INTEGRATED APPROACHES FOR THE DISCOVERY AND PRODUCTION OF

MARINE NATURAL PRODUCTS

by

Thomas E. Smith

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Medicinal Chemistry

The University of Utah

December 2016

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of   **Thomas E. Smith**

has been approved by the following supervisory committee members:

Eric W. Schmidt	, Chair	**10/17/2016** Date Approved
Chris M. Ireland	, Member	**10/17/2016** Date Approved
Amy M. Barrios	, Member	**10/17/2016** Date Approved
Louis R. Barrows	, Member	**10/17/2016** Date Approved
Ryan E. Looper	, Member	**10/17/2016** Date Approved

and by   **Kristen A. Keefe**   , Chair/Dean of

the Department/College/School of   **Pharmacy**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Natural products are structurally complex molecules and often exhibit intriguing biological activities. They are, however, notoriously difficult to supply, whether by chemical synthesis or isolation from a living organism. Recently, reconstitution of natural product biosynthetic pathways in a heterologous host has proved a successful strategy for producing natural products *in vivo*, but has rarely achieved the level of sustainability desired for medicinal applications. This approach is limited by our lack of understanding of biosynthesis and pathway expression. The focus of this dissertation is biosynthetic strategies for the supply of therapeutically-relevant natural products. Two examples are included, one involving a biosynthetic outlook of a known family of compounds with an uncharacterized metabolic origin, and a second uniting discovery, production, and biological characterization of a novel anti-HIV compound.

The adociasulfates are a family of marine sponge-derived meroterpenes known to inhibit kinesin, making them attractive anticancer drug leads. Despite difficulties in synthesizing adociasulfates, biosynthesis has never been investigated as a potential means of production. In Chapter 1, detailed consideration is given to the biosynthetic origin of these compounds, revealing a set of just four possible precursors for all sponge merotriterpenes. The mechanism of action of adociasulfates, addressed in Chapter 2, was shown to occur in a 1:1 interaction with kinesin, contrary to previous reports of microtubule-mimicking aggregates. Adociasulfates are thus shown to be valuable tools

for the study of kinesin and maintain potential therapeutic importance, making their production an ever more important goal.

The discovery, production, and biological characterization of an anti-HIV lanthipeptide, divamide A, is described in Chapter 3. The divamides were discovered from small tunicates from Papua New Guinea. By integrating structure- and genomics-based methodologies, we were able to elucidate the structure of a small amount of isolated material. This approach also provided us with a biosynthetic platform from which heterologous expression and sustainable production were achieved in *Escherichia coli*. The structure activity relationships of the divamides show that functional diversity is achieved by introducing minor structural changes to a conserved chemical scaffold Finally, an extended family of related peptides was identified that bears some of the hallmarks of known diversity-generating pathways.

Yeah this one right here goes out to all the baby's mamas, mamas
Mamas, mamas, baby mamas, mamas

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

CHAPTER 1

INTRODUCTION – A PRELIMINARY INVESTIGATION OF THE BIOSYNTHESIS

OF ADOCIASULFATES AND RELATED SPONGE MEROTERPENOIDS

1.1 Abstract

Kinesins are an attractive drug target for their role in both intracellular vesicle transport and cell division. While a number of kinesin inhibitors exist that target the site of ATP hydrolysis, only one family of compounds, the sponge-derived adociasulfates, inhibits kinesin competitively for microtubule-binding. Access to these compounds, however, is extremely limited. Thus, a sustainable means of producing adociasulfates is necessary in order to make full use of their potential. In this review, a biogenetic hypothesis is presented, highlighting several key structural features that provide clues to the metabolic origin of the adociasulfates. These clues suggest a bacterial source, although this has yet to be investigated. Additionally, the adociasulfates appear to be derived from at least four different hydroquinone hexaprenyl diphosphate precursors, each varying in number and position of epoxidations. A hypothetical biosynthetic pathway was reconstructed that may aid in the discovery of the authentic pathway from the sponge metagenome, with the intention of reconstituting this pathway for adociasulfate production. A similar rationale can be extended to other sponge

meroterpenoids, many of which display unique bioactivities, but all of which lack characterization of their respective biosynthetic pathways.

