotic bacteria can be manipulated *in vitro* in order to probe the functionalities of these enzymes from previously inaccessible sources.

3.4 Experimental Methods

3.4.1 Expression and Purification of Proteins.

Plasmids for the expression of CurB, JamF, SCO2389 (*Sc* FAS ACP), Bry FAS ACP and BryR were generated by amplification using PCR with LIC overhangs and inserted into either the vector pMCSG7 (CurB, JamF and Bry FAS ACP) or pMCSG9 (100). CurB was amplified from plasmid pDHS2412 and JamF from Jamf:pET20. BryR was amplified from cosmid MM5 and Bry FAS ACP from cosmid MM7. PCR fragments were inserted into the vectors via ligation independent cloning. All mutants were generated according to the Quikchange® site-directed mutagenesis protocol (Stratagene/Agilent). All DNA sequences were confirmed by sequencing. The expression construct for MacpC (pGTB340) was a gift from Prof. Christopher M. Thomas. pDHS278 (BryM3 ACP in pMCSG7) was published previously (75). Primers used are listed in Table 3-3.

Table 3-3. Primer list for protein expression.

Primer name	Primer sequence	Plasmid(s)
CurBLICF	TACTTCCAATCCAATGCC atg agc aaa gaa caa gta cta	pDHS9780
CurBLICR	TTATCCACTTCCAATGCTA caa ttt tgc tgc aaa taa atc	
JamFLICF	TACTTCCAATCCAATGCC atg agc aaa gaa caa gta ctc aaa cta a	pDHS9781
JamFLICR	TTATCCACTTCCAATGCTA taa ttt cgc cgc aaa taa atc agc	
BryFAS_ ACPLICF	TACTTCCAATCCAATGCC agc aac cca agc aac act ga	pDHS9812
BryFAS_ ACPLICR	TTATCCACTTCCAATGCTA atc tag gtg tgc gtt gat gta at	
*Sc*FAS_ ACPLICF	TACTTCCAATCCAATGCC gcc act cag gaa ga	pDHS9758
*Sc*FAS_ ACPLICR	TTATCCACTTCCAATGCTA ggc ctg gtg gtc gag gat gta	
BryRLICF	TACTTCCAATCCAATGCC agg tat att ggt ata gaa tca at	pDHS279
BryRLICR	TTATCCACTTCCAATGCTA att gat cca ctg ata ttc tct atg	
MacpC_ R15AF	tcg tca agc atg ccg cgg aaa ccc tcc ctg	pDHS9801
MacpC_ R15AR	cag gga ggg ttt ccg cgg cat gct tga cga	
MacpC_ E16RF	gtc aag cat gcc cgc aga acc ctc cct gag ct	pDHS9803
MacpC_ E16RR	agc tca ggg agg gtt ctg cgg gca tgc ttg ac	
MacpC_ E20AF	gcg aaa ccc tcc ctg cgc ttc aac aac cca t	pDHS9794
MacpC_ E20AR	atg ggt tgt tga agc gca ggg agg gtt tcg c	
MacpC_ Q22AF	gaa acc ctc cct gag ctt gca caa ccc atc gcc	pDHS9796
MacpC_ Q22AR	ggc gat ggg ttg tgc aag ctc agg gag ggt ttc	
MacpC_ Q22EF	ccc tcc ctg agc ttg agc aac cca tcg ccc g	pDHS9802
MacpC_ Q22ER	cgc gcg atg ggt tgc tca agc tca ggg agg g	
MacpC_ Q22RF	cga aac cct ccc tga gct tag aca acc cat cgc cc	pDHS9798
MacpC_ Q22RR	ggg cga tgg gtt gtc taa gct cag gga ggg ttt cg	
MacpC_ D29AF	cgc ccg gaa cgc ccg cct ggt gg	pDHS9797
MacpC_ D29AR	cca cca ggc ggg cgt tcc ggg cg	
MacpC_ V32RFor	gaa cga ccg cct gag gga cct ggg cgc c	pDHS9804
MacpC_ V32ERev	ggc gcc cag gtc cct cag gcg gtc ggt c	
MacpC_ G53RFor	gac ttt gag cgc cct tag gtt gcg cat gcc	pDHS9806
MacpC_ G53RRev	ggc atg cgc aac cta agg gcg ctc aaa gtc	

Plasmids encoding N-terminal His_6- or His_6/MBP- fusion protein tags were transformed into *E. coli* BL21(DE3) and grown at 37 °C (ACPs) or 30 °C (BryR) in TB medium to an OD_{600} of ~1.0 in 2L flasks. The cultures were cooled to 18 °C, and

isopropyl β-D-thiogalactopyranoside was added to a final concentration of 0.2 mM and grown 12-16 h with shaking. The cells were harvested by centrifugation and frozen at -20 °C. Cell pellets were thawed to 4 °C and resuspended in 5X volume of lysis buffer (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 1 mM $MgCl_2$, 0.7 mM Tris(2-carboxyethyl) phosphine (TCEP), ~100 mg CelLytic Express (Sigma-Aldrich)) before lysis via sonication. Centrifugation at 25,000*x*g for 30 min provided clarified lysates. Proteins were purified using Ni-Sepharose affinity chromatography on an Åkta FPLC. Briefly, after filtration of the supernatant through 0.45 μm membrane, the solution was loaded onto a 5-mL HisTrap nickel-nitrilotriacetic acid column. The column was washed with 10 column volumes of buffer A (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 0.7 mM TCEP) and eluted with a linear gradient of buffer B (20 mM HEPES, pH 7.8, 300 mM NaCl, 400 mM imidazole, 0.7 mM TCEP). For ACP purifications, fractions were pooled, concentrated, and loaded onto a HiLoad 16/60 Superdex 75 (GE Healthcare Life Sciences) column equilibrated with storage buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.7 mM TCEP). Fractions were combined, concentrated, frozen, and stored at -80 °C. Because some of the acyl carrier proteins lack amino acids with appreciable absorbance at 280 nm, protein concentrations were determined via the bicinchoninic acid (BCA) method using BSA as a standard. BryR purifications differed from ACP purifications in that all buffers contained 10% glycerol in addition to the components listed above. In addition, for SPR and FTICR-MS assays, His-MBP-tag removal was achieved by TEV protease incubation overnight at 4 °C in buffer A. TEV protease and the N-terminal His-MBP tag were removed by repassaging the solution over the HisTrap column. Flow-through fractions were pooled, concentrated, and loaded onto

a HiLoad 16/60 Superdex 200 column equilibrated with BryR storage buffer (10% glycerol, 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.7 mM TCEP). Fractions were combined, concentrated, frozen, and stored at -80 °C. Protein concentrations were determined using absorbance at 280 nm and calculated extinction coefficients (1 A_{280} = 1.2 mg/ml). ACPs were greater than 95% pure following the above purification. Typical yields for BryR batches were ~ 3 mg/L of cell culture. TEV-cleaved BryR was approximately 85-90% pure. Purity estimates are based on SDS-PAGE (Figure 3-13).

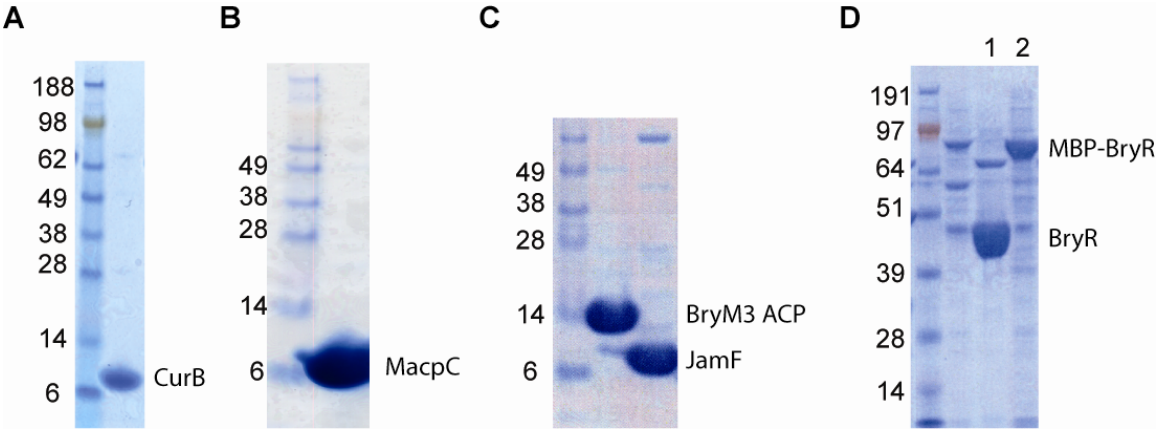


Figure 3-13. SDS-PAGE analysis of purified proteins. Apparent molecular weight of the SeeBlue Plus2 molecular weight marker (Invitrogen) is shown for reference. A-C) Proteins were run on a NuPAGE 12% SDS-PAGE gel using MES buffer. D) Proteins were run on a NuPAGE 4-12% SDS-PAGE gel using MOPS buffer. Lane 1 – MBP-BryR; Lane 2 – BryR after TEV protease cleavage.

3.4.2 Surface Plasmon Resonance Assays.

Sensor chips (CM-5) and HBS-P buffer were purchased from GE Healthcare Life Sciences. SPR experiments were performed on a BIAcore 3000 instrument. Running buffer for SPR was HBS-P+T (10 mM Hepes, pH 7.4, 0.15 M NaCl, 0.005% surfactant P20, 50 µM TCEP). The surface was prepared for immobilization of BryR by activating with 70 µL of a fresh mixture of 0.2 M 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) plus 0.05 M *N*-hydroxysuccinimide at 10 µL/min. BryR was freshly diluted to 20

μM in 10 mM phosphate/citrate buffer at pH 5.5 and loaded at 5 μL/min. Typically, 1000 – 8000 RU of BryR was immobilized. Activated carboxy groups were blocked with 1 M ethanolamine/HCl (70 μL at 10 μL/min). The surface was regenerated with 10 μL of 50 mM NaOH, 1 M NaCl after immobilization and between ACP binding cycles. To measure binding to BryR by SPR, solutions of ACPs in HBS-P+T were injected over the prepared surface as well as an ethanolamine treated control flow cell at a flow rate of 10 μL/min. Baseline subtraction was performed using a mock treated lane (activated with EDC/NHS and blocked with ethanolamine). Multiple injections (8 – 10 concentrations) were tested in duplicate or triplicate. Maximum testable concentrations for the ACPs were limited by their solubility. Data analysis was carried out using BIAevaluation software (GE Healthcare Life Sciences). Representative sensorgrams for apo-ACPs are shown as Figure 3-7. Nonlinear curve fitting of the equilibrium binding response was carried out using GraphPad Prism software.

3.4.3 Enzymatic Analysis of BryR via Radio-TLC

Radiolabeled and unlabeled acyl-CoA substrates were transferred onto the various ACPs using Svp, a phosphopantheine transferase (PPTase) from *Streptomyces verticillus* (67). Acyl-CoAs (500 μM) were combined with 75 μM ACPs (CurB, BryM3 ACP) and 5 μM Svp in a Tris buffer (pH 7.4) containing $MgCl_2$ (10 mM) and DTE (1 mM), and the reaction proceeded for 1 hr at room temperature. The substrate-bound ACPs were desalted, and utilized for experiments with BryR. The purified acylated donor (15 μM) and acceptor ACPs (30 μM) were incubated with BryR (10 μM) in 25 mM Tris buffer (pH 7.4) with DTE (1 mM) at room temperature for 5 minutes. Reactions were quenched by the addition of SDS-PAGE gel loading buffer. Samples were separated on

polyacrylamide gels by SDS-PAGE. The gels were first stained using SimplyBlue (Invitrogen), and were then exposed to Phosphoimager screens. The screens were scanned using a Typhoon Scanner (GE Healthcare), and analyzed using ImageQuant.

3.4.4 Identification of BryR Active Site Acetylation

BryR (10 μM) was reacted with acetyl-donor ACP (50 μM) and no acceptor ACP in 75 mM HEPES (pH 7.5) buffer, 1 mM TCEP. 1mg/ml TPCK trypsin (Pierce) was added to a final 1:100 ratio. Samples were incubated at 37 °C overnight. 20 μl of sample was injected onto an Jupiter C18 1x150mm 300 μm column (Phenomenex) using an Agilent 1100 LC system with a flow rate of 75 μl/min and a gradient of 2-98% solvent B (solvent A, water/0.1% fomic acid; solvent B, acetonitrile/0.1% formic acid) over 85 minutes. A divert valve was utilized for online desalting. The LC was coupled to an FTICR MS (APEX-Q with Apollo II ion source and actively shielded 7T magnet; Bruker Daltonics). Data was gathered from m/z 200–2,000 in positive ion mode. Electrospray was conducted at 2,600 V with 4 scans per spectra utilizing 0.33 s external ion accumulation in the hexapole prior to analysis in the FTICR using a loop value of 4. Collision cell pressure was reduced to 2.5e-6 torr. Data was analyzed using DECON2LC and VIPER (Pacific Northwest National Labs). The acetylated active site QA(Ac-**C)**YSGTAGFQMAINFILSR (2219.050 Da expected) was observed at 2219.045 Da representing a mass error of of -2 ppm uncalibrated. The same LC conditions were coupled to an LTQ Deca XP iontrap MS (Thermo). Online MS identified the same modified peptide, and online MS^2 allowed for confirmation that the modification occured on the active site cystiene (Cys114) (Table 3-4). Data was processed in Excaliber version 3.0 (Thermo).

