

Biochemical Studies of Polyketide Beta-Branching and Polyketide Synthase Module
Architecture

by

Samuel Thomas Slocum

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

Doctoral Committee:

Professor David Sherman, Co-Chair
Professor Janet Smith, Co-Chair
Associate Professor Gregory J. Dick
Professor Alexander J. Ninfa
Associate Professor Raymond C. Trievel

Samuel Thomas Slocum

stslocum@umich.edu

ORCID iD: 0000-0001-5942-4367

© Samuel Thomas Slocum 2017
All Rights Reserved

Dedicated to my parents, family, and loved ones that made this possible.

Acknowledgements

I would like to express my gratitude to my research advisors Dr. David H. Sherman and Dr. Janet L. Smith. They have provided excellent mentorship and scientific guidance from the beginning of my graduate career, and I cannot thank them enough for their continued support. I also must thank my thesis committee: Dr. Gregory J. Dick, Dr. Raymond C. Trievel, and Dr. Alexander J. Ninfa. Without the contributions of each committee helping move my thesis projects forward, I would not be where I am today.

I would also like to specifically thank a few colleagues that made significant contributions of time, energy, and advice at the bench. Dr. Joe Chemler was my first research advisor during my rotation in the Sherman lab, and I also received invaluable training from Dr. Eli Eisman, Dr. Andrew Lowell, and Dr. Ashu Tripathi. Additionally, the day-to-day contributions of every other Sherman and Smith lab member cannot be overstated. It has been a privilege to work in such a collaborative, cooperative, and rigorous atmosphere.

My esteemed collaborators have also contributed their effort and expertise to the success of our projects together. In particular, I would like to thank Dr. Phil McClory, Dr. Joseph Eschweiler, and Chunyi Zhao for their help in moving our work forward.

I am also grateful to those who helped start me down this path. My undergraduate education and mentorship has been invaluable, and I am thankful for the help of Dr. Mark Stocksdale, Dr. Michael Deibel, Dr. Corinne Deibel, Dr. John Howell, and Dr. Kalani Seu at Earlham as well as Dr. Hugh O'Neill at Oak Ridge National Labs.

Finally, I would not have been able to make it this far without the support of my family and friends. I can't thank them enough for their time, understanding, care, and compassion. I especially would like to thank Maryann Knowles, Tom Knowles, Tommy Knowles, Max Slocum, Tom Slocum, Ellen Kontowicz, Billy Link, and Carly Link for everything they have

done to help me through the difficult times. Lastly and most importantly, I am deeply grateful for Dr. Elizabeth Cogan, who is a source of inspiration and support in every aspect of my life.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Figures	vii
Abstract.....	ix
Chapter I: Introduction	1
1.1 Natural Product Medicines	1
1.2 Polyketide Natural Product Biosynthesis	3
1.3 Modular Architecture and Interactions	8
1.4 Beta Alkylation.....	12
1.5 Bryostatins	19
1.6 Summary	22
Chapter II: Polyketide Beta-Branching in Bryostatin Biosynthesis	24
2.1 Introduction	24
2.2 Results	42
2.3 Discussion	57
2.4 Summary	62
2.5 Conclusion	62
2.6 Materials and Methods.....	63
2.7 Supplemental Data	68
Chapter III: Structural Analysis of PKS Proteins by Ion-Mobility Mass Spectrometry	73
3.1 Introduction	73
3.2 Results	82
3.3 Discussion	88
3.4 Summary	91
3.5 Conclusion	91
3.6 Materials and Methods.....	92
3.7 Supplemental Data	95
Chapter IV: Intermodular Interactions of Polyketide Synthases	97
4.1 Introduction	97
4.2 Results	101
4.3 Discussion	106
4.4 Summary	107
4.5 Conclusion	108
4.6 Materials and Methods.....	108
Chapter V: General Discussion and Future Directions	112

5.1 Introduction and Overview	112
5.2 The Continuing Progress of Bryostatins	112
5.4 Conclusions	116
References	117

List of Figures

Figure 1: Schematic representation of initiation, iteration, and tailoring steps of natural product biosynthesis	4
Figure 2: Model polyketide synthase and fatty acid synthase systems	7
Figure 3: Beta alkylation strategies in polyketide synthase products	14
Figure 4: Diversity of structures installed by beta branching cassettes	16
Figure 5: The source and diversity of the bryostatin family	19
Figure 6: The bryostatin gene cluster	21
Figure 7: Bryostatin 1 and synthetic analogs	30
Figure 8: The life cycle of <i>Bugula neritina</i>	31
Figure 9: FISH microscopy of <i>B. neritina</i> larvae	32
Figure 10: Functional annotation of the bryostatin gene cluster from cosmid sequencing data	35
Figure 11: Shotgun sequencing data MDLC00000000 contig 14	41
Figure 12: Phylogenetic analysis of BryU	43
Figure 13: Mass spectrometric analysis of BryR HMGS activity	44
Figure 14: Phylogenetic analysis of BryT	45
Figure 15: Mass spectrometric analysis of BryT dehydration and BryA O-MT carboxyl methylation of HMG-ACP	47
Figure 16: Chemoenzymatic preparation of dideuterated 3-hydroxy-3-methylglutaryl-CoA	48
Figure 17: LC/MS analysis of γ-d₂-HMG-CoA dehydration by BryT	50
Figure 18: Acid-catalyzed mechanism of enoyl-CoA hydratases	51
Figure 19: NMR analysis of HMG- and 3MG-CoA generated by BryT	52
Figure 20: Phosphopantetheine ejection ions of HMG-ACP_A species treated with TMS-diazomethane	56
Figure 21: Updated model of bryostatin beta branching	60
Figure S1: SDS-PAGE of BryR, BryU, and BryA Mod3 ACP	68
Figure S2: SDS-PAGE of BryT with and without coexpressed chaperones	69
Figure S3: Comparison of HPLC traces of reactions of HMG-CoA with BryT, CurE, and JamI	70
Figure S4: Changes in deuteration of 3MG-CoA with different length reactions	71
Figure S5: Representative HPLC of preparative-scale BryT reaction with HMG-CoA	71
Figure S6: Deconvoluted intact protein masses of HMG-BryA Mod3 ACP with and without TMS-diazomethane treatment	72
Figure 22: Polyketide synthase extension and potential reductive domain architectures	74

Figure 23: Comparisons of the architecture of biosynthetic systems	77
Figure 24: Ion mobility mass spectrometry instrumentation	80
Figure 25: Pseudo-atomic resolution models of PikAIII catalytic states	81
Figure 26: Collision-induced unfolding (CIU) fingerprints of PikAIII catalytic states	83
Figure 27: De novo models generated from IM-MS data	84
Figure 28: Observed KS-AT species over time	87
Figure 29: Size exclusion chromatography of KS-AT over time	88
Figure S7: SDS-PAGE of PikAIII and KS-AT with dimerization element	95
Figure S8: SDS-PAGE of elution fraction of KS-AT size exclusion chromatography	96
Figure 30: Coupled PikAIII/PikIV assay system	101
Figure 31: Potential docking domain residues at the termini of bry PKS proteins	102
Figure 32: HPLC analysis of 10-dml/narbonolide production by PKS modules bearing heterologous docking domains	103
Figure 33: Fluorescence polarization of BryA/BryB docking domain peptides	105
Figure 34: Hypothetical introduction of an unnatural vinyl methyl ester	114

Abstract

Natural products are bioactive molecules produced across the tree of life and have been used medicinally by humans for millennia. Isolation and characterization of these molecules reveals an incredible diversity of structure, bioactivity, and biosynthetic mechanism. These biosynthetic systems have great potential for engineering to produce novel structures, however, to take advantage of these systems they must be studied to understand their natural behavior. Polyketide synthases (PKS) are a diverse class of enzymes that integrate distinct catalytic domains into modules that are repeated to form large molecular assembly lines. These PKS “megaenzymes” have intricate domain movements and protein-protein interactions that enable their functionality. Investigations of several key properties of these biosynthetic systems were carried out using diverse molecular techniques.

Ion-mobility mass spectrometry (IM-MS) integrates gas-phase structural data with mass detection and was used to investigate the domain movements and structure of PikAIII, a single module from the pikromycin pathway. *De novo* modeling was successfully carried out using these data, and IM-MS was validated as an investigative tool for the study of PKS structure and behavior. PikAIII and associated substrates were also used to investigate the docking interactions that facilitate the transfer of polyketide intermediates through the bryostatin pathway. By engineering bryostatin docking sequences into the established PikAIII chemoenzymatic system, these interactions were investigated.

Bryostatins are PKS-derived natural products that are potent PKC modulators and are under investigation for several clinical applications. Two pharmacologically important vinyl

methyl ester moieties are installed in the bryostatin structure by a beta-branching cassette, but several components were missing from this uncommon expansion of PKS machinery. Because bryostatins are produced by an obligate symbiont (*Candidatus Endobugula sertula*), investigation of their biosynthesis is difficult. However, BryT and BryU were identified in a metagenomic sequencing effort, and biochemical investigations of their activities were carried out. It was found that BryU is the donor ACP for the beta-branching cassette, BryT is the ECH₁ with an unexpected product, and the unusual carboxyl O-MT was also identified. These investigations bring the potential for leveraging biosynthetic enzymes to produce novel chemical structures.

Chapter I Introduction

1.1 Natural Product Medicines

Natural products are bioactive molecules produced by natural sources throughout the tree of life. The medical application of natural products stretches back through the majority, if not all, of written human history(1). So-called “traditional medicine” exemplifies early methods to identify species and processing techniques to produce medicines. For example, a Sumerian clay tablet dating to c. 3000 BC details a number of medicinal preparations using more than 200 plants, several of which are still in use today. Our advancement of pharmaceutical science has led from the isolation of quinine from the bark of *Cinchona officinalis*, long used by natives of the Amazon river, to the development of quinine-related compounds such as mefloquine and chloroquine as higher-efficacy analogs(2). Although the use of naturally-occurring medicines has existed for millennia, the ability to parse their complexities at the molecular level has only recently developed.

In spite of this, we have made an amazing amount of progress in an extremely short span of time. As medical science has progressed and the demand for new and better drugs has increased, the process of identifying active compounds and bringing those activities to the clinic have changed dramatically. The fundamental necessity of sampling from nature and identifying bioactivity has remained, but we are now able to identify the molecules responsible for biological activity, search for the target of that activity, and hypothesize the most likely biosynthetic mechanisms within the source organism. While one or more of these questions may

be intractable, biochemical tools and technology open many avenues and detours for drug development.

The modern approach to the development of natural products as drug leads follows a proven and well-worn path. The gold standard of assay-guided chemical extraction from natural sources continues to identify molecules with new, improved, or altered activities. Additionally, new methods to culture once-refractory microbiota and activate cryptic biosynthetic pathways during fermentation are continually improving(3, 4). While these methods may enable provide enough material for structural and biochemical characterization, the isolation of sufficient quantities of compound to pursue clinical development can be challenging(5). Total synthesis is one method to overcome this challenge. Synthetic chemistry provides an alternative to large-scale fermentation and may also allow for the development of derivative molecules that may engender improved efficacy and pharmacological properties. Though the value of these molecules as synthetic targets is often without question, these scaffolds tend to present significant challenges to chemists, often necessitating the simplification of the original natural product lead(6, 7). Derivatives can also be obtained through semisynthetic modification of the natural molecule, but this also depends on large-scale isolation, as is the case for azithromycin synthesis(8).

Alternatively, genetic information may be brought to bear. Through genomic or metagenomic analysis, the original biosynthetic cluster may be revealed(9, 10). Combined with a fermentable host, engineered biosynthetic pathways may result in higher yielding, new, or more diverse compounds than would be accessible by total or semi-synthesis. These outcomes depend upon a fundamental understanding of the enzymes themselves. The biochemical study of

natural product biosynthesis has grown by leaps and bounds, and a great deal remains to be discovered.

1.2 Polyketide Natural Product Biosynthesis

Polyketides are an important and highly diverse class of natural products biosynthesized by a large number of microbial, plant, and a few animal taxa. These molecules have a broad variety of both chemical structures and biological activities, including antitumor, antibacterial, antifungal, and immunomodulatory activities. Relatives of fatty acid synthase (FAS), large multi-peptide assembly lines termed polyketide synthases (PKS) are responsible for the biosynthesis of polyketides. They may act alone or in combination with non-ribosomal peptide synthetases (NRPS), another secondary metabolic system that condenses amino acid residues instead of malonyl derivatives. PKS and NRPS systems produce an amazing diversity of products and present an appealing target for engineering.

Polyketide synthases carry out iterative condensation reactions complemented by other activities to generate their products using a variety of protein architectures. The function of PKS is similar to that of FAS in that it requires an initiation step, repeated selections and condensations of activated extender units accompanied by modifications of the previous extension, a termination step, and tailoring to generate a mature molecule, represented schematically in **Figure 1**. Intermediates are carried as thioester conjugates of either coenzyme A or a phosphopantetheine prosthetic group covalently linked to a small protein domain. Initiation substrates vary greatly, and are derived from a number of metabolic pools. Extension equivalents are provided by specifically selected α -modified malonyl-CoAs. While these basic functions are common among PKS pathways and share similarities with FAS, a variety of domain groupings that diverge dramatically from FAS (and one another) have evolved. They

may be grouped into three major classes (type I, II, and III). Type I PKS uses a non-iterative assembly line of fused enzymatic domains in repeating groups referred to as modules, detailed in later sections. Type II PKS are not fused, but are rather made up of discrete domains that act iteratively. Type III PKS, interestingly, appear to have several activities of polyketide synthesis integrated into a single active site that eliminates much of the required machinery of type I and II systems. These diverse functional arrangements have contributed greatly to the diversity of PKS-derived natural products.

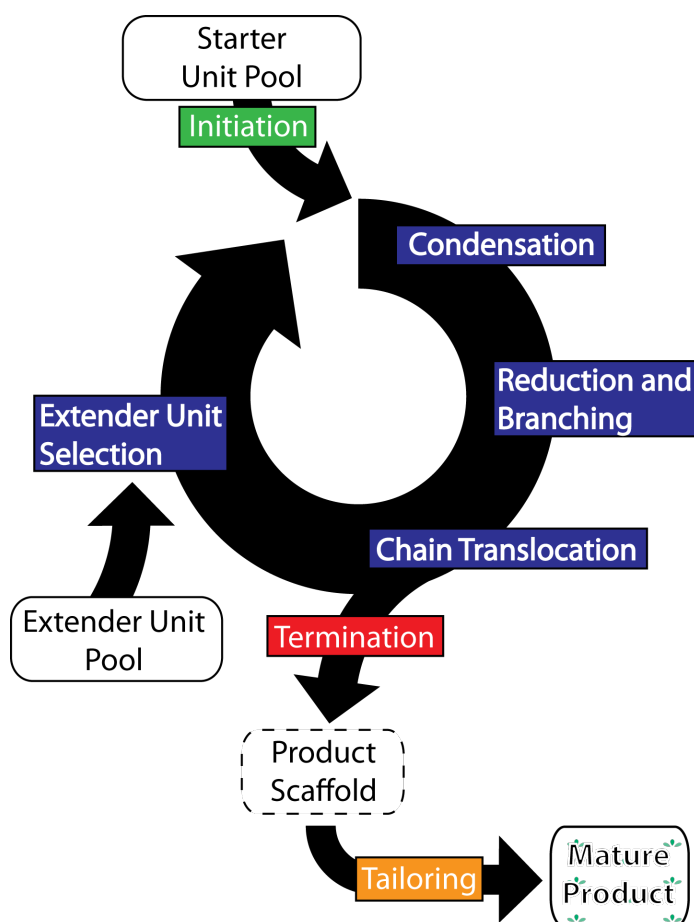


Figure 1. Schematic representation of initiation, iteration, and tailoring steps of natural product biosynthesis. This cycle may be carried out by one or more distinct modules spread over one or more gene products in a biosynthetic system.

1.2.1 Type I Polyketide Synthases

While the biosyntheses of natural products have been attributed to many interesting pathways that are often interrelated by functionality and chemistry, the primary focus of this thesis is the diversity of type I PKS systems. In type I PKS, the activities necessary for each cycle of elongation and modification of a nascent polyketide are provided by a set of domains grouped into a unit referred to as a module. The first module in the pathway initiates synthesis by loading a small molecule starter unit, a series of modules carry out extensions, and a terminal module offloads and often carries out cyclization of the product. The structure of the product of a type I PKS pathway is directly dependent upon the attendant domains within modules at each extension, as each module is typically utilized only once during extension. These properties have become very useful to researchers because they establish the logic of a biosynthetic system that may be leveraged for its predictability and ability to be modified.

Minimal type I PKS modules contain an acyltransferase (AT), ketosynthase (KS), and acyl carrier protein (ACP), which is post-translationally modified by the addition of a CoA-derived phosphopantetheine arm. The AT domain selects a malonyl extender unit to load onto the terminal thiol of the phosphopantetheine arm of the ACP, which then translocates to the KS active site. Here, the extender unit undergoes decarboxylative condensation with the nascent polyketide to yield a two carbon extension. Reductive domains then act sequentially on the newly-generated beta-keto substrate: ketoreductase (KR) domains stereoselectively reduce the ketone to an alcohol, dehydratase (DH) domains eliminate the alcohol leaving a double bond, and enoylreductase (ER) domains reduce the double bond to saturation. Modules may contain a KR, a KR with a DH, or all three reductive domains (**Figure 2**). Other modifying domains may also be found within specialized modules, including methyltransferases, halogenases, cyclases,

isomerases, among others. While similar to FAS, the biochemistry of PKS has the flexibility of a linear assembly line of diverse modules and the ability to recruit modifying and tailoring enzymes. These properties grant PKS the ability to synthesize an incredible diversity of chemical structures.

1.2.2 Trans-Acyltransferase Type I Polyketide Synthases

In addition to intra-modular domains, many pathways utilize standalone domains that transiently interact (thus, “in trans”) to complement the activities of the modules themselves. Most PKS pathways identified to date are “cis-AT” pathways which contain an AT domain within each module. Recently (and with increasing frequency) “trans-AT” pathways are being discovered and studied(11). These pathways contain one or more AT domains that are encoded separately from the large polymodular PKS genes and interact transiently with a specific subset of modules to load extender units (**Figure 2B**). Trans-AT pathways have several interesting properties that are indicative of a different evolutionary history from the many generations of vertical transmission hypothesized for cis-AT pathways. Rather, trans-AT pathways show evidence of horizontal transfer and recruitment of individual components of pathways from different sources. This results in the genetic structure of trans-AT pathways appearing much more fragmented and inconsistent than that of cis-AT pathways. While this complicates the application of conclusions from the study of cis-AT systems, it also provides a new set of biochemical capabilities to be understood and utilized.

In addition to differences in evolutionary history, another example of the difference in biosynthetic logic between cis- and trans-AT pathways is the domain organization of biosynthetic modules. Typically, cis-AT proteins contain an N-terminal KS domain and a C-terminal ACP domain, with relatively consistent alignment of gene and module boundaries. In

contrast, this pattern is exceedingly rare for trans-AT pathways. Almost any domain may be found at the N- or C-terminus of peptides, often splitting modules across two peptides. Without the docking domains found in cis-AT systems, it is unclear how intermediates are transferred from one peptide to the next. While specific examples have been studied(12), this question remains largely unanswered.

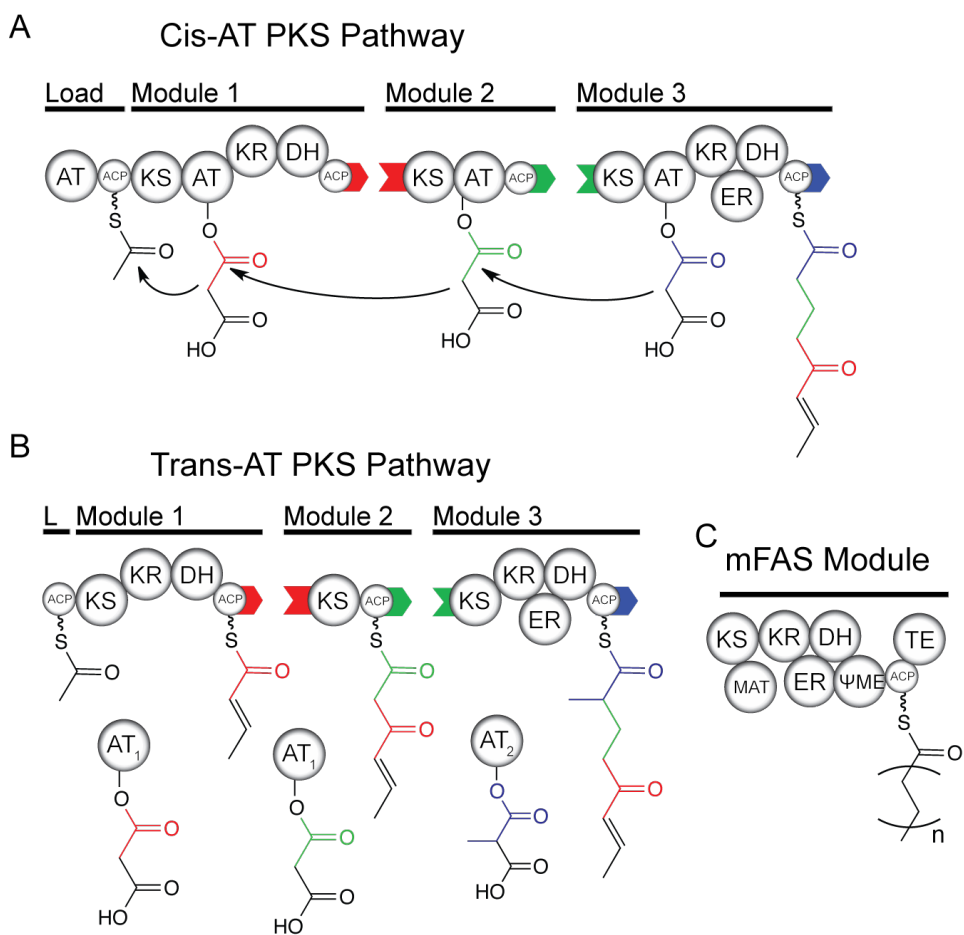


Figure 2. Model polyketide synthase and fatty acid synthase systems. A) Three protein, trimodular PKS pathway with three different reductive domain configurations. B) Model trans-AT PKS pathway with different extender unit selectivities. C) Domain configuration of mFAS “module” shows similar architecture to “fully reducing” PKS modules

The horizontal transfer of genetic components leads to trends in sequence identity of trans-AT pathways that have proven beneficial to researchers. Within cis-AT pathways, KS domains of the same pathway typically clade together, suggesting a long history of vertical transfer within the same lineage of an organism. In trans-AT pathways, the KS domains clade strongly based upon the chemical structure of their predicted substrate(13), indicative of horizontal transfer. This phylogenetic tendency provides predictive value from primary sequence patterns, similar to the patterns found in other common domains, such as the predictive motifs of KR(14). Trans-AT domains themselves and their ability to target specific extender units to specific ACPs within a pathway provide a target for pathway engineering(15, 16).

1.3 Modular Architecture and Interactions

In order to take advantage of the elegant modular PKS system, it is necessary to understand how domains are able to accommodate their unique substrates to work together as modules and how these modules are able to work together to form functional pathways. There are many steps during the catalytic cycle that contribute to the overall selectivity and throughput of a PKS pathway. Within a single module, both the ability of the ACP to interact productively with each domain as well as the ability of the modifying domains to accept the nascent polyketide as a substrate are key factors in synthetic efficiency. Between modules, many important protein-protein interactions have been described for several pathways(17-21). N- and C-terminal docking domains mediate interactions between PKS proteins and ensure correct procession order through the modules and have proven to be highly specific and very important for proper pathway function. These interactions, as well as those between KS and ACP of adjacent modules also appear to be significant. Recent experiments have also shed light on the relative positions of domains within modules as well as how multiple modules are arranged

within the producing organism(22, 23). Integrating the study of the large-scale architecture and atomic-scale interactions of PKS proteins will reveal the basis of polyketide synthase function. A deeper understanding of these phenomena will hopefully enable the application of PKS systems to the production of novel small molecules using modular functionality that is both natural and engineered.

1.3.1 Modular Superstructure

As PKS pathways share a common ancestor with mammalian fatty acid synthase (mFAS), several meaningful comparisons may be made to aid our understanding of PKS function. It appears that PKS modules arose through a series of gene duplication and deletion events, recently when compared to mFAS. The high levels of organization and structure of interdomain linkers(24, 25) within a FAS module contrasts strongly with the large number of disordered loops and linkers often found between domains in PKS pathways; a difference that is indicative of their different evolutionary lineages. PKS has been colloquially described as a naturally-occurring high-throughput screen using comprehensive permutations of mFAS modules as subunits in an assembly line, and the diversity of architectures certainly supports this description.

The addition or omission of domains within modules must certainly alter the structure of a PKS module relative to a FAS module, and a single reading frame may code for anywhere from one to thirteen modules. These variations lead to extremely large multidomain enzymes whose high-level architecture is largely mysterious. This is not to say that this high-level structure of PKS is entirely unknown. Several structural studies of full modules of PKS and NRPS have produced interesting contrasts and comparisons with the known structures of mFAS. The localization of the ACP during the PKS catalytic cycle is one important outcome of these

studies(22). Additional pieces of evidence have also been gleaned from smaller-scale structural studies of incomplete module constructs that describe important protein-protein interfaces. Experiments using KS-AT didomains(26), small-angle x-ray scattering experiments(27), and protein NMR(28) have allowed the deduction of the relative positions of domains in PKS modules through their catalytic cycles.

The movements of domains within modules occur within the context of entire pathways, which is an even more difficult structural question to answer. In the trans-AT subgroup of type I PKS, an interesting hypothesis has been proposed regarding the large-scale architecture formed by all of the PKS proteins in the bacillaene pathway(23). Using several pieces of structural data (primarily a conserved crystal contact) the authors propose that the entire multiprotein pathway forms a semi-stable organelle-scale complex that localizes to the plasma membrane of the cell. A set of structures from an NRPS pathway appears to indicate a degree of flexibility that would preclude this sort of long-range superstructure(29), which makes predictions regarding the numerous mixed PKS/NRPS pathways very difficult. These are very interesting questions about PKS and NRPS pathways that present great challenges to current techniques and thus remain largely mysterious.

New techniques, methodologies, and small molecules are always opening new avenues of study for these difficult problems. We have approached this landscape of questions with an array of constructs from the pikromycin pathway (a type I PKS) and computational analysis of ion mobility mass spectrometry (IM-MS). Using techniques that will be detailed in later sections, we have generated a model for the relative positions of domains within a PKS module during the catalytic cycle based upon collisional cross section and corresponding mass-to-charge information. While the resolution of this method is currently limited, it has the advantage of

expedience when compared to other structural techniques. This high throughput may allow IM-MS to play a critical role in the rapid screening of buffer, catalytic, and time-course effects to facilitate rapid analyses. Further development of IM-MS data acquisition and analysis may even see the technique become a competitor to other structural methods. Because the study of the structure, behavior, and biology of PKS pathways is so important to their engineering, there is a great deal of potential in the development of IM-MS for the study of these huge molecular machines.

1.3.2 Intra- and Inter-Modular Protein-Protein Interactions

Each step of PKS catalysis within a module requires movement of the ACP to a different domain. After condensation in the KS, the ACP must shuttle the intermediate to the remaining domains in the module before passing it to the KS of the next module. Active site recognition of the thioester substrate appears to confer some energetic preference to the localization of the ACP, but this hypothesis has alternatives, such as allosteric changes in domain structure to alter the most favored protein-protein interactions involving the ACP and other domains in the module. Further study of the protein-protein interactions of the ACP as well as modifications of the substrate will clarify the dramatic rearrangements of PKS modules. In particular, a combinatorial study of modules from related pathways using each of their substrates would be highly revealing. Several actinobacterial PKS pathways that produce the core macrolides for related macrolactone antibiotics have been studied. Specifically: erythromycin(30), tylosin(31), pikromycin(32), and juvenimicin(31). Their sequence identities are relatively high, but there are subtle differences in their native substrates that must be accommodated differently between each pathway. Synthetic intermediates have been prepared for several modules within these

pathways, putting them in prime position for combinatorial biochemical study. These modules have the potential to shed a great deal of light on the intricacies of PKS intra-module function.

Another interaction that is critical to the function of PKS pathways is the handoff of intermediates from one module to the next. As a PKS gene may code for a variable number of modules, many ACPs pass processed intermediates to downstream modules that are covalently linked on the same peptide. While this is more entropically favored than interacting with a separate protein, the ACP-KS interaction is still an interesting target for PKS engineering. When modules are not on the same peptide, short protein sequences termed docking domains provide the affinity and specificity for modules to associate and for substrate transfer to occur. These sequences appear to be related in unintuitive ways, with species that are distant phylogenetic relatives carrying PKS pathways with closely related docking domains(17). Structural studies have revealed a great deal about several of these types of docking domains(18-20, 33), and many studies have used them as engineering handles(34).

1.4 Beta Alkylation

The broad chemical diversity achieved by PKS pathways is augmented by several atypical extension strategies. Canonical PKS extension is achieved via decarboxylative Claisen condensation of a malonyl extender unit wherein the α -carbon of the malonyl unit attacks the thioester carbon of the polyketide intermediate. Atypical extensions in many pathways functionalize the beta-carbon and install a wide variety of chemical functionality that would not be achievable within the normal PKS catalytic framework. Most of these extensions are dependent upon the beta-ketone left by minimal PKS modules and use several other activities to finalize their products. The domains that carry out these activities are typically found outside the module and utilize specific protein-protein interactions.

After PKS extension, the beta carbonyl presents a viable target for nucleophilic attack if it is not reduced by modular domains. In the andrimid pathway(35), the nucleophile is found within the intermediate itself (**Figure 3A**), resulting in the formation of a 5-membered ring which is subsequently converted into the pharmacologically important methyl succinimide moiety found in the natural product. In another fascinating example of biosynthetic eccentricity, corallopyronin A is formed by two independently assembled beta-keto PKS products that are merged via cyclization by an as-yet unknown beta-ketone dependent mechanism(36) (**Figure 3B**). The rhizoxin pathway presents an alternative strategy wherein the beta carbon has been reduced to an sp^2 carbon that is attacked by a malonyl extender unit in an enzyme-catalyzed Claisen-like condensation with a subsequent lactonization in the context of a specialized module(37, 38) (**Figure 3C**).

Another form of beta functionalization uses hydroxymethylglutaryl synthase (HMGS) and an acetyl donor to generate a 3-hydroxy-3-methylglutarate-like intermediate (**Figure 3D**). This is then modified by several other enzymes that make up a beta branching cassette, such as those found in curacin(39), myxovirescin(40), bryostatin(41), and several other pathways. There are a handful of examples of beta branching cassettes in cis-AT pathways, though the vast majority are found in trans-AT systems. The downstream functionalization of this HMG-like intermediate can lead to some truly unusual chemical functionalities, which are described in later sections.

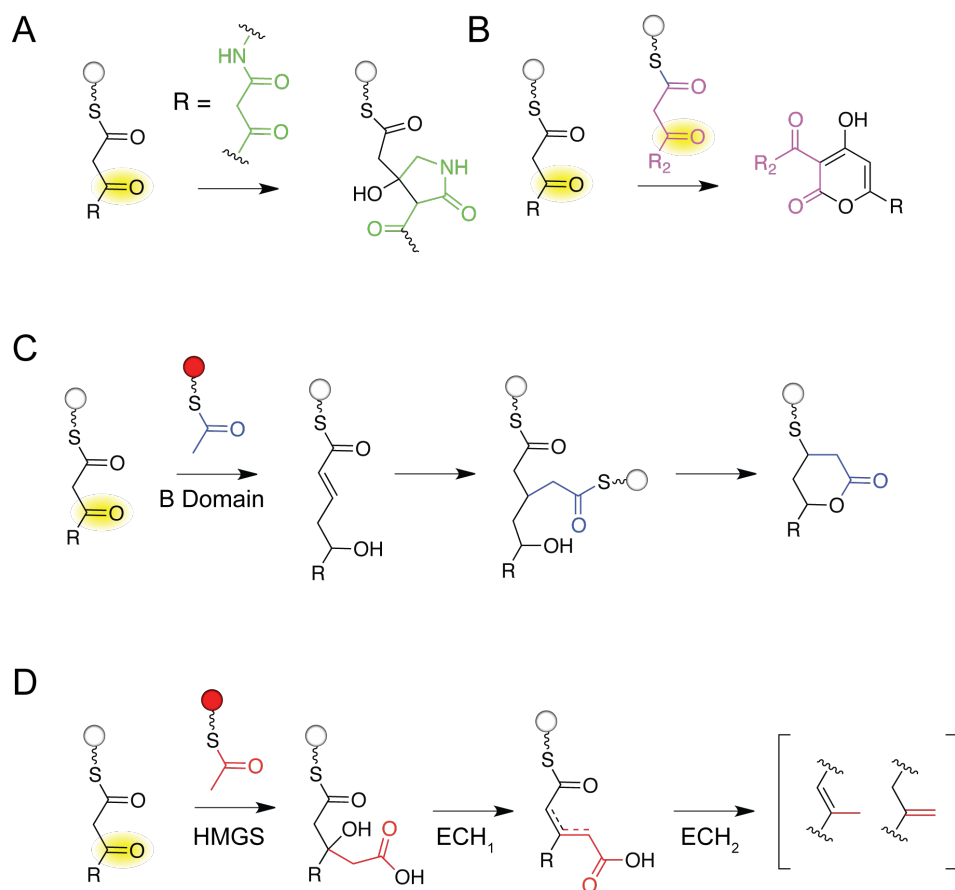


Figure 3. Beta alkylation strategies in polyketide synthase products. A) Andrimid biosynthesis uses an intra-molecular beta carbon attack. B) Corallopyronin is formed by beta attack dependent heterodimerization of two separate polyketides. C) Rhizoxin uses a so-called “B domain” to carry form a lactone branch. D) In a number of pathways, a relative of 3-hydroxy-3-methylglutaryl synthase generates an HMG-like intermediate that is dehydrated and decarboxylated.

1.4.1 Beta Branching Cassettes

HMGS is closely related to HMG-CoA synthase in isoprenoid biosynthesis in primary metabolism. The key differences being the necessity of protein-protein interaction with specific ACPs for the function of HMGS. In the first step of beta branching, HMGS interacts with a specific donor ACP(42, 43) and transfers a donor unit (typically acetyl, but may be propionyl(40)) to its active site cysteine sidechain. Acetyl-HMGS then interacts with a specific

acceptor ACP(44) and, through an enolate intermediate, the alpha carbon of the donor unit attacks the beta carbonyl of the intermediate tethered to the acceptor ACP. A water molecule then hydrolyzes the covalent HMGS intermediate, yielding an HMG-like product tethered to the acceptor ACP phosphopantetheine arm.