1.2 Introduction to kinesins

Kinesin motor proteins perform essential functions in eukaryotes, all involving the movement of motor domains along microtubule (MT) filaments. They regulate intracellular trafficking of vesicles and organelles, leading to roles in establishing cell polarity, development, and higher brain function.[1] They are also involved in cell division, where they are responsible for assembly and separation of the mitotic spindle.[2] As a result, inhibitors of kinesins are highly sought after, both as probes of molecular motor function and as drug leads. Kinesin superfamily (KIF) proteins each contain two motor domains, each of which has two regions considered to be druggable: the MT-binding domain and the motor domain. Kinesin motor domains couple the hydrolysis of ATP to conformational changes that translate into a stepping-like motion. Iterative rounds of ATP hydrolysis result in "walking" along an MT. The cycle of MT binding and ATP hydrolysis involves a weakly MT-binding state defined by bound ADP, and a strongly MT-binding state in which ATP is hydrolyzed.[3] Between these two states, binding of ADP-bound motor domains to MTs induces release of ADP and allows for the binding of ATP, while hydrolysis of ATP and release of phosphate allows the motor protein to assume the weakly MT-binding state once more. The cycles of hydrolysis of the two motor domains are staggered, such that an ATP-bound domain remains anchored to the MT while the ADP-bound domain releases the MT during the "step." The MT-binding region is not catalytic and simply serves as the binding site for MTs.

1.3 A mechanism of kinesin inhibition unique to adociasulfates

To date, adociasulfates remain the only known natural product kinesin inhibitors that compete with MTs for binding. All other known kinesin inhibitors, natural or synthetic, interfere with the ATPase function of kinesin either by direct competition with ATP or via an allosteric effect. Exceptions include the fluorescein analogue rose bengal lactone (RBL) and the polyoxometalate (POM) NSC 622124 ($[K_6Mo_{18}O_{62}P_2]^{6-}$) (**Figure 1.1**).[4-5] These inhibitors, though potent, display characteristic features of nonspecific inhibition, including a propensity to form aggregates, bind indiscriminately to positively charged protein surfaces, and inhibition of a variety of enzyme activities.

RBL was initially identified as a kinesin inhibitor in a computational docking screen for kinesin ligands, where it was shown to prevent MT binding by kinesin and inhibit both basal and MT-stimulated ATPase activity in the low micromolar IC_{50} range.[4] The RBL binding site was even mapped to the MT-binding domain of kinesin using docking and FRET techniques, implying its mode of binding was analogous to that of adociasulfate-2.[6] The designation of RBL as a kinesin inhibitor hit was later exposed as a false positive on the basis that RBL forms large, 300-500 nm diameter particles and acts as a nonspecific inhibitor of various enzyme activities.[8] RBL also blocks the binding of myosin to actin within the mid-micromolar range, calling into question its potential as a tool to probe molecular motor function.[4] Despite this, recent cell-based research has employed RBL to study kinesin *in vitro* and *in vivo*.[9-10] It is difficult to draw concrete conclusions from biological applications of RBL while controversies exist over its specificity, but perhaps more distressing is that these controversies go largely unnoticed by the broader scientific community.

NSC 622124 is a negatively-charged, achiral, nanometer-scale particle that inhibits both basal and MT-stimulated ATPase activity of a variety of kinesins with mid-nanomolar IC_{50} values.[5] The NSC 622124 binding site maps to two regions of kinesin, one that occurs within the MT-binding domain and the other nearby the nucleotide-binding domain. While NSC 622124 is encouragingly potent, a number of features suggest that it has limited usefulness. While adociasulfates and RBL can be classified as small molecules, NSC 622124 is large (MW >3000 Da), anionic and membrane impermeable.[7] NSC 622124 is a nanomolar inhibitor of a variety of unrelated proteins in addition to kinesin, including CK2, GSKß, and c-Src kinases, and DNA-binding by Sox-2.[7, 11] Charged surface residues mediate these interactions, without which NSC 622124 inhibition is impaired.[5, 7, 11] NSC 622124 is unstable and is readily hydrolyzed in aqueous environments, but the addition of nontarget protein enhances its lifetime and maintains its activity.[7] These observations suggest that off-target protein interactions do occur. There is also the potential for a large, achiral molecule to bind multiple copies of a target. Indeed, aggregations of NSC 622124 with target protein have been observed at low micromolar NSC 622124 concentrations.[7] While there is clear evidence that each POM exhibits selectivity for distinct targets and maintains unique structure-activity relationships, there is a propensity for NSC 622124, and likely for other POMs, to bind multiple targets with significant efficiencies. Collectively, use of NSC 622124 may yield misleading information with regard to target function *in vivo*.