Table 3-4. Ac-BryR Peptide Fragment Ions

Mass	Intensity	ID	dppm
480.18	62	a4	-28
463.13	30	a4-nh3	-65
567.19	18	a5	-67
550.24	28	a5-nh3	79
708.28	76	a7-nh3	19
753.19	51	b7	-132
735.22	451	b7-h20	-76
736.23	145	b7-nh3	-37
824.28	22	b8	-53
806.54	16	b8-h20	276
807.32	91	b8-nh3	28
732.35	17	y6-mh3	-70
749.30	21	y6-mh3	-171
845.34	20	y7-nh3	-172
1,064.52	18	y9	-63
1,047.46	35	y9-nh3	-104

3.4.5 Enzymatic Analysis of BryR via FTICR-MS

The preparation of acetyl-donor and acetoacetyl-acceptor ACPs was performed as above using Svp or Sfp PPTases (155). Acylated-ACPs were separated from CoA substrates via Zeba desalting columns (Pierce) or overnight dialysis in 3.5 kDa Slide-a-lyzer MINI dialysis units (Pierce) into 20 mM HEPES (pH 7), 150 mM NaCl. BryR (10 μM) was reacted with acetyl-donor ACP (50 μM) and acetoacetyl-acceptor ACP (80 μM) 75 mM HEPES (pH 7.5) buffer and 1 mM TCEP. After incubation for 60 min at room temperature, samples were acidified with 1% formic acid. Intact protein samples were desalted with Handee Microspin columns (Pierce) packed with 20 μl of 300 Å polymeric C4 resin (Vydac). Samples were loaded onto the columns and washed with 30 column

volumes of 0.1% formic acid prior to elution with 10 column volumes of 50% acetonitrile plus 0.1% formic acid. Intact protein samples were analyzed by an FTICR MS (APEX-Q with Apollo II ion source and actively shielded 7T magnet; Bruker Daltonics). Data was gathered from m/z 200–2,000 utilizing direct infusion electrospray ionization in positive ion mode. Electrospray was conducted at 3,600 V with 24–60 scans per spectra utilizing 0.5 s external ion accumulation in the hexapole prior to analysis in the FTICR using a loop value of 15. Collision cell pressure was reduced to 2.5e-6 torr. All IRMPD MS/MS was performed in the FTICR cell to avoid time of flight effects. Laser power was utilized at 40% with duration of 0.05 to 0.25 s. The entire mass range was fragmented, without any prior mass selection. Data was processed in Data Analysis (Bruker Daltonics) and Midas (NHMFL). All mass shifts shown were confirmed across all charge states for each ACP present. The most abundant charge state is used for all figures. All identified species were accurate to 20 ppm uncalibrated monoisotopic mass. All experiments were performed at least twice to verify the findings.

Notes:

This work has been submitted for publication as "Polyketide β-Branching in Bryostatin Biosynthesis: Identification of Surrogate Acetyl-ACP Donors for BryR, an HMG-ACP Synthase," Buchholz, T. J., Rath, C. M., Lopanik, N. B., Gardner, N. P., Håkansson, K., Sherman, D. H.

Author contributions:

Tonia Buchholz, Nicole Lopanik, Chris Rath and David Sherman designed the experiments;
Tonia Buchholz conducted the biochemical studies including cloning, protein purification, enzymatic assays and binding studies;
Chris Rath, Noah Gardner, and Kristina Håkansson performed and analyzed the mass spectrometry experiments.

Chapter 4

Gcs, a Type III PKS with an Acyl-ACP Substrate

4.1 Introduction

While the vast majority of studies regarding protein-protein interactions in PKS systems have been focused on the modular and iterative type I and iterative type II PKSs, recent work has led to a new appreciation for the homodimeric type III PKSs (50, 156, 157) that give rise to a range of aromatic compounds including the flavonoids and chalcones as well as to some α-pyrone acylketides. Pyrone-containing natural products exhibit a wide range of biological activities, a consequence of their structural diversity (Figure 4-1). Pyrones are synthesized by polyketide synthases (PKSs) through successive condensations of malonyl-CoA derived two carbon units followed by lactonization (Figure 4-2). The fungal metabolites fusapyrone (158) and D8646-2-6 (159) are most likely derived from iterative type I PKSs; enterocin is produced by a type II PKS (160). On the other hand, germicidins A, B and C (*41*) as well as the signaling molecule precursor, phlorocaprophenone (161) have been shown to be produced via PKS proteins from the type III family.

A number of characteristics make type III PKSs ideal candidates for engineering of artificial, enzyme-based systems for the creation of diverse pyrone-containing small molecules. The self-contained nature of the type III PKS opens up the potential exists for removing type III PKSs from their natural biosynthetic context and adapting them to

accept non-natural substrates in order to access a more structurally diverse chemical space (53). A single protein is capable of generating a complete natural product, as the same active site that catalyzes the Claisen condensations is also responsible for loading of the starter unit and cyclizing the polyketide intermediate. Additionally, rather than utilizing substrates that are covalently linked to acyl-carrier proteins, they are thought to use mainly acyl-CoA thioesters as substrates. Two notable exceptions to this have recently been reported and will be discussed below. Finally, the type III PKSs have demonstrated relaxed substrate specificity and have been shown to convert non-native acyl-CoA substrates to pyrones *in vitro* (53, 162, 163).

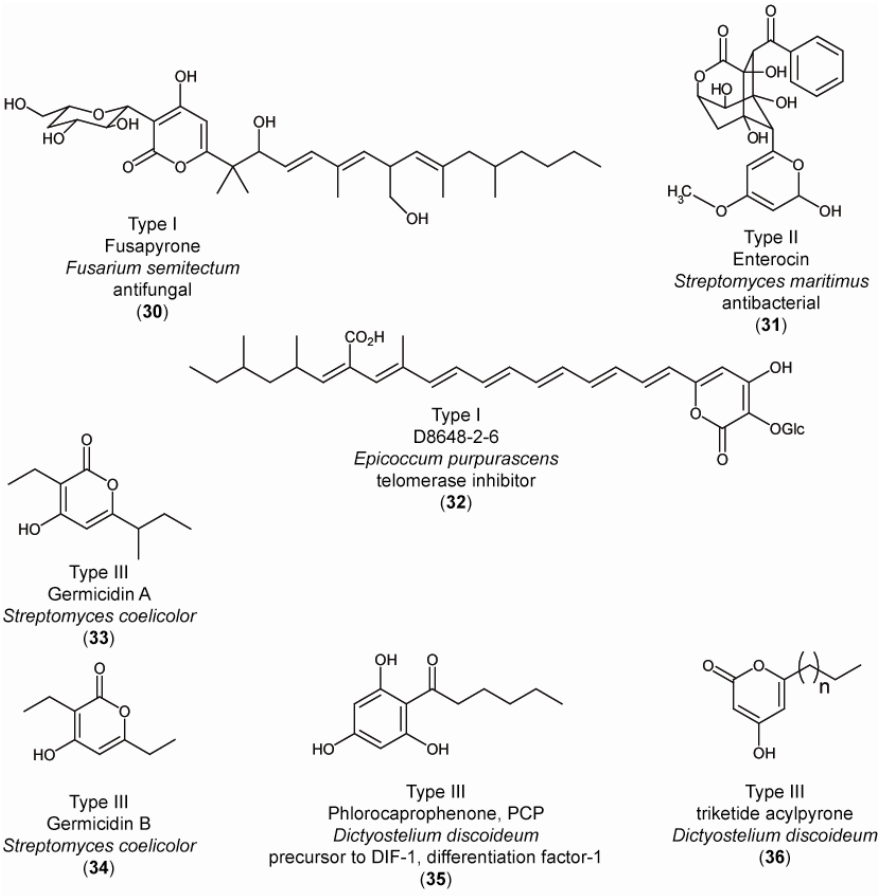


Figure 4-1. Structures of α-pyrone metabolites and closely related metabolites. Shown above are examples of natural products from type I, type II, and type III PKSs.

Streptomyces coelicolor hosts a wide array of genes homologous to various types of natural product biosynthetic enzymes, including type I modular polyketide synthases (PKSs), type II PKSs, type III PKSs, non-ribosomal peptide synthases, sesquiterpene synthases, siderophore synthetases, among others (164). Based upon predicted proteins encoded by its genome, *S. coelicolor* is thought to have the genetic capacity to produce more than 20 natural products (165, 166). Though the physiological roles for some of these gene clusters remain unclear, the Sherman laboratory sought to explore the capacity of the type III PKSs to create novel small molecules.

Of the three type III PKSs found in *S. coelicolor* (SCO1206, SCO7221, SCO7671), tetrahydroxynaphthalene synthase (THNS/SCO1206) has been extensively studied; its homolog from *Streptomyces griseus* was the first bacterial type III PKS to be identified and biochemically characterized (167-169). THNS is known to catalyze the sequential decarboxylative condensation, intramolecular cyclization, and aromatization of an oligoketide derived from five units of malonyl-CoA to give 1,3,6,8-tetrahydroxynaphthalene (THN) that spontaneously oxidizes to flaviolin (Figure 4-2B). In addition, the kinetics of THN formation and the crystal structure of this enzyme have been reported (161, 170). Based on previous studies that revealed substrate tolerance of THNS, this enzyme was explored for the biosynthesis of novel small molecules (53, 76). However, THNS prefers malonyl-CoA as a building block for both initiation and extension (Figure 4-2), which results in pyrone formation being secondary to flaviolin synthesis.

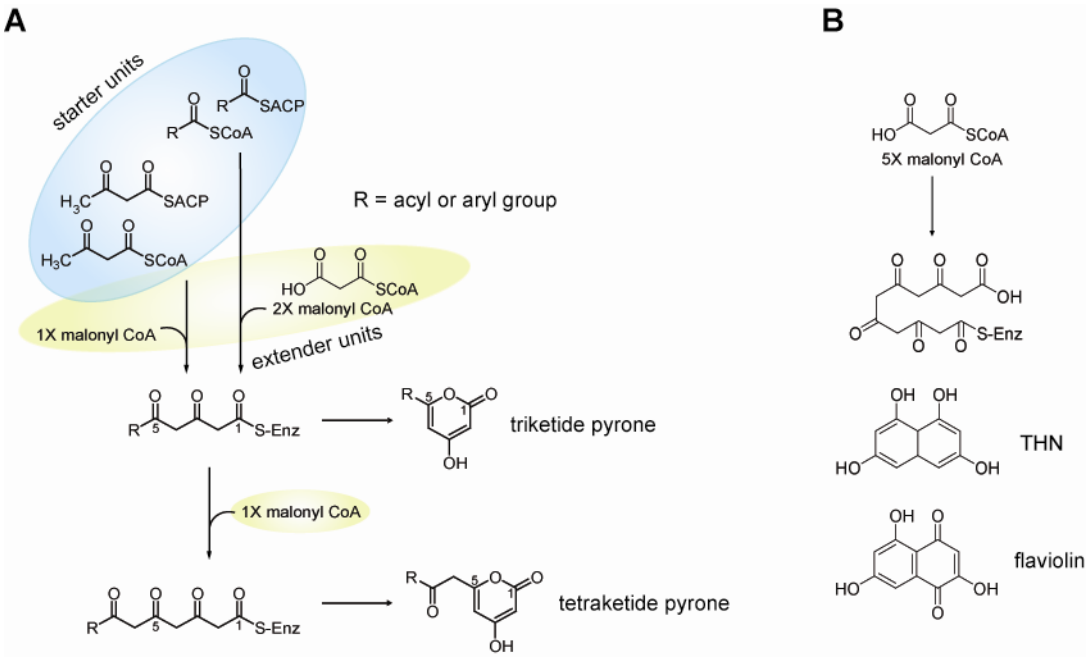


Figure 4-2. Type III PKS-catalyzed product formation. A) Reaction scheme for the formation of tri- and tetraketide pyrones. R groups that have been shown to function as type III PKS starter units include linear fatty acyl chains of various lengths, branched chains such as *iso*-butyryl-CoA, and even aromatic functionalities such as a benzoyl group. B) THNS catalyzed formation of THN, which is then oxidized to flaviolin.

The physiological role for SCO7221 has been reported and this type III PKS was designated Germicidin synthase (Gcs) to reflect its role in generating the germicidin series of natural product pyrones (Figure 4-1) (41). Though the physiological role for SCO7671 and its corresponding metabolite remain unidentified, we have confirmed that this enzyme belongs to the type III PKS class. We have recently described our *in vitro* investigation into the function, substrate specificity and product profile of Gcs and SCO7671 (*40*). Our results provided evidence that Gcs and SCO7671 type III PKSs are capable of accepting a variety of acyl-CoA starter units, extending them by one to three malonyl-CoA equivalents, and cyclizing the linear polyketide to form pyrone natural products. During the search for other type III PKSs whose starter unit specificity was orthogonal to extender unit preference and would, therefore, not compete with malonyl-

CoA, the remarkable finding was made that SCO7221(Gcs), has the ability to accept efficiently its starter unit as an acyl-ACP.

This work and other recent reports (40-42, 52, 68) have challenged the traditional view that all type III PKSs must use exclusively acyl-CoA starter units. Based upon sequence alignments, Austin and Noel suggested that perhaps bacterial type III PKSs, and not plant type III PKSs, might be able to accept starter units from ACPs (39). Evidence for this prediction came when the structure of a C-terminal type III PKS domain from the social amoeba, *Dictyostelium discoideum* was reported (*42*). In this system, the iterative type III PKS (Steely1) is comprised of a much larger polypeptide that includes upstream type I fatty acid synthase (FAS) domains.

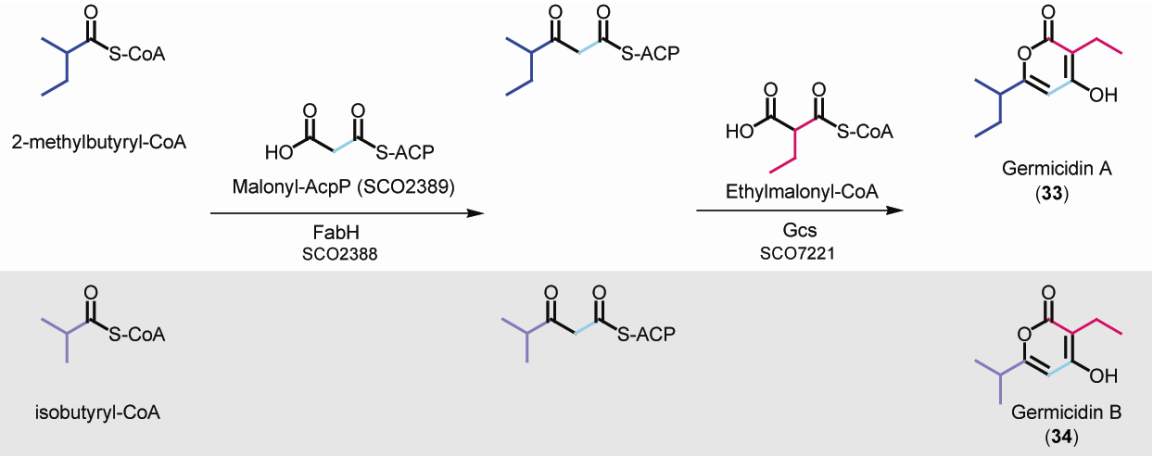


Figure 4-3. Proposed Gcs biosynthetic pathway. The proposed physiological pathway for germicidin A and B in *Streptomyces coelicolor* is shown. Starting with 2-methylbutyryl-CoA yields germicidin A, while condensation of one malonyl and one ethylmalonyl equivalent with isobutyryl-CoA produces germicidin B (shaded in gray).