The enzymes that modify this HMG-like intermediate are closely related to one another and open the door to an amazing diversity of chemical structures, a few examples of which are highlighted in **Figure 4**. In the majority of pathways, two enoyl-CoA hydratases (ECH) act sequentially to dehydrate (ECH₁) and then decarboxylate (ECH₂) the HMG-like product of HMGS. The activity of the ECHs may result in either a vinyl methyl (Δ^2 isomer) or exomethylene (Δ^3 isomer). This selectivity is likely governed by ECH₂ decarboxylation, which has been studied structurally in the and jamaicamide(45) pathway, which has different regiochemical product to the strikingly similar curacin ECH₂. ECH₁ is a standalone enzyme, and ECH₂ may either be standalone or integrated into a PKS modules. In some cases, pathways contain both standalone and modular ECH₂(46, 47), and correlation of their location in the pathway to the structure of the final product suggests that the standalone ECH₂ generate Δ^3 products while modular ECH₂ generate Δ^2 products in several cases.

The majority of beta branching cassettes process through the ECH₁-ECH₂ pair and are subsequently diversified by various transferases, oxidases, or reductases, leading to the products outlined in **Table 1**. However, bongkrelic acid(48), bryostatin(43), and leinamycin(49) are the only known beta branch-containing products that have structures that are not installed using ECH₂ at some stage of their beta branches.

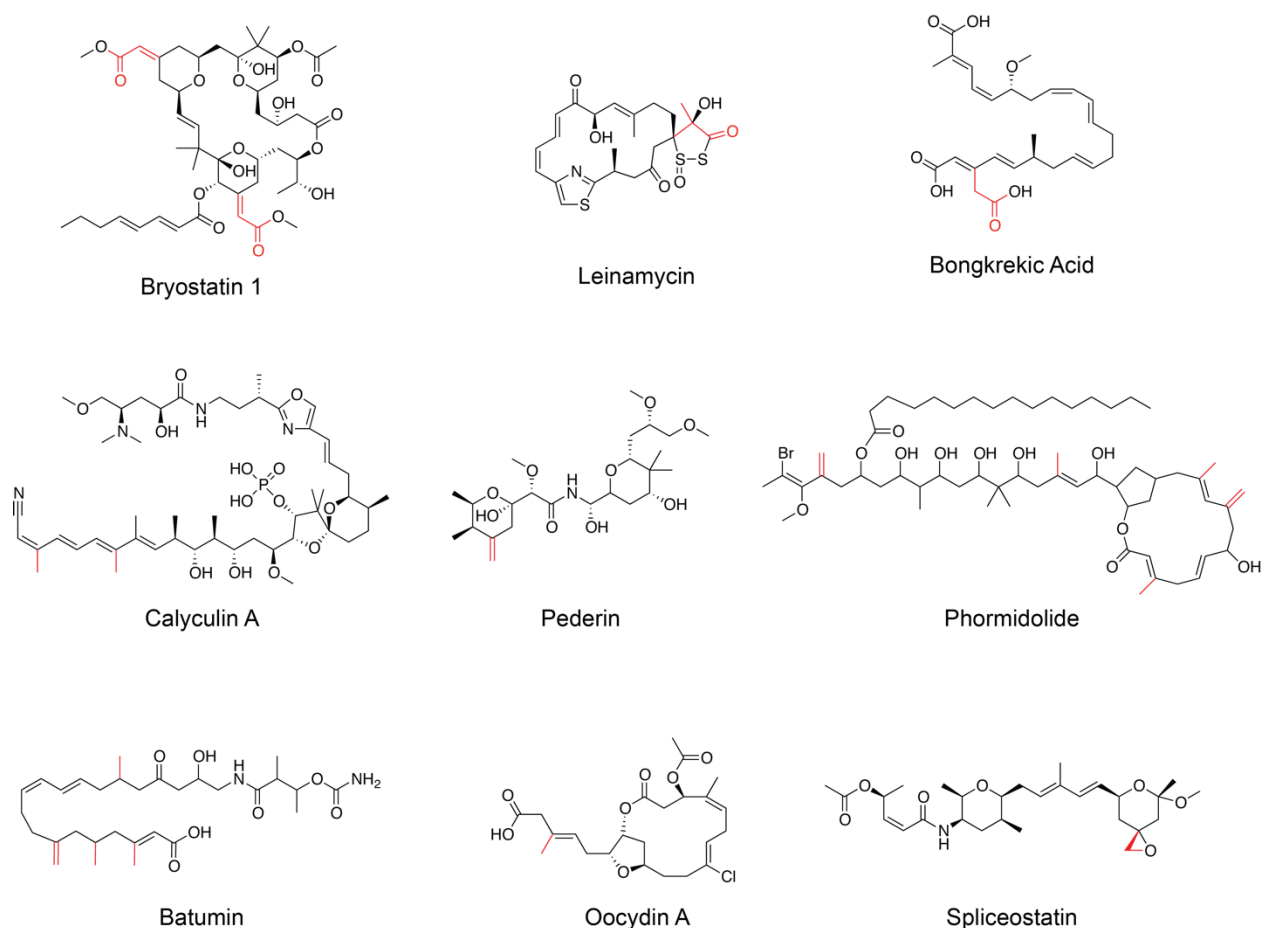


Figure 4. Diversity of structures installed by beta branching cassettes. Red atoms indicate functionality installed by beta branching cassettes.

The bongkreic acid pathway appears to have a selective ECH₂ that only acts during one of the two beta branching events in the pathway, leaving a Δ^3 carboxyl at the second branch position. Leinamycin biosynthesis suggests a truly remarkable series of modifications that include a propionyl-ACP beta branch donor generated by a bifunctional acyltransferase/decarboxylase, a cryptic L-cysteine ligase, and the invocation of an episulfonium ion to generate its characteristic thiazolypentadienone moiety. Bryostatin biosynthesis appears to be the only described pathway that contains a Δ^2 carboxyl installed by a beta branching

cassette. These vinyl methyl ester moieties also appear to require a previously unknown carboxyl O-methyltransferase, a unique transformation among beta branching cassettes. The unique enzymatic steps involved in bryostatin beta branching provide an opportunity to expand our understanding of these uncommon PKS extensions. Additionally, it provides the opportunity to understand the biosynthesis of the unique vinyl methyl ester structure, and add this biosynthetic machinery to the PKS engineering toolbox, thus expanding the potential for engineered systems.

Pathway	Pathway Acc.	HMGS Acc.	Path Type	AT Type	Final Structure	Branches	Organism
Croacin	KJ868728.1	ADI59527.1	PKSI/NRPS	Cis	1 a/b methyl	1	Chondromyces crocatus
Elansolid	HQ680975.1	CAG23985.1	PKSI	Trans	1 a/b methyl	1	Bacillus velezensis FZB42
Mupirocin	AF318063.3	ADC96649.1	PKSI	Trans	1 a/b methyl	1	Pseudomonas fluorescens
Oocydin	JX315604.1	AFX60304.1	PKSI	Trans	1 a/b methyl	1	Serratia plymuthica
Pristinamycin	FN563144.1	AGZ15448.1	PKSI/NRPS	Trans	1 a/b methyl	1	Streptomyces pristinaespiralis
SIAT7248	JQ269659.1	ADA82596.1	PKSI	Trans	1 a/b methyl	1	Streptomyces spp. A7248
Thiomarinol	NC_015708.1	AGN11885.1	PKSI	Trans	1 a/b methyl	1	Streptomyces pristinaespiralis
Virginiamycin	AB283030.2	CBK62727.1	PKSI/NRPS	Trans	1 a/b methyl	1	Streptomyces virginiae
Kalimantacin/Batumin	GU479979.1	AAS98779.1	PKSI/NRPS	Trans	1 a/b, 1 b/y, 2 saturated	4	Pseudomonas fluorescens
Diaphorin	CP003468.1	AFV96140.1	PKSI/NRPS	Trans	1 b/y methyl	1	Candidatus Proffella armatura
Difficidin	KY358224.1	AGS06834.1	PKSI	Trans	1 b/y methyl	1	Streptomyces anulatus
Nosperin	GQ979609.2	ABF92623.1	PKSI/NRPS	Trans	1 b/y methyl	1	Nostoc sp. 'Peltigera membranacea cyanobiont'
Onnamide	AY688304.2	ADA69245.1	PKSI/NRPS	Trans	1 b/y methyl	1	symbiont bacterium of Theonella swinhoei
Pederin	AH013687.2	AFX60327.1	PKSI/NRPS	Trans	1 b/y methyl	1	symbiont bacterium of Paederus fuscipes
Psymberin	FJ823461.1	CBW45695.1	PKSI/NRPS	Trans	1 b/y methyl	1	uncultured bacterium psy1
Jamaicamide	AY522504.1	ADH01492.1	PKSI	Cis	1 b/y vinyl chloride	1	Lyngbya majuscula
Apratoxin			PKSI/NRPS	Cis	1 saturated methyl	1	Lyngbya bouilonii
Carbamidocyclophane	KT826756.1	BAP05578.1	PKSI/III	Cis	1 saturated methyl	1	Nostoc sp. CAVN2
Cylindrocyclophane	JX477167.1	AEE88286.1	PKSI/III	Cis	1 saturated methyl	1	Cylindrospermum licheniforme UTEX B 2014
Calyculin	AB933566.1	ABM63533.1	PKSI	Trans	2 a/b methyl	2	symbiont bacterium of Theonella swinhoei
Corallopyronin	HM071004.1	AMB48448.1	PKSI/NRPS	Trans	2 a/b methyl	2	Coralloccoccus coralloides
Etnangien	AM746676.1	AEC04358.1	PKSI	Trans	2 a/b methyl	2	Sorangium cellulosum
Thailandamide	CP000085.1	AIC32697.1	PKSI	Trans	2 a/b methyl	2	Burkholderia thailandensis E264
Bryostatins	DQ889941.1	AFN27479.1	PKSI	Trans	2 b/y methyl ester	2	Candidatus Endobugula sertula
Phormidolide	KT727016.1	AAW33975.1	PKSI	Trans	3 a/b methyl, 2 b/y methyl	5	Leptolyngbya sp. ISBN3-Nov-94-8
Thailanstatin	JX307851.1	ABC34601.1	PKSI	Trans	1 b/y, multiple mods	1	Burkholderia thailandensis
Bacillaene	AJ634060.2	AJG34060	PKSI	Trans	a/b methyl	1	Bacillus velezensis FZB42
Bongkrekic Acid	JX173632.1	CAG23954.2	PKSI	Trans	a/b methyl, b/y? carboxyl	2	Burkholderia gladioli
Curacin	HQ696500.1	CBD77739.1	PKSI	Cis	cyclopropane	1	Moorea producens 3L
FR901464	HM047288.1	ADH01492.1	PKSI	Trans	epoxide	1	Pseudomonas sp. 2663
Spliceostatin	KJ461964.1	AIC32697.1	PKSI	Trans	epoxide	1	Burkholderia sp. FERM BP-3421
Myxovirescin	CP000113.1	AAM12922.1	PKSI	Trans	saturated ethyl, vinyl methylether	2	Myxococcus xanthus DK 1622
Leinamycin	AF484556.1	ADD82944.1	PKSI/NRPS	Trans	thiazolypentadienone	1	Streptomyces atroolivaceus

Table 1. Polyketides that contain beta branches.

1.5 Bryostatins

The bryostatins are a family of complex macrolactone molecules discovered in a screen of marine extracts originating in 1968 against murine P388 lymphocytic leukemia at the National Cancer Institute(50, 51). While the family now contains more than 20 members(52-58), the primary focus of clinical study has been bryostatin 1, isolated from a sessile marine bryozoan, *Bugula neritina* (**Figure 5A**). This molecule was found to be a potent modulator of protein kinase C (PKC) activity and a promising anti-cancer lead. Initial structural characterization and clinical trials were supported by a ten-month isolation of 18g of pure molecule from 14 tons of adult *B. neritina*(5).

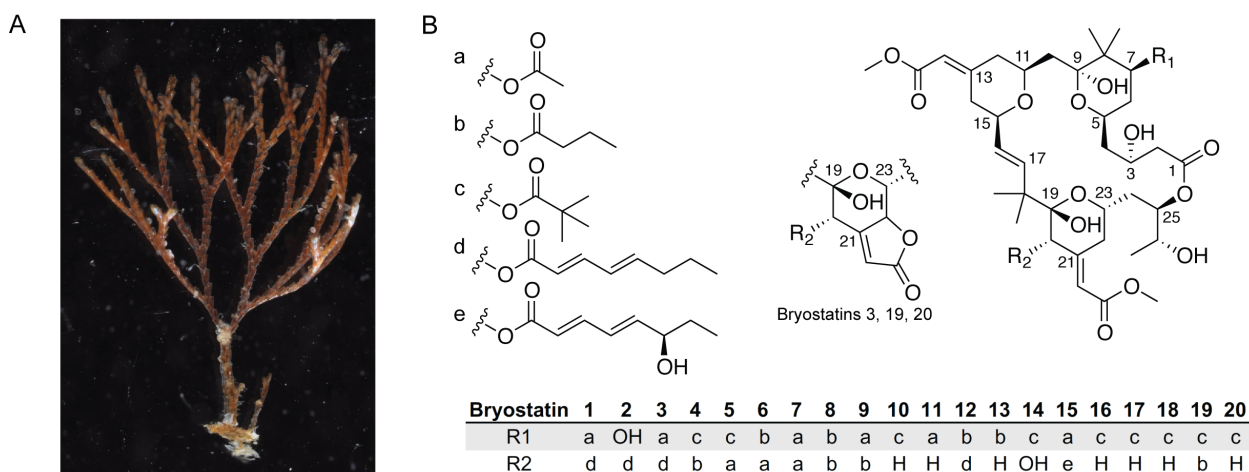


Figure 5. The source and diversity of the bryostatin family. The sessile marine bryozoan *Bugula neritina* (A) is the natural source of bryostatins 1-20 (B).

Bryostatin 1 has since been employed in a variety of phase I and II clinical trials for lymphoma(59, 60), renal cancers(61), Alzheimer's disease(62), and latent HIV(63, 64), with promising early work on Niemann-Pick disease(65), ischemic stroke(66-68), and traumatic brain injury(69).

Through a rigorous series of biochemical and microbiological experiments(70-74), bryostatin biosynthesis has been attributed to a gammaproteobacterial symbiont of *B. neritina*: *Endobugula sertula*. Likely providing a selective advantage to the vulnerable free-swimming larval stage of the *B. neritina* life cycle(75, 76), an interesting body of literature has developed describing several geographical and genomic variations of both host and symbiont(77-79). Unlike its cousin species *Teredinibacter turnerae*(80), *Ca. E. sertula* has remained refractory to laboratory culture. Several attempts have been made to culture adult *B. neritina* for the purposes of bryostatin isolation(81, 82), but investment in this technology has not surfaced. Several bryostatins have been synthesized(83-87), and many simplified and modified bryostatin scaffolds have been reported(6, 7, 88, 89), but high efficacy and economical sources of bryostatin activity remain elusive.

1.5.1 Bryostatin Biosynthesis

Concurrent with early clinical studies, work was carried out to study the natural source of bryostatins. The earliest information about bryostatin biosynthesis came from a feeding study that utilized radioactive precursors to probe the metabolic source of a handful of structural moieties(90). This study predates the discovery of *Ca. E. sertula*, and is perhaps even more impressive with our current understanding of the biology of this symbiosis and the low production of bryostatins in adult *B. neritina*. Microscopy, hybridization, and gene amplification experiments led to the proposal of *Ca. E. sertula* soon after. As *Ca. E. sertula* was (and remains) unculturable, alternative molecular approaches were pursued to study the details of bryostatin biosynthesis.

Because the abundance of *Ca. E. sertula* is so low in the host bryozoan, the abundance of symbiont DNA is very low, while large amounts of host DNA can be readily obtained. By

extracting the total genomic DNA from adult *B. neritina* samples and then enriching for symbiont DNA, one of the multimodular PKS genes responsible for bryostatin biosynthesis was cloned and studied(91). With further refinement of this process, a cosmid library carrying the entire bryostatin gene cluster was generated and sequenced(41). This development opened the door for the biochemical study of bryostatin biosynthesis without the need for further culturing or sampling of *B. neritina* or *Ca. E. sertula*.

The bryostatin cluster consists of five multimodular PKS genes (BryA, B, C, D, and X, **Figure 6**) and three smaller genes (BryP, Q, R). The large PKS genes contain exclusively type I trans-AT PKS modules, and the ensemble of modules appears to fit convincingly with the backbone structure of the bryostatins. The smaller genes make up a beta branching cassette, which was proposed to be responsible for the installation of the two unique vinyl methyl ester moieties found in all bryostatins. Removal of these moieties abolishes the vast majority of biological activity(88), thus their biosynthesis is of particular interest. Several components of the bryostatin pathway were studied using DNA cloned from the cosmid library: the HMGS(43), trans-AT(92), and an assortment of other proteins were cloned and expressed.

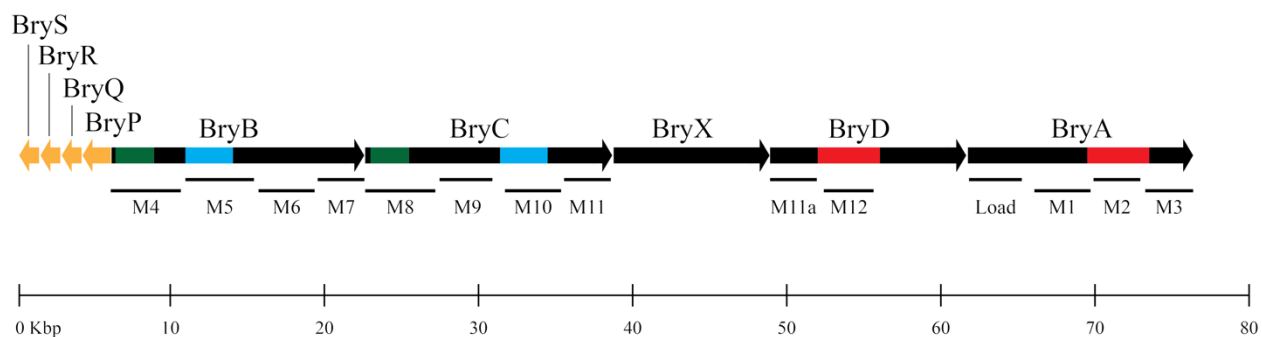


Figure 6. The bryostatin gene cluster. Multimodular PKS genes are shown in black, and beta branching cassette genes are shown in maize. Nucleotide repeat pairs are shown in green, blue, and red regions. PKS modules that are used in the current biosynthetic hypothesis are labeled under the genes.

While the majority of the cluster was found in the cosmid library, several components were missing. Specifically, a donor ACP for the HMGS, and a putative ECH₁ were not found in the assembled sequencing data. A second sequencing effort was undertaken using microdissection instead of DNA enrichment to obtain sufficient quantities of *Ca. E. sertula* DNA(10). This draft genome revealed two new genes a great distance from the main cluster that were annotated as a HMGS donor ACP and ECH₁. This genome assembly also revealed a large amount of information about the symbiosis and life cycle of *Ca. E. sertula* in the context of *B. neritina*, which is discussed in detail in chapter 2.

1.6 Summary

A great deal of progress has been made in the enzymology of PKS pathways. Structural studies of individual(45, 93-96) and multidomain(26, 38, 97, 98) PKS constructs have been carried out, including several reconstructions of an entire PKS module(22, 99). Protein-protein interactions of the different components of diverse PKS systems have been analyzed(12, 18, 19, 34, 98, 100). Mechanistic and substrate flexibility information has been obtained(101), and further, successful engineering of those flexibilities has been repeatedly demonstrated(19, 30, 102). Novel derivatives of PKS scaffolds have been isolated and characterized, and genetic manipulations have been used to great effect in fermentation. However, as is often the case, each answered question leads to new questions. The diversity of PKS systems across the tree of life is vast, and the biosynthesis of many clinically-promising natural products remains mysterious. Several rare chemical functionalities installed by PKS components do not have a clear biochemical basis. Additionally, many questions remain about the architecture of multidomain PKS enzymes.

Because of the continued importance of natural products to our development of new drugs, it is critical to expand our understanding of how these compounds are produced in nature. Expansion of our ability to modify existing natural products or produce new drug leads is dependent upon a deeper understanding of the biosynthesis of the original molecules. This thesis will explore several biophysical and biochemical properties of the biosynthetic machinery from the bryostatin and pikromycin, biosynthetic pathways.

Chapter II

Polyketide Beta-Branching in Bryostatin Biosynthesis

2.1 Introduction

Bryostatins, like many clinically important natural products, were first discovered in an exploratory screen of marine extracts against a cancer cell line. Clinical development focused on bryostatin 1 and revealed an amazing variety of biological activities and effects. Bryostatin 1 has become an important research tool as well as an active target for diverse clinical applications. In addition, the bryostatins have provided an interesting and unique target for expanding our understanding of polyketide biosynthesis. The bryostatin core structure contains several interesting motifs, including a pair of pyran rings with pendant vinyl methyl ester moieties. The importance of these structures to the activity of bryostatins has been discussed in structure-activity relationship (SAR) studies, and their biosynthesis is an interesting case study both as an academic exercise as well as a path toward pathway engineering and new natural products.

2.1.1 Discovery and Biological Activity of Bryostatins

The activities of bryostatins were first discovered in samples collected from the Gulf of Mexico in 1968 during a natural products identification project at the U.S. National Cancer Institute. Marine extracts were screened for bioactivity against a murine P388 lymphocytic leukemia cell line, and extracts from a bryozoan, *Bugula neritina*, were found to significantly inhibit tumor cell growth. A series of activity-guided fractionations led to the isolation of a total of 20 molecules that share the bryostatin core but differ in their side chain structures (**Figure 2B**). Concurrently, experiments connected the activity of bryostatins with that of phorbol esters,

which were soon thereafter linked to protein kinase C (PKC)(103-106). It was found that bryostatins bind to the C1 domains of PKC with single nanomolar affinity, but paradoxically cause downstream effects that are distinct from other PKC-activating molecules. Most notably, phorbol esters promote tumor growth in mouse skin cells, whereas bryostatin not only does not promote tumor growth but limits the tumor growth caused by phorbol esters when co-applied. This unique PKC-mediated response presented a novel target for cancer treatment, and a prescient prediction at the time suggested there would be many applications for a family of molecules with such activity. Indeed, promising preclinical data has led to many trials of bryostatin in the clinic for a surprising diversity of pathologies.

Bryostatin 1 has been involved in 41 clinical trials, 36 of which were for a variety of cancers. However, in clinical trials as a standalone treatment for cancers, bryostatin 1 has suffered from a combination of mediocre outcomes and dose-limiting side effects (60, 61, 107). Combination studies with proven anticancer drugs fared somewhat better, but the inherent complexity of the studies and the transfer of study findings into practice challenged the financial and medical viability of these regimens(108, 109). Often dose and dosing schedule had large and difficult to predict effects on trial results. For these reasons, bryostatin 1 has not found a home as an effective treatment for any cancer. In spite of these results, the use of bryostatin 1 as a tool in cell biology and the continued preclinical and clinical study of other pathologies have opened new doors for the molecule.

While the precise molecular mechanisms remain unclear, bryostatins appear to differentially modulate both the duration and intensity of the effects of PKC activation as well as differentially affecting PKC isozymes. In cells, activation of PKC is achieved by interaction with diacylglycerol (DAG) in combination with calcium ions (conventional isozymes: α , β I, β II,

γ) or by DAG alone (novel isozymes: δ , ϵ , η , and θ). Activation of PKC triggers translocation of the enzyme to the plasma membrane where a variety of downstream targets are phosphorylated, ultimately resulting in tissue-dependent effects such as altered membrane potential, transcription, and cell growth. After a matter of hours, PKC is then downregulated and proteasomally degraded. Even though this degradation happens relatively quickly, the downstream effects of PKC activation can persist for days to weeks, especially in nerve cells. Many of these effects are due to alterations of gene expression, which likely accounts for some of the differences between bryostatin and phorbol ester activities. Similar genes are transiently upregulated by bryostatin and phorbol esters, but the duration is much shorter for bryostatin. Molecules with intermediate biological effects had intermediate durations of upregulated gene expression(110), further suggesting that these observations are interrelated.

PKC is near the top of an enormous number of complex signaling pathways, and is involved in a large number of biological processes. Many effects of bryostatin 1 in the lab show similar preclinical promise to the tumor suppression activity that led to the original interest in the bryostatins. After blast-induced traumatic brain injury, bryostatin 1 appears to significantly influence the recovery of the blood-brain barrier through its effects on PKC isozymes(69). Additionally, several preclinical studies have indicated that bryostatin 1 may find a role as an emergency intervention for ischemic stroke that results in increased survival, neuroprotection, and expedited recovery of neural function, especially when used in combination with tissue plasminogen activator(66-68, 111). Another promising effect of bryostatin 1 is its ability to reactivate latent HIV in cells that are not actively producing virus. With modern antiretroviral treatment, many patients have undetectable viral load but retain reservoirs of HIV-infected cells in areas such as the central nervous system. These reservoirs of potential viral production can lie

dormant and appear to be cured for long periods of time only to become active and infectious again. Reactivation of these reservoirs is an important step towards the total cure of HIV, as activation of viral production allows cells that contain genomic integration of HIV to be identified and cleared. Bryostatin alone appears to activate HIV reservoirs via PKC(63), and studies of combination treatments with chromatin-remodeling histone deacetylase inhibitors show highly promising reactivation of latent HIV(112). These preclinical activities clearly warrant further study.

Other effects of bryostatin 1 have begun to find their way out of animal models and into human testing. One of the most significant of these is the treatment of Alzheimer's disease (AD). To date, the primary approach of drugs development to treat AD has been focused on clearing β -amyloid plaques. These drugs have largely failed to make significant clinical impact, and few other treatment mechanisms have been explored. Bryostatin 1 is one of only a handful of molecules in development that does not target amyloid as part of its mechanism, but instead targets different cellular signaling targets. At the time of this writing, phase II trials are underway and time will tell if this approach bears fruit.

2.1.2 Synthetic and Natural Sources of Bryostatins

With their myriad of activities and potential clinical applications, sourcing of bryostatins should be considered. As is discussed in other sections, adult *B. neritina* colonies contain miniscule amounts of bryostatins relative to their total mass, and the proliferation of *Ca. E. sertula* is mostly limited to the free-swimming larval stage of the life cycle. Several methods for the aquaculture and isolation of bryostatins were explored in a pilot project by CalBioMarine Technologies(82). This project developed a filter technology to isolate *B. neritina* larvae from seawater, however the bryostatin yield using this filtration technique was much lower than what

is typically achieved from larvae. While onshore tank growth was not economical, in-sea aquaculture of adult colonies was much more successful, providing a similar yield to adult colonies harvested from natural sources at a relatively similar cost. Unfortunately, these yields are still extraordinarily low and would require the same lengthy purification scheme developed for the first gram-scale isolation of bryostatin 1(5).

Total synthesis provides another avenue to clinical-scale abundance of bryostatins. To date, the total synthesis of bryostatin 1, 2, 3, 7, 9, and 16 have been reported(84-87, 113, 114). While the first reported schemes are not effectively scalable, progress has been made toward increasing the efficiency of syntheses of bryostatins that are most likely to be clinically relevant. A promising synthetic scheme has been devised by Wender and coworkers(115), which has been licensed for the production of material to be used in clinical trials.

In addition to total synthesis of structures discovered in nature, many simplified and substituted bryostatin structures have been synthesized and studied. These efforts have made promising progress toward molecules that have comparable activity to bryostatins but are far cheaper and more easily diversified for downstream drug development. These molecules have also greatly expanded our understanding of the structure-activity relationships of the bryostatins, and the complex biology they interact with. Using structural analogs, a hypothesis for the activity of bryostatin has been developed: the southern half (C-ring) of bryostatin confers tight binding of PKC while the northern hemisphere (A- and B-rings) somehow results in the activity of bryostatin that is distinct from phorbol esters. Interestingly, the unique vinyl methyl ester moieties that are installed by the beta branching cassette are involved in both hemispheres of the molecule (**Figure 7**). In an otherwise identical pair of bryostatin analogues, the removal of the methyl ester from the southern hemisphere beta branch resulted in a 20-fold decrease in PKC

affinity(89).

It appears that the southern hemisphere will tolerate some structural variation as long as the PKC binding is sufficiently tight, while the alteration of the northern hemisphere will result in the loss of bryostatin-specific activity and the conversion to tumor-promoting activity similar to phorbol esters. Neristatin is a sponge-derived natural that has bryostatin-like activity rather than phorbol ester-like tumor promotion(116), however no biosynthetic pathway has been connected to neristatin production to date. The structure of neristatin differs from bryostatin only in the southern hemisphere, and is still able to bind PKC and have bryostatin-like effects. Alteration of the northern hemisphere has been achieved in a variety of bryologue molecules, and generally results in different degrees of phorbol ester-like biological activity. It is important to note that the distribution of bryostatin-like to phorbol ester-like biological activity is by all accounts a continuum of responses that is strongly dependent upon cell type and biological readout. No single motif has proven to be the key to conferring bryostatin-like activity, although a great deal of information has been discovered that may lead to bryologues that are cheaper to manufacture at scale than bryostatin 1.

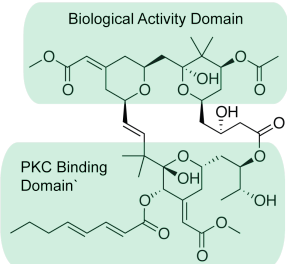
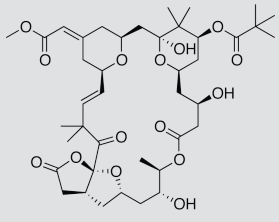
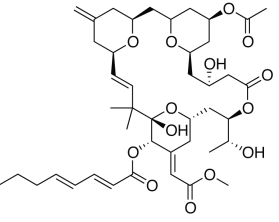
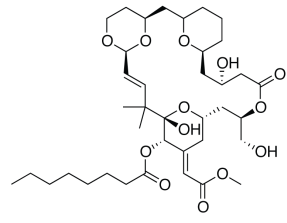
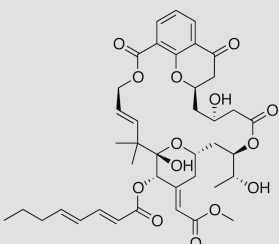
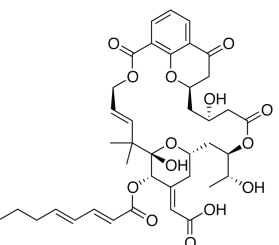
	Bryostatin 1	Neristatin	Merle 27
			
PKC α K _i	0.48 nM	21.2 nM	3.0 nM
Activity	Bryostatin 1-like	Bryostatin 1-like	Phorbol ester-like
	Picolog	WN-2	WN-4
			
PKC α K _i	3.2 nM	213.7 nM	3988 nM
Activity	Bryostatin 1-like	Phorbol ester-like	N/A

Figure 7. Bryostatin 1 and synthetic analogs. Modifications to the southern (PKC binding) and northern (biological activity determining) hemispheres have been explored. The vinyl methyl esters installed by the beta branching cassette appear to be key to both hemispheres.

Several unanswered questions remain about the more recently discovered activities of bryostatin and bryostatin-like molecules. For example: salicylate-derived analogs maintain a similar macrolactone core, but remove many of the pendant functional groups that are important for PKC affinity. These molecules have had the majority of their PKC binding abrogated, however some biological effects of bryostatins still persist, including cellular protection against chikungunya virus mediated cell death(117).

Whereas synthetic analogues have been studied in order to describe bryostatin activity and reduce this activity to its minimal structural basis, study of the natural source of these molecules can provide another fruitful avenue for the clinical development of bryostatins and

natural products in general. With an understanding of bryostatin biosynthesis, we may be able to facilitate the production of bryostatins from a heterologous host, generate new bryologues, confer bryostatin-like pharmacophores to other molecules, or generate completely novel structural scaffolds.

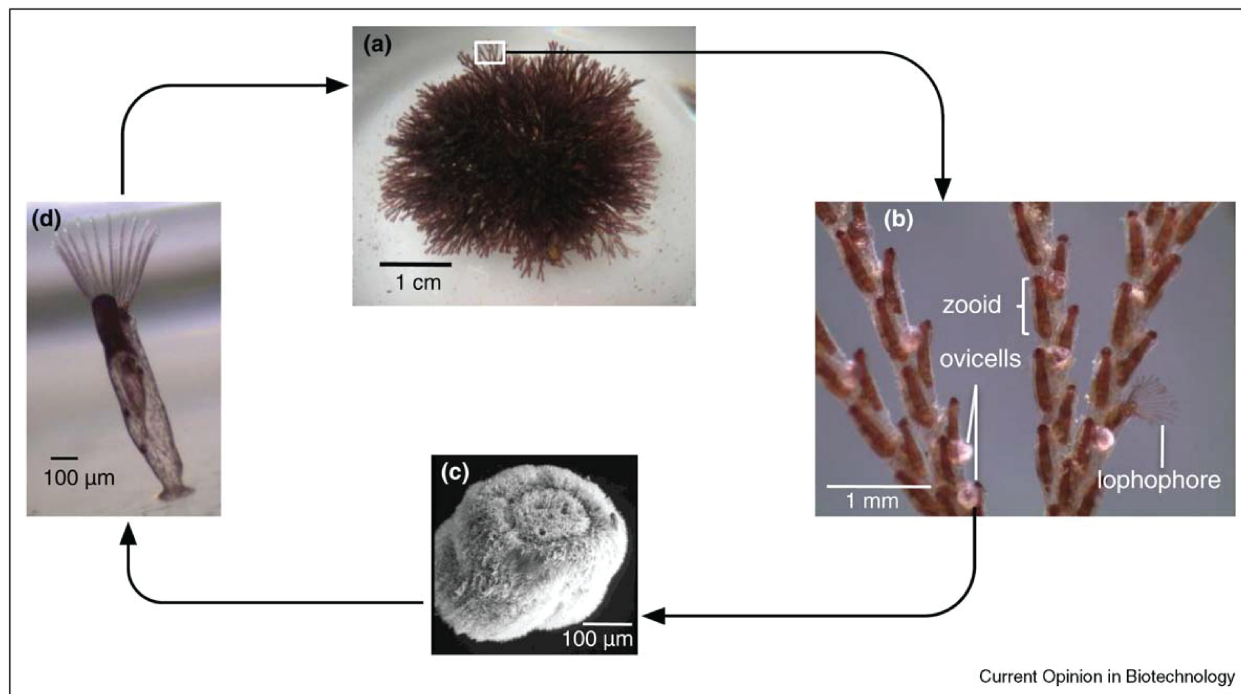


Figure 8. The life cycle of *Bugula neritina*. Branched adult colonies (a) reproduce by generating ovicells (b) that produce free-swimming larvae (c), which attach to suitable surfaces and undergo metamorphosis into a solitary adult colony (d), which then reproduces asexually to form adult colonies. (Reprinted from *Current Opinion in Biotechnology*, 21, Amaro E Trindade-Silva, Grace E Lim-Fong, Koty H Sharp, Margo G Haygood, *Bryostatins: biological context and biotechnological prospects*, pp 834-842, Copyright (2010), with permission from Elsevier.)