Until recently, adociasulfates were thought to form MT-mimicking aggregates, bringing into question their potential as drugs or mechanistic probes.[12] It is now understood that adociasulfates bind kinesin in a 1:1 interaction.[13] In light of these

findings, it is crucial to point out the unlikeliness of kinesin inhibitors RBL and NSC

622124 to behave as expected in biochemical or cell-based investigations. Adociasulfates

are the only experimentally validated inhibitors to compete with MTs for binding kinesin

at a single-molecule level. Thus, there exists some urgency to achieve sustainable

adociasulfate production. In this chapter, I have laid out a strategy to identify the

biosynthetic pathway responsible for adociasulfate production in the sponge holobiont

based on a proposed biosynthetic route, highlighting several distinguishing features of

these molecules. With knowledge of the adociasulfate biosynthesis available, production

might be achieved *in vitro* by pathway reconstitution or *in vivo* by heterologous

expression.

1.4 A proposed biosynthetic route for sponge merotriterpenoids

Adociasulfates are part of a large family of sponge-derived meroterpenoids.

Along with related triterpene hydroquinones, adociasulfates appear to be restricted to

sponges within the *Chalinidae* family, specifically members of the *Haliclona* and

Cladocroce genera (**Figure 1.2**).[6, 13-18] Merosesquiterpenoids encompass a much greater

diversity of compounds than merotriterpenoids and can be found in a diverse range of

sponge lineages, especially within the genus *Dysidea*.[19] Diterpene and sesterterpene

hydroquinones have also been isolated from sponges.[20-21] A recent and comprehensive

review of sponge meroterpenoids highlights the structural diversity and biological

activity of this extensive class of compounds.[19] Here, I will point out distinctive features

of the adociasulfate structure indicative of their biogenetic origin. These features have

broad implications for sponge meroterpenoid biosynthesis in general. As will become

evident in the following discussion, there are many overlapping structural features of

meroterpenoids that suggest a common metabolic origin. However, there is a lack of knowledge with regard to meroterpenoid biosynthesis by sponges or their associated bacteria. No pathways for such compounds have been described despite hundreds of known compounds. The characterization of one meroterpenoid pathway could shed light on other pathways responsible for the production of bioactive molecules, some of which have known mechanisms of action and exhibit unique activities that cannot be substituted using any readily available compounds.[13, 22-23] Adociasulfates provide an excellent starting point because a relatively simple biosynthetic hypothesis can be postulated, with few enzymatic steps leading to the construction of the carbon skeleton and a limited number of potential hydrobenzoquinoid precursors. In fact, it is conceivable that all sponge triterpene hydroquinols are derived from a single parent pathway. Below, I will discuss the features that unify the adociasulfates and their ilk, establish a biogenetic hypothesis, and provide a hypothetical, bacterial adociasulfate pathway based on these observations.

A defining feature of the adociasulfates is the requirement for a linear triterpene-diphosphate as opposed to a precursor derived from squalene. Prenyl diphosphates are typically formed by a head-to-tail condensation of isopentenyl diphosphate (IPP) with either dimethylallyl diphosphate (DMAPP) or the product of a previous such condensation, yielding linear terpenes that are extended by five carbons. Squalene, however, is made by the head-to-head condensation of two C_{15} farnesyl-diphosphates (FPP) to produce a symmetrical triterpene. The consequences of this are twofold. First, without the diphosphate, squalene is no longer activated for prenyl transfer to a hydrobenzoquinoid substrate (**Figure 1.3-A, B**). Second, cyclization of squalene gives a

characteristic arrangement of methyls that is not observed for adociasulfates or any other hydroquinone meroterpenoids (**Figure 1.3-C**). Occurrences of linear meroterpenoids in sponges have been noted, though not from sponges related to adociasulfate-producers.[19] Nonetheless, there is a precedent for prenyl transfer of linear triterpenes to quinones, as occurs in ubiquinone biosynthesis (**Figure 1.3-B**), while there is none for the equivalent transfer of squalene.