A further suggestion came from the recent work of Challis and colleagues on the physiological role for Gcs (*41*). Specifically, branched chain fatty acid metabolites were proposed as the priming units of Gcs (Figure 4-3). This complex starter unit is extended only once by Gcs *in vivo* using an ethylmalonyl thioester to produce germicidin. It is possible that type III PKSs might accept either the ACP-tethered acyl chain or the acyl-

CoA after off-loading from the ACP. More recently, it has been proposed that the bacterial type III PKSs implicated in the production of phenolic lipids (alkylresorcinols and alkylpyrones) in *Azotobacter vinelandii* (ArsB and ArsC) (68) and *Bacillus subtilis* (BpsA) (52) accept their starter units as acyl-ACPs from type I FAS systems. We, therefore, sought definitive biochemical evidence for acyl-ACP transfer in type III PKSs. We demonstrated that Gcs was able to accept an acyl starter unit directly from a hexanoyl-charged ACP, a model substrate for the acyl-ACP transfer reactions (*40*).

4.2 Results and Discussion

4.2.1 Germicidin Synthase (Gcs) Structure

It is notable that *in vitro* Gcs is capable of utilizing acyl groups carried by either the small molecule, CoA or the small protein, ACP. This is in contrast to the selectivity exhibited by BryR (Chapter 3) for acetyl-ACP_D over acetyl-CoA substrates. To gain fundamental information about ACP-type III PKS interactions, we first pursued the structural characterization of Gcs. Though the number of bacterial type III PKSs characterized to date continues to grow, there are still numerous family members with potentially novel activities yet to be described (157). As expected for bacterial type III PKSs, amino acid sequence similarities were in the 20-35% range when compared to other bacterial and plant chalcone-like synthases (Figure 4-4). Analysis of these sequences revealed that the only other known enzyme with significant similarity to SCO7221 (Gcs) is found in the pathogen *Streptomyces scabies* (www.sanger.ac.uk), with 86% sequence identity at the amino acid level. However, no physiological or metabolic role has been reported for this type III PKS homolog. Both of these enzymes contain a long insertion between helices 4 and 5 (Chalcone synthase nomenclature) not observed in

other type III PKSs (Figure 4-4). The role of this insertion is unclear based solely on sequence alignments, but may play a structural role.

A pET-28b-containing *sco7221* expression vector was used to over-express Gcs in *E. coli*. Gcs can be purified to greater than 90% in one step on Ni-NTA resin. Further purification over a size exclusion column provided crystallographic quality protein and was consistent with a dimeric form for the enzyme (see section 4.4.1). The Gcs crystal structure was solved by molecular replacement (MR) using the THNS structure (PDBid = 1U0M) as a homology model in combination with the automated MR server, BALBES (171). Though the N-terminal His-tag is disordered, the 2.9 Å crystal structure of Gcs includes the full protein sequence (residues 1-389).

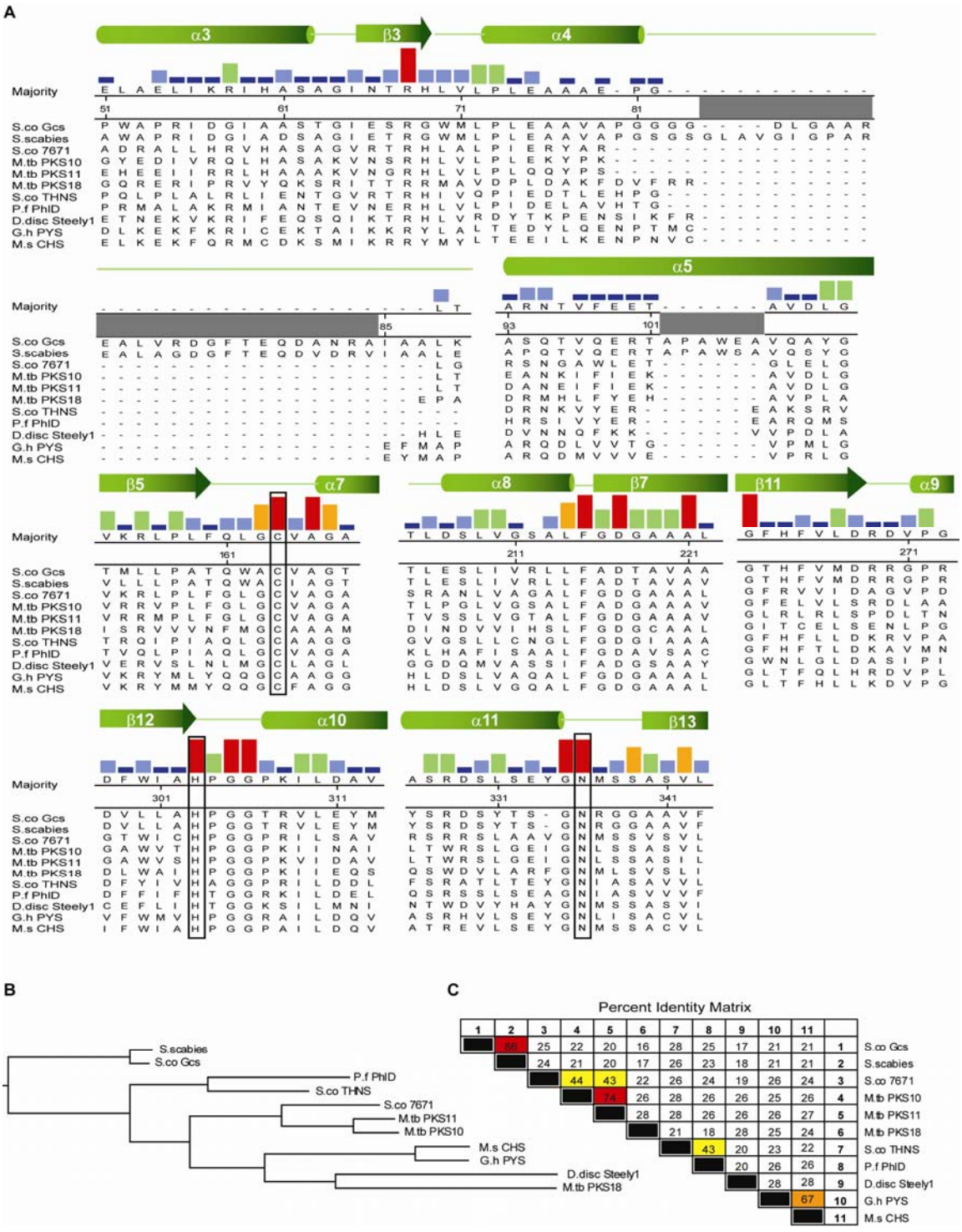


Figure 4-4. Sequence comparisons among selected type III PKSs. A) Sequence alignments are shown for selected residues only. Numbering and secondary structure assignment is based on Chalcone Synthase (CHS) from *Medicago sativa* (alfalfa). Catalytic triad residues are boxed. Similarity bars above the majority sequence are colored from red to blue for each position; red represents 100% identity. B) Cladogram of selected type III PKSs. C) Percent identities above 70% are highlighted in red, those between 50-69% in orange, and identities between 30-49% are colored yellow.

All previously reported type III PKS structures are dimeric, and our size exclusion chromatography data indicate that two Gcs molecules are associated to form a homodimer in solution. The asymmetric unit of Gcs contains a single protein molecule. However, the presumed physiologically relevant dimer is formed by a crystallographic 2-fold axis (Figures 4-5 and 4-7) and buries 2300 $Å^2$ of protein surface area per monomer (~13.6% of the monomer surface area). In our current model, residues 35-38 and 80-82 were disordered and remain unresolved. Due to the low resolution of the data and the extraordinarily high solvent content (80%), electron density was noisy in regions, but interpretable. Specifically, the density for the long insertion was poor, and absolute positioning of these residues is not possible with the current maps. At the current time, 46 residues are modeled as alanine residues (Figure 4-6).

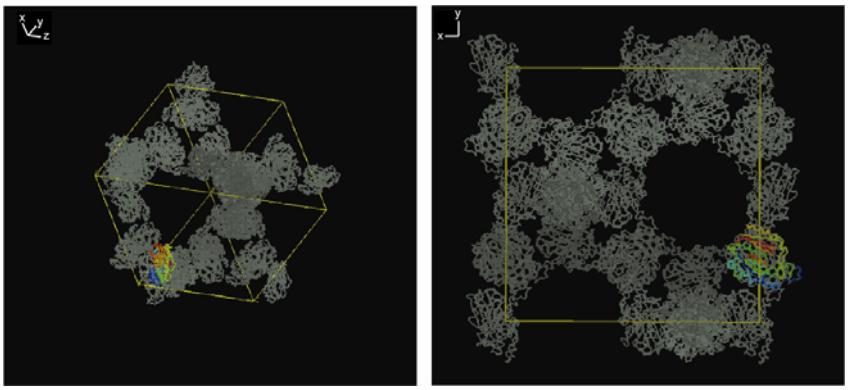


Figure 4-5. Packing of Gcs monomers in the crystal lattice. Two views of the crystal packing are shown; a single Gcs monomer is colored as a reference.

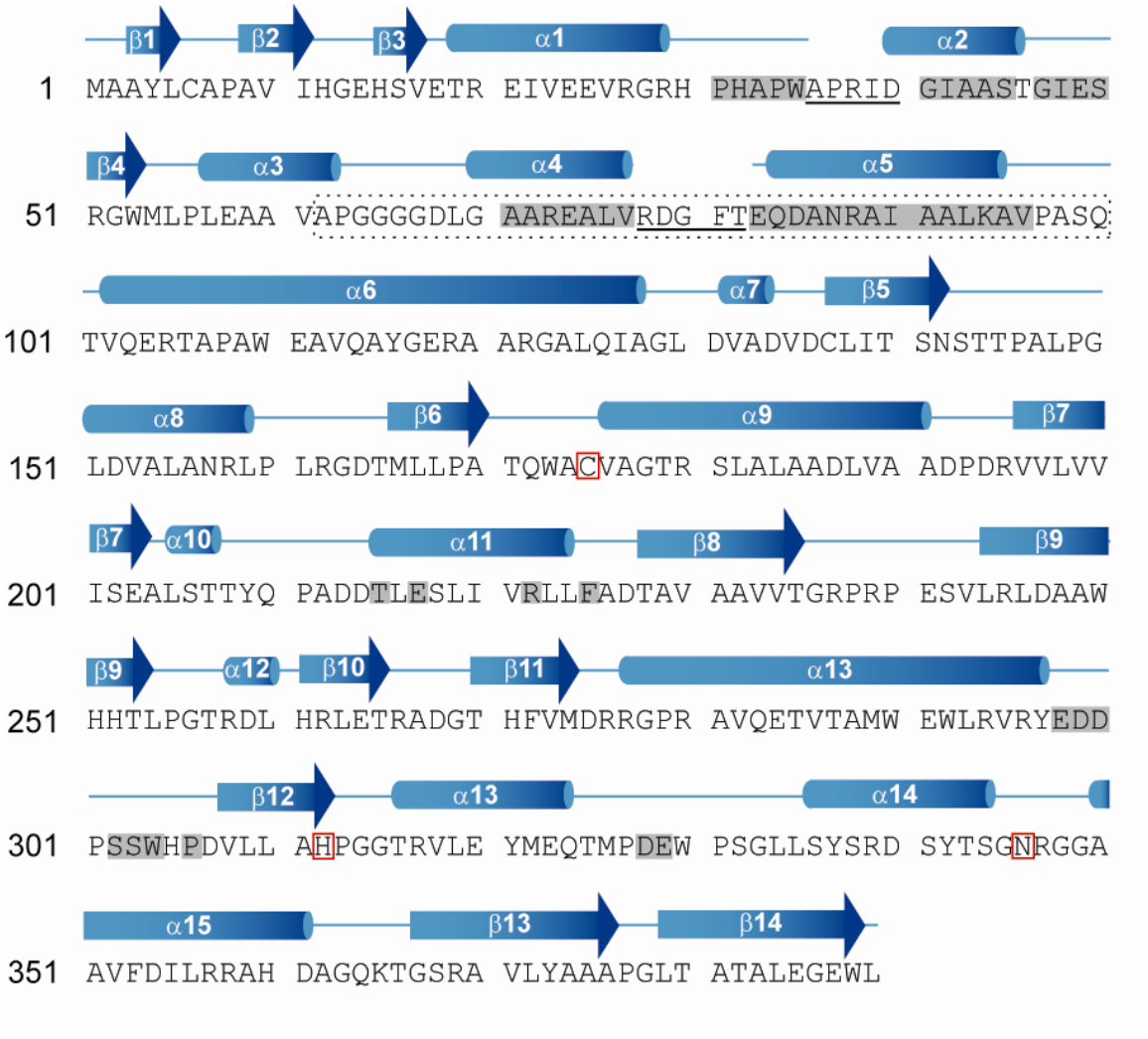


Figure 4-6. Secondary structure assignments of the current Gcs model. Residue numbers are shown at left for reference. Refinement of the model is still ongoing. Secondary structure assignment in blue is based on the Gcs structure shown in Figure 4-7. Disordered residues are underlined, and residues currently modeled as Ala are shaded gray. Members of the Cys-His-Asn catalytic triad are boxed in red. The long insertion ("basket") is in the dashed box.

Similar to other known type III PKS structures (39, 161), Gcs contains a Cys-His-Asn catalytic triad within a deep active site cavity that is accessible to the surrounding solvent. Apart from a few exterior loops, there are no drastic differences in the conserved αβαβα-fold or dimer interface as compared with the structure of the *S. coelicolor* THNS (Figure 4-7). The large insertion observed in the multiple sequence alignment (Figure 4-4) folds into a four-helix bundle made up of two helices from each monomer. The shape

of the helical bundle resembles a "basket" hanging down from the remainder of the protein. The basket residues are currently modeled as poly-alanine helices, as the side chains were not well resolved. Electron density is especially poor for the "basket" as it extends into the large solvent channels of the crystal and makes no contacts with other molecules within the lattice. Presumably, this insertion reinforces the dimer interface, though a role in ACP utilization has not been ruled out. As a physiological role for the basket is not apparent, we attempted to generate a "basket-less" Gcs protein by removing residues 62-99. The basket-less Gcs did not over-express in *E. coli*, so we were unable to assess its function biochemically. Further analysis of the Gcs crystal structure will be carried out following completion of the exploratory enzyme kinetics using the presumed physiological substrates. The fine-tuning of an HPLC-based pyrone detection assay to complete these studies is underway, and the synthesis of the native substrates is ongoing by a post-doctoral fellow in the Sherman laboratory. Based on our previous results using malonyl-CoA as an extender unit, acetoacetyl (Acac)-CoA and Acac-ACP can initially be used as model starter units. This assay will provide the means to compare CoA- to ACP-tethered substrates.