All 21 members of the bryostatin family of molecules were isolated from *Bugula neritina* gathered from temperate waters around the world. *Bugula neritina* is a species of filter-feeding marine invertebrate in the phyla bryozoa. These animals are commonly referred to as “moss animals,” due to the shape and appearance of their purple-orange mature colonies. These colonies are widely considered to be biofouling organisms due to their abundance, general lack

of discretion in growth surface, and concurrent growth with other opportunistic organisms that accelerate the accumulation of undesired biomass. Individual animals begin their life cycle as free-swimming larvae, which are ~0.5mm diameter cylindrical barrel-shaped structures with a circular sinus on one end of the barrel. These larvae affix themselves to a suitable surface and undergo metamorphosis into adults, which then feed via filtration and reproduce asexually. The colony develops into branched grouping of segments made up of individual animals, and eventually develops specialized structures called ovicells that produce new larvae which are released seasonally, thus continuing the cycle (**Figure 8**).

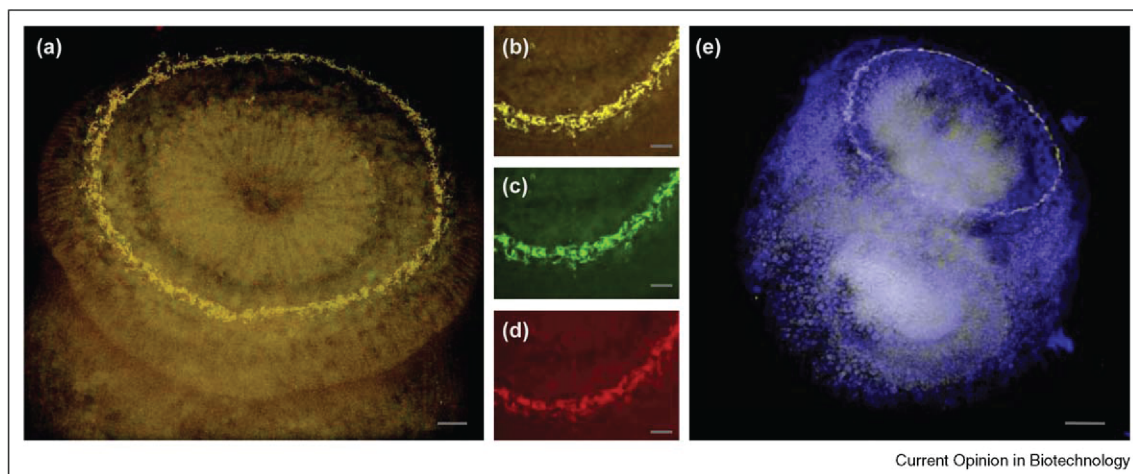


Figure 9. FISH microscopy of *B. neritina* larvae. Eubacterial probe CY5-EUB338 is shown in red (d), *Candidatus Endobugula sertula* specific probe is shown in green (c), and a merge of these two signals is shown in yellow (a) and (b). Localization of bryostatins on the surface of larva is revealed by protein kinase C-specific labeling, shown in blue (e). (Reprinted from Current Opinion in Biotechnology, 21, Amaro E Trindade-Silva, Grace E Lim-Fong, Koty H Sharp, Margo G Haygood, Bryostatins: biological context and biotechnological prospects, pp 834-842, Copyright (2010), with permission from Elsevier)

The concentration of bryostatins in adult *B. neritina* colonies is vanishingly small, but the consistent discovery of bryostatin-containing colonies around the world suggests these molecules confer a significant selective advantage. Further study revealed that bryostatins act as potent

chemical defenses that prevent predation by fish(118), though the source was still unclear. Microscopy(74) and rRNA sequencing revealed a near monoculture of a rod-shaped gram negative bacteria within the sinuses of larvae(73), leading to the proposal of *Candidatus Endobugula sertula* as the bryostatin-producing symbiont (**Figure 9**). Antibiotic treatment confirmed that bryostatins are produced by an uncultured bacterial symbiont engaged in an unusual relationship with its host(71). Most symbioses involve the exchange of rare or expensive metabolites and some mutual developmental benefit to both organisms. In this case, asymbiotic *B. neritina* colonies grow and develop just as well as symbiotic colonies, suggesting a different dynamic between symbiont and host. Indeed, this symbiosis resembles the development of chemical anti-predation defenses of terrestrial plants.

2.1.3 Previous Biochemical Study of the Bryostatin Pathway

In order to study the biosynthesis of bryostatins, the Sherman lab set out to identify the biosynthetic pathway. Because there is a high degree of sequence identity among polyketide synthase (PKS) ketosynthase (KS) domains, degenerate primers can be used to amplify PKS genes nonspecifically, and this technique has frequently been used to identify PKS gene clusters. Using genomic DNA isolated from adult *B. neritina*, a cosmid library was assembled and a novel four-module PKS gene, BryA, was identified(91). Using reverse transcriptase-PCR (RT-PCR), it was shown that RNA transcripts for this gene were present in preparations of *B. neritina*, demonstrating active gene transcription. The structure of this gene was very unusual among known type I PKS genes, with several deviations from typical domain architecture. Most noticeably, bryostatin is a trans-acyltransferase (trans-AT) PKS, requiring a standalone acyltransferase (AT) domain to load extender units instead of AT domains integrated into each biosynthetic module, as is most common for PKS pathways. As this sequencing effort identified

the first PKS gene of the pathway, particular focus was given to the initiation of biosynthesis. This first module was also unusual for PKS, and utilizes a domain termed FkbH that selects the glycolytic intermediate 1,3-bisphosphoglycerate, hydrolyses both phosphates, and attaches the resulting 3-carbon glyceryl unit to an ACP(*119*). This work motivated further cosmid library screening and sequencing of the complete bryostatin gene cluster (**Figure 10**).

The motivation for further study culminated in two assemblies derived from Sanger-sequencing of cosmids containing fragments of the ~77 kilobase bryostatin clusters from the previously described “deep” (California) and “shallow” (North Carolina) sibling *B. neritina* species(*41*). Again, RT-PCR confirmed that these assembled genes were all actively transcribed in *B. neritina* preparations. It was also demonstrated that these were the only large PKS clusters within the genomes of the two *B. neritina* isolates. The full cluster is made up primarily by five large multimodular PKS genes: BryA, B, C, D, and X. As can be readily appreciated, these genes violate the usual colinearity rules in PKS gene clusters and do not appear on the chromosome in the same order as they are predicted to function during bryostatin biosynthesis. In addition to the PKS genes, four other genes were annotated as being part of the biosynthesis: BryP, a didomain trans-AT; BryQ, a decarboxylative KS; BryR, a 3-hydroxy-3-methylglutaryl (HMG) synthase homologue; and BryS, a SAM-dependent methyltransferase (MT).

In the “deep” cluster, this cassette of smaller genes is not found immediately adjacent to the PKS genes. In both assemblies, the cluster is bordered by at least one transposase, suggesting the fragmentation of the cluster is likely due to transposition events. This fragmentation is not uncommon among PKS clusters, especially trans-AT pathways which tend to propagate via horizontal transfer more often than cis-AT pathways(13), resulting in varied cluster architectures for the same PKS product (e.g. the pederin family of molecules(120)). Additionally, symbiont genomes undergo a variety of genomic alterations due to their lifestyle, which has likely added to the tendency of cluster fragmentation. Despite these differences in cluster fragmentation, the overall sequence identity of the constituent genes is 98.8%.

Another unusual discovery within the assembled sequencing data are the large sections of almost identical sequence. Three pairs of these sections were found in the cluster, and are thought to be remnants of genetic recombination events, one of the fundamental evolutionary routes to PKS cluster diversification(121). Many clusters contain evidence of long histories of gene or module duplication, modification, or deletion. These repeats however contain much higher identity with one another than are typically observed, suggesting relatively recent origin. Also, these repeats are found on distant modules (modules 4/8, 5/10, 2/12) and may be of interest for the study of substrate specificity of PKS domains. Many of the types of catalytic domains covered by the repeats have been found to be discerning in their substrate selection, and it would be interesting to determine how these domains can cope with the disparate sizes of their predicted substrates. The KS domains covered in these repeats already have an interesting starting point, as several small sections of the primary sequence near the catalytic residues contain the only differences between the repeats, suggesting these residues must be different to accommodate their radically different substrates.

The previously described cosmid sequencing assemblies allowed for the analysis and preparation of a functional hypothesis of bryostatin biosynthesis. As described earlier, BryA carries out initiation from 1,3-bisphosphoglycerate and continues extension and modification through the activities of four PKS modules (modules 1-4) complemented by malonyl extender unit loading by BryP. BryB and BryC each also contain four trans-AT PKS modules (modules 5-11). It appears that BryD contains the reductive domains (and a second ACP) of module 11 (termed “module 11a”) in addition to two other modules (12 and 13) and a condensation domain (CD) reminiscent of those found in non-ribosomal peptide synthetase (NRPS) pathways. Macrolactonization and chain offloading is likely carried out either by this CD or by a thioesterase (TE) domain found in BryX. Outside of this TE domain, BryX does not fit into a logical biosynthetic scheme following our current understanding of PKS biochemistry. However, some evidence suggests that BryX is likely somehow involved in the biosynthesis and attachment of the variable sidechains that differentiate many bryostatin family members.

The first in vitro biochemical studies focused on confirming the function of BryP and BryR. To slightly differing extents, both AT domains within BryP were able to load both malonate and methylmalonate extender units onto a variety of ACPs, both excised and within modules(92). The lack of specificity shown by BryP may be explained by a lack of methylmalonate within *Ca. E. sertula* and therefore underdeveloped discretion in extender unit selection. Another set of experiments was carried out that attempted to complement a trans-AT knockout of a PKS pathway, mupirocin, within the producing organism, *Pseudomonas fluorescens*. By complementing the trans-AT knockout with different constructs of the two AT domains, it was determined that while the two domains appear to have different evolutionary origins, their functionality is actually fairly similar. In other pathways, the second AT domain of

a didomain trans-AT acts as an acyl hydrolase (AH), which is able to hydrolyze stalled or erroneous intermediates from ACPs, thus freeing those modules from dead-end covalent intermediates(122). Taken together, these experiments conclusively demonstrate that BryP is in fact a trans-AT that loads extender units onto PKS components.

An additional biochemical study of the bryostatin pathway focused on a key component of the beta branching cassette: BryR, an HMG-ACP synthase(43). The role of BryR as an HMGS in bryostatin biosynthesis was studied from the perspective of previously examined HMG-CoA synthase homologues in PKS pathways. Essential to the activity of HMGS is a donor ACP (ACP_D) and an acceptor ACP (ACP_A). A malonyl unit is loaded onto the phosphopantetheine arm of ACP_A, which is then decarboxylated by a specific KS. The resulting acetyl unit is then transferred to a specific cysteine sidechain on HMGS, which subsequently interacts with ACP_A, which has a nascent beta-ketone polyketide covalently linked to its phosphopantetheine arm. HMGS then catalyzes the aldol addition of the alpha carbon of the acetyl unit to the beta ketone carbon, resulting in a HMG-like product. Unlike HMG-CoA synthase that acts on CoA-linked substrates, the protein-protein interactions of HMGS with ACP_A and ACP_D have been found to be very specific, with little tolerance for substitution(43, 98). The specificity of HMGS stands to reason in the context of a PKS assembly line which would be greatly disrupted by the introduction of an erroneous beta branch in a growing polyketide. This specific ACP selectivity is an interesting target for pathway engineering and the intentional introduction of beta branches to generate novel PKS-derived compounds, but it also presents a challenge when the correct ACP is either unclear or unknown.

The functionality of BryR was demonstrated using a set of simplified and commercially available acyl-CoAs that had been used in the study of HMGS from the curacin and jamaicamide

pathways: namely, acetyl-, acetoacetyl-, and 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). The activity and role of BryR in the pathway was confirmed, but several questions about the bryostatin beta branching cassette persisted, as no donor ACP or ECH₁ were identified in the cosmid library(43). The search for suitable surrogate donor ACPs from beta branching cassettes of other pathways involved several interesting protein-protein interaction experiments(43). The data acquired in these experiments provided insight into the differences in donor ACP binding by different HMGS, but did not test the current biosynthetic hypothesis. Somehow, components of the pathway are able to convert the HMG-like product of BryR into a Δ^3 vinyl methyl ester found in the final product, and several hypotheses were proposed. The most appealing theories involved the use of an ECH₁ similar to those found in other beta branching cassettes in conjunction with a carboxyl O-methyltransferase. This hypothetical ECH₁, should it prove to exist, would be very interesting as it would be the only known ECH₁ with a dehydrated product that persists through the rest of the biosynthesis without decarboxylation by ECH₂ or other regiochemistry-altering modifications.

Additionally, the final vinyl methyl ester moieties in the bryostatin molecule are situated at two key sites of interaction with its biological target, PKC. Understanding the biosynthesis of these molecules and perhaps migrating this ability to other pathways may allow the portability of bryostatin-like PKC binding and biological activity. However, in order to more fully understand bryostatin biosynthesis, new insights into the *Ca. E. sertula* genome would be required.

2.1.4 Analysis of a Draft *Endobugula Sertula* Genome

Thanks in large part to the development of next-generation sequencing and accompanying analysis software, great advances have been made in metagenomic sequencing(123, 124). These technologies allow for samples containing DNA from the genomes

of multiple species to be analyzed directly without the need of cosmid assembly and time-consuming Sanger sequencing. Using software packages specifically developed for metagenomic analysis, sequencing reads can be pooled into groups of reads that originated from the same genome. This is an incredibly powerful technique in cases where natural products are produced by uncultured or low abundance symbionts. The ability to specifically isolate and analyze symbiont sequencing data in silico has led to several interesting results, including the genome of the tunicate symbiont that produces ET-743, *Candidatus Endoecteinascidia frumentensis*(125). Of course, these techniques have their limitations. Enrichment of the metagenomic source for symbiont DNA using specific tissue isolation, chemical separation, or other techniques makes a significant impact on the downstream data quality. Large levels of host or bacterial community DNA will inevitably add difficulty to the effective sorting and assembly of sequencing reads.

Using metagenomic DNA isolated from *B. neritina* larvae and targeted microdissections of adult ovicells, Miller and coworkers were able to achieve sufficient abundance of *Ca. E. sertula* DNA to carry out a shotgun sequencing effort(10). Subsequent analysis and careful pooling of reads from the symbiont genome resulted in a novel draft genome of *Ca. E. sertula*(123). Made up of 112 contigs with a total of 3.3 Mbp, this genome is much larger than would be expected for an obligate symbiont, and only bears some of the hallmarks of genome reduction, indicating a relatively recent symbiosis. The GC content and pseudogene abundance also suggest that *Ca. E. sertula* may have been able to live independently in the recent past. This is a key question for the isolation of large quantities of bryostatins, as independent *Ca. E. sertula* may be fermented to facilitate large-scale production rather than the difficult methods outlined in previous sections. While this is a desirable target, the genome also indicates several adaptations

for living within its host. Several chitinase genes that are most often associated with bacterial pathogens of insects were found to be highly expressed in *Ca. E. sertula* ovicells(10), suggesting they play a role in the vertical transfer of the symbiont from one generation of animal to the next. Additionally, *Ca. E. sertula* is likely dependent upon its host for nutrition, as several deficiencies in amino acid biosynthesis genes were found. This genome is a very interesting resource for the study of an obligate symbiont that provides an advantage to its host by providing a chemical defense.

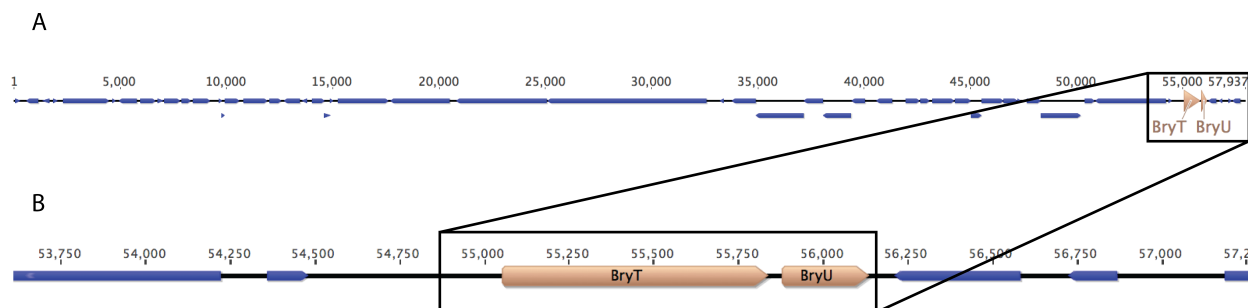


Figure 11. Shotgun sequencing data MDLC00000000 contig 14. BryT and BryU location within draft genome of *Candidatus Endobugula sertula*. Other potential open reading frames are shown in blue.

This genomic assembly also provides greater context for the bryostatin pathway than was previously available. It is another piece of evidence that there is no other PKS pathway in the *Ca. E. sertula* genome, and confirms the assembly of the bryostatin cluster generated from Sanger sequencing of the Sherman lab cosmid library. This sequencing effort also delivered two interesting genes, a FAS-like acyl carrier protein adjacent to a PKS enoyl-CoA hydratase (**Figure 11**). Because of their categorization being unique among genes identified in the assembly, these two genes, termed BryT and BryU, are strong candidates for the missing components from the bryostatin gene cluster. Additionally, the accompanying RNA sequencing provides strong evidence that these genes are actively transcribed while *Ca. E. sertula* is

proliferating, as is the entire bryostatin cluster.

2.2 Results

2.2.1 BryU is the Missing Donor ACP and Donates Acetyl Unit to BryR HMGS

BryU was identified by automated annotation within the draft genome of *Ca. E. sertula*. This ORF was 15bp away from the candidate ECH₁, BryT. Due to gaps in the genomic assembly and a lack of overlap with any of the earlier draft quality cosmid sequencing data, the distance from the bryostatin PKS genes can only be estimated to be >48kbp. No other hits were found resembling BryU or any other ACPD within the draft genome. BryU strongly resembles other FAS-like beta branch donor ACPs, including mAcpC from the mupirocin pathway(126) (40.5% identity) and CurB from the curacin pathway (43.0% identity), both of which were used in previous studies of BryR (**Figure 12**). Comparison of BryU and BryR to a cocrystal structure of CurB interacting with the curacin HMGS (CurD, 67.5% identity to BryR) shows that the key residues for HMGS-ACPD interaction are conserved in both pathways(98). This is perhaps to be expected given the relatively high sequence identity and binding affinity of CurB for BryR as a surrogate ACPD.

BryR, BryU and BryA module 3 ACP were expressed and purified using methods described previously(43), and activity was assessed by LC/MS using the phosphopantetheine ejection assay. Purity of the protein was assessed by SDS-PAGE (**Figure S1**) Hydrolysis of acetyl-BryU by BryR was observed in the absence of ACP_A, which is consistent with the generation of the covalent acetyl-HMGS intermediate in the mechanism of HMGS. When acetoacetyl-ACP_A (BryA module 3 ACP) was present, conversion to HMG-ACP_A was observed (**Figure 13**).

Analysis of beta branch acceptor ACPs revealed conserved residues that are specific to

that group and are not conserved among modular ACPs within PKS pathways(44). These residues are found in the ACPs of BryA module 3 and BryB module 7, the two predicted locations of beta branching in the bryostatin pathway. Interestingly, one ACP within BryX also contains these acceptor-specific residues, though it is difficult to rationalize what role a beta branch would play at this point in the pathway.

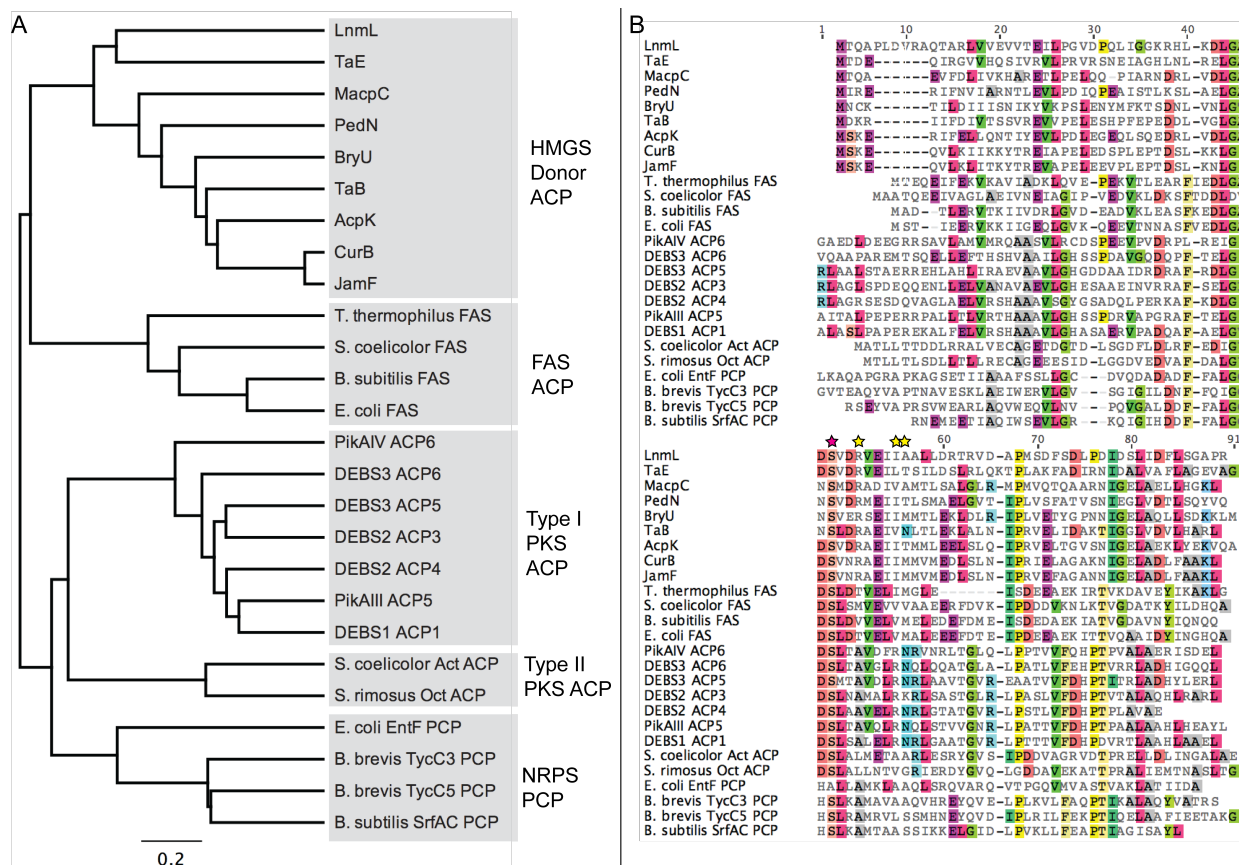


Figure 12. Phylogenetic analysis of BryU. Phylogenetic tree of BryU and other classes of acyl carrier protein suggest BryU is a beta branching cassette donor ACP (A). Sequence alignment of these ACPs shows this identity in detail (B). Conserved site of phosphopantetheinylation is indicated with a red star, residues indicated to be important to HMGS interaction in CurB are labeled with yellow stars.

Many cis-AT pathways show evidence of long evolutionary histories of vertical transmission, resulting in KS domains that tend to clade by pathway or organism of origin, regardless of their substrate specificity. In contrast, trans-AT PKS pathways often are

propagated and modified through horizontal transfer and genetic recombination of pathway components. The relatively short time KS domains of trans-AT pathways have spent within their host genomes results in the primary sequence of KS domains being predictive of their substrate specificity. This results in the useful tendency of trans-AT KS domains to clade together based on substrate specificity rather than pathway.

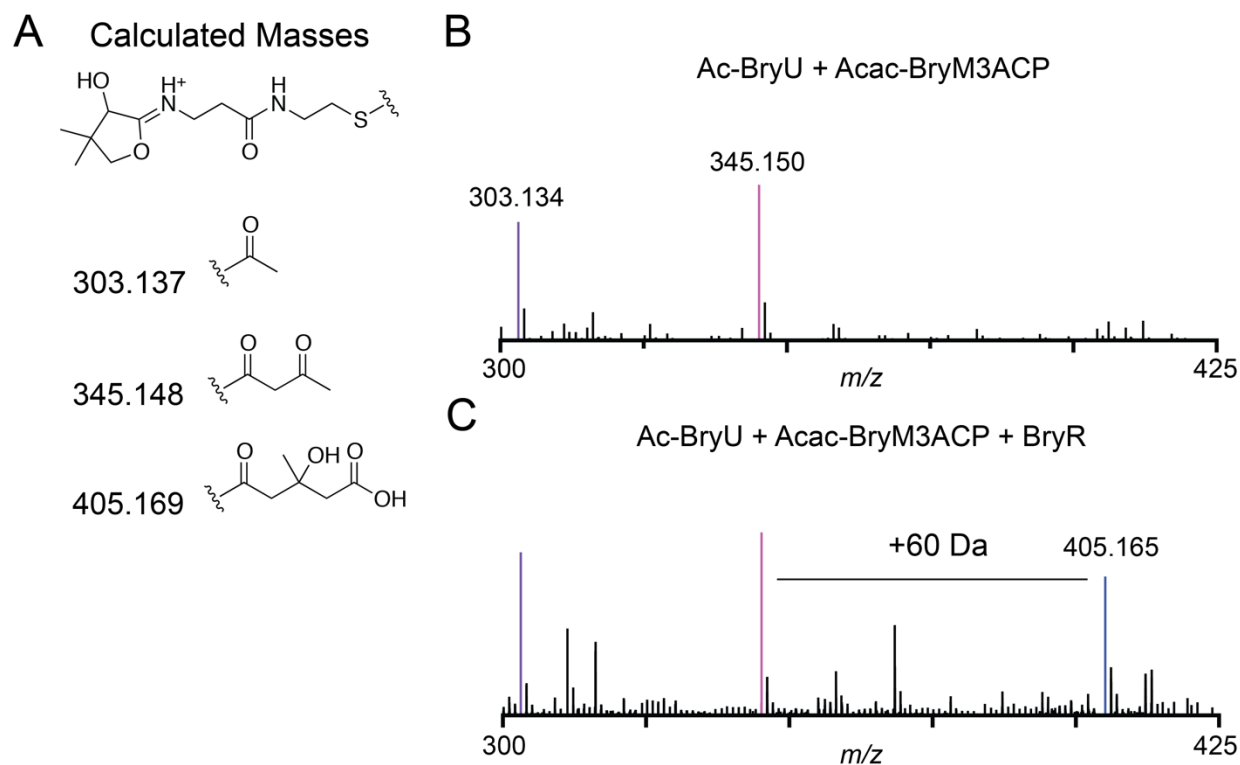


Figure 13. Mass spectrometric analysis of BryR HMGS activity. Predicted masses of phosphopantetheine ejection ions (A), negative control reaction containing only acetyl-BryU and acetoacetyl-BryA-M3-ACP, and the same reaction conditions with BryR added (C).

2.2.2 BryT is the Missing ECH₁ and Dehydrates HMG-ACP

BryT was found immediately adjacent to BryU in the draft genome of *Ca. E. sertula*. Automatically annotated as a polyketide synthesis ECH, phylogenetic analysis and sequence comparisons provided strong evidence that this was the missing bryostatin ECH₁. Because

several ECH₁ and ECH₂ from PKS clusters have been biochemically characterized(42, 127), primary sequence alone is enough to classify a protein as ECH₁, ECH₂, or primary metabolic enzyme. In sequence alignment with ECH₁, BryT has 38.3% identity with CurE and 37.9% identity with JamI. Compared to ECH₂, BryT has 15.7% identity with CurF(ECH₂) and 14.5% identity with JamJ(ECH₂), clearly indicating identity with ECH₁ (Figure 14).

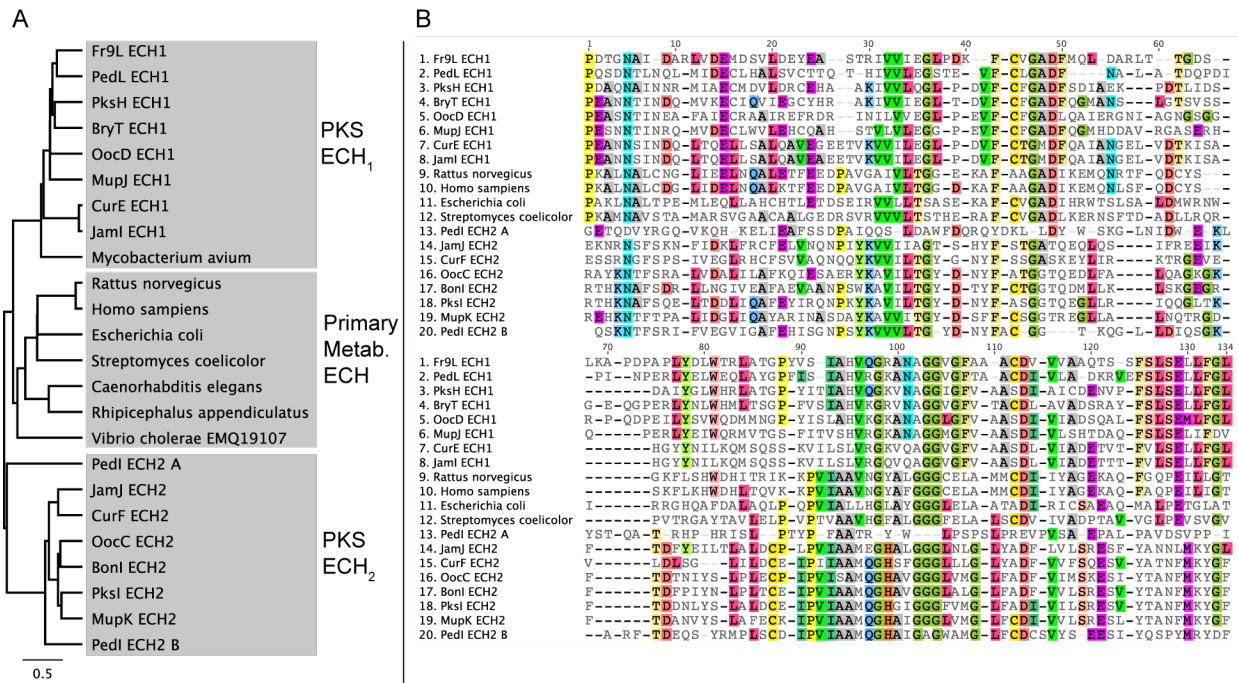


Figure 14. Phylogenetic analysis of BryT. Secondary metabolism ECH₁ and ECH₂ are compared by primary sequence to primary metabolism enoyl-CoA hydratases from a variety of organisms.

Rather than attempting to clone from metagenomic DNA, synthetic codon-optimized DNA coding for BryT was purchased and cloned into expression vectors. Coexpression with GroEL and GroES at relatively low growth temperatures was determined to be necessary, with low and inconsistent protein yields from other conditions (Figure S2). Protein purified under optimized conditions was >90% pure, but was sensitive to cooling below ~5C, and would semi-

reversibly precipitate if allowed to become too cold. Other ECH₁ from PKS pathways were also obtained for comparison: CurE and JamI from the curacin and jamaicamide pathways, respectively.

HMG-CoA was used as a substrate as previously described(127): both as a reagent used to charge ACPs and as a direct substrate without further modification. Using the phosphopantetheine ejection assay, a mass shift equivalent to a dehydration was observed after incubation with CurE, JamI, or BryT, indicating the generation of 3-methylglutaconyl-CoA (3MG-CoA)(**Figure 15**). Upon detailed review of the data, a small amount of dehydrated mass is still detectable in the absence of ECH₁. This was suspected to be ionization-mediated dehydration caused by the mass spectrometer, and further support for this theory is provided by data that is discussed later in this section.

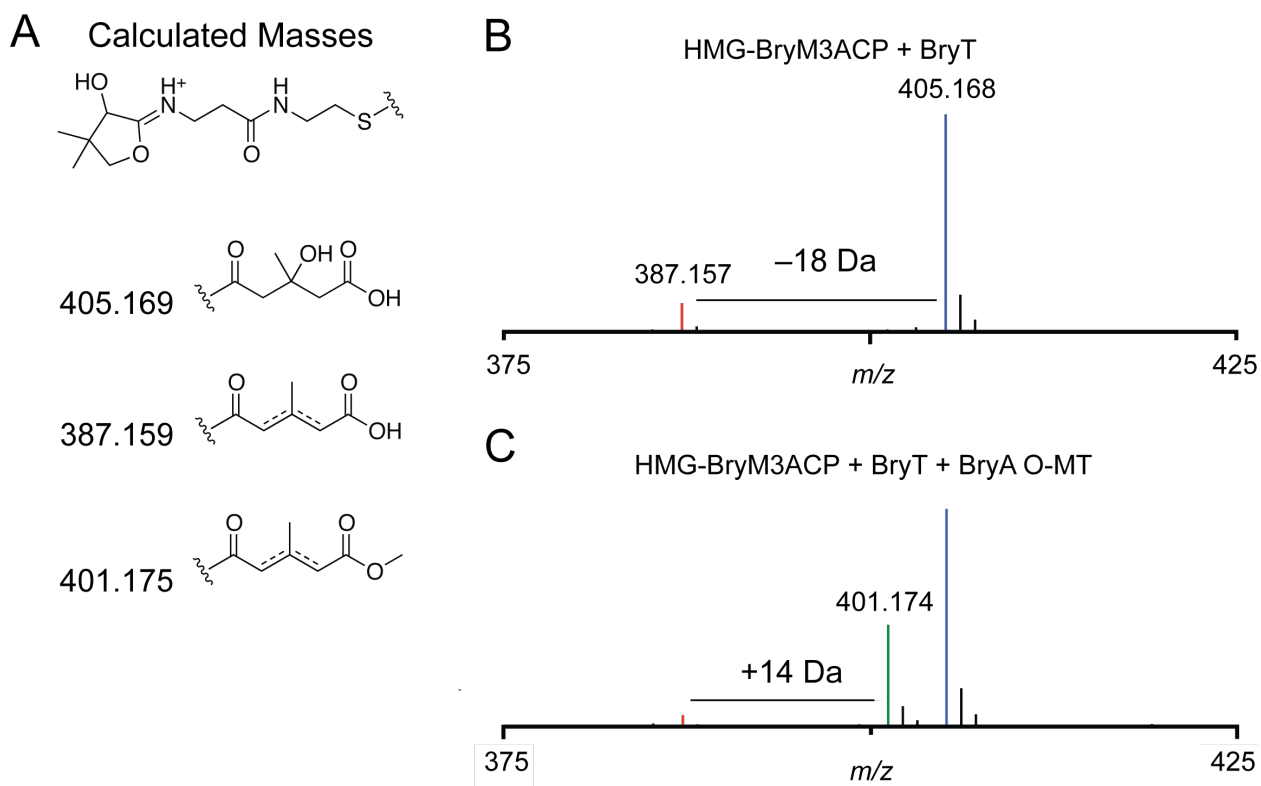


Figure 15. Mass spectrometric analysis of BryT dehydration and BryA O-MT carboxyl methylation of HMG-ACP. The calculated phosphopantetheine ejection masses (A) were detected in reactions of HMG-BryM3-ACP with BryT (B) and HMG-BryM3-ACP with BryT, BryA O-MT, and SAM (C).