All sponge merotriterpenoids can potentially be derived from a common series of linear precursors (**Figure 1.4**). These universal precursors, the products of aromatic prenylation by hexaprenyl diphosphate, would then be cyclized via a proton-initiated (type II) carbocation-mediated cyclization cascade. Most adociasulfates are hydroxylated at one (e.g. AS-2) or both (e.g. AS-3) carbons at positions corresponding to alkenes in hexaprenyl diphosphate. This suggests that epoxidation of the linear substrate occurs prior to cyclization. The number and location of epoxides in the cyclization substrate provides a convenient way to group biosynthetically related sponge merotriterpenoids. Thus, group I merotriterpenoids are epoxidized at position 10,11, group II at both positions 6,7 and 10,11, and group III at position 6,7, while group IV compounds are not epoxidized (**Figure 1.4**). All proposed cyclization schemes described herein are based on (*S,S*) epoxide configurations, given the observed stereochemistry of the putative epoxide-derived hydroxy-substituted carbons present in group I, II, and III adociasulfates.

The simplest hypothetical cyclization schemes involve the group I meroterpenoids. Compounds in this group likely undergo two independent cyclization cascades and exhibit few rearrangements. The initial cyclization of AS-1, -2, -5, -6, -7, and halicloic acid A would be identical for each compound, with epoxide opening to

form a hydroxyl group at C11, establishing the four rings of the "adociasulfate core" (**Figure 1.3-C**) with ring D fused to the hydrobenzoquinone moiety (**Figure 1.5-A**).[14-15, 24] The resulting carbocation would then be quenched by proton abstraction, restoring aromaticity. A second proton-initiated cyclization of the remaining two olefins would produce a fifth ring and a carbocation at position C6. Here, AS-1, -5, and -7 could differ from AS-2, -6, and halicloic acid A in the manner of base abstraction. In the former group, deprotonation would occur at C5 to introduce a new double bond, leaving the fifth ring independent of the core, while in the latter group a sixth, seven-membered ring could be formed by attack of the C11 hydroxyl on the carbocation. Postcyclization proton abstraction would then occur at the cyclic ether oxygen. AS-10 likely undergoes the same initial cyclization as the other group I members, but would then involve a hydride shift after the second proton initiation event, placing the carbocation on C7 instead of C6 and resulting in a six-membered heterocycle instead of seven (**Figure 1.5-B**).[17] The final structure of AS-2 would then be relatively flat with the terminal ring of AS-10 is twisted relative to the plane of the core. Halicloic acid B resembles AS-10, but its cyclization would involve an additional rearrangement: a methyl transfer following the hydride shift (**Figure 1.5-C**).[24] This would place the carbocation on C2. Deprotonation at C3 would then yield a tri-substituted alkene. Halicloic acids A and B and AS-10 may use an alternative aromatic prenyl acceptor to hydrobenzoquinone, as a glycolic acid moiety substitutes for the hydroxyl at the 5' position in these natural products. The final group I terpenes, toxicols A-C, may undergo a unique cyclization in which a rearrangement condenses the initial six-membered ring into a five-membered ring, resulting in an unstable secondary carbocation at C15.[25] Cyclization would continue with subsequent

attack by C19 (**Figure 1.5-D-i**). Alternatively, C14 instead of C15 could directly attack C10 in the initial epoxide opening, which would be sterically hindered by the two axial methyls of C10 and C14 (**Figure 1.5-D-ii**). A second cyclization step and proton abstraction would result in the final product, with two independent ring systems. Finally, adociasulfates and related meroterpenoids would be sulfated at either, none, or both hydrobenzoquinone hydroxyls, while 5' glycolic acids never appear to be modified further.