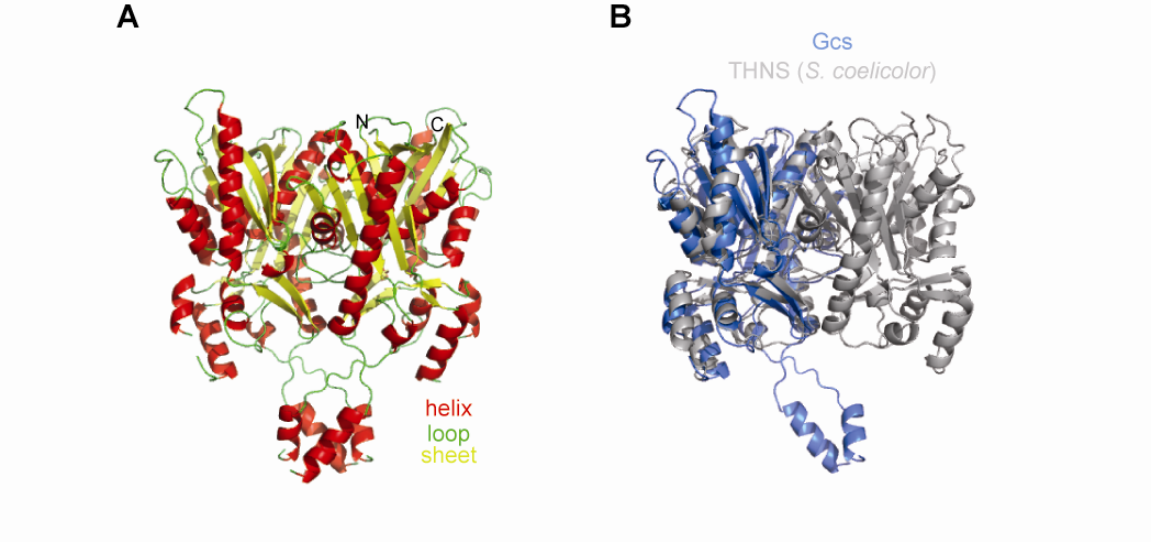


Figure 4-7. Overall structure of Gcs and related type III PKS from *S. coelicolor*. A) A view of the physiological dimer colored by secondary structure. B) Alignment of the Gcs structure with the THNS structure (PDBid: 1U0M).

4.2.2 Gcs Mutagenesis

We hypothesized that a patch of positively charged amino acids on the Gcs surface near the active site opening mediates binding to the negatively charged ACPs and faciliates usage of acyl-ACP starter units. To test this proposal, a series of mutations were designed to explore the effect of removing charge on a number of surface exposed residues near the substrate binding pocket opening (Figure 4-8 and Table 4-1). The cloning of expression vectors has been completed, and assessment of these proteins is ongoing. Characterization of these proteins awaits the results of the kinetic analysis of wild-type Gcs. In addition to generating mutant proteins to assess usage of acyl-ACP versus acyl-CoA starter units, we have designed and constructed vectors for the assessment of active-site flexibilty (Table 4-1). A small set of these proteins have been over-expressed and purified (C175S, C175A, L263A, T208S, T208C, F272A R317A). The proteins were soluble; Gcs mutant proteins L263A, T208S and T208C have been tested and shown to be active in an initial assay monitoring 4-hydroxy-3,6-dimethyl-2*H*-

pyran-2-one formation using Acac-CoA as a starter unit and methylmalonyl-CoA as the extender unit (data not shown).

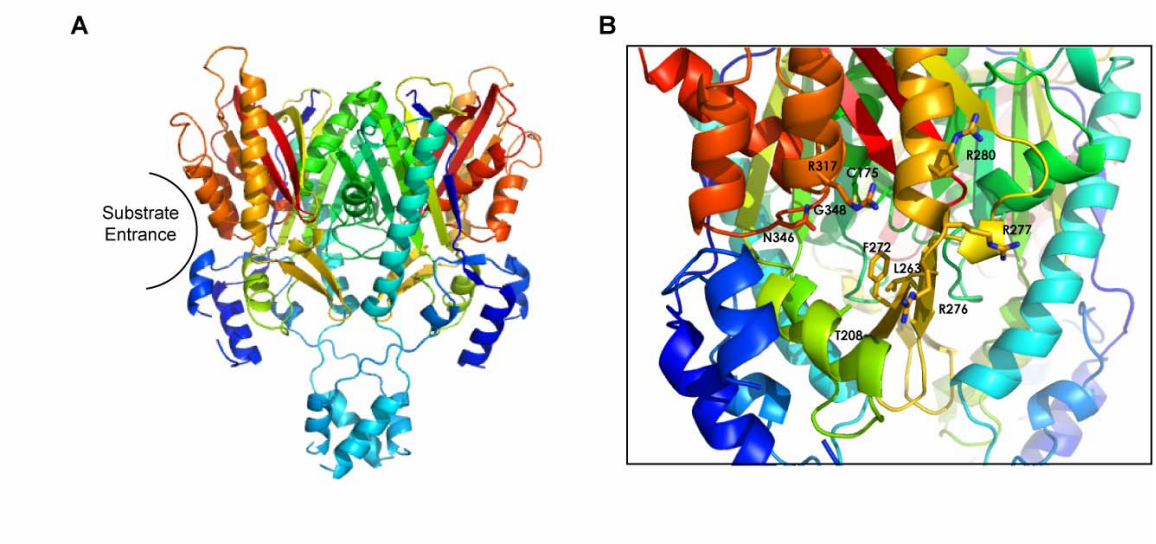


Figure 4-8. Gcs active site. A) A view of the Gcs dimer; individual monomers are colored blue to red from N- to C- terminus. Breaks in the chain are visible where the electron density maps are still too noisy to model amino acids as model refinement is still in progress. B) Close up of the active site. The Cys-His-Asn catalytic triad and planned mutations are labeled.

Table 4-1. Gcs mutant list

Mutant Type	Plasmid #	Mutant	Vector
Wild-type	pDHS9744		pMCSG7
Active Site	pDHS9745	C175S	pMCSG7
	pDHS9755	C175A	pMCSG7
Substrate Specificity	pDHS6650[a]	L263A	pET28b
	pDHS6651[a]	L263G	pET28b
	pDHS6652[a]	L263S	pET28b
	pDHS6653[a]	L263W	pET28b
	pDHS6654[a]	L263Y	pET28b
	pDHS9740	T208S	pET28b
	pDHS9742	T208A	pET28b
	pDHS9741	T208G	pET28b
	pDHS9739	T208C	pET28b
	pDHS9738	G348A	pET28b
	pDHS9784	F272A	pET28b
	pDHS9785[b]	F272T	pET28b
	pDHS9786	F272N	pET28b
	pDHS9787	F272L	pET28b
	pDHS9788	F272V	pET28b
Surface Residues	pDHS9753	R317A	pMCSG7
	pDHS9754	R317S	pMCSG7
	pDHS9807	R276A	pET28b
	pDHS9789	R276D	pET28b
	pDHS9790	R277A	pET28b
	pDHS9743	R277D	pET28b
	pDHS9791	R280A	pET28b
	pDHS9792	R280D	pET28b
"Basket-less"	pDHS9793	Δ63-98	pMCSG7

[a] Plasmids with changes at the L263 position were generated by Sabine Grüschow.
[b] F272T was a fortutious mutation identified through DNA sequencing of F272A colonies, therefore, no primers are listed in Table 4-2.

4.3 Summary

At the time of conception, the idea to engineer a pathway where a type I PKS intermediate could be shuttled to a type III PKS for incorporation of a pyrone had no literature precedence. There were no reported exceptions to the acyl-CoA starter preference of type III PKSs. Over the past three years the first biosynthetic pathways wherein an intermediate is passed off from a type I FAS to a type III PKS have been proposed (40, 42, 52, 68). The similarity between the type I FAS and the type I PKS polyproteins makes generation of a type I – type III PKS hybrid pathway more plausible. During our biochemical characterization of the type III PKSs in *S. coelicolor*, we have

identified robust conditions for over-expression, purification and crystallization of Gcs, an ACP-utilizing type III PKS from *S. coelicolor*. Complete refinement of the structural model of the Gcs protein is underway; our in-progress model is reported here. Though complete characterization of the binding and catalytic capabilities of Gcs is currently an ongoing project, our structural model represents a key step in enabling the design of a model type III PKS for use in an engineered pathway for the generation of complex pyrones.

4.4 Experimental Methods

4.4.1 Design of Expression Constructs

Generation of the Gcs expression vector, pDHS100019, was previously published (*40*). A plasmid for the expression of the Gcs with a TEV protease cleavable His-tag was generated by amplification using PCR with LIC overhangs and inserted into the vector pMCSG7 (100). Sequences in all capital letters represent the LIC overhangs necessary for insertion into the pMCSG7 vector. The "basket-less" Gcs insert was generated via sequential PCR amplification of 1) the left and right pieces of Gcs and 2) the fused construct from PCR amplification of the combined fragments using the same LIC primers listed for pDHS9744 construction. All single point mutants were generated according to the Quikchange® site-directed mutagenesis protocol (Stratagene/Agilent). Primers were designed using the online Quikchange primer generation software (Stratagene/Agilent). DNA sequences were confirmed by sequencing by the University of Michigan Sequencing Core. Primers used are listed in Table 4-2.

Table 4-2. Primer list for Gcs protein expression.

Primer name	Primer sequence	Plasmid(s)
GcsLICF	TACTTCCAATCCAATGCC atg gcc gca tac ctg tgc	pDHS9744, pDHS9793
GcsLICR	TTATCCACTTCCAATGCTA cag cca ctc ccc ttc cag a	
C175S_F	ccc agt ggg cca gcg tgg ccg gg	pDHS9745
C175S_R	ccc ggc cac gct ggc cca ctg gg	
C175A_F	acc cag tgg gcc gcc gtg gcc ggg ac	pDHS9755
C175A_R	gtc ccg gcc acg gcg gcc cac tgg gt	
T208S_F	gcg ctg agc acc tcc tac cag ccc g	pDHS9740
T208S_R	cgg gct ggt agg agg tgc tca gcg c	
T208A_F	gcg ctg agc acc gcc tac cag ccc g	pDHS9742
T208A_R	cgg gct ggt agg cgg tgc tca gcg c	
T208G_F	ggc gct gag cac cgg cta cca gcc cgc	pDHS9741
T208G_R	gcg ggc tgg tag ccg gtg ctc agc gcc	
T208C_F	ggc gct gag cac ctg cta cca gcc cgc	pDHS9739
T208C_R	gcg ggc tgg tag cag gtg ctc agc gcc	
G348A_F	gcg gca acc gcg ccg gcg ccg ccg t	pDHS9738
G348A_R	acg gcg gcg ccg gcg cgg ttg ccg c	
F272A_F	gga cgg cac cca cgc cgt gat gga ccg g	pDHS9784
F272A_R	ccg gtc cat cac ggc gtg ggt gcc gtc c	
F272N_F	gga cgg cac cca caa cgt gat gga ccg g	pDHS9786
F272N_R	ccg gtc cat cac gtt gtg ggt gcc gtc c	
F272L_F	cgg acg gca ccc act tag tga tgg acc g	pDHS9787
F272L_R	cgg tcc atc act aag tgg gtg ccg tcc g	
F272V_F	gga cgg cac cca cgt cgt gat gga ccg	pDHS9788
F272V_R	cgg tcc atc acg acg tgg gtg ccg tcc	
R317A_F	ccc ggc ggg acc gcg gtg ctg gag ta	pDHS9753
R317A_R	tac tcc agc acc gcg gtc ccg ccg gg	
R317S_F	ccc ggc ggg acc agc gtg ctg gag tac	pDHS9754
R317S_R	gta ctc cag cac gct ggt ccc gcc ggg	
R276A_F	ttc gtg atg gac gct cgc ggg ccg cg	pDHS9807
R276A_R	cgc ggc ccg cga gcg tcc atc acg aa	
R276D_F	act tcg tga tgg acg atc gcg ggc cgc ggg c	pDHS9789
R276D_R	gcc cgc ggc ccg cga tcg tcc atc acg aag t	

R277A_F	tga tgg acc ggg ccg ggc cgc ggg	pDHS9790
R277A_R	ccc gcg gcc cgg ccc ggt cca tca	
R277A_F	gtg atg gac cgg gac ggg ccg cgg gc	pDHS9743
R277D_R	gcc cgc ggc ccg tcc cgg tcc atc ac	
R280A_F	gcg cgg gcc ggc ggc ggt gca g	pDHS9791
R280A_R	ctg cac cgc cgc cgg ccc gcg c	
R280A_F	ccg gcg cgg gcc gga tgc ggt gca gga aa	pDHS9792
R280D_R	ttt cct gca ccg cat ccg gcc cgc gcc gg	
Gcs62Rev	TCCTGGACGGTCTGCGAcgcgacggcggcctccaggggagca	pDHS9793
Gcs99For	TGGAGGCCGCCGTCGCGtcgcagaccgtccaggagcgcaccgcc	

4.4.2 Protein Expression

The plasmid pDHS10019, which allows expression of SCO7221 (Gcs) as N-terminal His_6-fusion protein, was transformed into BL21(DE3) and cultured in LB media containing 50 μg/ml kanamycin until the OD_{600} reached 0.7 – 1.0. Protein expression was induced with 0.1 mM IPTG and culturing continued for 16-20 hours at 18 °C. Cells were harvested by centrifugation and frozen at -20 °C. Selenomethionyl protein was produced in a similar fashion using selenomethionine minimal medium (102). Cell pellets were thawed to 4 °C and resuspended in 5X volume of lysis buffer (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 1 mM $MgCl_2$, and ~100 mg CelLytic Express (Sigma-Aldrich)) before lysis via sonication. Centrifugation at 25,000*x*g for 30 min provided clarified lysates. Proteins were purified using affinity chromatography on an Åkta FPLC (GE Healthcare Life Sciences). Briefly, after filtration of the supernatant through a 0.45 μm membrane, the solution was loaded onto a 5 mL HisTrap nickel-nitrilotriacetic acid column. The column was washed with 10 column volumes of buffer A (20 mM HEPES, pH 7.8, 300 mM NaCl, 20 mM imidazole, 0.5 mM TCEP) and eluted

with a linear gradient of buffer B (20 mM HEPES, pH 7.8, 300 mM NaCl, 400 mM imidazole, 0.5 mM TCEP). Fractions were pooled based on A_{280} peak height, concentrated, and loaded onto a 120 mL HiLoad 16/60 Superdex 200 (GE Healthcare Life Sciences) column equilibrated with storage buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP). Fractions were combined, concentrated, frozen, and stored at -80 °C. Calibration of the column was performed with molecular weight markers from Sigma-Aldrich (St. Louis, MO). Gcs eluted as a single peak at 74 mLs which is consistent with a dimeric complex in solution (Figure 4-9). Approximately 50 mg Gcs could be purified from a 1 L culture. Protein concentrations were determined using absorbance at 280 nm and calculated extinction coefficient (1 A_{280} = 0.62 mg/ml). Proteins were estimated to be greater than 95% pure based on SDS-PAGE (Figure 4-10).