While the detection of a loss of water mass is compelling evidence for ECH₁ activity, it does not allow for the determination of the regiochemistry of the product. Several approaches were developed that take advantage of the ability of ECH₁ to use CoA-tethered substrates to unambiguously determine whether BryT was carrying out α/β (Δ^2 product) or β/γ (Δ^3 product). First, the similarity of PKS HMGS and HMG-CoA synthase in primary metabolism was leveraged to chemoenzymatically synthesize a specifically deuterated HMG-CoA (**Figure 16**). Fully deuterated acetic anhydride was used to generate d₃-acetyl-CoA, which was then used as a carbon donor for mvaS, the HMG-CoA synthase in the isoprenoid biosynthesis pathway from

Staphylococcus aureus. This reaction series generated γ -d₂-HMG-CoA, which when dehydrated by ECH₁, would result in either a +1 Da or +2 Da difference from the natural isotopic abundance. Detection of the product masses was achieved through HPLC separation of the reaction components and MS confirmation of the individual masses.

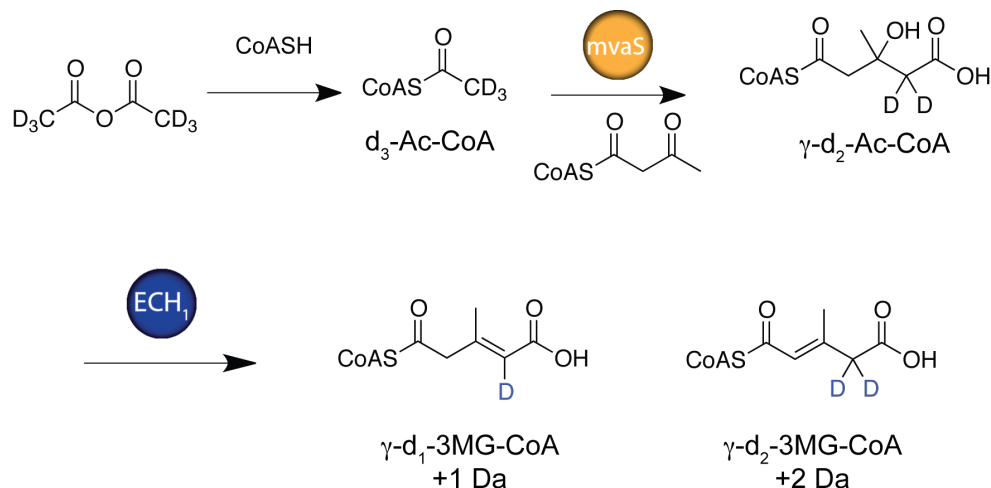


Figure 16. Chemoenzymatic preparation of dideterated 3-hydroxy-3-methylglutaryl-CoA. Deuterated acetic anhydride rapidly converts coenzyme A into trideuterated acetyl-CoA, which can be used as a carbon source for HMG-CoA synthase. After reaction with BryT, the double bond configuration can be inferred by mass.

The addition of BryT to this reaction system resulted in the generation of primarily d₂-3MG-CoA, strongly indicating α/β dehydration and a Δ^2 product (**Figure 17**). This is somewhat surprising, as no other obvious enzymatic components exist within the cluster to carry out further transformations of this double bond to generate the configuration found in the final product. Analysis of other ECH₁ using this assay system generate similar results (**Figure S3**). Because of proton exchange with solvent, the timing of the reaction has a large impact on the isotope distribution detected in the products. However, proton exchange must be balanced with the generation of sufficient product to be reliably separated and detected. In the attempt to balance

these demands, it was found that shorter reaction times led to an even higher ratio of $d_2:d_1$ 3MG-CoA (**Figure S4**), further indicating that the dideuterated product is the preferred product of BryT.

While analyzing these data, it is important to consider the possibility of a kinetic isotope effect altering the outcome. Because ECH₁ (and most of the crotonase family) proceed through an enolate intermediate stabilized by an oxyanion hole, the rate-limiting step is likely either the formation of the enolate or the first deprotonation of the substrate at the α -carbon (**Figure 18**). In ECH₁, this deprotonation is catalyzed by a conserved active site glutamate, although mutagenesis data suggest this residue is critical to efficiency but is not essential. When this glutamate is replaced with a glycine or alanine, the low level of activity may be attributed to a water molecule acting as the catalytic base. Because the abstraction of a γ -deuteron is highly unlikely to be the rate-limiting step of the reaction, a kinetic isotope effect is highly unlikely to alter the product distribution drastically enough to explain the definitive majority of γ -d₂-3MG-CoA detected in these reactions.

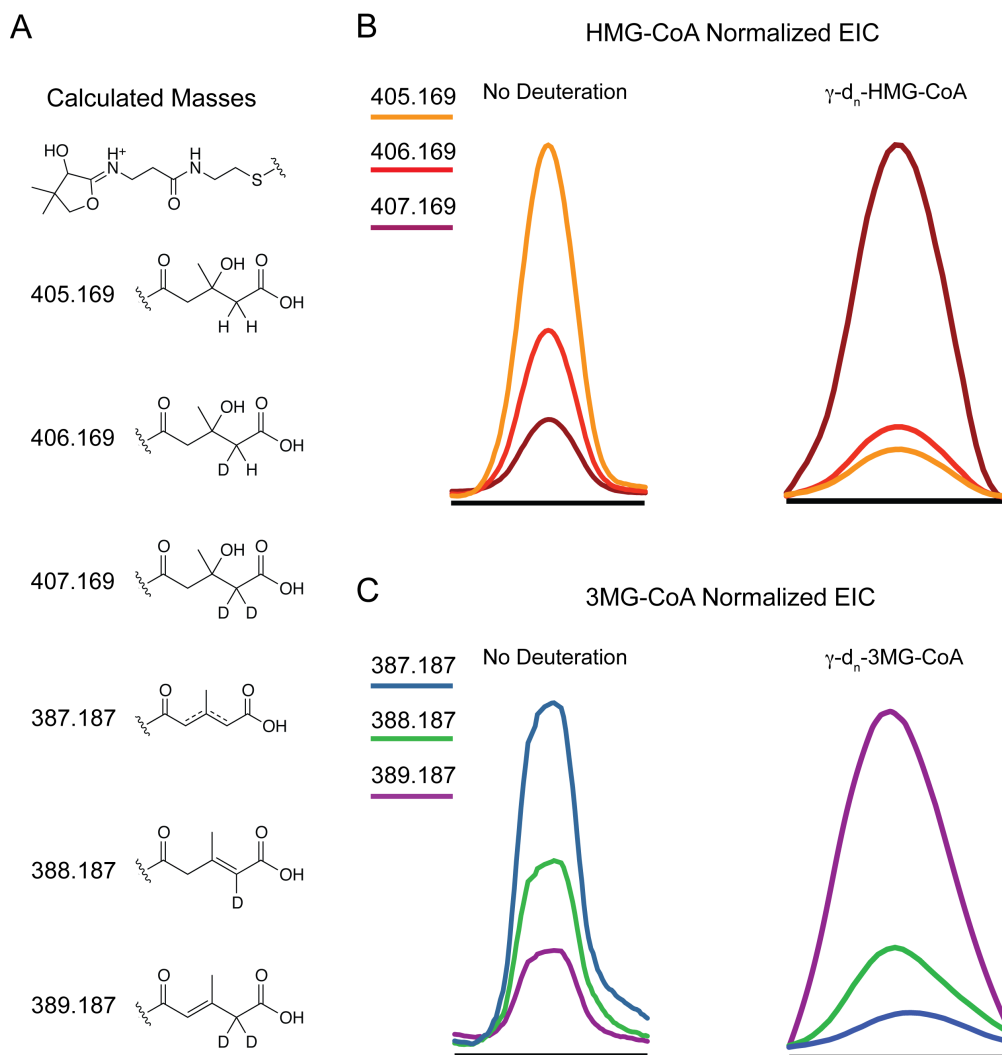


Figure 17. LC/MS analysis of γ -d₂-HMG-CoA dehydration by BryT. A) Expected masses from variable deuteration. B) Normalized extracted ion chromatograms for 406.169±1 Da, truncated to show only HMG-CoA elution peaks. C) Normalized extracted ion chromatograms for 388.167±1 Da, truncated to show only 3MG-CoA elution peaks.

In order to corroborate these data, a large scale reaction and purification effort was undertaken. Fresh BryT was purified and immediately added to a 10mg-scale reaction of HMG-CoA. Because BryT catalyzes both dehydration and the reverse reaction, and also prefers a single stereoisomer of the ~50:50 mixture present in commercial HMG-CoA, the yields of these reactions are typically modest (~10%). Representative HPLC traces are shown in (Figure S6)

Through careful method development and scaling, a rapid workup and HPLC purification protocol was established that minimized the amount of time that the 3MG-CoA product spent at room temperature. Even so, the low yields and degradation were persistent problems. Eventually, NMR spectra were obtained using samples of sufficient concentration and purity. Initial proton NMR data showed conclusively that a methylene is lost and a new methine is gained, indicating a new double bond. To enable analysis of the double bond installed by BryT, a partial assignment of commercial HMG-CoA starting material was carried out. By comparing this assignment to the 3MG-CoA spectra, the location lost methylene could be correlated on HMG-CoA (**Figure 19**). This analysis concluded that the methylene signal was lost from the α -carbon of 3MG-CoA, indicating an α/β dehydration. This agrees with the mass shifts in the deuteration experiments and strongly supports that BryT carries out α/β dehydration activity.

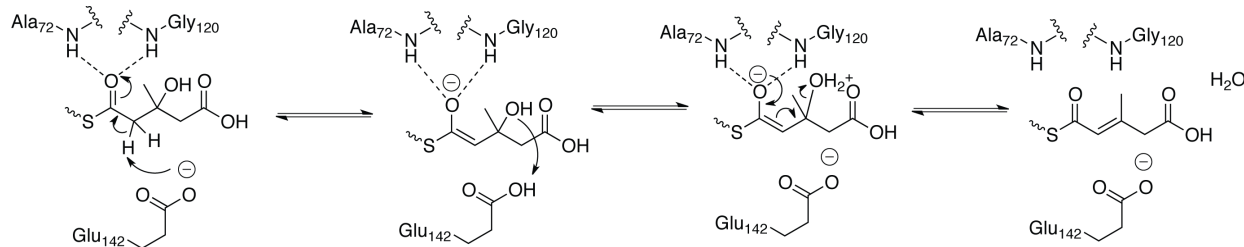


Figure 18. Acid-catalyzed mechanism of enoyl-CoA hydratases. Stabilization of an oxyanion is a common feature of the crotonase superfamily. The rate limiting step of the reaction is likely the initial deprotonation. Residue numbering refers to BryT.

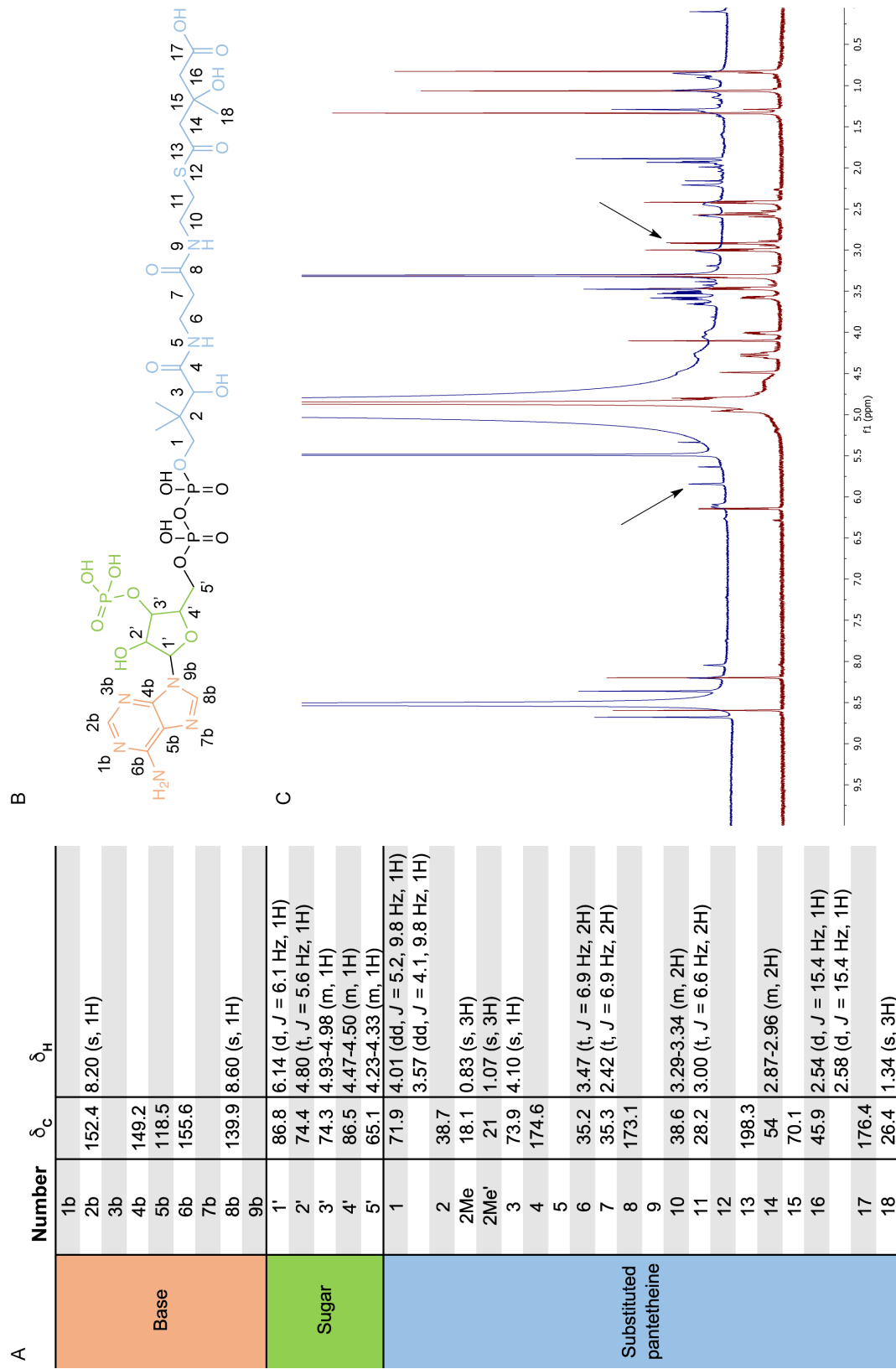


Figure 19. NMR analysis of HMG- and 3MG-CoA generated by BryT. Signals were assigned using commercial HMG-CoA (A), (B). By comparing the 3MG-CoA spectrum (C, blue) to HMG-CoA (C, red), the lost methylene and gained methine (labeled with arrows) are apparent. Some impurities are present in the product, most significantly ammonium formate (8.5), water (4.8), and methanol (3.3).

2.2.3 BryA Module 3 MT is the Beta Branching Cassette Carboxyl O-MT

Previous attempts to identify the carboxyl O-MT that generates the bryostatin methyl esters were not met with success. Several candidate SAM-dependent methyltransferases were identified within the sequencing data: BryA module 3 and BryB module 4 both contain MTs, and a standalone MT was identified near the other beta branching cassette genes. None of these domains were able to catalyze methylation of the substrate mimics in initial tests, and an impasse was reached that required additional information. However, revisiting this search after the discovery of BryT proved fruitful.

After publication of the draft genome and accompanying transcriptomic data, BryS could be confidently removed as a candidate for O-methylation. In a BLAST search, many close homologues (>50% identity) returned as tRNA C-methyltransferases. It is notoriously difficult to predict the substrate of methylation for class I MTs by sequence alone, but the large number of BLAST hits for C-MTs raised suspicion. In addition, while transcripts were detected for the rest of the pathway, no mRNA transcripts of BryS were detected in the transcriptomic investigation carried out by the Kwan group. Taken together, these results provide strong evidence that BryS is not the carboxyl O-MT, and may in fact not be involved in bryostatin biosynthesis at all.

Computational analysis and BLAST searches of other candidate methyltransferases against the MiBIG database(128) provided some guidance. BryA module 3 MT returned hits for a few N- and O- standalone methyltransferases with modest sequence identity (~26%). BryB module 7 MT, at 65% identity to BryA module 3 MT, returned similar hits. The other pair of bryostatin MTs had different results. At 95% identity, BryB module 4 MT and BryC module 9 MT are almost identical and return a large number of module-integrated PKS C-methyltransferases with a typical sequence identity of ~45%. Considering that BryA module 3

MT and BryB module 7 MT are located within the modules that are predicted to accept beta branches, these BLAST results strongly suggest that these are the carboxyl O-MTs of the beta branching cassette.

With the function of BryT confirmed, it was hypothesized that the carboxyl O-MT may require dehydration before being able to methylate its substrate. To test this, BryA module 3 was cloned from the cosmid library, expressed and purified. Additionally, standalone BryA module 3 ACP was purified and loaded with HMG. After incubating this HMG-ACP with BryT, BryA module 3, and S-adenosyl methionine (SAM), a mass shift corresponding to dehydration and methylation was detected on the phosphopantetheine ejection ion (**Figure 20**). Several constructs of the excised O-MT domain were then cloned and purified, greatly improving the specific activity and total turnover of the system. Both in the context of the module and as a standalone domain, the O-MT acted with a strict dependence upon BryT dehydration, ACP tethering, and supplemental SAM, suggesting that it does not copurify with SAM.

Much like the dehydration catalyzed by BryT, mass spectrometry alone cannot determine the location of the added methyl mass. Because of BryA O-MT was unable to utilize CoA-tethered substrates, scale-up and isolation was not an option. So, we turned again to organic chemistry to solve this problem. Diazomethane (DM) is a reagent that is used to convert carboxylic acids into methyl esters with a high degree of selectivity. Although it is acutely toxic, carcinogenic, and explosive, these risks can be almost completely ameliorated by using a trimethylsilyl-DM (TMS-DM) derivative. The covalently linked TMS group is hydrolyzed in situ to generate small, safer amounts of diazomethane, which then react with carboxylic acids in the same way as DM alone.

This reaction can be used a control for O-methylation by coupling TMS-DM treatment

with enzymatic methylation. By treating HMG-ACP with BryT and BryA O-MT prior to chemical treatment with TMS-DM and assessing the outcome by phosphopantetheine ejection, the location of methylation may be tested. If BryA O-MT is, in fact, a carboxyl O-MT, it will preclude methylation by DM, resulting in a mass shift equivalent to one methylation (14 Da). Enzymatic methylation at any other position will not preclude methyl ester formation, resulting in a mass shift of two methylations (28 Da).

TMS-DM is commercially available as a solution in toluene, which presents a challenge for use with aqueous solutions of biomolecules. Initial experiments of TMS-DM treatment alone resulted in a bimodal distribution of methylation of HMG-ACP, with approximately equal populations of ~120 methylations and 0 methylations, with almost no methylation of the phosphopantetheine ion detected. This is perhaps because the only contact made between HMG-ACP and TMS-DM was at the aqueous-organic interface, resulting in the extensive methylation of the fraction of protein in close proximity to the interface and no methylation of protein that was deeper into the aqueous layer. It is interesting that even in the population of extensively methylated HMG-ACP, there was almost no conversion of HMG-phosphopantetheine to methyl ester. This may be due to sequestration of the phosphopantetheine arm away from solvent, a phenomenon that has been studied in detail in type-II FAS and PKS ACPs(129, 130), but remains poorly studied in type-I PKS. The lack of study may be because of the entropic advantage of covalently linked catalytic domains enjoyed by type-I systems, likely causing substrate sequestration to be a much less significant factor than in type-II systems.

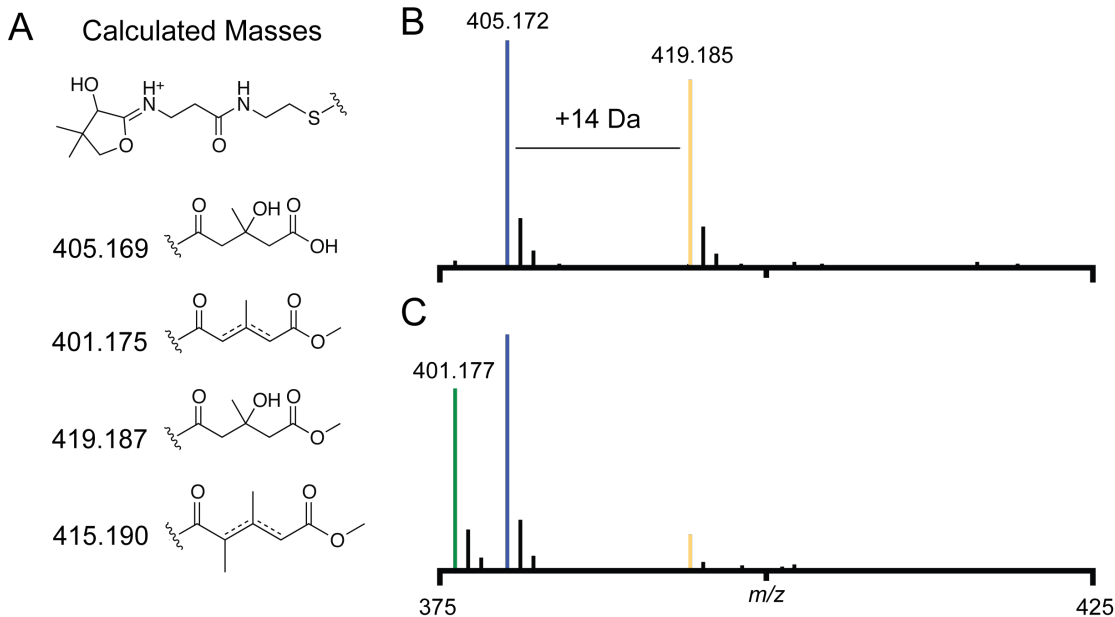


Figure 20. Phosphopantetheine ejection ions of HMG-ACP_A species treated with TMS-diazomethane. Calculated masses (A), HMG-ACP_A alone (B), and HMG-ACP_A pretreated with BryT and BryA O-MT prior to TMS-diazomethane treatment (C).

Serial dilutions of TMS-DM from toluene into methanol resolved this problem and altered the methylation pattern. HMG-ACP is soluble in methanol at concentrations up to ~70%, and the addition of dilute TMS-DM to protein that has been pre-diluted in methanol results in the generation of diverse methylation states of the protein (**Figure S6**). Using this method, ~90% of the protein is methylated at least once, with distinct masses corresponding to 1, 2, 3... and so on methylations up to ~20. The abundance of these species roughly approximates a Gaussian distribution centered at 11 methylations. Interestingly, the frequency of methylation of the HMG moiety on the phosphopantetheine arm is still much lower than that of the protein as a whole, with roughly 60% of the original HMG-phosphopantetheine ion still intact. This, like the previously described experiments, also implies some sort of sequestration of the phosphopantetheine arm away from solvent.

Most importantly, only one methylation was detected on the phosphopantetheine arm of HMG-ACP that had been treated with BryT and BryA O-MT prior to TMS-DM treatment. This indicates that the only site competent for methylation by TMS-DM has been blocked enzymatically by the O-MT. Other control samples confirmed that TMS-DM was only able to methylate in one location of the phosphopantetheine arm. Incubation of HMG-ACP with BryT or no enzyme prior to TMS-DM treatment also resulted in masses corresponding to a single methylation of the phosphopantetheine arm. This confirms the role of BryA O-MT as the carboxyl O-MT of the beta branching cassette.

2.3 Discussion

Taken together, the presented results support a much more complete picture of beta branching in bryostatin biosynthesis (**Figure 21**). The donor ACP for BryR, the missing ECH₁, and missing methyltransferase are conclusively identified, and the regiochemistries of these products were confirmed using a variety of chemoenzymatic techniques. The insights into this unique beta branching cassette contributes to potential applications engineered PKS pathways. Additionally, these results provide insight into the abilities of beta branching cassettes in general. The techniques used to obtain these data are also generalizable and provide additional chemical tools for the study of polyketide synthases.

2.4.1 BryT Acts on CoAs as well as ACP While O-MT Requires ACP

The existence of a specific donor ACP that is accepted by BryR has been confirmed. Every beta branching cassette found in type I PKS pathways has been identified with an associated donor ACP that brings an acetyl (or propionyl(40)) unit to HMGS for incorporation into the polyketide, and bryostatin is no exception. HMGS has been shown in several pathways, including bryostatin, to have a high degree of selectivity for both donor and acceptor ACPs. The

basis of this selectivity has been described structurally in the curacin pathway, and the residues that are key to this binding mode appear to be conserved in both BryR (D214, D222, E225, and R266, CurD numbering) and BryU (R42, I46, M47, CurB numbering)(98). While the affinity of the curacin donor ACP is far tighter for its native HMGS than for BryR, the measured affinity still suggests that many of these donor ACPs have high degrees of interoperability(43). If future attempts are made to transfer HMGS functionality to other pathways, the matching of donor ACP for native HMGS will likely not be a crucial consideration.

The role of BryT as the bryostatin ECH₁ has also been confirmed biochemically. This beta branching component has long been hypothesized to carry out α/β dehydration and set up the polyketide intermediate for pyran ring closure. Our experiments have demonstrated that this ECH₁, much like enoyl-CoA hydratases in primary metabolism, specifically catalyzes the reversible hydration/dehydration of α/β double bonds, not β/γ . Barring the discovery of truly unusual chemistry, it appears likely that all PKS ECH₁ share this product selectivity. The high sequence identity (48.9% across 13 pathways from a variety of organisms), conserved mechanism, and conserved active site pocket residues indicate that all PKS ECH₁ carry out α/β dehydration.

The activity of BryA O-MT provides further evidence that BryT natively carries out α/β dehydration. In experiments with jamaicamide and curacin ECH₁, BryA O-MT was able to methylate the substrate just as effectively as with BryT, suggesting the same configuration of double bond is produced by both of these ECH₁. Previous studies concluded from UV/Vis data that the curacin and myxovirescin ECH₁ carry out α/β dehydration(42, 127). While UV/Vis provides weak evidence for these conclusions, the present analysis shows that the previous conclusions were correct. Additionally, double bonds impart a large degree of steric rigidity, and

thus it is unlikely that Δ^2 and Δ^3 isomers would be accommodated by the same active site. Such accommodation is especially unlikely for BryA O-MT, which has stringent substrate specificity as demonstrated by its ability to discriminate very effectively between HMG-ACP and 3MG-ACP.

The sequential activities of BryT and BryA O-MT provide interesting insight into the order of events of the bryostatin beta branching cassette and the parallels it has with other pathways. In other beta branching cassettes, the reversibility of ECH₁ dehydration is prevented by the relative irreversibility of the next step. In 43 of the 47 beta branches found in pathways in the MiBIG database(128), dehydration is followed immediately by ECH₂ decarboxylation, thus preventing rehydration of the intermediate and pushing the equilibrium forward. Two of the remaining four branches are found in the bryostatin pathway, which does not contain ECH₂. Instead, the steric bulk of carboxyl O-methylation prevents the substrate from accessing the BryT active site, preventing BryT from catalyzing rehydration. The relative scarcity of dehydrated product compared to the large accumulation of dehydrated methyl ester product is indicative of the reversibility of dehydration and irreversibility of O-methylation. Additionally, no hydrated methyl ester product was detected without the addition of TMS-DM, demonstrating that BryT cannot catalyze rehydration after the methyl ester has been installed.

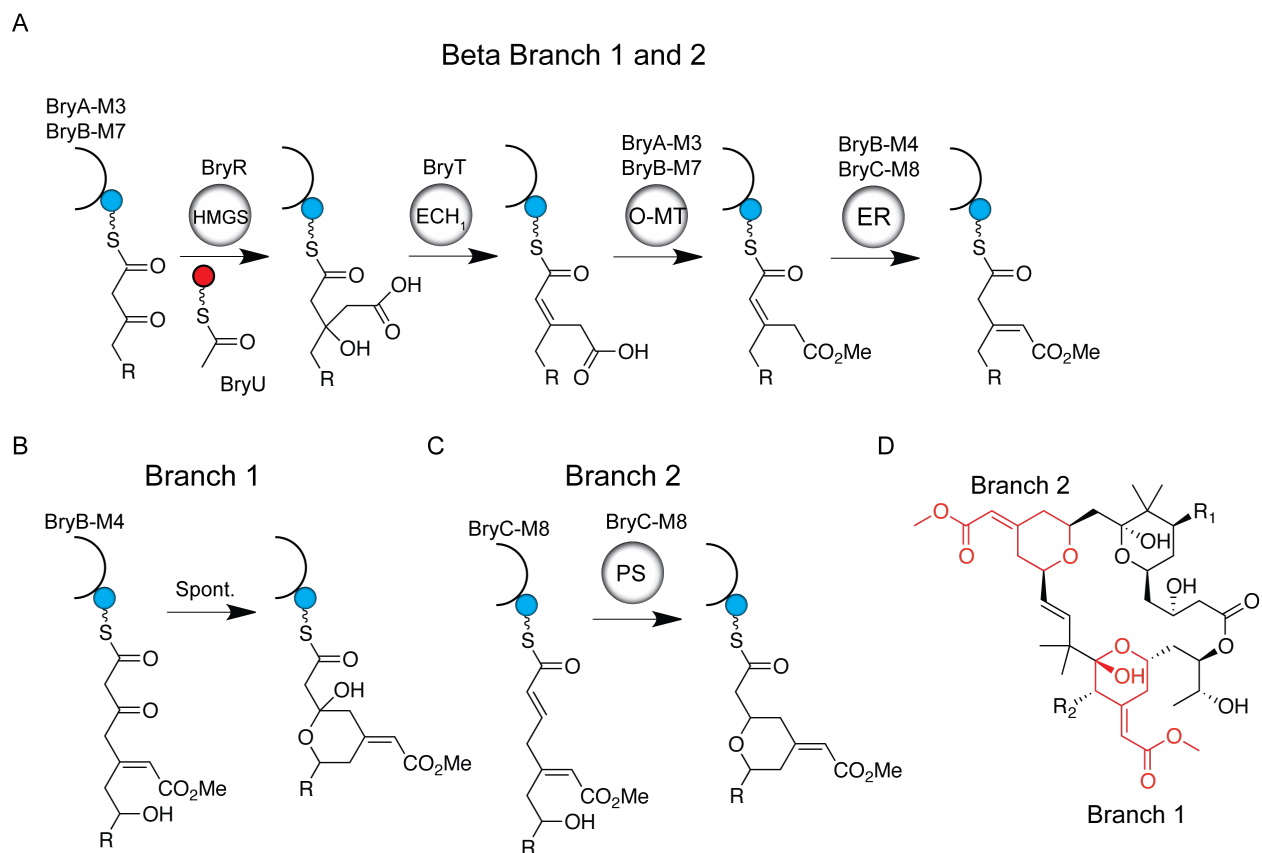


Figure 21. Updated model of bryostatin beta branching. Both beta branches depend on the activities of BryR, BryT, and intramodular O-MT within BryA and BryB (A). We propose double bond isomerization prior to cyclization is carried out by ER domains within BryB and BryC. Cyclization of branch 1 likely occurs spontaneously (B), and branch 2 cyclization is catalyzed by a PS domain within BryC (C).

Unexpectedly, the combined activities of BryR, BryT, and BryA O-MT do not account for the complete set of enzymatic conversions necessary to obtain the complete structure of the distinctive “bryopyran” rings with their unique vinyl methyl esters. However, the structure of their product verifies several previous hypotheses and supports the proposal of a modified model of bryostatin biosynthesis. After beta branching by BryR facilitated by acetyl-BryU, BryT dehydrates the beta position to a Δ^2 product (α/β). The carboxylate of this 3-methylglutaconate-like product is then O-methylated. Because the Δ^2 double bond is trans, it must be isomerized

before the geometry will allow pyran ring closure. The predicted activities of several downstream domains are the most likely sources of these conversions.

Double bond isomerization may be catalyzed by an as-yet unidentified enzyme in the genome or by the N-terminal ER-like domains found in both BryB and BryC immediately downstream from the beta branch modules. These domains are part of the nucleotide repeat regions in the gene cluster and have sequence identities >97% (both nucleotide and amino acid). It is unclear if this identity has been maintained due to a functional pressure or if it is simply an artifact of recent genetic recombination. The myxovirescin gene cluster also contains a beta branch associated ER domain that is 75.3% identical to the bryostatin ER domains, but this domain has an obvious function in the reduction of the α/β double bond installed by a ECH₁/ECH₂ beta branch. While no alteration of the double bond configuration was detected using our γ -d₂-HMG-CoA system, it is quite possible that these domains require interaction with an ACP in order to properly process their substrates.

Pyran formation is likely caused by different mechanisms for each beta branch due to the different reductive domains present in the module immediately downstream of the beta branch. In the first branch, spontaneous hemiketalization is likely to occur due to the presence of a ketone and alcohol bookending the beta branched carbon. This reaction has been frequently observed and is a well-known obstacle for synthetic chemists that attempt to synthesize linear polyketide intermediates. The second beta branch cannot undergo hemiketalization as the reductive domains present in BryC module 8 would result in an α/β double bond, an unsuitable target for uncatalyzed nucleophilic attack. Instead, this module contains a pyran synthase (PS) domain, which likely catalyzes ring closure via a Michael addition as has been demonstrated by PS domains from the pederin(131) and ambrutricin(132) pathways. The third pyran ring in

bryostatin is not associated with a beta branch, but also fits into the logic of the other two rings: there is no PS domain, and the predicted structure of the linear polyketide is configured for spontaneous hemiketalization.

2.4 Summary

We have successfully confirmed the roles of BryU, BryT, and BryA O-MT in the construction of the unique vinyl methyl ester moieties found in the structures of the bryostatins. These enzymes act in a strict order during the beta branching of the polyketide intermediate, beginning with donation of an acetyl unit by BryU to beta branching, then dehydration by BryT, and then O-methylation by BryA O-MT. We have also confirmed the regiochemistry of both dehydration and O-methylation steps using novel chemoenzymatic tools. The generation and processing of γ -d₂-HMG-CoA revealed α/β (Δ^2) position of the double bond resulting from BryT dehydration. Chemical generation of the methyl ester of HMG- and 3MG-ACP_A combined with enzymatic methylation by BryA O-MT conclusively demonstrated carboxyl O-methylation with a strict dependence upon α/β dehydrated 3MG-CoA. These results support the proposal of a modified model of bryostatin beta branching that includes post-beta branching double bond isomerization prior to spontaneous hemiketalization to generate the unique vinyl methyl ester pyran rings found in all bryostatins.

2.5 Conclusion

In the majority of beta branching cassettes, the double bond configuration of ECH₁ products is obfuscated by downstream modifications. In most cases, this means ECH₂ decarboxylation, which has been shown to be the determinant of product regiochemistry in the cases where it is used(45). In newly discovered pathways, the knowledge that ECH₁ catalyzes

α/β dehydration can be used to inform biosynthetic hypotheses about intermediates and biosynthetic strategies.