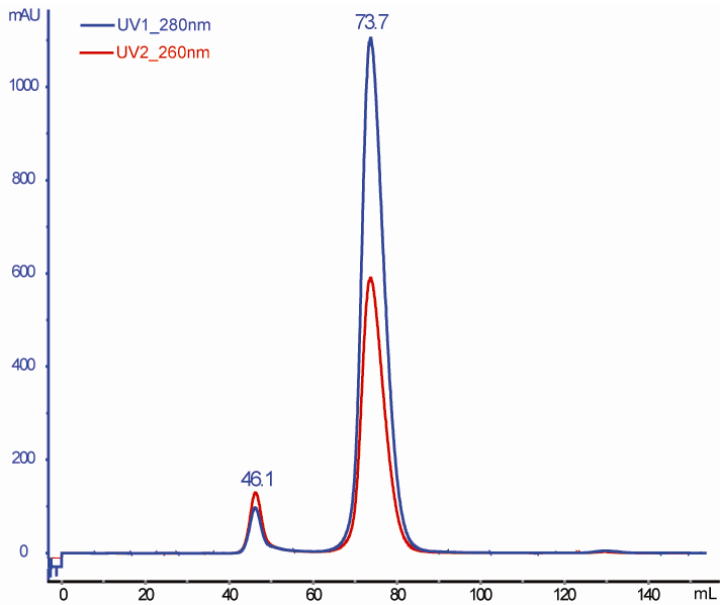


Figure 4-9. Size exclusion chromatography of Gcs in 20 mM HEPES, pH 7.4, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP. The peak around 45 mLs represents material eluting near the void volume. The major protein peak eluted at 74 mLs. To calculate the apparent molecular weight, this elution volume was fit to the equation, mLs = -12.447(LnMW) + 129.52 (MW in kDaltons). An apparent molecular weight of 90,000 Daltons is consistent with a dimer for the His-tagged protein whose calculated molecular weight is 44,000 Daltons.

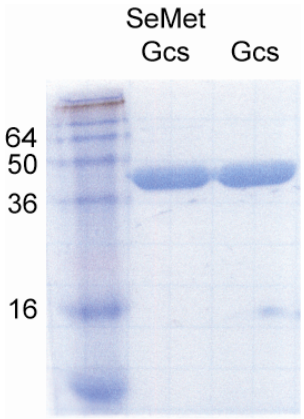


Figure 4-10. SDS-PAGE analysis of Gcs and selenomethionyl-Gcs. Apparent molecular weight of the SeeBlue Plus2 molecular weight marker (Invitrogen) is shown for reference. Proteins were run on a 12% Tris-Glycine SDS-PAGE gel.

4.4.3 Crystallization, Data Collection and In Progress Refinement of the Gcs Structure

Initial screening with Gcs produced crystals of cubic morphology under buffered isopropanol conditions containing acetate. For crystal growth, an equal volume of protein solution (10 mg/ml, freshly dialyzed into 20 mM HEPES, pH 7.5, 1 mM TCEP from storage buffer) was mixed with mother liquor containing 21-25% isopropanol, 300 – 400 mM sodium (or ammonium acetate), 0.1 M Tris buffer, pH 8.5 – 9.0. The best-diffracting crystals grew in 3-5 days at 4 °C using hanging-drop vapor diffusion techniques. Selenomethionyl Gcs did not crystallize under conditions similar to those used for native Gcs. Twice the volume of glycerol was added to the drop before the crystals were harvested in loops and frozen in liquid N_2. Attempts to obtain co-crystals with a coenzyme A substrate (hexanoyl- or Acac-) were unsuccessful. Attempts to soak in hexanoyl- or Acac-CoA into Gcs crystals after they had formed resulted in dissolution of the crystals within seconds.

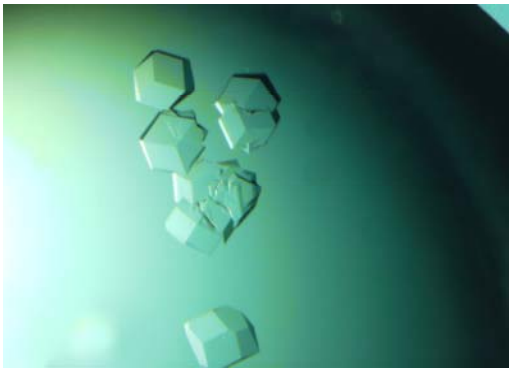


Figure 4-11. Gcs crystals. 10 mg/ml Gcs protein in 20 mM HEPES, pH 7.5, 1 mM TCEP were mixed 1:1 with 23% isopropanol, 300 mM NH_4Oac, 0.1 M Tris, pH 8.5. Crystals averaged 70 to 100 μm across.

Diffraction data were collected at 100 K on GM/CA-CAT beamline 23ID-D at the Advanced Photon Source in the Argonne National Laboratory (Argonne, IL). The data were processed using the HKL2000 suite (104); data were indexed, integrated and scaled in the cubic space group $P4_132$. Initial phasing by molecular replacement using THNS (PDBid - 1U0M) as a homology model identified a single molecule in the asymmetric unit (Vm – 6.1) and extraordinarily high (80%) solvent content in the unit cell. Phasing was completed using the automated MR at the BALBES server (171). The presumed physiologically relevant dimer forms on a crystallographic 2-fold axis (Table 4-3). Modeling was completed manually using COOT (107). The current model was refined against the 2.9 Å native dataset using REFMAC5 of the CCP4 suite (108-110). (Table 4-4). Electron density was noisy, but interpretable. Specifically, the "basket" density was poor, and absolute positioning of these residues is not possible with the current maps.

Table 4-3. Diffraction data for Gcs

Parameter	Native
Space Group	$P4_132$
Dimensions (Å) *a,b,c*	182.70
X-ray source	23ID-D
Wavelength λ (Å)	0.97934
d_{min}(Å) [a]	2.90 (3.00-2.90)
Unique observations	23,767
R_{merge} (%) [a,b]	6.4 (59.1)
$\langle I/\sigma \rangle$ [a]	32.3 (4.1)
Completeness (%) [a]	99.9 (100)
Avg. redundancy [a]	9.6 (9.7)

[a] Values in parenthesis are for outer shell
[b] $R_{merge} = \Sigma|I_i-\langle I\rangle|/\Sigma I_i$, where I_i is the intensity of the *i*th observation and $\langle I\rangle$ is the mean intensity

Table 4-4. Current state of refinement (non-final)

	Gcs
Date range	39.9-2.90
R/R_{free} [a,b]	0.298/0.323
RMSD bond length (Å)	0.013
RMSD bond angle (°)	0.546
Avg. Protein B-factor (Å^2)	63.3
Wilson B (Å^2)	88.8
Ramachandran plot[c]	
Favored	94.2
Allowed	2.7
Disallowed	3.1
Protein atoms	2752

[a] $R = \Sigma||F_o|-|F_c||/\Sigma|F_o|$ where F_o is the observed structure factor and F_c is the calculated structure factor used in the refinement
[b] $R_{free} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$ where F_o is the observed structure factor and F_c is the calculated structure factor from 5% of reflections not used in the refinement
[c] From output of MOLProbity

4.4.4 Sequence and Structure Analysis.

Alignments, cladogram and percent identities were calculated using the ClustalW method using Lasergene MegAlign software from DNASTAR. Sequences: S.co Gcs,

S.co 7671, S.co THNS = *Streptomyces coelicolor* ORFs *sco7221, sco7671, sco1206,* www.sanger.ac.uk; S.scabies = *Streptomyces scabies*, www.sanger.ac.uk; M.tb PKS10, M.tb PKS11, M.tb PKS18 = *Mycobacterium tuberculosis* CHS-like proteins, accessions CAB06631, CAB09101, A70958; P.f PhlD = *Pseudomonas fluorescens* PhlD, accession AAB48106; D.disc Steely1 = *Dictyostelium discoideum* C-terminal Type III PKS domain of Steely1, pdb 2H84; G.h PYS = *Gerbera hybrida* (daisy) 2-pyrone synthase, accession CAA86219; M.s CHS = *Medicago sativa* (alfalfa) chalcone synthase, accession P30074 (Figure 4-4).

Notes:

Research on the Gcs enzyme is currently ongoing in the Sherman laboratory by a post-doctoral fellow. To date, the following people have contributed to the work.

Author contributions:

Tonia Buchholz, Sabine Grüschow, Todd Geders, Janet Smith and David Sherman designed the experiments;
Tonia Buchholz, Sabine Grüschow cloned the expression constructs;
Tonia Buchholz conducted the protein expression, purification, and crystallization;
Frank Bartley, III and Todd Geders set and harvested crystals;
Todd Geders solved the x-ray crystal structure;
Tonia Buchholz, Todd Geders, David Sherman and Janet Smith analyzed the x-ray crystal structure.

GM/CA CAT has been funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract DE-AC02-06CH11357.

Chapter 5

Conclusion

Since the first report of the genetic programming of a type I PKS pathway in the early 1990s, the promise of simple "mix-and-match" reprogamming for the generation of novel polyketides has loomed over the natural products field. The ultimate goal is to design and build hybrid PKS systems utilizing heterologous module pairs in a combinatorial fashion to achieve more efficient polyketide production and generate novel drug-like products. To reach this goal, we will be required to combine the lessons learned for optimizing key protein-protein interactions at the inter-polypeptide interface and acyl transfer positions with those related to identification of catalytic domains capable of processing non-native substrates. Towards these goals, we have studied protein-protein interactions involved in acyl group transfer in three separate biosynthetic pathways.

An increased understanding of the key steps involved in PKS docking domain-mediated intermodular acyl group transfer provides a number of exciting opportunities. As the number of orphan non-colinear biosynthetic clusters rises with the completion of microbial genome sequencing projects, the ability to sequentially order the polypeptides via prediction of docking domain compatibility could enable more facile prediction of core polyketide structures. While we have reported progress in decoding the selectivity filters for the H1-T1 and H2-T2 subclasses of type I PKS docking domains, a number of PKS docking domain structures remain obscure. For example, N- and C-terminal amino

acid sequences of the PKS polypeptides from the *bry* pathway do not fall within the current classification system.

Moreover, while the static pictures of type I PKS docking interactions determined by structural biology efforts thus far have proven valuable for understanding the type I PKS systems, we still have no model for the orientation of the full ACPdd (helices 1-2 relative to 3) or the position of the ACPdd within a full module (Figure 5-1). Likewise, it is tempting to speculate that ACPdd's upstream helices (1-2) are not only important for dimerization (85), but that they may interact with portions of the PKS outside of the ACPdd.

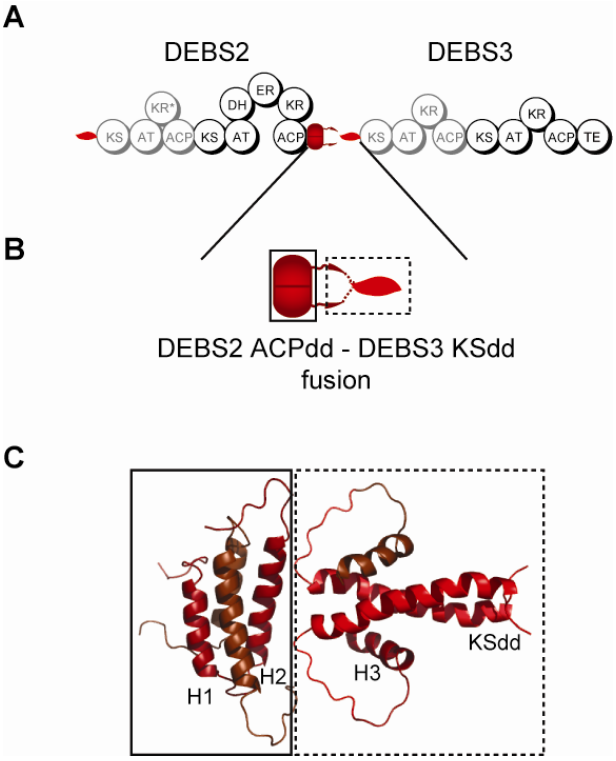


Figure 5-1. Full ACPdd Model from DEBS2-DEBS3. A) Domain organization of DEBS2 and DEBS3 from the erythromycin PKS. B) Cartoon representation of the DEBS2-DEBS3 fusion protein used to solve the NMR structure of the docking domain complex. C) One possible orientation of the full ACPdd-KSdd complex (from PDBids 1pzr and 1pzq) (85) The ACPdd helices (H1, H2, H3) and the KSdd coiled coil are labeled.

Multiple pieces of circumstantial evidence support a broader role for helices 1-2 of the ACPdd. In the work presented in chapter 2, the presence of downstream domains in the longer PikAIV constructs resulted in an increased affinity for the ACPdd. While a likely explanation is that these larger PikAIV constructs simply stabilize the productive binding conformation of the PikAIV KSdd, it is possible that additional protein-protein contacts exist between the upstream ACPdd and the downstream KS-AT region of the module. Additionally, in our version of the intermodular transfer and elongation assay (Figure 5-2), we observe a slightly different product profile when comparing competition with discrete PikAIII ACPdd (helices 1-3) to deletion of the PikAIII ACPdd terminal helix. Specifically, the addition of discrete PikAIII ACPdd to the reaction is able to compete with the PikAIII full module for binding to PikAIV, thereby decreasing production of both narbonolide and 10-dml in a dose-dependent manner (Figure 5-2B). However, upon deletion of only the terminal helix of the PikAIII ACPdd from the full PikAIII module, narbonolide formation is completely abolished, but 10-dml formation still occurs (albeit at a significantly decreased rate) (Figure 5-2C) (26). Our extension and elongation assay data suggest a role for the PikAIII ACPdd helices 1 and 2 in mediating the interaction of PikAIII with PikAIV TE. Alternately, the TE may be interacting directly with the PikAIII ACP. Using similar surface plasmon resonance or fluoresence polarization assay conditions as described in chapter 2, we may be able to decipher the remaining molecular details of the direct PikAIV TE-catalyzed 10-dml formation observed in the pikromycin/methymycin system.