These results have implications in the future of PKS pathway engineering. Because beta branch acceptor ACPs can be readily identified by a primary sequence motif, one can envision the modification of a normal modular ACP to transform it into a beta branch acceptor. This method would result in the introduction of a novel chemical structure at that step in polyketide synthesis and ultimately a novel polyketide product. Of course, this depends on the processing of this novel intermediate by the remaining downstream domains of the pathway, but a great deal of progress is being made in this area. The addition of a vinyl methyl ester would have large implications for the activity of a novel polyketide, and the combined activities of the bryostatin beta branching cassette can install this functionality.

2.6 Materials and Methods

Protein Expression and Purification. Open reading frames encoding BryT and BryU were identified in the genomic sequence of *Endobugula sertula*(10). The coding sequence for *S. aureus* mvaS was obtained from GenBank. The coding sequence for BryM3-OMT was identified within the full BryA gene. These sequences were codon-optimized for expression in *E. coli* and linear DNA was purchased from Integrated DNA Technologies. Overexpression plasmids were obtained by amplifying the genes from the linear DNA and inserting into pMCSG7 through ligation independent cloning (LIC(133)). BryR was expressed from a previously described pMCSG7 construct(43). BryM3-ACP was amplified from a cosmid containing a fragment of the *Endobugula sertula* genome(41) and cloned into pMCSG7 using LIC. Soluble expression and purification of BryR and BryT was achieved by coexpression with GroEL and GroES from pGro7 (Takara(134)). ACPs, mvaS, and BryM3-OMT were expressed

in *E. coli* BL21 (DE3) for apo ACP, and in *E. coli* Bap1 (which contains a genomic integration of *sfp* phosphopantetheinyl transferase(135), for holo ACPs and PKS modules. Transformed cells were grown at 37°C in 1L of Terrific Broth (TB) in 2L baffled Fernbach flasks until OD₆₀₀ ~ 1. For BryR and BryT, transformed cells were grown at 24°C, and overexpression of GroEL/GroES was induced by addition of 2g/L arabinose at OD₆₀₀ ~ 0.4. Cells were cooled to 20°C and overexpression of protein of interest was induced by addition of 0.2mM isopropyl b-D-thiogalactopyranoside (GoldBio). Cells were grown for an additional 12-16 hours with shaking before harvesting by centrifugation and freezing at -80°C. Cell pellets were subsequently thawed at 4°C and resuspended with 7mL/gram buffer A [300mM NaCl, 10% glycerol, 50mM HEPES pH 8.0]; 0.8mg/mL lysozyme and 5mM MgCl₂ were added. Cells were lysed by sonication and clarified by centrifugation. Cell free lysate preparation was completed by flash freezing and storing at -80°C. For purified proteins: clarified lysates were passed over a 0.45 µm filter and imidazole (pH 7.5) was added to 45mM before application to a 5mL HiPrep FF NiNTA (GE Healthcare) column. The column was washed with 100mL buffer A with 50mM imidazole added and proteins were eluted with 20mL buffer A with 200mM imidazole. High concentration eluted fractions were pooled and exchanged into storage buffer [100mM NaCl, 20mM HEPES 7.8, 10% glycerol] using PD10 columns (GE Healthcare), flash frozen in liquid nitrogen and stored at -80°C.

ACP Loading and Enzyme Activity Assay. Purified apo-ACPs were covalently modified *in vitro* using acyl-CoAs and a phosphopantetheinyl transferase cloned from *Streptomyces verticillus*(136) and purified in *E. coli*. Briefly, ACP was incubated with 0.1 eq svp, 10eq acyl-CoA (Mal-, Ac-, Acac-, HMG-CoAs, Sigma), 10mM MgCl₂, 50mM HEPES pH 7.0, 10%

glycerol for 1.5 hours at 30°C. Loaded ACPs were purified by gel filtration chromatography on a Superdex S200 equilibrated with storage buffer. Elution fractions were concentrated to ~250μM, flash frozen in liquid nitrogen and stored at -80°C.

BryT reactions were carried out by adding 30μM purified BryT to 100μM loaded ACP in 150mM NaCl, 50mM HEPES 7.8, 10% glycerol and incubating at 30°C for 1 hour. BryR reactions were carried out using the same method, although with 50μM each of acyl-BryU and acyl-BryM3-ACP. Methyltransferase reactions were carried out in the same buffer with 5mM SAM and 20μM purified BryM3-OMT added. Reactions were quenched with 8 volumes of 50% MeOH containing 5% formic acid and centrifuged at 17,000 rcf for 20 minutes before assessment by LC-MS using the phosphopantetheine ejection assay(137). Proteins were separated from other reagents using reverse phase chromatography (Waters Acquity BEH300 C4 column 1.7μm, 2.1x50mm) using a gradient of 15-100% B (A, H₂O with 0.1% formic acid; B, 95% acetonitrile in H₂O with 0.1% formic acid) at a flow rate of 0.4mL/min.

Deuteration Reactions. Deuterated d₃-acetyl-CoA was generated by adding four equivalents of d₆-acetic anhydride (Sigma) to a 100mM solution of CoASH (trilithium salt, Sigma) in 1M HEPES pH 7.0. The reaction was incubated at 23C for five minutes and the resulting d₃-ac-CoA was used without further purification to minimize deuterium exchange. Acetylation was confirmed by comparison with Ac-CoA standards and by LC-MS using reverse phase chromatography (xBridge C18 column 3.6μm, 2.1x250mm, Waters) using a gradient of 7-15% B (A, 100mM ammonium formate pH 5.0; B, acetonitrile) at a flow rate of 1.5mL/min. Deuterated γ-d₂-HMG-CoA was generated by incubation of 5mM d₃-Ac-CoA with 5mM Acac-CoA (sodium salt, Sigma) and 100uM purified mvaS in 100mM NaCl, 20mM HEPES pH 7.0, 1mM TCEP,

and 10mM MgCl₂ for varying times (10 minutes to 2 hours) at 25C. Reactions were quenched with 3 volumes 100% MeOH, spun at 17000 rcf for 30 minutes, and assessed by LC-MS using the HPLC method described above. Deuterated γ -d₂-HMG-CoA was dehydrated in situ by adding 30uM BryT to the mvaS reaction. Analysis of these reactions followed the same protocol.

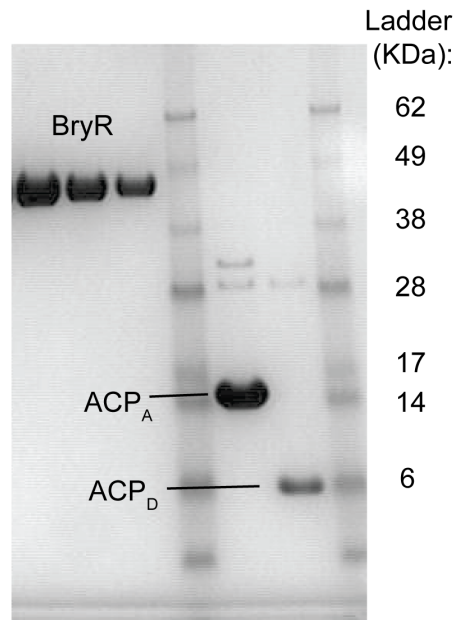
Large Scale Purification and NMR of 3MG-CoA. Large scale reactions of BryT with HMG-CoA were carried out by purifying BryT using the previously described protocol and immediately adding HMG-CoA to 10mM to the buffer exchanged protein. Reactions were incubated at 25C for 2 hours before being quenched by 4 volumes cold methanol and centrifuged. Persistent precipitated material was removed via filtration over a cellulose membrane. The resulting material was dried on a Biotage V10 evaporator system to a volume of ~1.5mL. Repeated injections of 0.5mL were purified using reverse phase chromatography (Luna C18 5 μ , 21.2x250mm, Phenomenex) using a gradient of 7-11% B (A, 100mM ammonium formate pH=5.0; B, acetonitrile) at a flow rate of 8mL/min. Fractions were collected based on absorbance at 254nm, evaporated using the V10 system, and immediately frozen using liquid nitrogen.

Methyl Ester Synthesis with TMS-Diazomethane. HMG-BryA module 3 ACP was prepared as described above. TMS-diazomethane (Acros) was drawn via syringe from a sealed bottle under positive pressure of nitrogen and transferred to a 1.5mL tube. This material was either added directly to the protein sample or serially diluted with methanol. Each step of the serial dilution was ~8 fold and was accompanied with careful but thorough mixing by pipette. Protein

samples were prepared by careful mixing of protein stocks (~150uM HMG-ACPA) with 80% methanol to a variety of final protein and methanol concentrations. Reactions of neat and diluted TMS-DM with protein stocks were prepared by slowly adding the prepared TMS-DM to the prepared protein samples and carefully mixing before allowing the reaction to run for varying lengths of time (10 seconds to 2 hours). Samples were carefully quenched (nitrogen gas evolves quickly from unreacted TMS-DM) using 15% formic acid in methanol, centrifuged for 20 minutes, and then analyzed using the LC-MS phosphopantetheine ejection assay described above.

2.7 Supplemental Data

Purified BryR, ACP_D, and ACP_A
SDS-PAGE



ACP_D - BryU
ACP_A - BryA Mod3 ACP

Figure S1. SDS-PAGE of BryR, BryU, and BryA Mod3 ACP.

Purified BryT SDS-PAGE

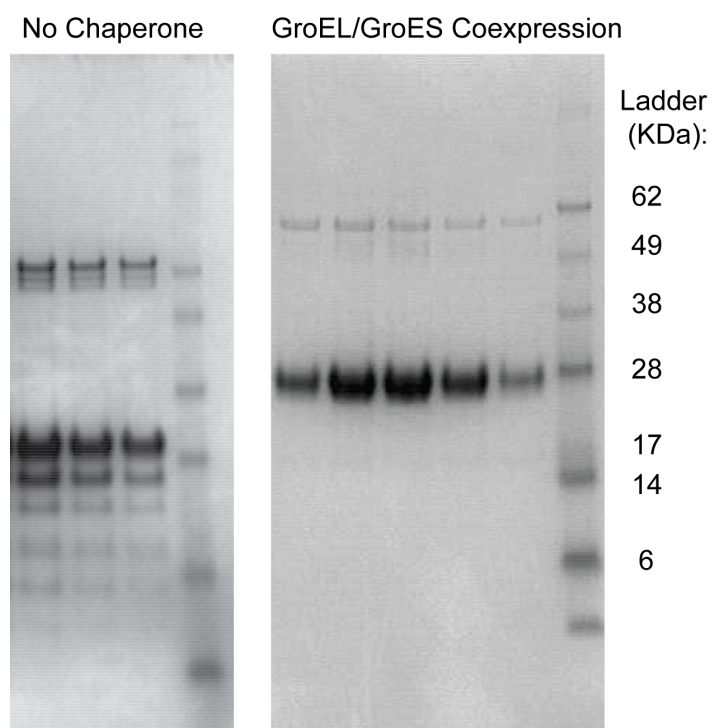


Figure S2. SDS-PAGE of BryT with and without coexpressed chaperones. Expected mass of BryT with his-tag fusion is 31 KDa

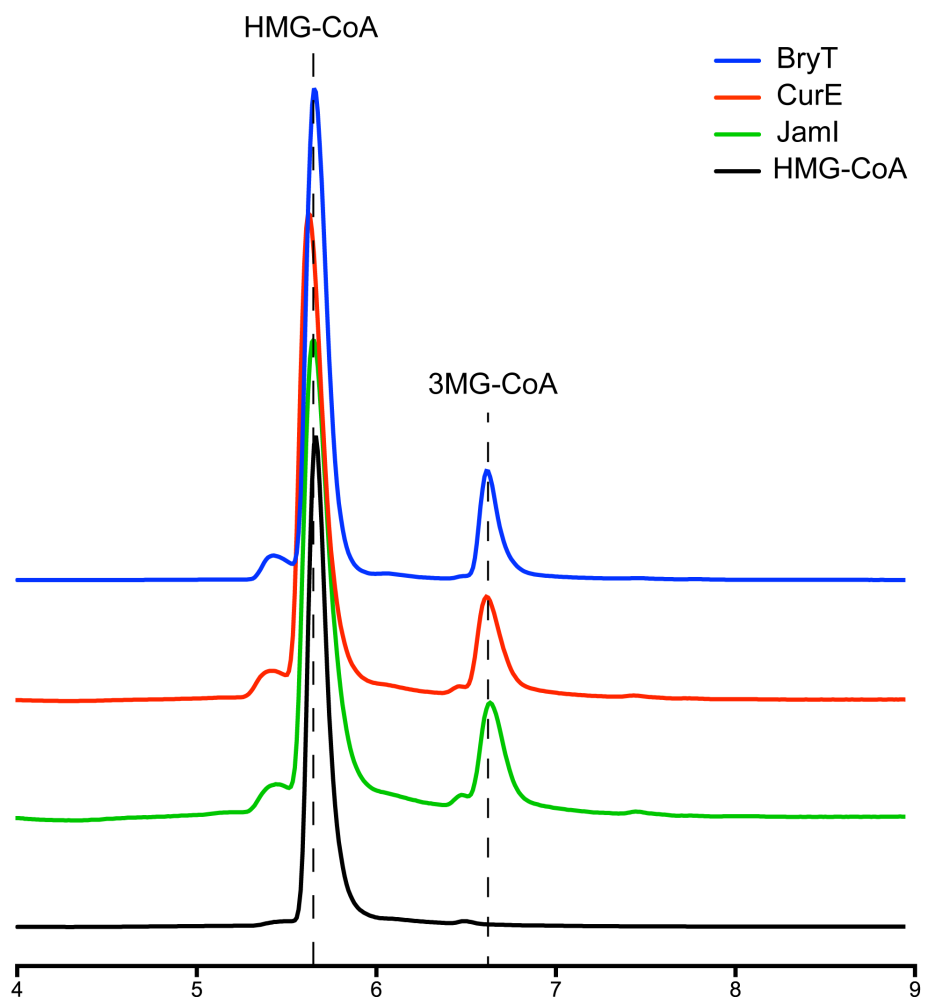


Figure S3. Comparison of HPLC traces of reactions of HMG-CoA with BryT, CurE, and JamI. Reactions were prepared using 200uM HMG-CoA and 40uM protein.

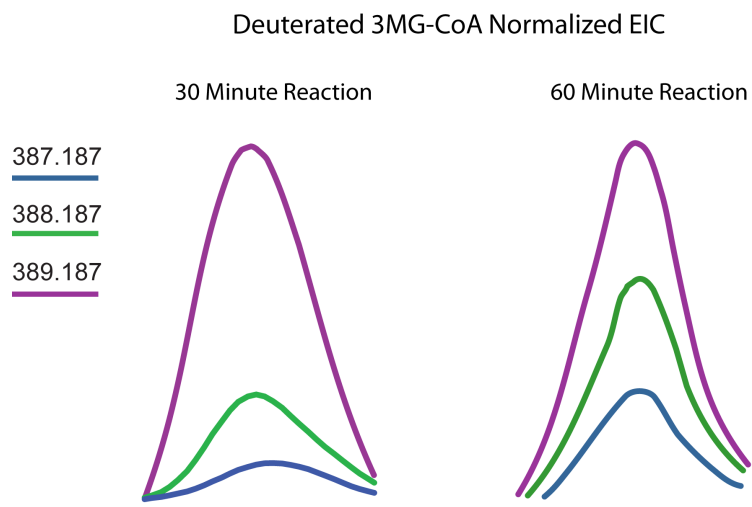


Figure S4. Changes in deuteration of 3MG-CoA with different length reactions. With a 30-minute reaction of γ -d₂-HMG-CoA with BryT (left), the abundance of d₂-3MG-CoA is much higher than that of a 60-minute reaction (right), likely due to deuterium exchange.

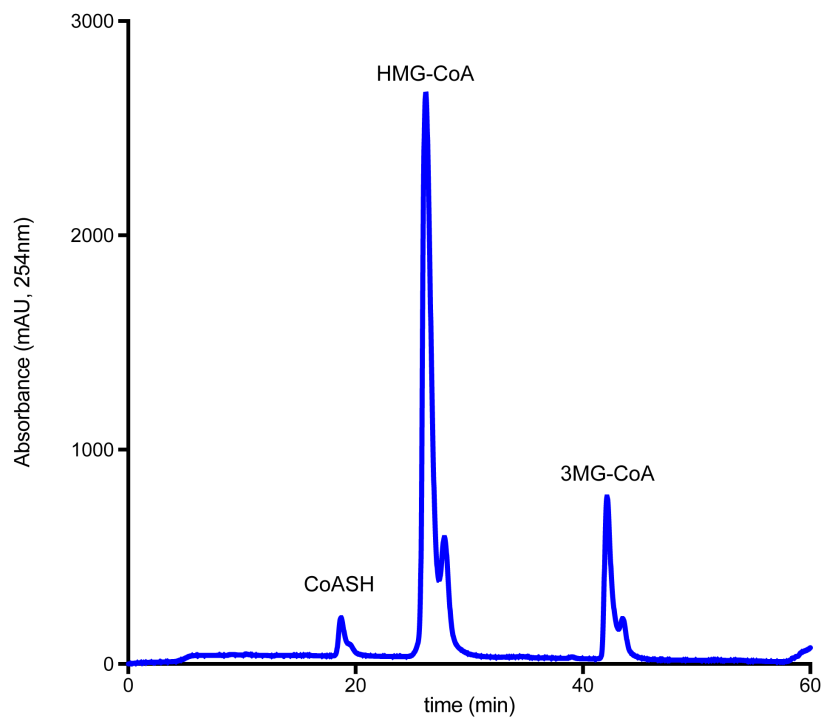


Figure S5. Representative HPLC of preparative-scale BryT reaction with HMG-CoA. Baseline separation was achieved at the preparative scale using 10mg of HMG-CoA starting material and 40uM BryT. 3MG-CoA peak was isolated and used in NMR analysis.

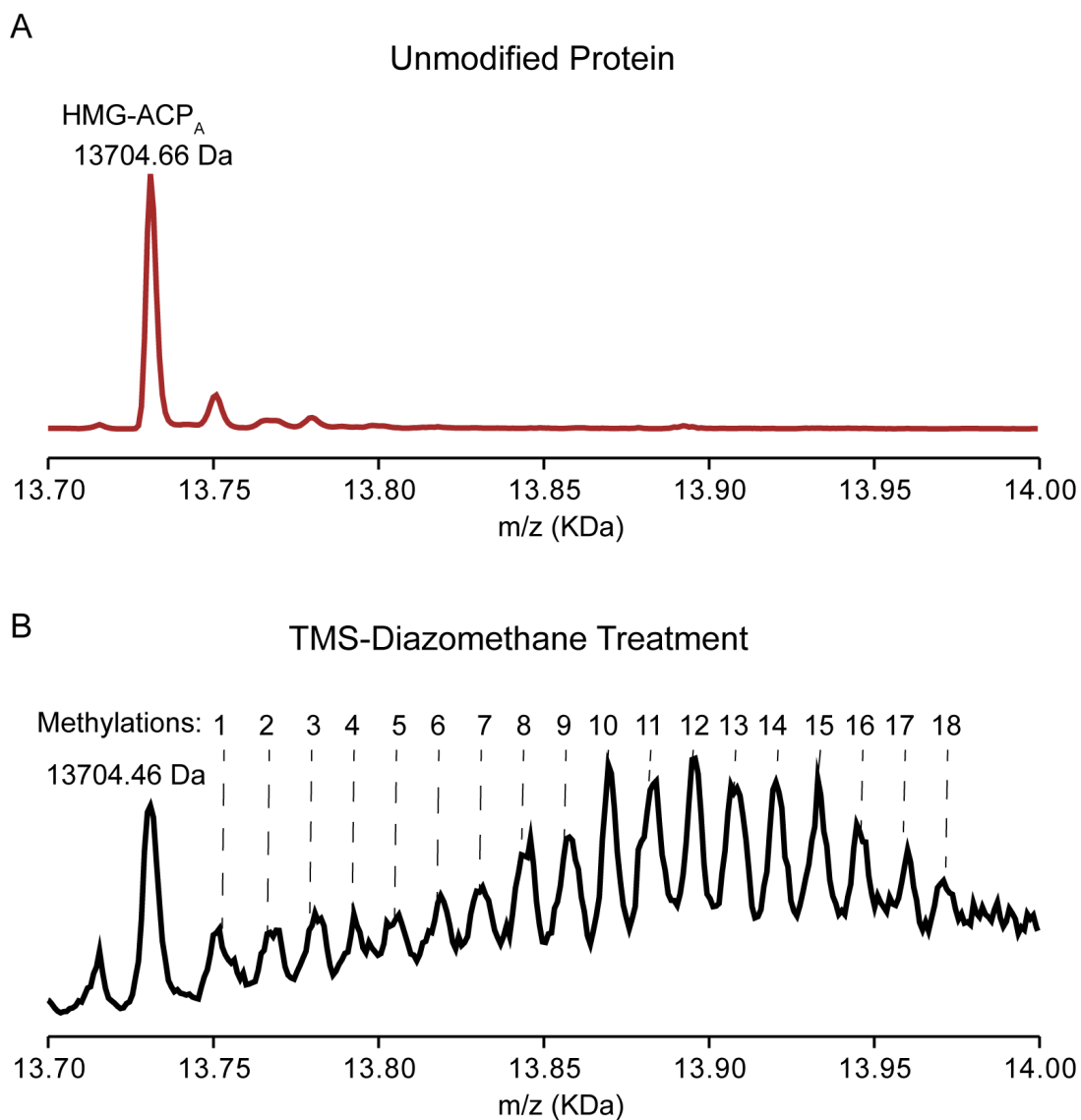


Figure S6. Deconvoluted intact protein masses of HMG-BryA Mod3 ACP with and without TMS-diazomethane treatment. A) Untreated protein shows a single dominant species at 13.7 KDa. B) Diazomethane treatment results in individually detected species corresponding to 1 through 18 methylations of the starting material.

Chapter III

Structural Analysis of PKS Proteins by Ion-Mobility Mass Spectrometry

3.1 Introduction

The architecture of type I polyketide synthases (PKS) is a complex and technically challenging research target. The variability in peptide length, module makeup, domain functionality, and organism of origin confer unique physical and chemical properties to each PKS protein. These properties add to the challenge and to the reward in understanding their biochemistry, as is the case in many intricate biological systems. The most promising reward lies in the ability of PKS to synthesize a huge diversity of chemical structures, including a catalog of known bioactive compounds and a hypothetical multitude of rationally designed molecules. In some cases, these hypothetical products have been achieved by modifying the PKS(138-140). Many of the product structures, regiochemistries, and functional groups achieved by PKS are difficult or impossible for synthetic chemistry. These abilities make PKS extremely appealing biocatalysts, and with a deeper understanding of their function, more of their hypothetical abilities will be become reality.

Polyketide assembly is achieved by non-iterative biosynthetic modules made up of distinct domains that each play a role in the catalytic cycle of extension. The minimal activity of a PKS module requires ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP)(**Figure 22A**). The ACP serves to carry intermediates on a phosphopantetheine prosthetic

group, covalently attached by a phosphopantetheinyltransferase. At the start of a modular catalytic cycle, the ACP of the upstream module carries the extension intermediate to the active site of the current module KS, where it is transferred to a specific cysteine side chain. The AT domain then selectively transfers a CoA-derived extender unit (typically an α -substituted malonate residue) to the thiol of the phosphopantetheine arm of the ACP. The ACP brings this extender unit to the KS active site, where it undergoes a decarboxylative Claisen condensation with the upstream polyketide intermediate, thus resulting in a two-carbon extension of the nascent polyketide, which is now linked to the phosphopantetheine arm and ready for transfer to the downstream KS active site.

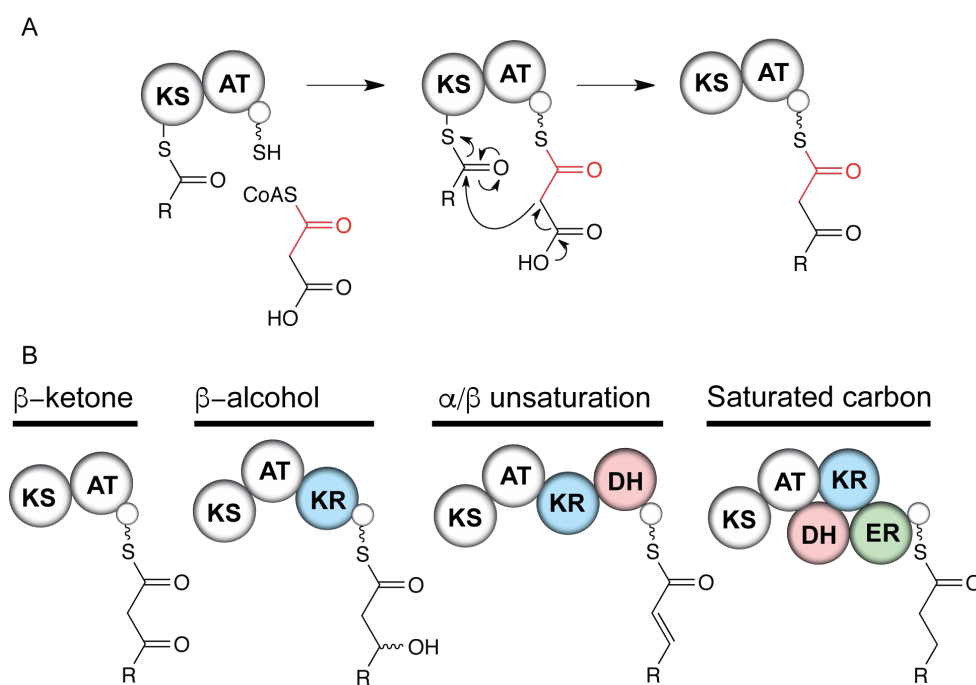


Figure 22. Polyketide synthesis extension and potential reductive domain architectures. Polyketide extension is effected by the decarboxylative condensation of a malonyl extender unit selected by an acyltransferase domain (AT) catalyzed by the ketosynthase (KS) domain (A). The beta ketone installed by the previous round of extension is modified by the reductive domains of the current module (B). Ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains may be present in a given module.

After KS extension, a variety of activities may be carried out on the intermediate prior to transfer to the next module, depending upon the presence of the corresponding domain. Most commonly, reductive domains like ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) act in series to stereoselectively reduce the beta ketone (installed by the previous module) to an alcohol, olefin, or saturated carbon (**Figure 22B**). Other domains may also be present such as methyltransferases, cyclases, and halogenases.

By assessing the domains that are present within a module, it is often possible to accurately predict the chemical outcome of the extension of that module. Consolidating the predictions of every module in a pathway allows for the prediction of the product of an entire biosynthetic cluster, thereby alluding to the possibility of rationally designing a series of modules to assemble arbitrary polyketide structures. However, the attempts to construct unnatural pathways have rarely achieved the levels of activity of their naturally-occurring counterparts. Even modest rearrangements of biosynthetic components often result in large losses in productivity(141-143). Somehow, natural selection has resulted in the recombination and rearrangement of PKS systems in such a way as to facilitate the efficient biosynthesis of diverse polyketide natural products, but rational design still falls short. This suggests that PKS modules are not truly “modular” and a deeper understanding of their function is needed.

PKS have been studied through the lens of structural biology with many advances in understanding, but integrating these incremental advances into large-scale models has proven difficult. Precious few full-module structures exist, though many structures of individual and pairs of domains have been solved. These structures have rationalized the substrate preferences, stereochemical outcomes, and protein-protein interactions of several domains, greatly improving the predictive power of PKS primary sequence. Given the relatively low affinities that different

PKS domains have for one another without covalent links, it appears likely that many protein-protein interactions within PKS modules are directed by the affinity of the intermediate for a particular active site. These atomic-resolution details have been effectively described in many cases by single-domain structures, greatly contributing to the understanding of the detailed biochemistry of individual steps of PKS functionality(144). Additionally, insights provided by these smaller structures can be usefully combined into models of full modules, but there are many pitfalls to this in practice that can only be avoided with additional structural information. Further structural study of different domain combinations and modules with different attendant domains is necessary to understand how PKS pathways can maintain their functionality and efficiency with such diverse constituent modular architecture.

The variable structure of PKS is thought to have its origins with a common ancestor shared with fatty acid synthase. Mammalian fatty acid synthase (mFAS) is an essential component of primary metabolism whose central function is the synthesis of palmitic acid from malonyl- and acetyl-CoA building blocks. Natively a homodimeric multidomain multifunctional complex, mFAS uses many enzymatic components that have homologues in PKS, and appears to share some fundamental structural principles. Each mFAS monomer is made up of seven domains: KS, malonyl/acetyltransferase (MAT), KR, DH, ER, thioesterase (TE), and ACP. Based on structural data acquired using cryo-electron microscopy (cryo-EM) and crystallography, mFAS monomers appear to form head-to-head dimers, with two distinct structural units connected by a relatively flexible linker(24, 145). These units are made up of the KS and MAT of both chains in one unit and the remaining domains of both chains in the other. Along the central axis of the dimer are KS, DH, and ER, with other domains not forming direct

dimers. This arrangement forms two separate reaction chambers within which each mobile ACP can access all of the necessary active sites.

Because of their apparent evolutionary relationship, many useful comparisons can be made between mFAS and PKS modules (**Figure 23**). Several crystal structures of KS-AT didomains excised from PKS modules have been solved(19, 21, 26), and the configuration appears to closely match the KS-MAT structural unit of mFAS, with the KS dimer forming the core of the structure and the ATs extended outward perpendicular to the axis of symmetry on opposite sides. Additional single domain structures of KR, DH, and ER also fit into the mFAS-like structural model. This has led to mFAS being thought of as a fully-reducing PKS module of sorts, with their key differences in function being the added inter-module handoff in PKS and the iteration of mFAS.

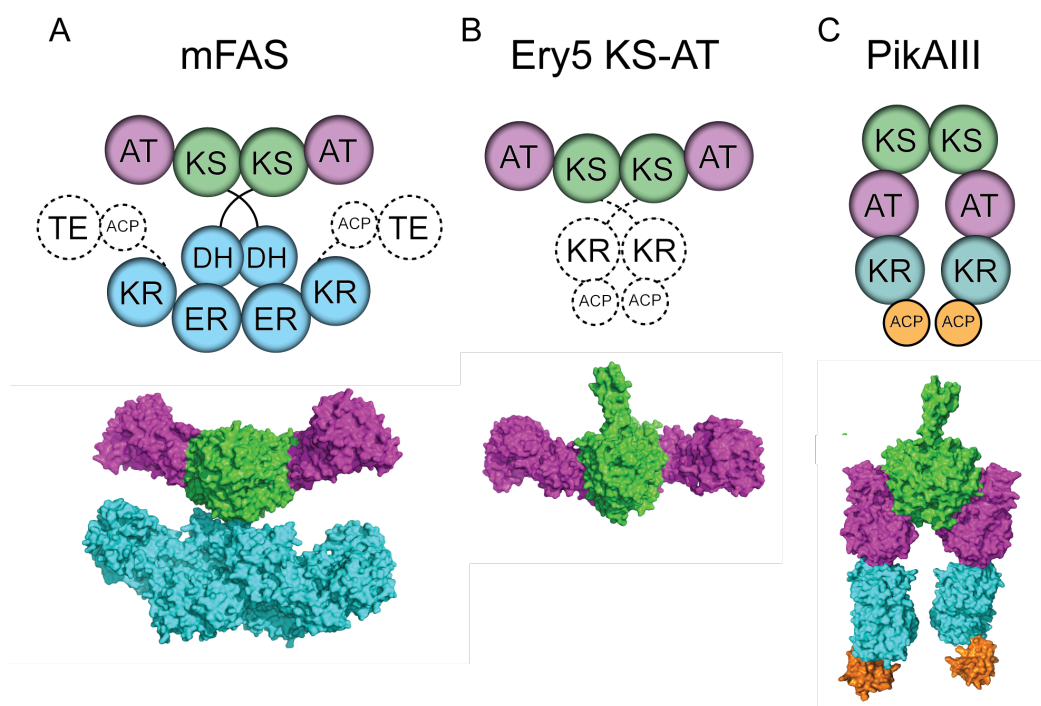


Figure 23. Comparisons of the architecture of biosynthetic systems. Crystal structure and domain architecture of mFAS (A), Ery5 KS-AT from the DEBS PKS (B), and a cryo-electron microscopy reconstruction of PikAIII, a monomodular PKS protein from the pikromycin PKS.

The architectural hypothesis based on structural identity between PKS and mFAS has been complicated by the structural reconstructions obtained from cryo-electron microscopy of a module from the pikromycin pathway(22, 99). Pikromycin is a macrolide antibiotic, and the macrolactone core is produced by four PKS genes: PikAI – PikAIV(146). Using PikAIII, this work produced reconstructions of a KS-AT-KR-ACP module during various stages of catalysis by including or withholding different substrates and cofactors. PikAIII is a PKS gene that contains only one module, so no unnatural gene excisions or truncations were required during subcloning. In all of the catalytic states, the overall appearance of the module was quite disparate from the KS-AT structures and mFAS. The reconstruction quality allowed for the fitting of previously-solved mono- and di-domain crystal structures into the EM envelope, creating an approximation of an atomic-resolution structure. In each of the catalytic states, the KS dimer sits atop an arch-like structure made up by the AT domains and KR domains, with the ACPs in a variety of positions, depending on the added substrates. The monomers are still roughly symmetrical, but instead of forming two reaction chambers distal to the axis of symmetry, a single chamber centered upon the axis of symmetry is formed.

Within this chamber, several large domain movements occur during the catalytic cycle. Most significantly, the ACPs appear to assume several different preferred positions depending on the modification of the phosphopantetheine arm. In the absence of any substrates, the ACPs appear to be situated near the AT domains. Extender unit loading of the phosphopantetheine arm is a prerequisite for chain extension, so this confirmation stands to reason. When methylmalonyl-CoA is present, the AT is able to charge the phosphopantetheine and the ACPs shift to the face of the KS dimer that is accessible from the reaction chamber. When a thiophenol-conjugated pentaketide (the native substrate for PikAIII) is added with the extender

unit, the ACPs appear to move to interact with the KR domains. In this catalytic state, the KR domains appear to flip end-to-end relative to their positions in all other catalytic states. Finally, when NADPH is added with the other substrates, KR-catalyzed reduction occurs and the ACPs migrate dramatically along the axis of symmetry to the ends of the “archway” distal to the KR. This position is hypothesized to facilitate handoff of the processed intermediate to the KS of the next module.

These reconstructions and previously solved crystal structures raise many questions about domain mobility, ACP positioning, and protein-protein interactions within each module. The available mFAS structures reveal fairly structured linker regions that contrast with the nonconserved sequences and secondary structures found among the numerous types and subtypes of PKS modules. Perhaps a symptom of shorter and more dynamic evolutionary variation, these linkers may simply have had less time to arrive at a primary sequence that allows for the necessary domain movements while minimizing extraneous flexibility. Moreover, the domain attendance in mFAS has likely been consistent for much longer than that of any given PKS module. On the other side of the same coin, this ability to accommodate different domains and retain function is what makes PKS modules such interesting subjects. Dynamic motion and domain rearrangements are necessary for the function of mFAS as much as they are for PKS modules.

Obtaining structural information about these proteins is difficult. The variable domains, linker structures, and flexibility that are so vital to their diverse functions adds difficulty to the study of PKS modules. Ion mobility mass spectrometry (IM-MS) can provide a complementary approach to crystallography and cryo-EM. This technique utilizes so-called native mass spectrometry to analyze high molecular weight, intact, and heterogeneous protein

complexes(147). Acquisition of IM-MS data uses traditional mass detection coupled with gas-phase collisional cross section data, thus connecting target identity with three-dimensional structure(148). To obtain these data, protein complexes with different collisional cross sections are rapidly separated in a neutral gas collision cell (**Figure 24**). Additionally, small differences between domain configurations can be revealed through collision-induced unfolding (CIU). In CIU experiments, variable collision energies are imparted to the target complex, and changes in the detected drift time generate data reminiscent of a thermal melting curve, which can be analyzed in similar ways(149). Applying these methods to PKS proteins provides a novel and interesting route to understanding PKS architecture.

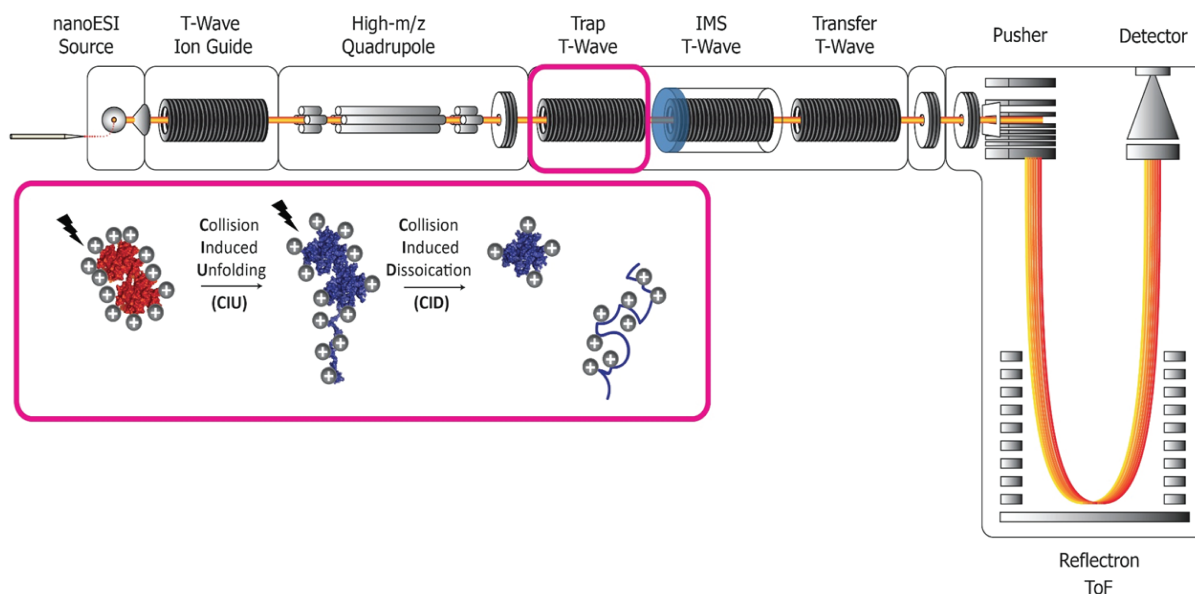


Figure 24. Ion mobility mass spectrometry instrumentation. Ions are generated by nanoelectrospray ionization, and then directed by a T-wave ion guide. The ion trap T-wave ion guide accumulates and pulses ions into the traveling wave ion mobility cell and initiates collision induced dissociation (CID). Manipulation of acceleration voltage in the ion trap can be tuned to cause protein unfolding (CIU) and dissociation (CID). The ions are then separated by ion mobility and analyzed by mass in a time-of-flight system.

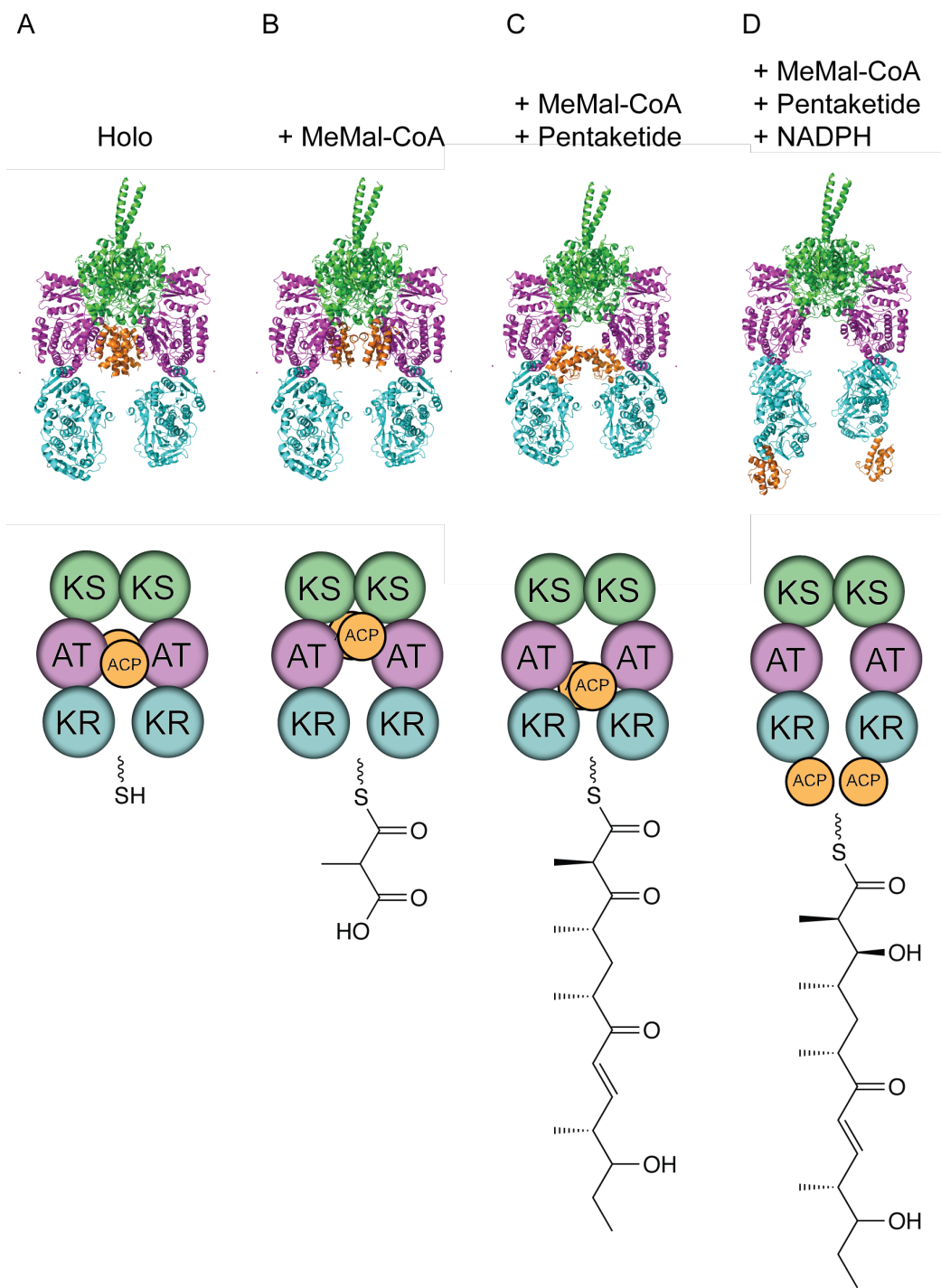


Figure 25. Pseudo-atomic models of PikAIII catalytic states. Crystal structures fit into cryo-EM reconstructions yield the models reported by Whicher, et. al. (22)4, represented here. The addition of small molecules produced four independent domain configurations: A) Holo-PikAIII, B) methylmalonyl-PikAIII, C) hexaketide-PikAIII, and D) reduced hexaketide-PikAIII.

We decided to examine the well-established model system of PikAIII and PikAIV to examine the behavior of PKS modules using this new methodology. Each of the substrates and necessary cofactors have been previously characterized, and the cryo-EM reconstructions provide a very useful set of structures for comparison(22, 99). The same conditions used to generate different catalytic states of PikAIII may be used to generate PKS modules for IM-MS analysis (**Figure 25**). Using IM-MS experiments and *de novo* modeling techniques, we attempted to analyze these PKS modules and validate the effectiveness of gas-phase structural analysis of PKS modules.

3.2 Results

3.2.1 PikAIII Catalytic Cycle IM-MS and Modelling

Purified PikAIII was prepared using the same methods that were used to prepare the catalytic states identified by cryo-EM. Namely, holo PikAIII, holo PikAIII + methylmalonyl-CoA, and holo PikAIII + methylmalonyl-CoA + thiophenol-pentaketide + NADPH; termed holo, methylmalonyl, and hexaketide, respectively. The condition without NADPH used in the cryo-EM studies was omitted due to the similarities in architecture to holo-PikAIII and low resolution of IM-MS methods. The catalytic states of the module were confirmed by phosphopantetheine ejection, and the detected masses matched with homogeneous populations of each catalytic state. Masses detected using IM-MS corresponded with both monomeric and dimeric PikAIII, each in two different conformational states with distinct mass to charge ratios and calculated collisional cross sections. The higher mass to charge ratio species had a higher collisional cross section that was close to the cross section calculated from the cryo-EM reconstruction, suggesting a natively folded protein. The lower mass to charge species had a much lower collisional cross section,

suggesting a major conformational change. We therefore termed these states “folded” and “unfolded”.

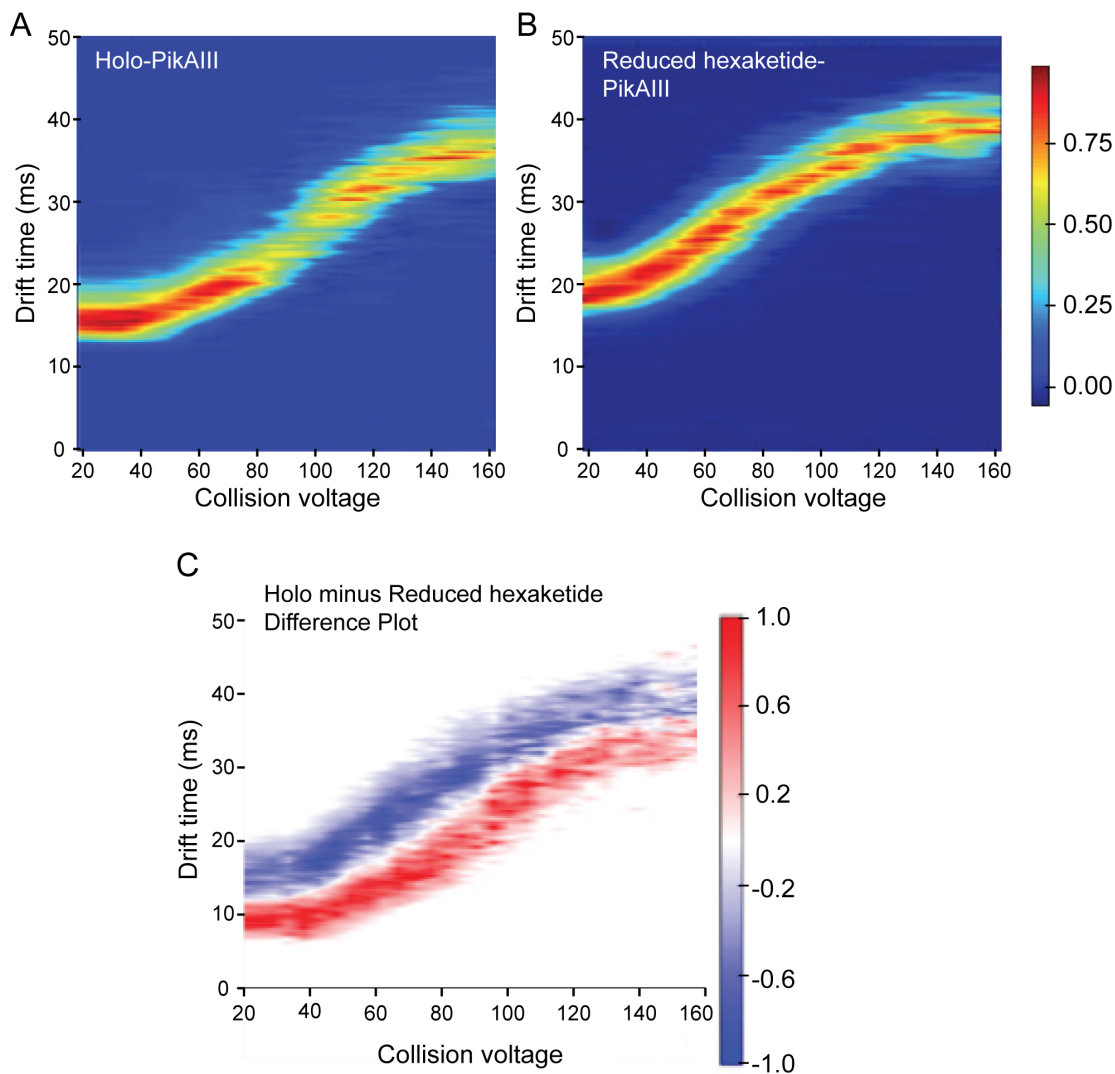


Figure 26. Collision-induced unfolding (CIU) fingerprints of PikAIII catalytic states. Increasing voltage correlates to increased collision energy, inducing a conformation-dependent unfolding. Holo-PikAIII (A) generates a unique CIU “fingerprint” from reduced hexaketide-PikAIII (B). This difference can be quantitated by subtracting hexaketide-PikAIII from holo-PikAIII and plotting the resulting values (C). Signals that increased in abundance are shown in shades of red, while signals that decreased are shown in shades of blue.

Collisional cross sections (CCS) were calculated for each oligomeric state, catalytic state, and for each conformational state within each conditions. These cross sections show several interesting trends. In every case, any dimer species had higher CCS than any monomeric species. Within each oligomeric state, the unfolded species had higher CCS than the folded species, as described above. Additionally, the hexaketide-PikAIII species consistently had a higher CCS than the holo-PikAIII species for every detected conformation and oligomeric state. Interestingly, the largest difference in CCS between holo- and hexaketide-PikAIII was found in the dimeric, folded species. The difference in CCS was much smaller between the holo- and hexaketide-PikAIII for unfolded dimers and for both folded and unfolded monomers.

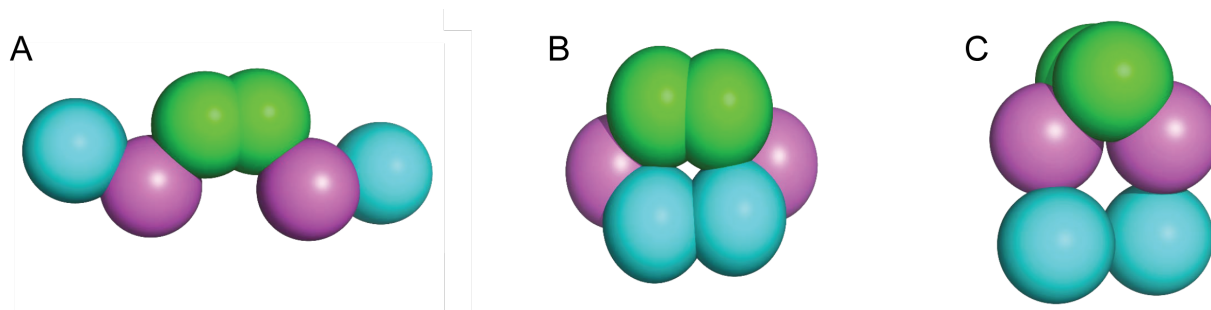


Figure 27. *De novo* models generated from IM-MS data. Model of the “unfolded” conformation (A) does not closely resemble other structures. The “folded” conformation (B) has more agreement with previous structures, but still has significant differences from structures based on cryo-electron microscopy experiments (C).

Collisional-induced unfolding experiments were conducted to compare the two states with the largest CCS differences: holo- and hexaketide-PikAIII (**Figure 26A, B**). The pseudo melting curve generated from these experiments can be most easily analyzed using the same method one would use for a thermal melting curve and identifying the inflection point of each curve. The inflection point of holo-PikAIII curve is higher than that of hexaketide-PikAIII,

indicating that much more collisional energy is required to unfold holo-PikAIII than hexaketide-PikAIII. This difference is made clear in the difference plots generated by subtracting one dataset from the other and highlighting changes in red (increased abundance) and blue (decreased abundance) (**Figure 26C**). The calculated collisional voltage RMSD of 40.35V between the two datasets also validates the significant difference in CIU behavior of the two species.

Low-resolution gas phase structures were modeled from these CCS data, CIU measurements, and other detected mass-to-charge abundances (**Figure 27**). These models were then compared to the low-resolution adapted models generated from cryo-EM reconstructions. Models of unfolded protein did not closely resemble the cryo-EM model, and the calculated CCS values were not in agreement. Models of folded PikAIII were somewhat closer to the cryo-EM model, although the positioning of AT and KR domains was quite different.

3.2.2 PikAIII and PikAIV KS-AT Didomain IM-MS Analysis

PikAIII and PikAIV KS-AT didomain proteins were purified to homogeneity using affinity chromatography. Due to protein behavior and for data consistency, PikAIII KS-AT was fused with a dimerization element that is found in full-length PikAIII between the KR and ACP domains. This element has been found to aid the dimerization of the full PKS module and also greatly improves the expression and purification behavior of the PikAIII KS-AT didomain (**Figure S7**). Masses were selected by quadrupole filter prior to CIU experiments. Analysis of the CIU data revealed distinct populations of monomers and dimers that undergo different unfolding paths as collisional energy is increased. Within both monomer and dimer populations, two subgroups can be identified that we have termed “extended” and “compact” based on their

relative mass to charge ratios. A third subgroup, “intermediate,” within the monomeric population can also be identified.

After exchange into the MS-compatible ammonium acetate buffer (pH 7.4), the protein was analyzed by IM-MS and size exclusion chromatography to assess baseline unfolding and/or aggregation (**Figure 29**). The size exclusion data suggest a transition from >90% native dimer to ~60% high molecular weight soluble aggregate over the course of four hours. Consistent with the expected results of using the dimerization element, no monomer was detected in solution. Over the same time period, a drop in IM-MS drift time was also detected indicative of a transition from an extended species to compact monomers and dimers. Additionally, the ratio of monomers to dimers shifted dramatically toward monomeric protein, with both compact and extended dimers almost disappearing. Some intermediate monomers were left at the longer timepoints, but the vast majority of the protein was detected as compact monomeric species (**Figure 28**).

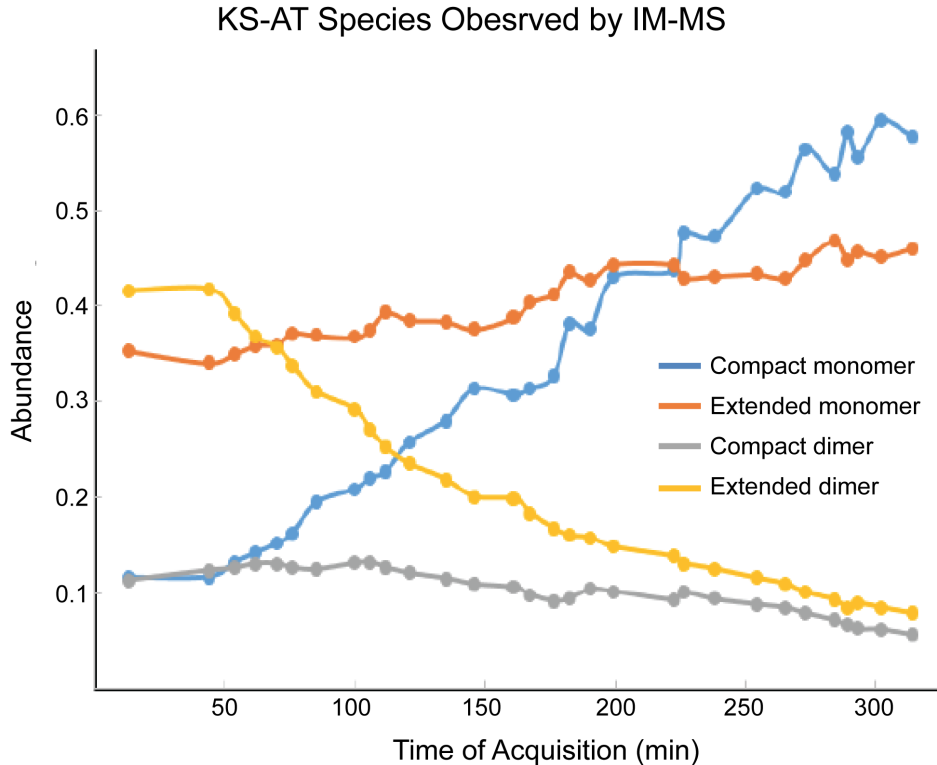


Figure 28. Observed KS-AT species over time. Detected dimeric species (gray and yellow) decrease as incubation times prior to analysis are increased. Monomeric species (orange and blue) both increase, but the “compact” monomeric form (blue) appears to be the eventual endpoint of this pathway.

Collision-induced unfolding experiments showed distinct unfolding fingerprints for each of the four species. These experiments were carried out at different time points to most easily analyze specific folding states of the KS-AT. The compact species of each oligomer required more energy to unfold than extended species. The difference between extended and compact KS-AT dimers was much larger than the change between extended and compact monomers, as shown in the pseudo-melting CIU curves. Paradoxically, the monomeric species appear to unfold more easily than the dimeric species, even though the abundance of dimer ions detected decreases over time after buffer exchange into ammonium acetate.

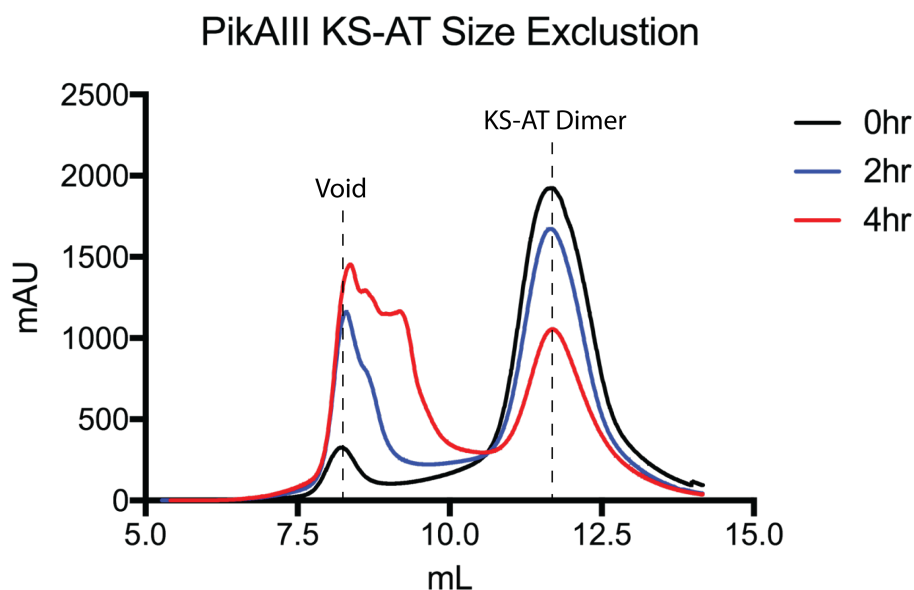


Figure 29. Size exclusion chromatography of KS-AT over time. After buffer exchange, longer incubation times prior to size exclusion analysis resulted in a decrease in KS-AT dimer abundance and an increase in material eluting in the void volume.

3.3 Discussion

3.3.1 Examination of the PikAIII Catalytic Cycle by IM-MS

Thus far, the IM-MS data and modeling for PikAIII reveal interesting biophysical processes, some of which are relevant to PKS biology. The detection of folded dimers of holo-PikAIII is a result that validates the gas-phase collision methodology and supports the use of this technique. The change in CCS between holo- and hexaketide-PikAIII matches the structural rearrangements observed in cryo-EM experiments, and the followup CIU experiments also stand to reason, given that a larger CCS will result in more collisions and lower CIU “melting point” within the IM-MS instrument. The precise mechanism of ACP migration out of the central reaction chamber warrants further study, but an explanation can be hypothesized from the substrate-directed ACP positioning concept described in the introduction of this chapter. In the

case of hexaketide-PikAIII, the fully processed substrate that is ready to be passed to the next module is not accepted by the active sites of any domain within PikAIII and the affinity of the ACP itself for any domain is relatively low, leaving the ACP with nowhere to go except out of the reaction chamber. Intuitively, this elongated, larger CCS structure would be less stable in the context of an IM-MS collision cell, as this structure will receive more collisional energy than a lower CCS structure, thus resulting in our observed CIU data.

The spherical model reconstructions of PikAIII allude to the potential of IM-MS as a structural technique. The folded, dimeric species has distinct deviations from the cryo-EM reconstructions, but still shows promise. In the cryo-EM reconstruction, the AT and KR domains are related through a 2-fold axis of symmetry defined by the central axis of the KS dimer. Whereas in the IM-MS model, the KS and KR domains are coplanar, while the AT domains are out of this plane in the same direction, generating a structure reminiscent of a boat conformation. It is unlikely that the post-KS linker region would be able to accommodate this arrangement, but this result is encouraging nonetheless. The addition of restraints from other biophysical techniques or structures will improve subsequent models. Most importantly, this result was obtained from IM-MS data alone. Using only the knowledge that PikAIII is a dimer made up of two tetradomain peptides as guidance, a novel model of the three-dimensional structure of a PKS module was generated.

3.3.2 KS-AT Didomains as a Simplified Model System

The analysis of KS-AT didomain constructs was undertaken as another experiment to compare IM-MS results to x-ray crystallography results. Specifically, we sought to examine the position of the AT domains. The contradiction in AT position between the first KS-AT crystal structures and the cryo-EM reconstructions is a large physical difference and an important

conceptual question, as it applies to the vast majority of PKS modules. Unfortunately, the CCS of the “AT out” (crystal structure) conformation is very close to that of “AT down” (cryo-EM), precluding the ability of IM-MS to tell them apart. However, in the interest of future development of this technique, several methodological questions were pursued such as buffer conditions, instrument and ionization settings, and complementary analytical techniques to validate the IM-MS observations.

The detection of monomeric and partially unfolded species is likely irrelevant to the biological function of PikAIII, but is an interesting biophysical phenomenon nonetheless. In solution, PikAIII KS-AT didomain constructs without the PikAIII dimerization element present as a mix of monomeric and dimeric species with greatly reduced expression yield and solubility. This suggests that monomeric PikAIII is unfavorable, likely unstable, and most likely irrelevant to the domain rearrangements during the catalytic cycle. Additionally, the size exclusion chromatography data imply two potential sources for these species: soluble aggregates eluting in the void volume of the column or artifacts of ionization and the IM-MS collision cell. However, it is curious that no higher-order oligomers of KS-AT were detected by IM-MS. If the unfolded and monomeric KS-ATs were indeed generated from soluble aggregates, one would expect at least some of these high molecular weight complexes to be detected. Thus, it appears most likely that these aggregates are filtered out of the sample prior to analysis, and the breakdown of dimers and unfolding of protein is caused by thermal instability of KS-AT dimers over time combined with desolvation, ionization, and gas collision during IM-MS analysis.

The detected species can be rationalized as discrete steps in a classic protein misfolding pathway over time. The initial population is natively-folded dimers, and the ultimate endpoint is highly compact monomers. The dissociation of dimeric species into folded monomers happens

simultaneously with the much slower conversion from folded to misfolded dimers. Monomeric species rapidly misfold in a two-step process, adopting conformations that result in a lowered CCS and mass-to-charge ratio, suggesting compact conformations that imply misfolding of one domain to form “intermediate” monomers or both domains to form fully compact monomers. Misfolded dimers appear to dissociate into either intermediate or compact monomers. There does not appear to be a distinct population of intermediate misfolded dimers, perhaps because their dissociation into monomers is accelerated with each additional misfolded domain.

3.4 Summary

These experiments have used a small subset of the capabilities of IM-MS applied for the first time to a PKS module. We have demonstrated that IM-MS is capable of detecting the rearrangements of the ACPs during the PikAIII catalytic cycle through both CCS calculations as well as a shift in CIU fingerprint. We have also shown the potential for de novo modeling of PKS modules from IM-MS data by generating a novel spherical approximation structure of a PikAIII dimer. The detection of misfolded protein was not greatly beneficial to the analysis of PKS biology, but does demonstrate the ability of IM-MS to be used as an analytical tool for oligomer dissociation and protein unfolding over time.

3.5 Conclusion

We have demonstrated with these data a proof of concept series of experiments that show that the application of IM-MS techniques to the structural biology of PKS modules carries a great deal of promise. With further development of the modeling capabilities and complementary experiments, de novo structures of arbitrary PKS modules may be realized. This would allow rapid access to architectural information without time-consuming and costly high-resolution structural techniques. There are many different types of PKS modules that have been

described but very few experiments that test our current models of the variability of their structures. These techniques may also be used to detect large domain rearrangement during catalysis and has the added advantage of being able to detect multiple conformations simultaneously. Additionally, the native gas-phase analysis used by IM-MS can keep relatively weak protein-protein associations intact, which may open the door to direct observation of the many transient interactions used by PKS such as trans-AT, beta branching, and intermodule docking domains. As a complement to other structural techniques such as cryo-electron microscopy and small-angle scattering, IM-MS has the ability to be a rapid screening technique as well as an analytical workhorse. IM-MS will make a powerful complementary technique to the structural study of PKS modules and their interactions.

3.6 Materials and Methods

Protein Expression and Purification. PikAIII, PikAIV, PikAIII KS-AT, PikAIV KS-AT, PikAIII Δ ACP, and DEBS3 plasmids have been previously described. Proteins were expressed in *E. coli* BL21 (DE3) when phosphopantetheinylation was not required and BAP1 for holo protein(135). Transformed cells were grown with shaking at 37°C in 0.5L of Terrific Broth in 2L baffled Fernbach flasks until OD₆₀₀ ~ 0.8. Cultures were cooled to 18°C and isopropyl b-D-thiogalactopyranoside (GoldBio) was added to 0.2mM to induce expression. Growth was continued for 12-14 hours before harvesting by centrifugation and flash freezing. To purify protein, cell pellets were thawed at 4°C and resuspended with 7mL/gram buffer A [300mM NaCl, 10% glycerol, 50mM HEPES pH 7.5]; 0.8mg/mL lysozyme (Sigma), 5mM MgCl₂, and 50U/mL benzonase (Millipore) were added. Cells were lysed by sonication and clarified by centrifugation. Lysates were then filtered using 0.45 μ m membranes and imidazole (pH 7.5) was

added to 50mM. Affinity purification was achieved using a 5mL HiPrep FF NiNTA (GE Healthcare) using a stepwise elution of buffer A and buffer B (buffer A supplemented with 300mM imidazole). After loading the column with lysate, the column was washed with 150mL 15% buffer B before elution with 80% buffer B. The highest concentration fractions were collected and either prepared for further handling or used directly in mass spectrometry.

Size exclusion chromatography was carried out either in buffer A or a variety of concentrations of ammonium acetate (100-750mM). Elution fractions from affinity purification were filtered with a 0.2 μ m membrane and 100-500 μ L were injected onto either a Superose 6 Increase 10/300 GL or Superdex S200 10/300 GL column (GE Healthcare). Protein was collected using an isocratic elution method. Reactions of PikAIII were carried out using a previously described method(22). Catalytic turnover was assessed using the phosphopantetheine ejection assay. Due to instrumental incompatibilities, only volatile salts could be used with the described mass spectrometer. Protein that was eluted in any buffer other than ammonium acetate was exchanged into ammonium acetate using [40K cutoff desalting resin in spin column format]

Ion Mobility Mass Spectrometry Experiments. Data were acquired on a G2 ion mobility mass spectrometry platform (Waters) in tandem-MS (quadrupole selection) mode. Ions were selected by the quadrupole mass filter for m/z from 4000-9000. For full module and KS-AT disassembly pathways, spectra were collected over 8-12 hours and the abundance and drift time distribution of 23⁺ (monomer) and 33⁺ (dimer) ions were tracked. Collision-induced unfolding (CIU) experiments were carried out using a voltage bias prior to the ion mobility cell to modulate collision energies. Collision voltage was increased by 5 V increments from 10 V to 180 V and

the drift time distributions for 21^+ (monomer) and 29^+ (dimer) ions were tracked to generate CIU fingerprints. Difference plots and RMSD calculations were carried out using CIUSuite(*149*).

Collisional cross section (CCS) measurements were calibrated using standards with known CCS. Several abundant and reproducible ions observed in experiments using PikAIII were tracked for comparison between different catalytic states of PikAIII, namely 42^+ , 43^+ , 44^+ , and 45^+ . Projection approximation CCS values were calculated from the pseudo-atomic resolution structures of PikAIII catalytic states generated by fitting the crystal structures of individual domains as rigid bodies into a lower-resolution envelope generated by experimental cryo-electron microscopy data. Gas-phase structures of PikAIII were generated using a coarse-grained model with experimental CCS data using previously described methods(*150*).

3.7 Supplemental Data

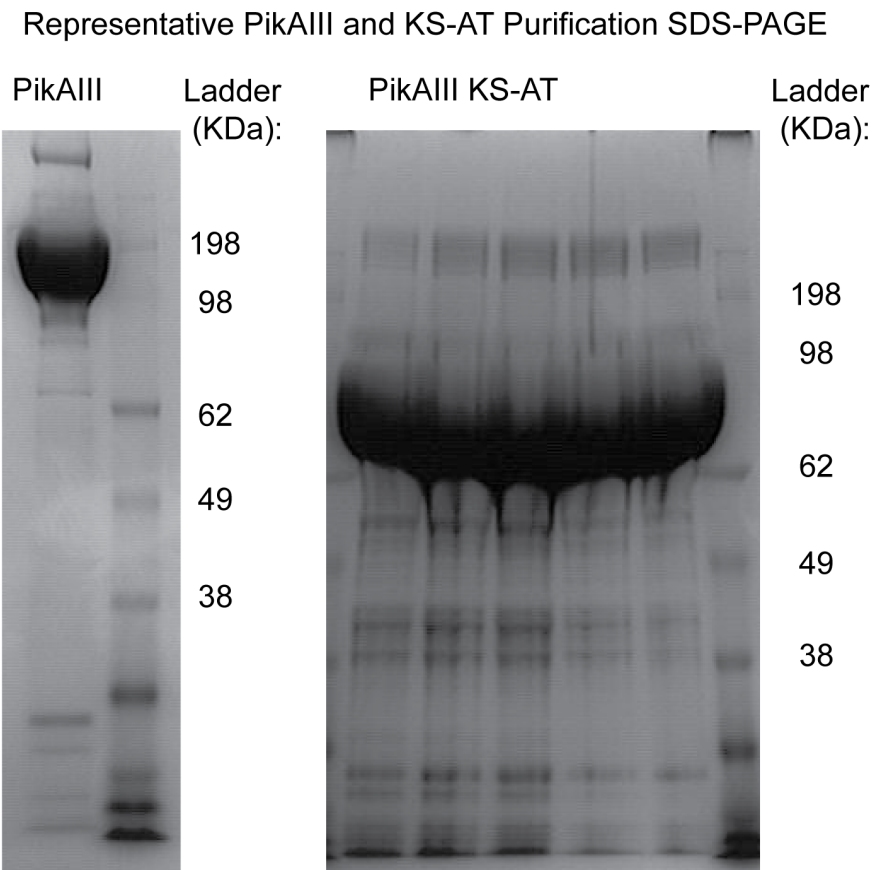


Figure S7. SDS-PAGE of PikAIII and KS-AT with dimerization element. With the added dimerization element from PikAIII, the KS-AT didomain construct purifies cleanly as a dimer with using NiNTA affinity methods.