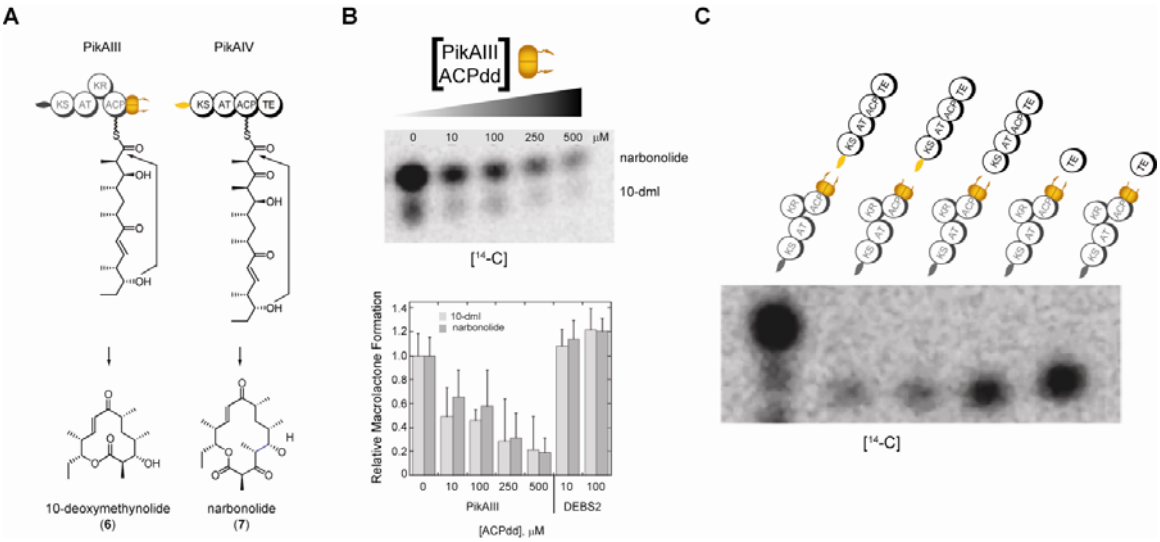


Figure 5-2. PikAIII - PikAIV transfer and elongation assay with full length substrates. A) Cartoon representation of the assay system. Synthetic pentaketide *N*-acetyl cysteamine thioester (SNAc) is loaded onto PikAIII, transferred to, extended and cyclized by PikAIV to form narbonolide. 10-dml is formed after loading onto PikAIII and cylclization by PikAIV. B) Radio-TLC image and quantitation of the reaction products produced when wild-type PikAIII and PikAIV were treated with increasing amounts of PikAIII ACPdd in the presence of the pentaketide SNAc and 2-[^{14}C]-methylmalonyl-CoA. C) Radio-TLC reaction products of truncated pikromycin modules. Figure modified from (26).

To explore the potential interactions of the complete ACPdd, we have begun the structural characterization of a larger portion of the PikAIII-PikAIV interface (Figure 5-1). In collaboration with Dr.Janet Smith's laboratory, initial crystallization conditions have been identified for the PikAIV KSdd-KS-AT tridomain, and the solution of the structure is underway. Co-crystallization studies with the PikAIII ACP-ACPdd could then be combined with a binding analysis of ACPdd mutants to gain a further understanding into role of the upstream ACPdd helices (1 and 2) in the larger type I PKS organization.

A subset of type I modular PKSs (and hybrid NRPS/PKS megasynthases) recently have been identified that contain multiple enzymes acting in *trans* during the traditional linear assembly-line process to incorporate methyl or methylene groups (or functional groups derived from such groups) at the β-position. Though the acyl transfer steps

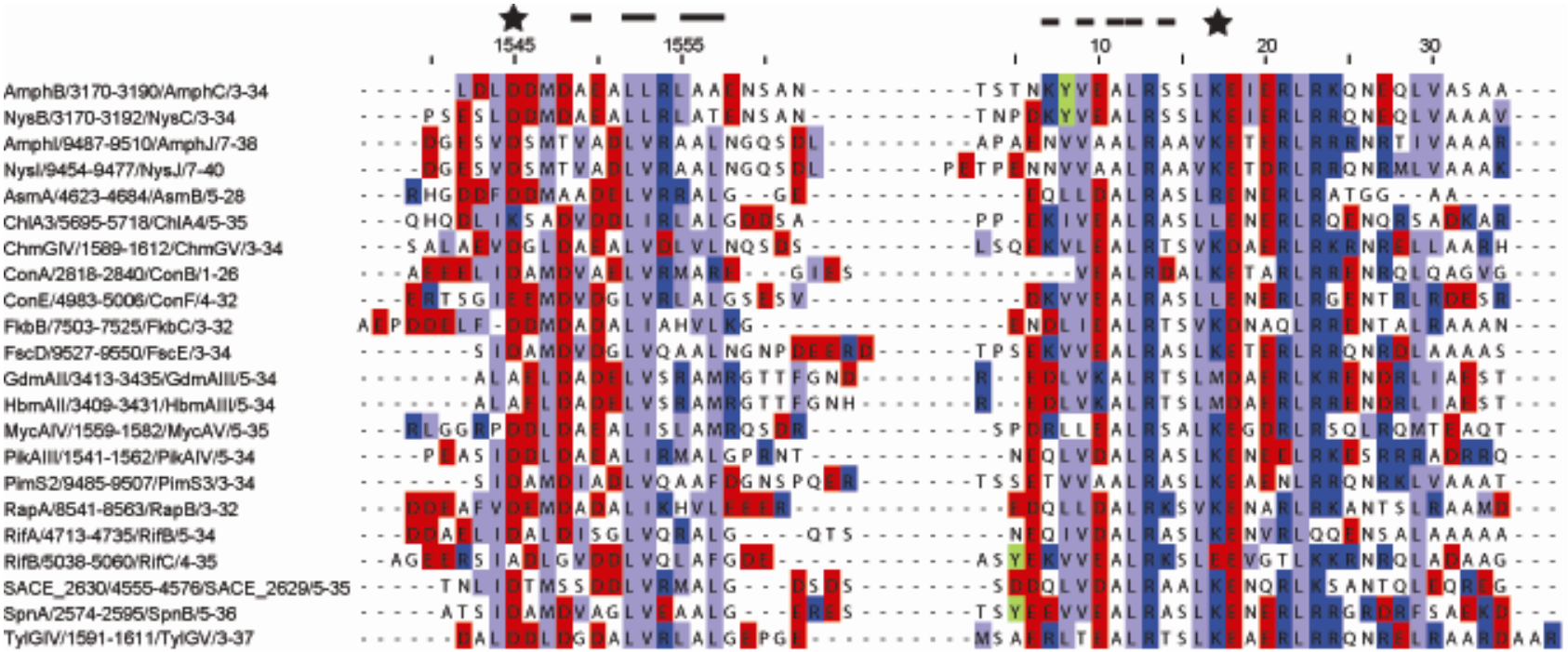
involving discrete ACPs from FAS and type II PKS systems have been well studied, much less is known about the details of this new emerging subclass of ACPs found in the HMGS cassettes (Figure 3-4 and Table 3-1). Though no ACP_D has been identified in the *bry* cluster, BryR, an HMG-ACP synthase from the bryostatin pathway, is able to catalyze HMG-ACP formation when presented with a surrogate acetyl-ACP substrate. For BryR, we have demonstrated binding specificity for three members of this sub-group of ACPs (CurB, JamF, MacpC). We have also observed a weaker affinity for the native acceptor ACP (BryM3 ACP).

Despite testing a small set of MacpC residues to identify key BryR - ACP_D contact points, no single amino acid residues of the MacpC protein were identified that significantly impacted binding affinity. The major limitation to this set of experiments was generating high micromolar concentrations of the mutant proteins in the current assay conditions. Therefore, to map the BryR - ACP_D interface, a different strategy must be taken. The ideal situation would be to identify the native ACP_D from the bryostatin pathway and probe the molecular mechanisms of binding for this ACP – enzyme pair. To date, no structural insights have been reported for the interaction of HMGS cassette enzymes with partner ACP_Ds. Structural data will be important to determine the nature of BryR's ACP binding selectivity. This will require significant improvement in the purity of the recombinant BryR, as the current method of production only yields ~80% pure protein. Coexpression with additional HMGS cassette member proteins may increase the solubility of BryR if these protein do exist as a large, megacomplex inside the cell as in the bacillaene case (99). Alternatively, one could explore the binding preferences and

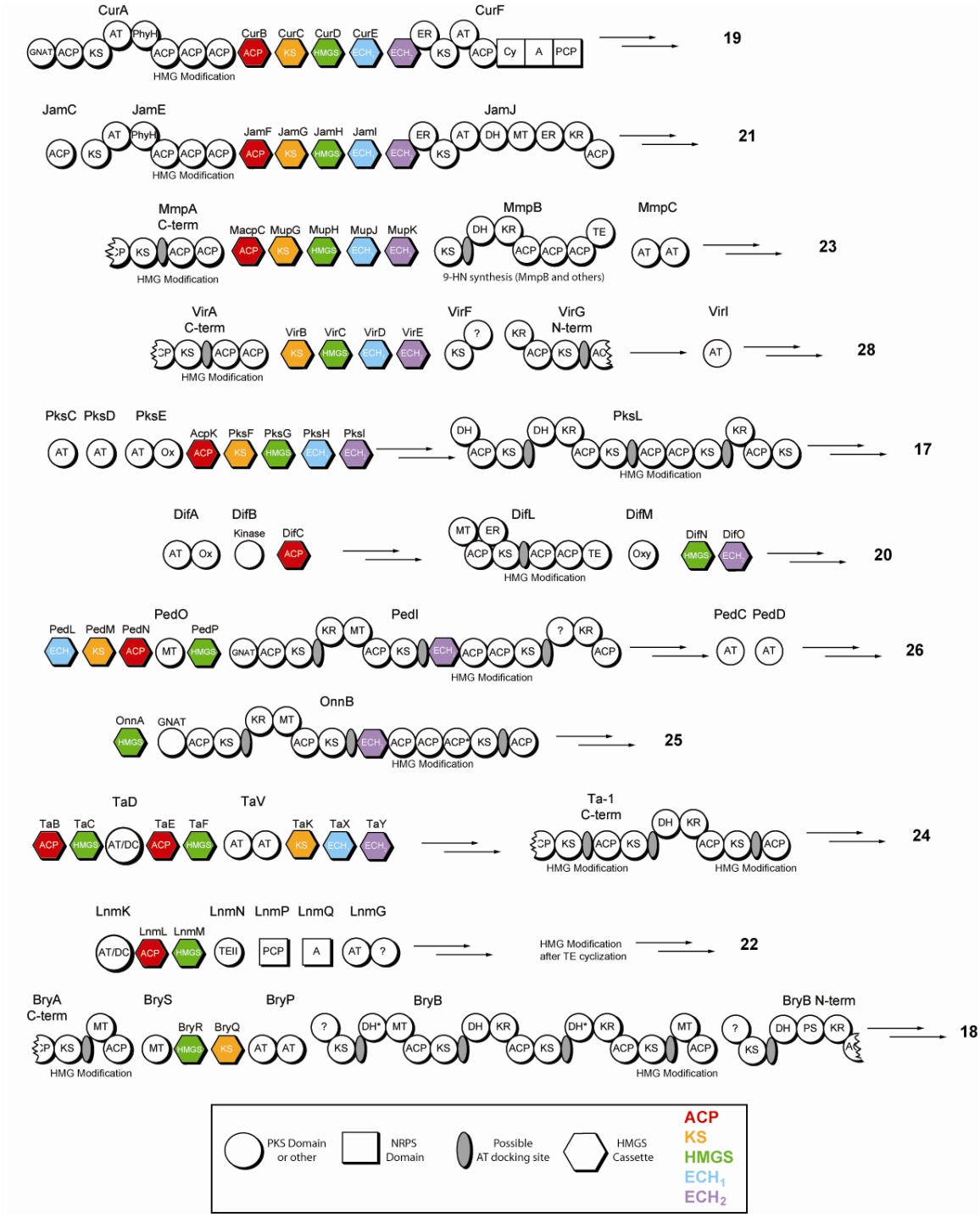
structural characterization of the HMGS enzymes from the myxovirescin system (where two orthogonal HMGS/ACP_D pairs exist).

Finally, the reports that a small number of type III PKSs accept their starter units from acyl-ACPs instead of acyl-CoAs has prompted initial exploratory studies into the molecular mechanisms involved in this capability. A complete biochemical study comparing the K_m and k_{cat} values between acyl-CoAs and acyl-ACPs has yet to be completed for either the hexanoyl- or acetoacetyl- series of starter units with Gcs, an ACP-utilizing type III PKS from *S. coelicolor*. However, to the best of our knowledge, this is the first time that acyl transfer from ACP to type III PKS has been demonstrated *in vitro*. Additionally, our structural characterization of Gcs has allowed for the construction of a number of surface residue mutant proteins. Following the kinetic characterization of wild-type Gcs, these proteins will be used to probe the surface for residues that allow Gcs to access substrates tethered to large macromolecules as well as small molecules. Furthermore, application of this information to the design of hybrid type I – type III PKS pathways will open the door to generate further diversity using combinatorial biosynthesis of polyketides.

APPENDICES



Appendix A. Alignment of H2-T2 type PKS docking domains. The numbering shown matches the P3P4dock construct. Multiple sequence alignment generated using the ClustalX method within JalView software. Acidic residues are colored red, basic residues are in dark blue, hydrophobics are light blue, and tyrosine residues are colored light green. NCBI accession numbers: AmphB - AAK73513, AmphC - AAK73514, NysB - AAF71775, NysC – AAF71776, AmphI – AAK73501, AmphJ – AAK73502, NysI – AAF71766, NysJ – AAF71767, AsmA – AAM54075, AsmB – AAM54076, ChlA3 – AAZ77696, ChlA4 – AAZ77697, ChmGIV – AAS79462, ChmGV – AAS79462, ConA – AAZ94386, ConB – AAZ94387, ConE – AAZ94390, ConF – AAZ94391, FkbB – AAF86393, FkbC – AAF86392, FscD – AAQ82568, FscE – AAQ82567, GdmAII – AAO06917, GdmAIII = AAO06918, HbmAII – AAY28226, HbmAIII – AAY28227, MycAIV – BAC57031, MycAV – BAC57032, PikAIII – AAC69331, PikAIV – AAC69332, PimS2 – CAC20921, PimS3 – CAC20920, RapA – CAA60460, RapB – CAA60459, RifA – AAC01710, RifB – AAC01711, RifC – AAC01712, SACE_2630 – CAM01914, SACE_2629 – CAM01913, SpnA – AAG23264, SpnB – AAG23265, TylGIV – AAB66507, TylGV – AAB66508.



Appendix B. Architecture of HMGS cassettes within their biosynthetic pathways. Only portions of the PKS or PKS/NRPS pathways are shown. The pathway for Psymberin (**19**) is not shown above, as the sequences are not yet publicly available. *Abbreviations*: A – Adenylation, ACP – acyl carrier protein, AT – acyltransferase, C – condensation, Cy – cyclization, DH – dehydratase, ER – enoyl reductase, FkbH – FkbH homolog, GNAT – GCN5-related *N*-acetyltransferase, KS – ketosynthase, KR – ketoreductase, MT – methyltransferase, PCP – peptidyl carrier protein, PhyH – phytanoyl-CoA dioxygenase, PS – pyrone synthase, TE – thioesterase, unk – unknown function, * - inactive domain.