KS-AT Size Exclusion SDS-PAGE
4 Hour Incubation

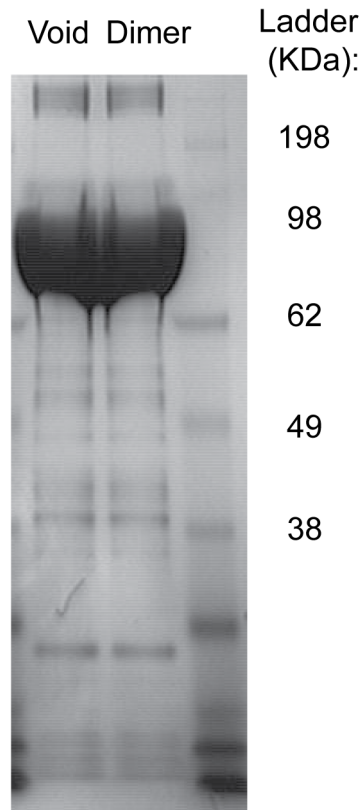


Figure S8. SDS-PAGE of elution fraction of KS-AT size exclusion chromatography. No obvious proteolysis or degradation of the KS-AT didomain is evident in the gel, leaving aggregation as one of the only remaining possibilities.

Chapter IV

Intermodular Interactions of Polyketide Synthases

4.1 Introduction

The protein-protein interactions of polyketide synthases are critical to their ability to produce their products. There are many transiently-binding domains that complement the functionality of PKS modules such as trans-AT domains, beta branching enzymes, and a variety of oxygenases and reductases. Another key protein-protein interaction that is absolutely essential to PKS proteins is their ability to bind to the correct upstream and downstream PKS polypeptides and pass intermediates in the correct direction. In the most well-studied cases, these interactions are mediated by terminal residues that are structurally independent from their adjacent catalytic domain and mediate the interaction of PKS polypeptides. These sequences have been termed “docking domains”.

Docking domains are typically short (<60) residue sequences found at the N- and C-termini of PKS proteins that provide specific affinities for the upstream and downstream neighbors in the functional order of the pathway. This role makes docking domains a very powerful and appealing target for engineering, which was recognized very early in the DEBS pathway(21, 139), and has since been leveraged to produce modified polyketides(151, 152). Biochemical, structural, and phylogenetic analyses of these docking domains have been carried out in order to more deeply understand and modify docking domain structure and function. These studies have led to several broad structural classes: 1a, 1b, 2, and 3(12, 19, 153).

Class 1a and 1b docking domains incorporate a dimerization element into their C-terminal docks and adopt a predominantly helical secondary structure which, when docked, forms a parallel four-helix bundle. In this bundle, the dimeric C-terminal docks form a central coiled coil(153-155). Class 2 docks adopt a different helical bundle structure made up of eight helices and suggests a distinct positioning for the upstream ACP(19). Despite their differences in sequence and structure, these docking domains all depend upon a similar size interface to impart high specificity with relatively weak affinity (5-50 μ M). Selectivity of the docking interaction was found to be largely dependent upon a handful of charged residues, and switching docking domains from one class to another within the same biochemical system had significant functional outcomes(19).

Expanding upon these specific examples, computational analysis of many pathways from diverse source organisms found that most docking domains can be sorted into three groups that correlate loosely with the structural classes described above(17). Interestingly, two of the three groups are frequently found not only in the same phyla, but within peptides of the same PKS pathways, further indicative of the frequent horizontal transfer and recombination events discussed in previous sections. The third group is primarily found within myxo- and cyanobacterial PKS pathways. These experiments and analysis have greatly improved our understanding of PKS interactions, but many pathways do not fit into these models.

Trans-AT pathways are an example of an expanding class of PKS systems that do not seem to fit well into the previous understanding of docking domain function. The peptide divisions of trans-AT PKS often do not match the divisions between modules, resulting in modules being split across different peptides. One outcome of these unusual divisions is N- and C-terminal catalytic domains that are rarely found at the termini of cis-AT PKS peptides. In

some cases, domains themselves are split across two peptides, which appears to be a novel docking strategy in and of itself(18). In other cases, docking domain sequences can be identified at the termini of the catalytic domain, regardless of its identity. These docking domains do not appear to be related by sequence to the previously characterized classes of docking domains, but do appear to also form a helical bundle. Several recent studies have shed light onto the structure and function of these docking domains(18, 19, 156).

Recently, the docking sequence involved in one of the few ACP-KS interpeptide interactions in a trans-AT pathway was identified in a SAXS model of a virginiamycin PKS module(27). The docking domains were examined for docking domain behavior, and a novel binding mode was described. This docking domain interaction involved intrinsically disordered terminal sequences that adopted ordered conformations when docked with their partner, the N-terminal of which vaguely resembles (~12% identity) class 2 docking domains(12). This docked complex also forms a four helix bundle, but with a structure that is distinct from the other docking domain classes. Also, the C-terminal dock was observed to engage in a degree of self-association. Additionally, this study identified several other trans-AT pathways that contain similar docking domain pairs, including the bryostatin pathway.

The work to date on trans-AT PKS docking domains demonstrates that some trans-AT pathways function through analogous mechanisms to the cis-AT docking domains that have been previously described. Further, it appears likely that at least two of these trans-AT-type interactions enable bryostatin biosynthesis. However, many details about PKS interactions remain unclear, and the bryostatin pathway is especially mysterious. Using the PikAIII/PikAIV system adapted from a study of docking domains in the curacin pathway, we set out to study these interactions in more detail and to examine the interactions between the multimodular PKS

proteins in the bryostatin pathway. In this enzymatic system, the native pentaketide substrate (“Pik pentaketide”) is conjugated to a thiophenol activating group that allows the diffusive labeling of the KS of PikAIII. PikAIII then carries out extension and reduction, generating the native hexaketide substrate of PikAIV. With correctly configured docking domains, PikAIII ACP will interact with PikAIV KS and transfer the hexaketide from the phosphopantetheine arm to the active site cysteine (**Figure 30**). PikAIV will carry out extension and the TE will then cyclize and offload the heptaketide as the 14-membered macrolactone narbonolide(157). This system may also produce a 12-membered macrolactone, 10-dml(32). While the structural mechanism is still under investigation, it has been shown that PikAIV TE may cyclize the hexaketide directly from PikAIII ACP, skipping the final two-carbon extension. This reaction has also been shown to be dependent upon functional docking domains, and so either 10-dml or narbonolide may be used as a readout for docking domain interaction.

Using this coupled PikAIII-PikAIV coupled system in conjunction with more direct measurements of the interactions of small docking domain peptides, we investigated the critical protein-protein interactions between bryostatin PKS proteins. These methods are complementary, in that the production of 10-dml and/or narbonolide does not give a complete picture of docking domain function. Observed K_D measurements between fluorescently labeled peptides and specific bryostatin proteins provide an additional source of information outside of the context of a catalytic system.

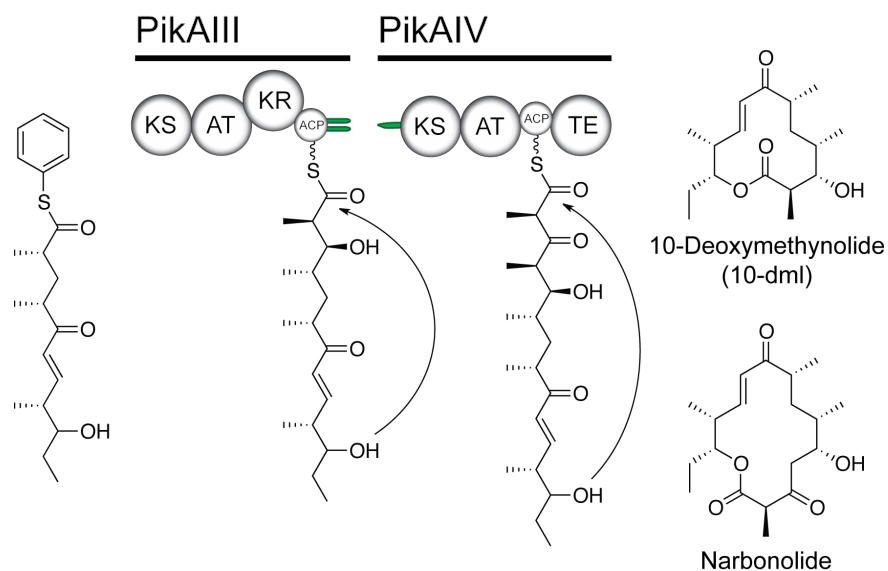


Figure 30. Coupled PikAIII/PikAIV assay system. Thiophenol pentaketide is loaded, extended, and reduced by PikAIII. With functional docking domains, PikAIII interacts with PikAIV to either immediately offload 10-dml, or to transfer the intermediate for extension by PikAIV and offloading as narbonolide.

4.2 Results

4.2.1 10-dml and Narbonolide Production

Purified and cell-free lysates of PikAIII, PikAIV and PikAIII-TE were prepared as previously described(101). 10-dml and narbonolide production was determined by HPLC using authentic standards. In the first tests, residues were chosen from each of the bryostatin PKS proteins, as indicated in **Figure 31**. These residues were used to replace the docking domains of PikAIII and PikAIV. Proteins with swapped docking domains were prepared for enzymatic reactions using the same methods. Expression and purification yields were significantly reduced for both N- and C-terminal docking domain swaps, and this effect was most significant for PikAIV with replaced N-terminal docking domains. Individual and combinatorial pairing of each PikAIII and PikAIV docking domain swap construct did not result in the production of

detectable levels of 10-dml or narbonolide except in the case of BryA-ddC and BryB-ddN. Using this pairing, ~50% of wild type 10-dml production and no narbonolide production was observed. In light of these results and the identity with other characterized trans-AT PKS docking domains, the BryA-BryB pairing was targeted for further examination.

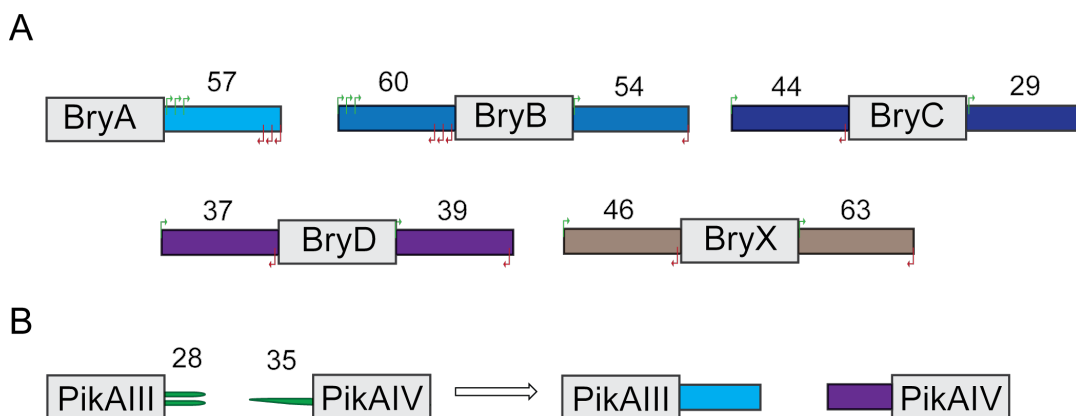


Figure 31. Potential docking domain residues at the termini of bry PKS proteins. Green and red arrows in (A) represent primers used to amplify docking domain sequences and prepare chimeric PikAIII and PikAIV constructs (B).

To examine the interaction of BryA and BryB in more detail, PikAIII and PikAIV constructs bearing truncated bryostatin docking domains were prepared. Six constructs were prepared by shortening each dock from the terminus by steps of ten residues, in order to generate ddN- and ddC- Δ 10, Δ 20, and Δ 30. Additional constructs were prepared by maintaining the native termini but removing residues from the bryostatin docking domains that would be closest to the juncture of Bry and Pik sequence. Six more constructs were prepared in this way (three PikAIII, three PikAIV).

Again, combinatorial reactions were carried out using each chimeric PikAIII and PikAIV both as cell-free lysates and as affinity-purified material. **Figure 32** shows representative data of cell-free lysate reactions. In a handful of cases, 10-dml production was detected at similar levels

to the original 60-residue BryA-BryB docking domain pair. The only modification that improved the yield of 10-dml was the pairing of BryAddC- Δ 10 and BryBddN- Δ 10, which resulted in 10-dml production within 20% of PikAIII-TE. In every other pairing, either no product or a trace amount of 10-dml was detected. No pairing was able to generate detectable levels of narbonolide.

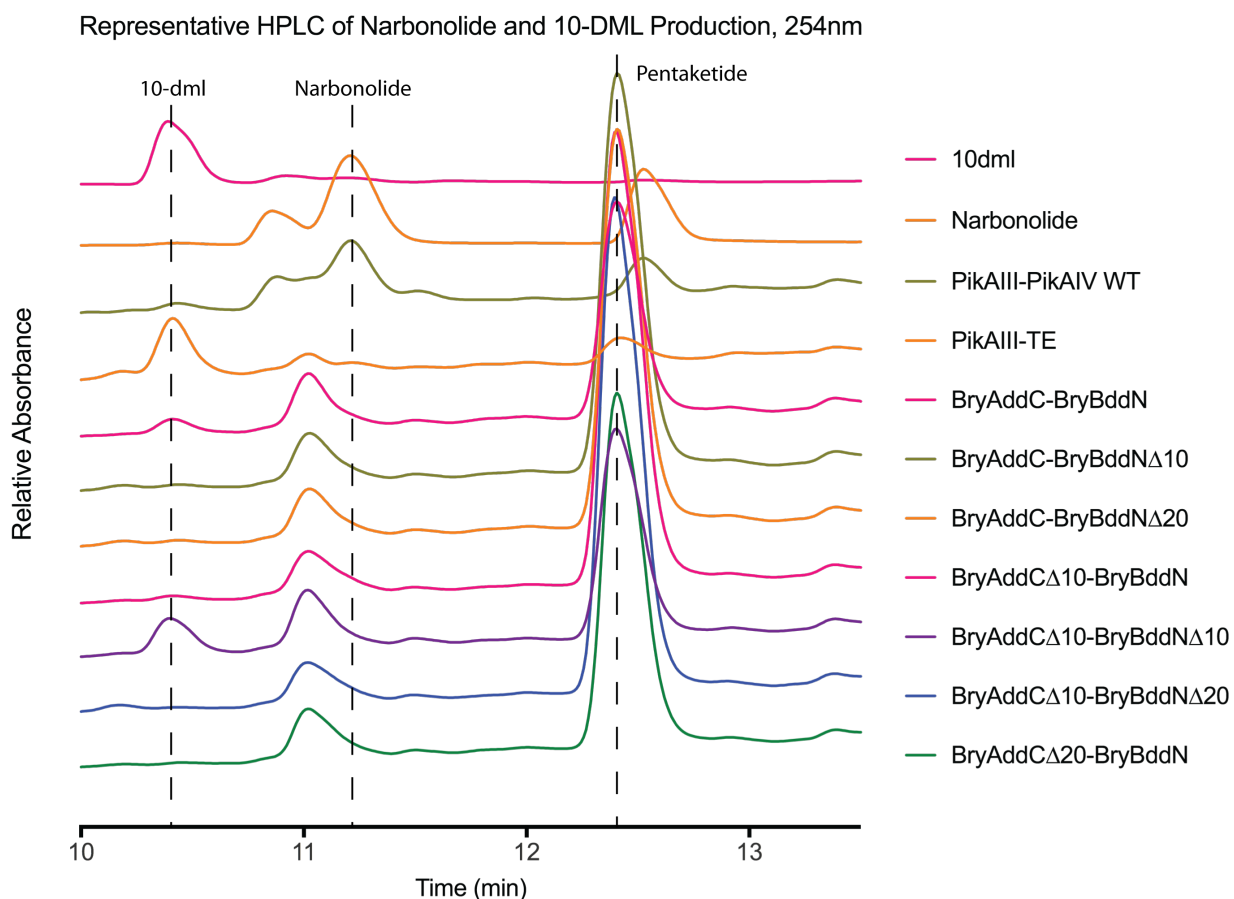


Figure 32. HPLC analysis of 10-dml/narbonolide production by PKS modules bearing heterologous docking domains. Few docking domain combinations resulted in productive interactions. The amount of residual starting material in these reactions suggests problems with protein behavior.

The protein behavior for the docking domain swap constructs appears to be greatly altered. Rough estimates of expression yield from SDS-PAGE of crude cell lysates show

between 20 and 50% lowered expression, and purification yields are similarly lowered. Additionally, the removal of the C-terminal docking sequence from expression constructs of BryA-module 3 and BryA-M3 ACP result in greatly improved protein expression and purification yields, suggesting this sequence may have deleterious effects on protein expression or stability. Phosphopantetheine ejection of docking domain swap proteins showed the same profile of phosphopantetheine adducts as wild-type PikAIII, and these constructs were all able to carry out extension and reduction to generate reduced hexaketide, further indicating that the docking heterologous docking domains did not alter catalytic efficiency, but instead altered protein stability.

4.2.2 Fluorescence Polarization

Fluorescent peptides were generated synthetically as 16- and 26-mers corresponding to the C- and N-termini of BryA and BryB, respectively. The peptide sequence was linked C-terminally with either a PEG or a beta-alanine linker to fluorescein. Standalone ACP constructs were prepared from BryA-module 3 with and without the C-terminal 60 residues of the native sequence. Purified ACPs were then used in fluorescence polarization experiments with the synthetic peptides.

Fluorescence Polarization of BryB^{NDD} Peptides with BryA Module 3 ACPs

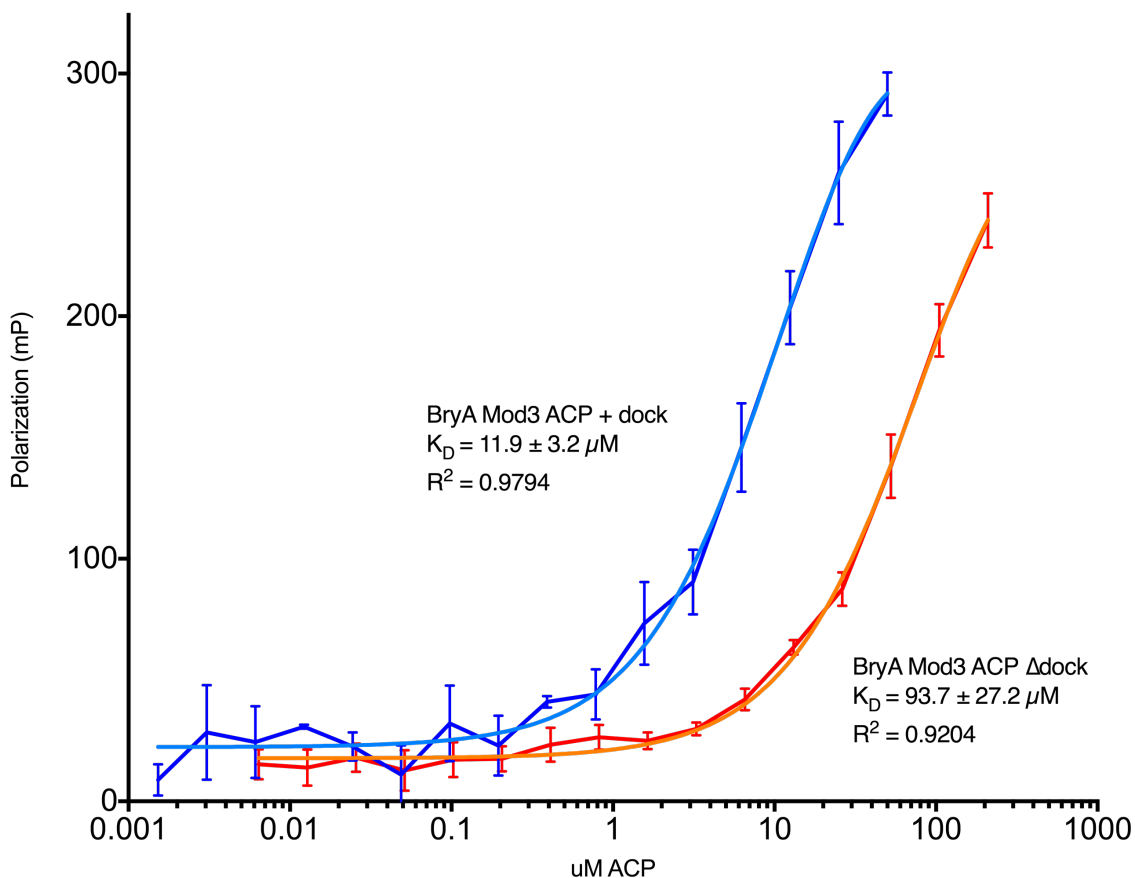


Figure 33. Fluorescence polarization of BryA/BryB docking domain peptides. BryA module 3 ACP constructs with (blue) and without (red) docking domains were titrated against fluorescein labeled synthetic docking domain peptides.

Significant changes in polarization were detected using the 16-mer BryB peptides with a PEG linker (**Figure 33**). The apparent K_D of this peptide with full-length BryM3-ACP was $11.9 \pm 3.2 \mu\text{M}$, while the K_D with truncated ACP was approximately nine-fold lower: $96.7 \pm 27.2 \mu\text{M}$. Full length ACP K_D was $25.8 \pm 6.6 \mu\text{M}$ and truncated ACP K_D was 36.2 ± 13.2 . As described in the previous section, the docking domain led to lowered protein yields and stability, resulting

in a limited range of concentrations for full-length BryA-module 3 ACP in these experiments. Small changes in polarization were detected using the 26-mer peptides, but these results were inconsistent and largely nonspecific.

4.3 Discussion

4.3.1 Pik Chimeras and 10-dml/Narbonolide Production

10-dml/narbonolide production by coupled reactions of PKS modules (PikAIII and PikAIV) on separate peptides has been used as an effective test bed for docking domain function in the past(19). However, details of the normal function of this system appear to complicate analysis of heterologous docking domain function. In native pikromycin biosynthesis, the product of PikAIII can be passed directly to the thioesterase at the C-terminus of PikAIV, skipping the last round of extension and cyclizing to form the 12-membered macrolactone 10-dml instead of the 14-membered product. Even in the native system, the mechanism of this module skipping remains somewhat mysterious, and it has been previously reported that altering docking domains can drastically shift product distribution(19).

With several curacin docking domains, a two-fold increase in 10-dml production was reported(19), suggesting the altered docking domain structure makes module skipping somehow more favorable. Likewise, the BryA-BryB docking domain pair appears to shift the product profile completely to 10-dml production. Curiously, every combination of docking domain construct prepared from the native BryA-BryB docks resulted in varying levels of 10-dml production exclusively. This may be due to the different structure of bryostatin docking domains resulting in PikAIII ACP being positioned unfavorably for chain transfer to PikAIV KS and favorably positioned for interaction with PikAIV TE. This becomes a limitation in analyzing

data from this system, as the affinity of the docking domains does not necessarily correlate with enzyme turnover.

More direct examination of the affinity between docking domains allows for the structural effects of different binding modes to be better understood. The affinity of 16-mer BryBddN for BryA module 3 ACP is within the typical range for docking domain interactions, indicating that the first 16 residues of BryB are involved in the docking interaction. However, the highest levels of 10-dml production resulted from bryostatin docking domains that are truncated by 10 residues from each terminus. This suggests that the residues that contribute to docking domain affinity include the termini of both BryA and BryB, but inclusion of these residues in the chimeric constructs does not result in interaction of PikAIII ACP with downstream domains and no chain translocation to PikAIV KS or TE occurs. The complete lack of narbonolide production may be also be a result of this structural incompatibility between PikAIII/PikAIV function and the structure of bryostatin docking domains.

4.4 Summary

The results described in this section describe some of the protein-protein interactions that facilitate bryostatin biosynthesis. While the interaction between BryA and BryB is likely analogous to that between VirA and VirFG in the virginiamycin pathway the functional interaction between other bryostatin PKS proteins must rely upon a different mechanism. The ability of chimeric PikAIII-BryAddC and PikAIV-BryBddN to interact productively indicate properties that are on par with other docking domains, but the lack of productivity from any other chimeric protein suggests other methods must be used. The observed K_D values of bryostatin dock peptides further support that BryA/BryB interactions are analogous to VirA/VirFG. Unfortunately, these studies fall victim to the induced-fold mechanism of these

docking pairs, which results in chimeric proteins that likely self-associate and are difficult to study. Overall, we have described a small subset of the docking interactions that underlie bryostatin biosynthesis and have laid the groundwork for future understanding of trans-AT PKS docking domains.

4.5 Conclusion

While we have made progress with these experiments, the docking interactions between bryostatin PKS proteins remain largely mysterious. The interaction between BryA and BryB has been demonstrated to be of a similar nature to VirA/VirFG, MlnB/MlnC, etc. as described in previous studies of trans-AT PKS docking domains. These docks use an induced-fold mechanism that results in much lower sequence identity among docking domains than is apparent in other classes of docks. This may be partially responsible for the destabilizing effects of bryostatin docking domains which appear to be specific to this pathway. Because of these effects and their ability to confound functional assays rather than biophysical readouts, a greater focus on synthetic peptides would likely be a more fruitful approach. Further experimentation may be able to isolate the terminal regions of BryC, BryD, and perhaps even BryX that mediate their sequential interactions.

4.6 Materials and Methods

Protein Expression and Purification. PikAIII, PikAIV, and PikAIII-TE expression plasmids have been previously described(101). Bryostatin docking domain sequences were identified for each PKS protein in the pathway by sequence and structural homology. N- and C-terminal sequences of each PKS were prepared, and searches of the NCBI Conserved Domain Database(158, 159) and the RCSB Protein Data Bank were carried out. By identifying the boundaries of the terminal structural domains, excess sequence near the termini with helical

character could be identified in most cases, indicating the presence of docking domains. Docking domain protein sequences were reverse translated in silico and linear DNA was purchased (Integrated DNA Technologies). Docking domain DNA was amplified using KOD polymerase prior to restriction digestion and ligation into their target vectors using T4 ligase (New England Biolabs). Docking domain swap constructs were prepared by traditional restriction site cloning. For C-terminal docking domains, a silent NheI restriction site was introduced into PikAIII pET28b between the dimerization helices and the C-terminal docking domain. Sequences coding for bryostatin docking domains were ligated into this region using NheI and HindIII restriction sites. A stop codon was included in the bryostatin docking domain sequence. For N-terminal docking domains, a previously-constructed PikAIV plasmid carrying a silent MfeI restriction site was obtained(19), and sequences coding for bryostatin docking domains were ligated into this vector using NdeI and MfeI restriction sites. Truncations of BryA ddC and BryB ddN were prepared by PCR amplification of subsections of the respective DNA and cloning using the same methods. Plasmid sequences were confirmed by Sanger sequencing.

Proteins were expressed in *E. coli* BAP1 (DE3), which contains a genomic integration of a phosphopantetheinyltransferase *sfp*. Transformed cells were grown in 2L Luria Broth in a 2.5L baffled fernbach flask at 37°C until OD600 ~0.7. Cultures were cooled to 18°C and isopropyl β-D-thiogalactopyranoside (GoldBio) was added to 0.4mM to induce expression. Growth was continued for 12-14 hours before harvesting by centrifugation and flash freezing of cell pellets, which were stored at -80°C. Cell pellets were later thawed at 4°C and resuspended with 7mL of buffer A per gram of cells. Buffer A: 300mM NaCl, 50mM HEPES, 10% glycerol pH 7. Additionally, 0.8mg/mL lysozyme and 5mM MgCl₂ were added. Cells were lysed by sonication

and clarified by centrifugation. Cell free lysate preparation was completed by flash freezing and storing at -80°C. For purified proteins: clarified lysates were passed over a 0.45 µm filter and imidazole (pH 7.5) was added to 50mM before application to a 5mL HiPrep FF NiNTA (GE Healthcare) column. The column was washed with 100mL buffer A (with added 50mM imidazole) and proteins were eluted with 20mL buffer A with 200mM imidazole. High concentration eluted fractions were pooled and exchanged into storage buffer (100mM NaCl, 20mM HEPES 7.8, 10% glycerol, 2mM EDTA) using PD10 columns (GE Healthcare), flash frozen in liquid nitrogen and stored at -80°C.

Analysis of Enzymatic 10-dml and Narbonolide Production. Reactions were initiated by the addition of ~2µM PKS protein either as affinity purified material or cell-free lysate (PKS concentration estimated by SDS-PAGE). A pre-mixed reaction buffer was prepared that included a NADPH recycling system (ref): 400mM phosphate buffer, 20% glycerol, 2.5mM glucose-6-phosphate, 0.1mM NADP⁺, 0.5U/mL glucose-6-phosphate dehydrogenase, 20mM methylmalonyl-N-acetylcysteamine, and 1mM thiophenol-Pik pentaketide. After incubation at 23°C for 4 hours, reactions were quenched with 3 volumes cold methanol and clarified by centrifugation prior to analysis by HPLC.

HPLC analysis was carried out using reversed phase chromatography on a Luna C18 5µ C18 250 x 4.6mm column (Phenomenex) and a linear, 20 minute gradient from 5-100% B. A: H₂O with 0.1% formic acid, B: 95% acetonitrile with 0.1% formic acid. UV/Vis was monitored at 210, 254, and 280nm.

Synthesis of Bryostatin Docking Domain Peptides. CLEAR amide resin (Peptides International) was used to synthesize peptides with previously described coupling conditions(160). Peptides were purified using reversed phase chromatography on a Poroshell C18 column (Agilent) and varied lengths of linear gradient of 0-100% B. A: 0.1% ammonium acetate, B: acetonitrile. Fluorescein labeled peptide concentration was determined by serial dilution into a 10mM phosphate pH 7.4 buffer and UV/Vis absorbance at 495nm using $\epsilon = 72,000 \text{ M}^{-1}\text{cm}^{-1}$, as described by the manufacturer (Pierce). Peptide purity was assessed by analytical HPLC and masses were confirmed by mass spectrometry.

Fluorescence Polarization Measurements. Binding assays were performed in a final volume of 25uL in fully opaque, low binding, black 384 well plates (Corning) in triplicate. Wells were prepared with a single concentration (25nM) of fluorescein-labelled peptide in each well and a series of 2-fold serial dilutions of protein of interest (ACP with or without docking domains, control proteins, no protein) set up in triplicate across the columns. Running buffer (same as protein of interest storage buffer: 100mM NaCl, 25mM HEPES, 10% glycerol, pH 7.5) was added first, followed by the serial dilutions of protein of interest, followed by fluorescent peptide. Wells were mixed thoroughly by pipetting and allowed to equilibrate for 30 minutes at room temperature prior to analysis. Analysis was carried out using a PHERAstar plate reader with 485nm polarized excitation and 535nm emission intensity measured through a parallel and perpendicularly polarized filter. Data processing was carried out using Prism and a nonlinear one-site binding model accounting for peptide depletion fit to the polarization data as a function of protein concentration.

Chapter V

General Discussion and Future Directions

5.1 Introduction and Overview

In the previous sections of this thesis, the importance of polyketide natural products as drug leads has been outlined, as has the significance of understanding how these compounds are biosynthesized. A deeper understanding of the mechanisms that enable polyketide synthase pathways to produce the intricate structures of their products will allow the production of new bioactive molecules. Evolutionary time scales were required for the natural development of polyketide synthases that produce bioactive molecules, and by understanding the outcomes of that evolution we can engineer pathways in a fraction of the time. To this end, it is important to continue to study the catalytic mechanisms and biosynthetic logic of polyketide synthases, as well as their large-scale architecture and domain rearrangements. These enzymatic complexes are very interesting pieces of biochemical machinery, and have demonstrated amazing potential for modification and the production of novel polyketide scaffolds.

5.2 The Continuing Progress of Bryostatins

5.2.1 Large-Scale Production for Clinical Applications

Preclinical and clinical data about the use of bryostatin 1 have been produced at an increasing pace in recent years. It has taken some time after the stalled clinical trials for cancer treatment, but it appears highly likely that more work will be done to investigate the other biological effects of bryostatins. Promising preclinical (and occasionally clinical) data have been

generated for Alzheimer's disease, latent HIV reactivation, Chikungunya virus protection, Niemann-Pick type C, and neural recovery after ischemic stroke. All of these applications will require significant amounts of bryostatin, and sourcing of the molecule may become a limiting factor. To overcome this, the study of *Ca. E. sertula* is a promising avenue. From the draft genome, several deficiencies in amino acid biosynthesis were identified, and proper supplementation may result in viable growth conditions for the organism. If ideal growth conditions for free-living *Ca. E. sertula* can be ascertained, large-scale fermentation and isolation of any of the members of the bryostatin family will be possible. A deep biochemical understanding of the organism is critical to this goal. Understanding the bryostatin biosynthetic pathway specifically is also important, and carries other potential benefits.

5.2.2 Realizing the Potential of Beta Branching Cassettes

An in-depth understanding of bryostatin biosynthesis is not only an interesting academic pursuit, but contributes meaningful information to aid in the engineering of polyketide synthases. Progress toward the goal of novel polyketide products from modified biosynthetic pathways has had mixed results and is a developing process. Like learning a new language, each newly understood concept has exceptions and limitations, but also opens new doors to deeper and more significant conversation. Each newly understood concept from unique PKS pathways has exceptions and limitations, but also great biochemical potential when properly utilized. Beta branching cassettes in general, and the bryostatin cassette specifically, are excellent examples of this. Through the substitution of a few residues in an otherwise standard PKS acyl carrier protein, a productive interaction with HMGS may be imparted. If the nascent polyketide has a beta ketone at that point in the assembly line, the door will be unlocked to the chemical diversity imparted by beta branching cassettes. The bryostatin cassette adds to this potential by installing

a vinyl methyl ester structure that is unique among beta branching cassette outcomes (**Figure 34**). One can envision the use of beta branching cassette machinery from other characterized pathways being used to install a variety of nonnative functionalities including cyclopropyl, vinyl chloride, epoxide, and vinyl methyl ester groups at arbitrary positions during polyketide biosynthesis.

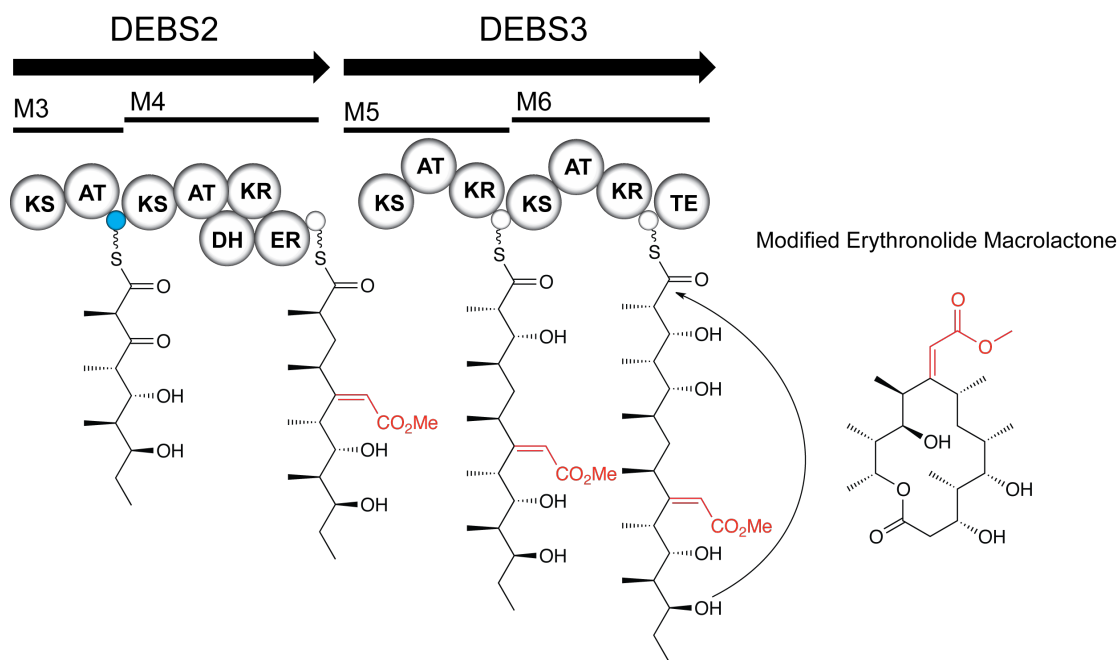


Figure 34. Hypothetical introduction of an unnatural vinyl methyl ester. The DEBS pathway natively synthesizes the polyketide backbone of erythromycin, erythronolide. With modified ACPs, beta branching machinery may be recruited to introduce novel functionality.

5.3 Future Structural Understanding of Polyketide Synthase Modules

Fundamental to the modification of polyketide synthase systems is a detailed model of their multidomain, multimodule structure. Comparisons with mFAS are useful in developing structural concepts, but the variability in organization and domain content of PKS modules causes these comparisons to break down in many cases. Modules that are split between peptides, have unusual catalytic domains, or depend on transient protein-protein interactions present more

complicated situations that do not currently have many structural relatives to generate useful comparisons. Larger-scale structural questions that do not have parallels in other systems also defy analysis by homology. It is a difficult exercise to rationalize how these PKS modules are able to operate A) as individual units, with all the domain rearrangements required for a single extension and B) as an integrated complex, with each module doing its duty in a correct and timely manner and passing that intermediate down the pathway to ultimately form a specific product. Integrating our understanding of this process from individual active sites to megadalton protein complex is the key to understanding and engineering polyketide synthases.

The techniques described in chapters 3 and 4 have great potential to be applied to a number of PKS systems. First, the study of docking domains in polyketide synthases is essential. The ability of PKS proteins to interact with one another in a specific order is essential to the function of the pathway as a whole. However, while this unidirectional interaction is essential, it is not sufficient. Numerous experiments have demonstrated that functional docking domains do not necessarily result in functional polyketide extension. These specificities are largely dictated by the active sites of the catalytic domains in the module, and are also being studied for their potential to be modified. Assuming the catalytic domains are accommodating, specificity and binding mode of docking domains has also been shown to be essential. An engineered pathway requires specific and yet low-affinity docking interactions that position the upstream ACP so it may directly interact with the downstream KS and pass the intermediate to the KS active site. Further work will reveal the structural basis for the differences in docking domain classes that result in shifted product distribution in the PikAIII/PikAIV assay system. These questions are essential to the assembly of functional and optimized docking domain interactions.

Larger-scale structural questions about domain rearrangements of individual and groups of modules are also very important to our ability to modify PKS pathways. The changes in the architecture of a PKS module through its catalytic cycle has been examined in a single case, but a multitude of types of PKS modules remain to be studied. Additionally, the handoff of intermediates between PKS modules on the same or different peptides has been examined only a handful of times. At the largest scale, the structural configuration of entire PKS pathways within their host organism remains controversial. Hypotheses have been proposed, and high-resolution structural techniques are being applied to these questions, but there is still a great deal of room for other approaches to tackle these questions. We have demonstrated in chapter 3 that gas-phase examination of PKS modules is a rapid and viable approach to probe their structure, and that meaningful information may be gleaned from modeling the data from these experiments. This technique may be developed and applied to any number of PKS module types and perhaps even multiple modules in sequence to examine their behavior during polyketide synthesis.

5.4 Conclusions

This thesis has described the elucidation of missing steps in the biosynthesis of the polyketide natural product bryostatin. Additionally, we have examined some of the key structures and interactions of polyketide synthase modules. The information obtained from these studies improves our understanding of how these bioactive natural products are produced, and shows a great deal of promise for expanded future studies. The techniques used to carry out these studies may be applied to great effect to other natural product biosynthetic systems, and ultimately may result in the production of novel natural products and novel drug leads.

References

1. B. B. Petrovska, Historical review of medicinal plants' usage. *Pharmacogn Rev.* **6**, 1–5 (2012).
2. G. M. Cragg, D. J. Newman, Biodiversity: A continuing source of novel drug leads. *Pure and Applied Chemistry.* **77**, 7–24 (2005).
3. K. Scherlach, C. Hertweck, Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753–1760 (2009).
4. N. Robbins *et al.*, Discovery of Ibomycin, a Complex Macrolactone that Exerts Antifungal Activity by Impeding Endocytic Trafficking and Membrane Function. *Cell Chemical Biology.* **23**, 1383–1394 (2016).
5. D. E. Schaufelberger *et al.*, The Large-Scale Isolation of Bryostatin 1 from *Bugula neritina* Following Current Good Manufacturing Practices. *J. Nat. Prod.* **54**, 1265–1270 (1991).
6. P. A. Wender *et al.*, Design, synthesis, and evaluation of potent bryostatin analogs that modulate PKC translocation selectivity. *Proceedings of the National Academy of Sciences.* **108**, 6721–6726 (2011).
7. D. O. Baumann *et al.*, Synthesis and Biological Evaluation of Several Bryostatin Analogues Bearing a Diacylglycerol Lactone C-Ring. *J. Org. Chem.* **81**, 7862–7883 (2016).
8. Miguel Bayod-Jasanada, R. J. C. and, Fernando López-Ortiz, Synthesis of 9-deoxy-9a-aza-9a-homoerythromycin A 11,12-Hydrogen Borate and Azithromycin 11,12-Hydrogen Borate. A New Procedure to Obtain Azithromycin Dihydrate. *J. Org. Chem.* **62**, 7479–7481 (1997).
9. C. M. Rath *et al.*, Meta-omic Characterization of the Marine Invertebrate Microbial Consortium That Produces the Chemotherapeutic Natural Product ET-743. *ACS Chem. Biol.* **6**, 1244–1256 (2011).
10. I. J. Miller, N. Vanev, S. S. Fong, G. E. Lim-Fong, J. C. Kwan, Lack of overt genome reduction in the bryostatin-producing bryozoan symbiont, "Candidatus *Endobugula sertula*". *Appl. Environ. Microbiol.* (2016), doi:10.1128/AEM.01800-16.

11. E. J. N. Helfrich, J. Piel, Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep.* **33**, 231–316 (2016).
12. J. Dorival *et al.*, Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *Journal of the American Chemical Society*, jacs.5b13372 (2016).
13. T. Nguyen *et al.*, Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
14. A. T. Keatinge-Clay, Stereocontrol within polyketide assembly lines. *Nat Prod Rep* (2016).
15. E. M. Musiol *et al.*, The AT₂ domain of KirCI loads malonyl extender units to the ACPs of the kirromycin PKS. *Chembiochem.* **14**, 1343–1352 (2013).
16. E. M. Musiol *et al.*, Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the trans-AT PKS. *Chemistry & Biology.* **18**, 438–444 (2011).
17. M. Thattai, Y. Burak, B. I. Shraiman, The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **3**, 1827–1835 (2007).
18. J. Zeng *et al.*, Portability and structure of the four-helix bundle docking domains of trans-acyltransferase modular polyketide synthases. *ACS Chem. Biol.*, acschembio.6b00345 (2016).
19. J. R. Whicher *et al.*, Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chemistry & Biology.* **20**, 1340–1351 (2013).
20. T. J. Buchholz *et al.*, Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
21. S. Y. Tsuji, D. E. Cane, C. Khosla, Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry.* **40**, 2326–2331 (2001).
22. J. R. Whicher *et al.*, Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature.* **510**, 560–564 (2014).
23. D. C. Gay *et al.*, The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. *Journal of Structural Biology* (2015), doi:10.1016/j.jsb.2015.12.011.
24. T. Maier, Architecture of Mammalian Fatty Acid Synthase at 4.5 Å Resolution. *Science.* **311**, 1258–1262 (2006).

25. I. B. Lomakin, Y. Xiong, T. A. Steitz, The Crystal Structure of Yeast Fatty Acid Synthase, a Cellular Machine with Eight Active Sites Working Together. *Cell* (2007).
26. Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11124–11129 (2006).
27. J. Davison *et al.*, Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem. Sci.* **5**, 3081–3095 (2014).
28. A. Busche *et al.*, Characterization of Molecular Interactions between ACP and Halogenase Domains in the Curacin A Polyketide Synthase. *Journal of the American Chemical Society.* **7**, 378–386 (2012).
29. M. J. Tarry, A. S. Haque, K. H. Bui, T. M. Schmeing, X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* (2017), doi:10.1016/j.str.2017.03.014.
30. C. D. Reeves *et al.*, Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations †. *Biochemistry.* **40**, 15464–15470 (2001).
31. A. N. Lowell *et al.*, Chemoenzymatic total synthesis and structural diversification of tylactone-based macrolide antibiotics through late-stage polyketide assembly, tailoring, and C-H functionalization. *Journal of the American Chemical Society* (2017), doi:10.1021/jacs.7b02875.
32. J. D. Kittendorf, B. J. Beck, T. J. Buchholz, W. Seufert, D. H. Sherman, Interrogating the molecular basis for multiple macrolactone ring formation by the pikromycin polyketide synthase. *Chemistry & Biology.* **14**, 944–954 (2007).
33. K. J. Weissman, The structural basis for docking in modular polyketide biosynthesis. *Chembiochem.* **7**, 485–494 (2006).
34. J. Yan, S. Gupta, D. H. Sherman, K. A. Reynolds, Functional dissection of a multimodular polypeptide of the pikromycin polyketide synthase into monomodules by using a matched pair of heterologous docking domains. *Chembiochem.* **10**, 1537–1543 (2009).
35. M. Jin, M. A. Fischbach, J. Clardy, A biosynthetic gene cluster for the acetyl-CoA carboxylase inhibitor andrimid. *Journal of the American Chemical Society.* **128**, 10660–10661 (2006).
36. Ö. Erol *et al.*, Biosynthesis of the myxobacterial antibiotic coralopyronin A. *Chembiochem.* **11**, 1253–1265 (2010).

37. B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, Polyketide-Chain Branching by an Enzymatic Michael Addition. *Angew. Chem. Int. Ed.* **48**, 5001–5004 (2009).
38. T. Bretschneider *et al.*, Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature*. **502**, 124–128 (2013).
39. L. Gu *et al.*, Metamorphic enzyme assembly in polyketide diversification. *Nature*. **459**, 731–735 (2009).
40. V. Simunovic *et al.*, Myxovirescin A Biosynthesis is Directed by Hybrid Polyketide Synthases/Nonribosomal Peptide Synthetase, 3-Hydroxy-3-Methylglutaryl-CoA Synthases, and trans-Acting Acyltransferases. *Chembiochem*. **7**, 1206–1220 (2006).
41. S. Sudek *et al.*, Identification of the putative bryostatin polyketide synthase gene cluster from “Candidatus Endobugula sertula,” the uncultivated microbial symbiont of the marine bryozoan Bugula neritina. *J. Nat. Prod.* **70**, 67–74 (2007).
42. C. T. Calderone, D. F. Iwig, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, Incorporation of Nonmethyl Branches by Isoprenoid-like Logic: Multiple β -Alkylation Events in the Biosynthesis of Myxovirescin A1. *Chemistry & Biology*. **14**, 835–846 (2007).
43. T. J. Buchholz *et al.*, Polyketide β -branching in bryostatin biosynthesis: identification of surrogate acetyl-ACP donors for BryR, an HMG-ACP synthase. *Chemistry & Biology*. **17**, 1092–1100 (2010).
44. A. S. Haines *et al.*, A conserved motif flags acyl carrier proteins for β -branching in polyketide synthesis. *Nat. Chem. Biol.* **9**, 685–692 (2013).
45. T. W. Geders *et al.*, Crystal Structure of the ECH2 Catalytic Domain of CurF from *Lyngbya majuscula*: INSIGHTS INTO A DECARBOXYLASE INVOLVED IN POLYKETIDE CHAIN -BRANCHING. *Journal of Biological Chemistry*. **282**, 35954–35963 (2007).
46. T. Wakimoto, Y. Egami, I. Abe, Calyculin: Nature's way of making the sponge-derived cytotoxin. *Nat Prod Rep.* **33**, 751–760 (2016).
47. M. J. Bertin *et al.*, The Phormidolide Biosynthetic Gene Cluster: A trans-AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide. *Chembiochem*. **17**, 164–173 (2016).
48. N. Moebius *et al.*, Biosynthesis of the Respiratory Toxin Bongkrekic Acid in the Pathogenic Bacterium *Burkholderia gladioli*. *Chemistry & Biology*. **19**, 1164–1174 (2012).
49. G.-L. Tang, Y. Q. Cheng, B. Shen, Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and

- nonribosomal peptide synthetase. *Chemistry & Biology*. **11**, 33–45 (2004).
50. G. R. Pettit, J. F. Day, J. L. Hartwell, H. B. Wood, Antineoplastic components of marine animals. *Nature* (1970).
 51. G. R. Pettit, C. L. Herald, D. L. Doubek, Isolation and structure of bryostatin 1. *Journal of the American Chemical Society* (1982).
 52. G. R. Pettit, C. L. Herald, Y. Kamano, Antineoplastic agents. 93. Structure of the *Bugula neritina* (marine bryozoa) antineoplastic component bryostatin 3. *J. Org. Chem.* **48**, 5354–5356 (1983).
 53. G. R. Pettit, Y. Kamano, C. L. Herald, Antineoplastic agents. 119. Isolation and structure of bryostatins 10 and 11. *J. Org. Chem.* **52**, 2848–2854 (1987).
 54. G. R. Pettit *et al.*, Antineoplastic agents. 126. Isolation and structure of bryostatins 12 and 13. *J. Org. Chem.* **52**, 2854–2860 (1987).
 55. G. R. Pettit, Y. Kamano, C. L. Herald, M. Tozawa, Isolation and structure of bryostatins 5–7. **63**, 1204–1208 (1985).
 56. G. R. Pettit, F. Gao, P. M. Blumberg, C. L. Herald, Antineoplastic Agents. 340. Isolation and Structural Elucidation of Bryostatins 16– 18 1. *Journal of natural ...* (1996).
 57. G. R. Pettit, F. Gao, D. Sengupta, J. C. Coll, C. L. Herald, Isolation and structure of bryostatins 14 and 15. *Tetrahedron ...* (1991).
 58. N. Lopanik, K. R. Gustafson, N. Lindquist, Structure of Bryostatin 20: A Symbiont-Produced Chemical Defense for Larvae of the Host Bryozoan, *Bugulaneritina*†. *J. Nat. Prod.* **67**, 1412–1414 (2004).
 59. T. Weber *et al.*, Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety. *Chemistry & Biology*. **15**, 175–188 (2008).
 60. F. H. Blackhall *et al.*, A phase II trial of bryostatin 1 in patients with non-Hodgkin's lymphoma. *Br. J. Cancer.* **84**, 465–469 (2001).
 61. L. Pagliaro *et al.*, A Phase II trial of bryostatin-1 for patients with metastatic renal cell carcinoma. *Cancer.* **89**, 615–618 (2000).
 62. L. M. Schrott *et al.*, Acute oral Bryostatin-1 administration improves learning deficits in the APP/PS1 transgenic mouse model of Alzheimer's disease. *Curr Alzheimer Res.* **12**, 22–31 (2015).
 63. R. Mehla *et al.*, Bryostatin Modulates Latent HIV-1 Infection via PKC and AMPK Signaling but Inhibits Acute Infection in a Receptor Independent Manner. *PLoS*

- ONE*. **5** (2010), doi:10.1371/journal.pone.0011160.
64. C. Gutiérrez *et al.*, Bryostatin-1 for latent virus reactivation in HIV-infected patients on antiretroviral therapy. *AIDS*. **30**, 1385–1392 (2016).
 65. Y. A. Ioannou, L. Altstiel, D. R. Crockford, METHODS AND COMPOSITIONS FOR TREATMENT OF LIPID STORAGE DISORDERS. *US Patent ...* (2015).
 66. Z. Tan *et al.*, Bryostatin Improves Survival and Reduces Ischemic Brain Injury in Aged Rats After Acute Ischemic Stroke. *Stroke*. **44**, 3490–3497 (2013).
 67. M.-K. Sun, J. Hongpaisan, D. L. Alkon, Postischemic PKC activation rescues retrograde and anterograde long-term memory. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14676–14680 (2009).
 68. I. D. Peña, C. Borlongan, G. Shen, W. Davis, Strategies to Extend Thrombolytic Time Window for Ischemic Stroke Treatment: An Unmet Clinical Need. *J Stroke*. **19**, 50–60 (2017).
 69. B. P. Lucke-Wold *et al.*, Bryostatin-1 Restores Blood Brain Barrier Integrity following Blast-Induced Traumatic Brain Injury. *Mol Neurobiol*. **52**, 1119–1134 (2014).
 70. K. H. Sharp, S. K. Davidson, M. G. Haygood, Localization of “Candidatus Endobugula sertula” and the bryostatins throughout the life cycle of the bryozoan *Bugula neritina*. *ISME J*. **1**, 693–702 (2007).
 71. S. K. Davidson, S. W. Allen, G. E. Lim, C. M. Anderson, M. G. Haygood, Evidence for the Biosynthesis of Bryostatins by the Bacterial Symbiont “Candidatus Endobugula sertula” of the Bryozoan *Bugula neritina*. *Appl. Environ. Microbiol*. **67**, 4531–4537 (2001).
 72. S. K. Davidson, M. G. Haygood, Identification of sibling species of the bryozoan *Bugula neritina* that produce different anticancer bryostatins and harbor distinct strains of the bacterial symbiont “Candidatus Endobugula sertula.” *The Biological Bulletin*. **196**, 273–280 (1999).
 73. M. G. Haygood, S. K. Davidson, Small-subunit rRNA genes and in situ hybridization with oligonucleotides specific for the bacterial symbionts in the larvae of the bryozoan *Bugula neritina* and proposal of “Candidatus endobugula sertula”. *Appl. Environ. Microbiol*. **63**, 4612–4616 (1997).
 74. R. M. Woollacott, Association of bacteria with bryozoan larvae. *Mar. Biol*. **65**, 155–158 (1981).
 75. N. Lopanik, N. Lindquist, N. Targett, Potent cytotoxins produced by a microbial symbiont protect host larvae from predation. *Oecologia*. **139**, 131–139 (2004).

76. N. B. Lopanik, N. M. Targett, N. Lindquist, Ontogeny of a symbiont-produced chemical defense in *Bugula neritina* (Bryozoa). *Mar. Ecol. Prog. Ser.* **327**, 183–191 (2006).
77. J. Linneman, D. Paulus, G. Lim-Fong, N. B. Lopanik, Latitudinal variation of a defensive symbiosis in the *Bugula neritina* (Bryozoa) sibling species complex. *PLoS ONE.* **9**, e108783 (2014).
78. G. E. Lim-Fong, L. A. Regali, M. G. Haygood, Evolutionary Relationships of “Candidatus Endobugula” Bacterial Symbionts and Their *Bugula* Bryozoan Hosts. *Appl. Environ. Microbiol.* **74**, 3605–3609 (2008).
79. T. M. McGovern, M. E. Hellberg, Cryptic species, cryptic endosymbionts, and geographical variation in chemical defences in the bryozoan *Bugula neritina*. *Mol. Ecol.* **12**, 1207–1215 (2003).
80. J. C. Yang *et al.*, The Complete Genome of *Teredinibacter turnerae* T7901: An Intracellular Endosymbiont of Marine Wood-Boring Bivalves (Shipworms). *PLoS ONE.* **4**, e6085 (2009).
81. D. Mendola, Aquacultural production of bryostatin 1 and ecteinascidin 743. *Drugs from the Sea* (2000).
82. D. Mendola, Aquaculture of three phyla of marine invertebrates to yield bioactive metabolites: process developments and economics. *Biomol. Eng.* **20**, 441–458 (2003).
83. S. Manaviazar, K. J. Hale, Total Synthesis of Bryostatin 1: A Short Route. *Angew. Chem. Int. Ed.* **50**, 8786–8789 (2011).
84. K. Ohmori *et al.*, Total Synthesis of Bryostatin 3. *Angew. Chem. Int. Ed.* **39**, 2290–2294 (2000).
85. P. A. Wender, A. J. Schrier, Total synthesis of bryostatin 9. *Journal of the American Chemical Society.* **133**, 9228–9231 (2011).
86. Y. Lu, S. K. Woo, M. J. Krische, Total synthesis of bryostatin 7 via C-C bond-forming hydrogenation. *Journal of the American Chemical Society.* **133**, 13876–13879 (2011).
87. David A Evans *et al.*, Total Synthesis of Bryostatin 2. *Journal of the ...* (1999).
88. G. E. Keck *et al.*, Substitution on the A-Ring Confers to Bryopyran Analogues the Unique Biological Activity Characteristic of Bryostatins and Distinct From That of the Phorbol Esters. *Org. Lett.* **11**, 593–596 (2009).
89. J. M. Ketcham *et al.*, Evaluation of Chromane-Based Bryostatin Analogues Prepared via Hydrogen-Mediated C-C Bond Formation: Potency Does Not Confer Bryostatin-

- like Biology. *Journal of the American Chemical Society*. **138**, 13415–13423 (2016).
90. R. G. Kerr, J. Lawry, K. A. Gush, In vitro biosynthetic studies of the bryostatins, anti-cancer agents from the marine bryozoan *Bugula neritina*. *Tetrahedron letters* (1996).
 91. M. Hildebrand *et al.*, *bryA*: An Unusual Modular Polyketide Synthase Gene from the Uncultivated Bacterial Symbiont of the Marine Bryozoan *Bugula neritina*. *Chemistry & Biology*. **11**, 1543–1552 (2004).
 92. N. B. Lopanik *et al.*, In vivo and in vitro trans-acylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. *Chemistry & Biology*. **15**, 1175–1186 (2008).
 93. D. L. Akey *et al.*, Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure*. **18**, 94–105 (2010).
 94. L. Gu *et al.*, GNAT-like strategy for polyketide chain initiation. *Science*. **318**, 970–974 (2007).
 95. D. C. Gay *et al.*, A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* (2014), doi:10.1016/j.str.2013.12.016.
 96. M. A. Skiba *et al.*, Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol.* (2016), doi:10.1021/acschembio.6b00759.
 97. D. C. Gay, P. J. Spear, A. T. Keatinge-Clay, A double-hotdog with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* (2014), doi:10.1021/cb500459b.
 98. F. P. Maloney, L. Gerwick, W. H. Gerwick, D. H. Sherman, J. L. Smith, Anatomy of the β -branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proceedings of the National Academy of Sciences* (2016), doi:10.1073/pnas.1607210113.
 99. S. Dutta *et al.*, Structure of a modular polyketide synthase. *Nature*. **510**, 512–517 (2014).
 100. A. T. Keatinge-Clay *et al.*, Catalysis, Specificity, and ACP Docking Site of *Streptomyces coelicolor* Malonyl-CoA:ACP Transacylase. *Structure*. **11**, 147–154 (2003).
 101. D. A. Hansen *et al.*, Biocatalytic Synthesis of Pikromycin, Methymycin, Neomethymycin, Novamethymycin, and Ketomethymycin. *Journal of the American Chemical Society*. **135**, 11232–11238 (2013).
 102. K. Bravo-Rodriguez *et al.*, Substrate Flexibility of a Mutated Acyltransferase

- Domain and Implications for Polyketide Biosynthesis. *Chemistry & Biology*. **22**, 1425–1430 (2015).
103. R. L. Berkow, A. S. Kraft, Bryostatin, a non-phorbol macrocyclic lactone, activates intact human polymorphonuclear leukocytes and binds to the phorbol ester receptor. *Biochem. Biophys. Res. Commun.* **131**, 1109–1116 (1985).
 104. E. A. Tallant, J. B. Smith, R. W. Wallace, Bryostatins mimic the effects of phorbol esters in intact human platelets - ScienceDirect. *Biochimica et Biophysica Acta (BBA)- ...* (1987).
 105. I. L. Dale, A. Gescher, Effects of activators of protein kinase C, including bryostatins 1 and 2, on the growth of A549 human lung carcinoma cells. *Int. J. Cancer*. **43**, 158–163 (1989).
 106. D. J. de Vries, C. L. Herald, G. R. Pettit, Demonstration of sub-nanomolar affinity of bryostatin 1 for the phorbol ester receptor in rat brain - ScienceDirect. *Biochemical ...* (1988).
 107. P. A. Philip *et al.*, Phase I Study of Bryostatin 1: Assessment of Interleukin 6 and Tumor Necrosis Factor Induction In Vivo. *JNCI Journal of the National Cancer Institute*. **85**, 1812–1818 (1993).
 108. R. J. Morgan *et al.*, Phase II trial of bryostatin-1 in combination with cisplatin in patients with recurrent or persistent epithelial ovarian cancer: a California cancer consortium study. *Invest New Drugs*. **30**, 723–728 (2012).
 109. F. Nezhat *et al.*, Phase II trial of the combination of bryostatin-1 and cisplatin in advanced or recurrent carcinoma of the cervix: a New York Gynecologic Oncology Group study. *Gynecol. Oncol.* **93**, 144–148 (2004).
 110. N. Kedei *et al.*, Comparison of transcriptional response to phorbol ester, bryostatin 1, and bryostatin analogs in LNCaP and U937 cancer cell lines provides insight into their differential mechanism of action. *Biochem. Pharmacol.* **85**, 313–324 (2013).
 111. Z. Tan *et al.*, Bryostatin extends tPA time window to 6 h following middle cerebral artery occlusion in aged female rats. *Eur. J. Pharmacol.* **764**, 404–412 (2015).
 112. B. J. Albert *et al.*, Combinations of isoform-targeted histone deacetylase inhibitors and bryostatin analogues display remarkable potency to activate latent HIV without global T-cell activation. *Sci Rep.* **7**, 7456 (2017).
 113. G. E. Keck, Y. B. Poudel, T. J. Cummins, A. Rudra, J. A. Covel, Total synthesis of bryostatin 1. *Journal of the American Chemical Society*. **133**, 744–747 (2010).
 114. B. M. Trost, G. Dong, Total synthesis of bryostatin 16 using atom-economical and chemoselective approaches. *Nature*. **456**, 485–488 (2008).

115. P. A. Wender *et al.*, Scalable synthesis of bryostatin 1 and analogs, adjuvant leads against latent HIV. *Science*. **358**, 218–223 (2017).
116. N. Kedei *et al.*, Neristatin 1 Provides Critical Insight into Bryostatin 1 Structure–Function Relationships. *J. Nat. Prod.* **78**, 896–900 (2015).
117. D. Staveness *et al.*, Inhibition of Chikungunya Virus-Induced Cell Death by Salicylate-Derived Bryostatin Analogues Provides Additional Evidence for a PKC-Independent Pathway. *J. Nat. Prod.* **79**, 680–684 (2016).
118. N. Lindquist, M. E. Hay, Palatability and Chemical Defense of Marine Invertebrate Larvae. *Ecological Monographs*. **66**, 431–450 (1996).
119. B. J. Carroll *et al.*, Identification of a set of genes involved in the formation of the substrate for the incorporation of the unusual “glycolate” chain extension unit in ansamitocin biosynthesis. *Journal of the American Chemical Society*. **124**, 4176–4177 (2002).
120. A. Kampa *et al.*, Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3129–E3137 (2013).
121. E. Shelest, N. Heimerl, M. Fichtner, S. Sasso, Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans. *BMC Genomics*. **16**, 1015 (2015).
122. M. Jenner *et al.*, Acyl hydrolases from trans-AT polyketide synthases target acetyl units on acyl carrier proteins. *Chem. Commun. (Camb.)*. **52**, 5262–5265 (2016).
123. I. J. Miller, T. R. Weyna, S. S. Fong, G. E. Lim-Fong, J. C. Kwan, Single sample resolution of rare microbial dark matter in a marine invertebrate metagenome. *Sci Rep.* **6**, 34362 (2016).
124. M. M. Schofield, D. H. Sherman, Meta-omic characterization of prokaryotic gene clusters for natural product biosynthesis. *Current Opinion in Biotechnology*. **24**, 1151–1158 (2013).
125. M. M. Schofield, S. Jain, D. Porat, G. J. Dick, D. H. Sherman, Identification and analysis of the bacterial endosymbiont specialized for production of the chemotherapeutic natural product ET-743. *Environ. Microbiol.* **17**, 3964–3975 (2015).
126. A. K. El-Sayed *et al.*, Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chemistry & Biology*. **10**, 419 (2003).
127. L. Gu *et al.*, Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair.

Journal of the American Chemical Society. **128**, 9014–9015 (2006).

128. M. H. Medema *et al.*, Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
129. A. C. Mercer, M. D. Burkart, The ubiquitous carrier protein—a window to metabolite biosynthesis. *Nat Prod Rep.* **24**, 750–773 (2007).
130. R. W. Haushalter *et al.*, Binding and pKa Modulation of a Polycyclic Substrate Analog in a Type II Polyketide Acyl Carrier Protein. *ACS Chem. Biol.* **6**, 413 (2011).
131. P. Pöplau, S. Frank, B. I. Morinaka, J. Piel, An Enzymatic Domain for the Formation of Cyclic Ethers in Complex Polyketides. *Angew. Chem. Int. Ed.* **52**, 13215–13218 (2013).
132. G. Berkhan, F. Hahn, A Dehydratase Domain in Ambruticin Biosynthesis Displays Additional Activity as a Pyran-Forming Cyclase. *Angew. Chem. Int. Ed.* **53**, 14240–14244 (2014).
133. L. Stols *et al.*, A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* **25**, 8–15 (2002).
134. K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi, T. Yura, Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in *Escherichia coli*. *Applied and ...* (1998).
135. B. A. Pfeifer, Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science*. **291**, 1790–1792 (2001).
136. C. Sánchez, L. Du, D. J. Edwards, M. D. Toney, B. Shen, Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide–polyketide antitumor drug bleomycin. *Chemistry & Biology*. **8**, 725–738 (2001).
137. D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet, P. C. Dorrestein, Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorganic & Medicinal Chemistry Letters*. **18**, 3107–3111 (2008).
138. A. Thamchaipenet, Multiple Domain Substitutions of Erythromycin Polyketide Synthase to Produce a Combinatorial Library. *日本放線菌学会誌*. **13**, 113–119 (1999).
139. R. S. Gokhale, S. Y. Tsuji, D. E. Cane, C. Khosla, Dissecting and exploiting

- intermodular communication in polyketide synthases. *Science*. **284**, 482–485 (1999).
140. C. C. Ladner, G. J. Williams, Harnessing natural product assembly lines: structure, promiscuity, and engineering. *J. Ind. Microbiol. Biotechnol.* **43**, 371–387 (2016).
141. R. McDaniel *et al.*, Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1846–1851 (1999).
142. R. McDaniel, M. Welch, C. R. Hutchinson, Genetic Approaches to Polyketide Antibiotics. 1. *Chem. Rev.* **105**, 543–558 (2005).
143. K. J. Weissman, P. F. Leadlay, Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*. **3**, 925–936 (2005).
144. A. T. Keatinge-Clay, The structures of type I polyketide synthases. *Nat Prod Rep.* **29**, 1050–1073 (2012).
145. T. Maier, M. Leibundgut, N. Ban, The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science*. **321**, 1315–1322 (2008).
146. Y. Xue, L. Zhao, H.-W. Liu, D. H. Sherman, A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12111–12116 (1998).
147. T. L. Pukala *et al.*, Subunit Architecture of Multiprotein Assemblies Determined Using Restraints from Gas-Phase Measurements. *Structure/Folding and Design*. **17**, 1235–1243 (2009).
148. S. Niu, J. N. Rabuck, B. T. Ruotolo, Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development. *Curr Opin Chem Biol.* **17**, 809–817 (2013).
149. J. D. Eschweiler, J. N. Rabuck-Gibbons, Y. Tian, B. T. Ruotolo, CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Analytical Chemistry*. **87**, 11516–11522 (2015).
150. A. Politis *et al.*, Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS ONE*. **5**, e12080 (2010).
151. H. G. Menzella *et al.*, Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).
152. H. G. Menzella, J. R. Carney, D. V. Santi, Rational design and assembly of synthetic trimodular polyketide synthases. *Chemistry & Biology*. **14**, 143–151 (2007).

153. R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, The structure of docking domains in modular polyketide synthases. *Chemistry & Biology*. **10**, 723–731 (2003).
154. D. Gahloth *et al.*, Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis. *Nat. Chem. Biol.* **13**, 975–981 (2017).
155. R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, Structure of fused docking domains from the erythromycin polyketide synthase (DEBS), a model for the interaction between DEBS 2 and DEBS 3: The A domain (2004), doi:10.2210/pdb1pzq/pdb.
156. T. J. Buchholz *et al.*, Crystal structure of fused docking domains from PikAIII and PikAIV of the pikromycin polyketide synthase (2009), doi:10.2210/pdb3f5h/pdb.
157. C. C. Aldrich, B. J. Beck, R. A. Fecik, D. H. Sherman, Biochemical investigation of pikromycin biosynthesis employing native penta- and hexaketide chain elongation intermediates. *Journal of the American Chemical Society*. **127**, 8441–8452 (2005).
158. A. Marchler-Bauer *et al.*, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*. **45**, D200–D203 (2017).
159. A. Marchler-Bauer *et al.*, CDD: NCBI's conserved domain database. *Nucleic Acids Research*. **43**, D222–6 (2015).
160. C. Y. Majmudar, B. Wang, J. K. Lum, K. Håkansson, A. K. Mapp, A High-Resolution Interaction Map of Three Transcriptional Activation Domains with a Key Coactivator from Photo-Cross-Linking and Multiplexed Mass Spectrometry. *Angew. Chem. Int. Ed.* **48**, 7021–7024 (2009).