BIBLIOGRAPHY

1. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **2009,** 48, 4688-4716.
2. Hill, A. M., The biosynthesis, molecular genetics and enzymology of the polyketide-derived metabolites. *Nat. Prod. Rep.* **2006,** 23, 256-320.
3. Newman, D. J.; Cragg, G. M., Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **2007,** 70, 461-477.
4. Newman, D. J.; Cragg, G. M.; Snader, K. M., Natural products as sources of new drugs over the period 1981-2002. *J. Nat. Prod.* **2003,** 66, 1022-1037.
5. Newman, D. J.; Cragg, G. M.; Snader, K. M., The influence of natural products upon drug discovery. *Nat. Prod. Rep.* **2000,** 17, 215-234.
6. Wilson, Z. E.; Brimble, M. A., Molecules derived from the extremes of life. *Nat. Prod. Rep.* **2009,** 26, 44-71.
7. Piel, J., Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals. *Curr. Opin. Med. Chem.* **2006,** 13, 39-50.
8. Kumar, K.; Waldmann, H., Synthesis of Natural Product Inspired Compound Collections. *Angew. Chem. Int. Ed. Engl.* **2009,** 48, 3224-3242.
9. Kennedy, J., Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat. Prod. Rep.* **2008,** 25, 25-34.
10. Cuevas, C.; Pérez, M.; Martin, M. J.; Chicharro, J. L.; Fernández-Rivas, C.; Flores, M.; Francesch, A.; Gallego, P.; Zarzuelo, M.; de La Calle, F.; Garcia, J.; Polanco, C.; Rodríguez, I.; Manzanares, I., Synthesis of ecteinascidin ET-743 and phthalascidin Pt-650 from cyanosafracin B. *Org. Lett.* **2000,** 2, 2545-2548.
11. Cuevas, C.; Francesch, A., Development of Yodelis (travectedin, ET-743). A semisynthetic process solves the supply problem. *Nat. Prod. Rep.* **2009,** 26.
12. Weissman, K. J.; Leadlay, P. F., Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **2005,** 3, 925-936.
13. Reeves, C. D.; Rodriquez, E., Genetic engineering to produce polyketide analogues. *Meth. Enzymol.* **2009,** 459, 295-318.
14. Giraldes, J. W.; Akey, D. L.; Kittendorf, J. D.; Sherman, D. H.; Smith, J. L.; Fecik, R. A., Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat. Chem. Biol.* **2006,** 2, 531-536.
15. Smith, S.; Tsai, S. C., The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **2007,** 24, 1041-1072.
16. Weissman, K. J., Introduction to polyketide biosynthesis. *Meth. Enzymol.* **2009,** 459, 3-16.
17. Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E., Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* **2007,** 76, 11.1-11.27.
18. 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.* **2007,** 16, 2093-2107.

19. Khosla, C., Structures and mechanisms of polyketide synthases. *J. Org. Chem.* **2009,** 74, 6416-6420.
20. Khosla, C.; Kapur, S.; Cane, D. E., Revisiting the modularity of modular polyketide synthases. *Curr. Opin. Chem. Biol.* **2009,** 13, 135-143.
21. Akey, D. L.; Kittendorf, J. D.; Giraldes, J. W.; Fecik, R. A.; Sherman, D. H.; Smith, J. L., Structural basis for macrolactonization by the pikromycin thioesterase. *Nat. Chem. Biol.* **2006,** 2, 537-542.
22. Tsai, S. C.; Ames, B. D., Structural enzymology of polyketide synthases. *Meth. Enzymol.* **2009,** 459, 17-47.
23. Kim, C. Y.; Alekseyev, V. Y.; Chen, A. Y.; Tang, Y.; Cane, D. E.; Khosla, C., Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. *Biochemistry* **2004,** 43, 13892-13898.
24. Siskos, A. P.; Baerga-Ortiz, A.; Bali, S.; Stein, V.; Mamdani, H.; Spiteller, D.; Popovic, B.; Spencer, J. B.; Staunton, J.; Weissman, K. J.; Leadlay, P. F., Molecular basis of Celmer's rules: Stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. *Chem. Biol.* **2005,** 12, 1145-1153.
25. Chen, A. Y.; Cane, D. E.; Khosla, C., Structure-based dissociation of a type I polyketide synthase module. *Chem. Biol.* **2007,** 14, 784-792.
26. Kittendorf, J. D.; Beck, B. J.; Buchholz, T. J.; Seufert, W.; Sherman, D. H., Interrogating the molecular basis for multiple macrolactone ring formation by the pikromycin polyketide synthase. *Chem. Biol.* **2007,** 14, 944-954.
27. Beck, B. J.; Aldrich, C. C.; Fecik, R. A.; Reynolds, K. A.; Sherman, D. H., Substrate recognition and channeling of monomodules from the pikromycin polyketide synthase. *J. Am. Chem. Soc.* **2003,** 125.
28. Beck, B. J.; Aldrich, C. C.; Fecik, R. A.; Reynolds, K. A.; Sherman, D. H., Iterative chain elongation by a pikromycin monomodular polyketide synthase. *J. Am. Chem. Soc.* **2003,** 125, 4682-4683.
29. Weissman, K. J.; Müller, R., Protein-protein interactions in multienzyme megasynthetases. *Chembiochem* **2008,** 9, 826-848.
30. Hopwood, D. A.; Sherman, D. H., Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **1990,** 24, 37-66.
31. Shen, B., Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003,** 7, 285-95.
32. Cronan, J. E.; Thomas, J., Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Meth. Enzymol.* **2009,** 459, 395-433.
33. Schweizer, E.; Hofmann, J., Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol. Mol. Biol. Rev.* **2004,** 68, 501-517.
34. Kittendorf, J. D.; Sherman, D. H., The methymycin/pikromycin pathway: a model for metabolic diversity in natural product biosynthesis. *Bioorg. Med. Chem.* **2009,** 16, 2137-2146.
35. Wilson, D. J.; Xue, Y.; Reynolds, K. A.; Sherman, D. H., Characterization and analysis of the PikD regulatory factor in the pikromycin biosynthetic pathway of Streptomyces venezuelae. *J. Bacteriol.* **2001,** 183, 3468-3475.

36. Xue, Y.; Zhao, L.; Liu, H. W.; Sherman, D. H., A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** 95, 12111-12116.
37. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A., Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **2007,** 24, 162-190.
38. Marrakchi, H.; Zhang, Y. M.; Rock, C. O., Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochem. Soc. Trans.* **2002,** 30, 1050-1055.
39. Austin, M. B.; Noel, A. J. P., The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **2003,** 20, 79-110.
40. Grüschow, S.; Buchholz, T. J.; Seufert, W.; Dordick, J. S.; Sherman, D. H., Substrate profile analysis and ACP-mediated acyl transfer in Streptomyces coelicolor Type III polyketide synthases. *Chembiochem* **2007,** 8, 863-868.
41. Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udwary, D. W.; Austin, M. B.; Noel, J. P.; Challis, G. L., Type III polyketide synthase β-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by Streptomyces coelicolor genome mining. *J. Am. Chem. Soc.* **2006,** 128, 14754-14755.
42. Austin, M. B.; Saito, T.; Bowman, M. E.; Haydock, S.; Kato, A.; Moore, B. S.; Kay, R. R.; Noel, J. P., Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat. Chem. Biol.* **2006,** 2, 494-502.
43. Kittendorf, J. D.; Sherman, D. H., Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **2006,** 17, 597-605.
44. Rodriquez, E.; McDaniel, R., Combinatorial biosynthesis of antimicrobials and other natural products. *Curr. Opin. Microbiol.* **2001,** 4, 526-534.
45. Gaisser, S.; Kellenberger, L.; Kaja, A. L.; Weston, A. J.; Lill, R. E.; Wirtz, G.; Kendrew, S. G.; Low, L.; Sheridan, R. M.; Wilkinson, B.; Galloway, I. S.; Stutzman-Engwall, K.; McArthur, H. A.; Staunton, J.; Leadlay, P. F., Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of Streptomyces avermitilis ATCC31272. *Org. Biomol. Chem.* **2003,** 1, 2840-2847.
46. Menzella, H. G.; Carney, J. R.; Santi, D. V., Rational design and assemby of synthetic trimodular polyketide synthases. *Chem. Biol.* **2007,** 14, 143-151.
47. Menzella, H. G.; Reeves, C. D., Combinatorial biosynthesis for drug development. *Curr. Opin. Microbiol.* **2007,** 10, 238-245.
48. 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.* **2005,** 23, 1171-1176.
49. Chandran, S. S.; Menzella, H. G.; Carney, J. R.; Santi, D. V., Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of ketosynthase domains. *Chem. Biol.* **2006,** 13, 469-474.
50. Haapalainen, A. M.; Merilainen, G.; Wierenga, R. K., The thiolase superfamily: condensing enzymes with diverse reaction specificities. *TRENDS Biochem. Sci.* **2006,** 31, 64-71.

51. Jiang, C.; Kim, S. Y.; Suh, D.-Y., Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol. Phylogenet. Evol.* **2008,** 49, 691-701.
52. Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S., Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in Bacillus subtilis. *J. Bacteriol.* **2009,** 191, 4916-4923.
53. Jeong, J. C.; Srinivasan, A.; Grüschow, S.; Bach, H.; Sherman, D. H.; Dordick, J. S., Exploiting the reaction flexibility of a type III polyketide synthase through in vitro pathway manipulation. *J. Am. Chem. Soc.* **2005,** 127, 64-65.
54. Gu, L.; Jia, J.; Liu, H.; Håkansson, 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.* **2006,** 128, 9014-9015.
55. Lee, T. S.; Khosla, C.; Tang, Y., Orthogonal protein interactions in spore pigment producing and antibiotic producing polyketide synthases. *J. Antibiot.* **2005,** 58, 663-666.
56. Tang, Y.; Lee, T. S.; Kobayashi, S.; Khosla, C., Ketosynthases in the initiation and elongations modules of aromatic polyketide synthases have orthogonal acyl carrier protein specificity. *Biochemistry* **2003,** 42, 6588-6595.
57. Wong, H. C.; Liu, G.; Zhang, Y. M.; Rock, C. O.; Zheng, J., The solution structure of acyl carrier protein from Mycobacterium tuberculosis. *J. Biol. Chem.* **2002,** 277, 15874-15880.
58. Mercer, A. C.; Burkart, M. D., The ubiquitous carrier protein-a window to metabolite biosynthesis. *Nat. Prod. Rep.* **2007,** 24, 750-773.
59. Holak, T. A.; Nilges, M.; Prestegard, J. H.; Gronenborn, A. M.; Clore, G. M., Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry. *Eur. J. Biochem.* **1988,** 175, 9-15.
60. Byers, D. M.; Gong, H., Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem. Cell Biol.* **2007,** 85, 649-662.
61. Pfleger, B. F.; Lee, J. Y.; Somu, R. V.; Aldrich, C. C.; Hanna, P. C.; Sherman, D. H., Characterization and analysis of early enzymes for petrobactin biosynthesis in Bacillus anthracis. *Biochemistry* **2007,** 46, 4147-4157.
62. Zhou, Z.; Lai, J. R.; Walsh, C. T., Directed evolution of aryl carrier proteins in the enterobactin synthetase. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** 104, 11621-11626.
63. Qiao, C.; Wilson, D. J.; Bennett, E. M.; Aldrich, C. C., A Mechanism-Based Aryl Carrier Protein/Thiolation Domain Affinity Probe. *J. Am. Chem. Soc.* **2007,** 129, 6350-6351.
64. Quadri, L. E.; Sello, J.; Keating, T. A.; Weinreb, P. H.; Walsh, C. T., Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. . *Chem. Biol.* **1998,** 5, 631-645.
65. Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T., A new enzyme superfamily - The phosphopantetheinyl transferases. *Chem. Biol.* **1996,** 3, 923-936.
66. Koglin, A.; Walsh, C. T., Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep.* **2009,** 26, 987-1000.

67. Sanchez, 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.* **2001,** 8, 725-738.
68. Miyanaga, A.; Funa, N.; Awakawa, T.; Horinouchi, S., Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** 105, 871-876.
69. Van Bambeke, F.; Harms, J. M.; Van Laethem, Y.; Tulkens, P. M., Ketolides: pharmacological profile and rational positioning in the treatment of respiratory tract infections. . *Expert Opin. Pharmacother.* **2008,** 9, 267-283.
70. Roberts, M. C., Update on macrolide-lincosamide-streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS Microbiol. Lett.* **2008,** 282, 147-159.
71. Xue, Y.; Wilson, D. J.; Sherman, D. H., Genetic architecture of the polyketide synthases for methymycin and pikromycin series macrolides. *Gene* **2000,** 245, 203-211.
72. Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H., Synthesis and biochemical analysis of complex chain-elongation intermediates for interrogation of molecular specificity in the erythromycin and pikromycin polyketide synthases. *J. Am. Chem. Soc.* **2009,** 131, 15784–15793.
73. Fortman, J. L.; Sherman, D. H., Utilizing the power of microbial genetics to bridge the gap between the promise and the application of marine natural products. *Chembiochem* **2005,** 6, 960-978.
74. 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.* **2007,** 70, 67-74.
75. Lopanik, N. B.; Shields, J. A.; Buchholz, T. J.; Rath, C. M.; Hothersall, J.; Haygood, M. G.; Håkansson, K.; Thomas, C. M.; Sherman, D. H., In vivo and in vitro trans-acylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. *Chem. Biol.* **2008,** 15, 1175-1186.
76. Ku, B.; Cha, J.; Srinivasan, A.; Kwon, S. J.; Jeong, J. C.; Sherman, D. H.; Dordick, J. S., Chip-based polyketide biosynthesis and functionalization. *Biotechnol. Prog.* **2006,** 22, 1102-1107.
77. Maier, T.; Jenni, S.; Ban, N., Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **2006,** 311, 1258-1262.
78. 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.* **2006,** 103, 11124-9.
79. Fischbach, M. A.; Walsh, C. T., Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006,** 106, 3468-3496.
80. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **2001,** 40, 15464-15470.

81. Tsuji, S. Y.; Cane, D. E.; Khosla, C., Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **2001,** 40, 2326-2331.
82. Worthington, A. S.; Hur, G. H.; Meier, J. L.; Cheng, Q.; Moore, B. S.; Burkart, M. D., Probing the compatibility of type II ketosynthase-carrier protein partners. *Chembiochem* **2008,** 9, 2096-2103.
83. Tang, L.; Fu, H.; McDaniel, R., Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chem. Biol.* **2000,** 7, 77-84.
84. Floss, H. G., Combinatorial biosynthesis - potential and problems. *J. Bacteriol.* **2006,** 124, 242-257.
85. Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. F.; Leadlay, P. F.; Weissman, K. J., The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **2003,** 10, 723-731.
86. Kumar, P.; Li, Q.; Cane, D. E.; Khosla, C., Intermodular communication in modular polyketide synthases: structural and mutational analysis of linker mediated protein-protein interactions. *J. Am. Chem. Soc.* **2003,** 125, 4097-4102.
87. Weissman, K. J., Single amino acid substitutions alter the efficiency of docking in modular polyketide biosynthesis. *Chembiochem* **2006,** 7, 1334-42.
88. Wu, N.; Cane, D. E.; Khosla, C., Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **2002,** 41, 5056-5066.
89. Wu, N.; Tsuji, S. Y.; Cane, D. E.; Khosla, C., Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* **2001,** 123, 6465-6474.
90. Weissman, K. J., The structural basis for docking in modular polyketide biosynthesis. *Chembiochem* **2006,** 7, 485-494.
91. Thattai, M.; Burak, Y.; Shraiman, B. I., The origins of specificity in polyketide protein interactions. *PLoS Comput. Biol.* **2007,** 3, 1827-1835.
92. Myszka, D. G., Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. *Meth. Enzymol.* **2000,** 323, 325-340.
93. Papalia, G. A.; Leavitt, S.; Bynum, M. A.; Katsamba, P. S.; Wilton, R.; Qiu, H.; Steukers, M.; Wang, S.; Bindu, L.; Phogat, S.; Giannetti, A. M.; Ryan, T. E.; Pudlak, V. A.; Matusiewicz, K.; Michelson, K. M.; Nowakowski, A.; Pham-Baginski, A.; Brooks, J.; Tieman, B. C.; Bruce, B. D.; Vaughn, M.; Baksh, M.; Cho, Y. H.; Wit, M. D.; Smets, A.; Vandersmissen, J.; Michiels, L.; Myszka, D. G., Comparative analysis of 10 small molecules binding to carbonic anhydrase II by different investigators using Biacore technology. *Anal. Biochem.* **2006,** 359, 94-105.
94. Richter, C. D.; Nietlispach, D.; Broadhurst, R. W.; Weissman, K. J., Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **2008,** 4, 75-81.
95. Nieba, L.; Nieba-Axmann, S. E.; Persson, A.; Hämäläinen, M.; Edebratt, F.; Hansson, A.; Lidholm, J.; Magnusson, K.; Karlsson, Å. F.; Plückthun, A.,

BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Anal. Biochem.* **1997,** 252, 217-228.
96. Owicki, J. C., Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *J. Biomolec. Screen.* **2000,** 5, 297-306.
97. Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C., Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **1999,** 6, 117-125.
98. Yan, J.; Gupta, S.; Sherman, D. H.; Reynolds, K. A., Functional dissection of a multimodular polypeptide of the pikromycin polyketide synthase into monomodules using a matched pair of heterologous docking domains. *Chembiochem* **2009,** 10, 1537-1543.
99. Straight, P. D.; Fischbach, M. A.; Walsh, C. T.; Rudner, D. Z.; Kolter, R., A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** 104, 305-310.
100. 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.* **2002,** 25, 8-15.
101. Delproposto, J.; Majmudar, C. Y.; Smith, J. L.; Brown, W. C., Mocr: A novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expression Purif.* **2009,** 63, 40-49.
102. Guerrero, S. A.; Hecht, J.-J.; Hormann, B.; Biebl, H.; Singh, M., Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl. Microbiol. Biotechnol.* **2001,** 56, 718-723.
103. 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* **2001,** 291, 1790-1792.
104. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Macromolec. Crystallogr. Pt A* **1997,** 276, 307-326.
105. Sanishvili, R.; Nagarajan, V.; Yoder, D.; Becker, M.; Xu, S.; Corcoran, S.; Akey, D. L.; Smith, J. L.; Fischetti, R. F., A 7 micron mini-beam improves diffraction data from small or imperfect crystals of macromolecules. *Acta Crystallogr. D Biol. Crystallogr.* **2008,** 64, 425-35.
106. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **2002,** 58, 1948-54.
107. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004,** 60, 2126-32.
108. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994,** 50, 760-3.
109. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **1997,** 53, 240-55.

110. Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **2004,** 32, W615-W619.
111. Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J., The Jalview Java alignment editor. *Bioinformatics* **2004,** 20, 426-427.
112. González-Lergier, J.; Broadbelt, L. J.; Hatzimanikatis, V., Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J. Am. Chem. Soc.* **2005,** 127, 9930-9938.
113. Nelson, T. J. A., D.L., Neuroprotective versus tumorigenic protein kinase C activators. *TRENDS Biochem. Sci.* **2009,** 34, 136-145.
114. Banarjee, S.; Wang, Z.; Mohammad, M.; Sarkar, F. H.; Mohammad, R. M., Efficacy of selected natural products as therapeutic agents against cancer. *J. Nat. Prod.* **2008,** 71, 492-496.
115. Sun, M.-K.; Hongpaisan, J.; Alkon, D. L., Postischemic PKC activation rescues retrograde and anterograde long-term memory. *Proc. Natl. Acad. Sci. USA* **2009,** 106, 14676-14680.
116. Khan, T. K.; Nelson, T. J.; Verman, V. A.; Wender, P. A.; Alkon, D. L., A cellular model of Alzheimer's disease therapeutic efficacy: PKC activation reverses Abeta-induced biomarker abnormality on cultured fibroblasts. *Neurobiol. Dis.* **2009,** 34, 332-339.
117. Nelson, T. J.; Cui, C.; Luo, Y.; Alkon, D. L., Reduction of beta-amyloid levels by novel PKCepsilon activators. *J. Biol. Chem.* **2009**, epub 2009 Oct 22 DOI:.
118. Singh, R.; Sharma, M.; Joshi, P.; Rawat, D. S., Clinical status of anti-cancer agents derived from marine sources. *Anticancer Agents Med. Chem.* **2008,** 8, 603-617.
119. Cheng, Y.-Q.; Tang, G.-L.; Shen, B., Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** 100, 3149-3154.
120. Tang, G.-L.; Cheng, Y.-Q.; Shen, B., Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem. Biol.* **2004,** 11, 33-45.
121. Wu, K.; Chung, L.; Revill, W. P.; Katz, L.; Reeves, C. D., The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **2000,** 251, 81-90.
122. Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J., The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** 104, 1506-1509.
123. 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.* **2006,** 103, 8977-8982.
124. Chen, X. H.; Vater, J.; Piel, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsi, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A. W.; Gottschalk, G.; Süssmuth, R. D.; Borriss, R., Structural and functional characterization of three polyketide

synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* **2006,** 188, 4024-4036.
125. El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E. R.; Simpson, T. J.; Thomas, C. M., Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **2003,** 10, 419-430.
126. Pulsawat, N.; Kitani, S.; Nihira, T., Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **2007,** 393, 31-42.
127. Simunovic, V.; Müller, R., 3-Hydroxy-3-methylglutaryl-CoA-like synthases direct the formation of methyl and ethyl side groups in the biosynthesis of the antibiotic myxovirescin A. *Chembiochem* **2007,** 8, 497-500.
128. Simunovic, V.; Müller, R., Mutational analysis of the myxovirescin biosynthetic gene cluster reveals novel insights into the functional elaboration of polyketide backbones. *Chembiochem* **2007,** 8, 1273-1280.
129. 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* **2006,** 7, 1206-1220.
130. Piel, J.; Hui, D.; Wen, G.; Butzke, D.; Platzer, M.; Fusetani, N.; Matsunaga, S., Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** 101, 16222-16227.
131. Piel, J., A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** 99, 14002-14007.
132. Piel, J.; Wen, G.; Platzer, M.; Hui, D., Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *Chembiochem* **2004,** 5, 93-98.
133. Fisch, K. M.; Gurgui, C.; Heycke, N.; van der Sar, S. A.; Anderson, S. A.; Webb, V. L.; Taudien, S.; Platzer, M.; Rubio, B. K.; Robinson, S. J.; Crews, P.; Piel, J., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **2009,** 5, 450-452.
134. Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J.; Sherman, D. H.; Gerwick, W. H., Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2004,** 67, 1356-1367.
135. Gu, L.; Wang, B.; Kilkarni, 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* **2009,** 459, 731-735.
136. 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*. *Chem. Biol.* **2004,** 11, 817-833.

137. Liu, T.; Huang, Y.; Shen, B., Bifunctional acyltransferase/decarboxylase LnmK as the missing link for beta-alkylation in polyketide biosynthesis. *J. Am. Chem. Soc.* **2009,** 131, 6900-6901.
138. Lange, B. M.; Rugan, T.; Martin, W.; Croteau, R., Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** 97, 13172-13177.
139. 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.* **2007,** 14, 835-846.
140. 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* **2006,** 45, 12756-12766.
141. 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 ECH_2 catalytic domain of CurF from *Lyngbya majuscula*: Insights inot a decarboxylase involved in polyketide chain β-branching. *J. Biol. Chem.* **2007,** 282, 35954-35963.
142. Campobasso, N.; Patel, M.; Wilding, I. E.; Kallender, H.; Rosenberg, M.; Gwynn, M. N., *Staphylococcus aureus* 3-hydroxy-3-methylglutaryl-CoA synthase. *J. Biol. Chem.* **2004,** 279, 44883-44888.
143. Steussy, C. N.; Robison, A. D.; Tetrick, A. M.; Knight, J. T.; Rodwell, V. W.; Stauffacher, C. V.; Sutherlin, A. L., A structural limitation on enzyme activity: the case of HMG-CoA synthase. *Biochemistry* **2006,** 45, 14407-14414.
144. Steussy, C. N.; Vartia, A. A.; Burgner, J. W., II; 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* **2005,** 44, 14256-14267.
145. Theisen, M. L.; 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.* **2004,** 101, 16442-16447.
146. 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.* **2006,** 103, 11491-11496.
147. Calderone, C. T., Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* **2008,** 25, 845-853.
148. Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J., The Jalview Java alignment editor. *Bioinformatics* **2004,** 20, 426-427.
149. 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* **2007,** 16, 2093-2107.

150. Arthur, C. J.; Williams, C.; Pottage, K.; Płoskoń, E.; Findlow, S. C.; Burston, S. G.; Simpson, T. J.; Crump, M. P.; Crosby, J., Structure and malonyl CoA-ACP transacylase binding of streptomyces coelicolor fatty acid synthase acyl carrier protein. *ACS Chem. Biol.* **2009,** 4, 625-636.
151. Gu, L.; Geders, T. W.; Wang, B.; Gerwick, W. H.; Håkansson, K.; Smith, J. L.; Sherman, D. H., GNAT-like strategy for polyketide chain initiation. *Science* **2007,** 318, 970-974.
152. DelProposto, J.; Majmudar, C. Y.; Smith, J. L.; Brown, W. C., Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.* **2009,** 63, 40-49.
153. Bahnson, B. J., An atomic-resolution mechanism of 3-hydroxy-3-methylglutaryl-CoA sythase. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** 101, 16399-16400.
154. Dorrestein, P. C.; Blackhall, J.; Straight, P. D.; Fischbach, M. A.; Garneau-Tsodikova, S.; Edwards, D. J.; McLaughlin, S.; Lin, M.; Gerwick, W. H.; Kolter, R.; Walsh, C. T.; Kelleher, N. L., Activity screening of carrier domains within nonribosomal peptide synthetases using complex substrate mixtures and large molecule mass spectrometry. *Biochemistry* **2006,** 45, 1537-1546.
155. Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T., A new enzyme superfamily - The phosphopantetheinyl transferases. *Chemistry & Biology* **1996,** 3, 923-936.
156. Austin, M. B.; Noel, A. J. P., The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports* **2003,** 20, 79-110.
157. Gross, F.; Luniak, N.; Perlova, O.; Gaitatzis, N.; Jenke-Kodadma, H.; Gerth, K.; Gottschalk, D.; Dittmann, E.; Muller, R., Bacterial type III polyketide sythases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **2006,** 185, 28-38.
158. Altomare, C.; Perrone, G.; Zonno, M. C.; Evidente, A.; Pengue, R.; Fanti, F.; Polonelli, L., Biological characterization of fusapyrone and deoxyfusapyrone, two bioactive secondary metabolites of Fusarium semitectum. *J. Nat. Prod.* **2000,** 63, 1131-1135.
159. Kanai, A.; Kamino, T.; Kuramochi, K.; Kobayashi, S., Synthetic studies directed toward the assembly of the C-glycoside fragment of the telomerase inhibitor D8646-2-6. *Org. Lett.* **2003,** 5, 2837-2839.
160. Piel, J.; Hertweck, C.; Shipley, P. R.; Hunt, D. M.; Newman, M. S.; Moore, B. S., Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase. *Chem. Biol.* 7, 943-955.
161. Austin, M. B.; Izumikawa, M.; Bowman, M. E.; Udwary, D. W.; Ferrer, J. L.; Moore, B. S.; Noel, J. P., Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. *J. Biol. Chem.* **2004,** 279, 45162-45174.
162. Abe, I.; Watanabe, T.; Noguchi, H., Enzymatic formation of long-chain polyketide pyrones by plant type III polyketide synthases. *Phytochem.* **2004,** 65, 2447-2453.

163. Morita, H.; Takahashi, Y.; Noguchi, H.; Abe, I., Enzymatic formation of unnatural aromatic polyketides by chalcone synthase. *Biochem. Biophys. Res. Comm.* **2000,** 279, 190-195.
164. Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002,** 417, 141-147.
165. Challis, G. L.; Hopwood, D. A., Synergy and contigency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** 100, 14555-14561.
166. Nett, M.; Ikeda, H.; Moore, B. S., Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **2009,** 26, 1362 - 1384.
167. Funa, N.; Ohnishi, Y.; Ebizuka, Y.; Horinouchi, S., Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in *Streptomyces griseus*. *J. Biol. Chem.* **2002,** 277, 4628-4635.
168. Funa, N.; Ohnishi, Y.; Ebizuka, Y.; Horinouchi, S., Alteration of reaction and substrate specificity of a bacterial type III polyketide synthase by site-directed mutagenesis. *Biochem. J.* **2002,** 367, 781-789.
169. Funa, N.; Ozawa, H.; Hirata, A.; Horinouchi, S., Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in Azotobacter vinelandii. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** 103, 6356-6361.
170. Izumikawa, M.; Shipley, P. R.; Hopke, J. N.; O'Hare, T.; Xiang, L. K.; Noel, J. P.; Moore, B. S., Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from Streptomyces coelicolor A3(2). *J. Indust. Microbiol. Biotechnol.* **2003,** 30, 510-515.
171. Long, F.; Vagin, A. A.; Young, P.; Murshudov, G. N., BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **2008,** 64, 125-132.