Adociasulfates and related meroterpenoids of group II are likely derived from a diepoxy precursor (**Figure 1.6**). Three of five members of this group exhibit a 5' glycolic acid substitution akin to AS-10 and the halicloic acids.[13, 16, 26] The primary proton-initiated cyclization event of AS-9 may mirror that of AS-2 from group I, with the opening the 10,11-epoxide and establishment of the adociasulfate core. The second cyclization could involve the opening of the 6,7-epoxide by back-side attack of the C11 hydroxyl at the more-substituted C6 position in a typical acid-catalyzed epoxide opening. This would result in the formation of a seven-membered ring and an inversion of C6 stereochemistry. Assumption of a *pro*-chair conformation would position C6 into a pro-(*R*) configuration relative to the C11 hydroxyl attack, resulting in an axial terminal olefin. For group I compounds, the lack of the 6,7-epoxide may allow for inclusion of the 2,3 terminal alkene in the second cyclization event (**Figure 3.4-A**), whereas all group II compounds display a free terminal olefin. This may reflect the ease of protonating an epoxide over an alkene.

Group III merotriterpenoids are likely derived from a 6,7-epoxy precursor (**Figure 1.7**). This group is characterized by a lack of fusion to the aromatic ring and by

implied quenching with water. In the proposed cyclization of AS-3 and -4, initiation by protonation would result in a bicyclic drimane-like skeleton that undergoes rearrangement before deprotonation by an active-site base, yielding a highly stable tetra-substituted double bond and unique configurations of methylated carbons (**Figure 1.7-A**).[14] The bond formed between C14 and C19 differs from those from groups I and II in that C19 would attack the pro-(*S*) face of the alkene as opposed to the pro-(*R*) face. Protonation of the 6,7-epoxide would initiate the second cyclization in AS-4 with the 10,11 olefin, while hydrolysis of the epoxide would lead to AS-3. Both hydrolysis and cyclization would result in an inversion of C5 stereochemistry, similar to the 6,7-epoxide opening of AS-9. The first cyclization event of shaagrockol C would also produce a bicyclic system, though deprotonation would occur prior to any rearrangement, yielding a tetra-substituted alkene (**Figure 1.7-B**).[27] The second cyclization would be similar to that of AS-4 but would occur from the opposite face of the 10,11 alkene, such that C11 has an (*R*) configuration instead of (*S*). A rearrangement would then follow immediately after ring formation and water would attack the C11 carbocation with inversion of stereochemistry. The net result of these dramatically different cyclization routes is that the newly formed ring of shaagrockol C would incorporate an axial hydroxyl in place of a proton, as in AS-4. Thus, shaagrockol C and AS-4 display the same relative configuration about C11, despite differing absolute configurations. Additionally, an attack from the pro-(*R*) face of the 10,11 alkene on the 6,7 protonated epoxide would place a large alkyl group in an unfavorable axial position. An attack from the pro-(*S*) face, which would place the same large group in the equatorial position, should be more favorable, but the occurrence of the less favorable pro-(*R*) route may allow for the addition of water over

deprotonation. Shaagrockol B, isolated together with shaagrockol C, is the oxidation

product of C about the 22,23-alkene and is likely not enzymatic in origin.[27]

The remaining six known sponge merotriterpenoids of group IV may be derived

from a substrate lacking epoxidation. The majority of these compounds undergo complex

rearrangements, as evidenced by their atypical methyl positions. Like the group III

compounds, none of the group IV members exhibit fused rings with the aromatic moiety,

suggesting that aromatic ring fusion requires the presence of the 10,11-epoxide. Another

common feature between groups III and IV is the absence of 5' glycolic acid substitution.

For AS-11 and -12, the proposed initial cyclization would again yield only a two-ring

system, but would differ from groups I, II, and III in the attack of C23 on C18 from the

pro-(*S*) face of the 22,23-alkene (**Figure 1.8-A**).[18] In the previous merotriterpenoid

groups, this attack likely occurs from the pro-(*R*) face. The bond between C14 and C19

would be formed as in groups I and II. The second cyclization step of AS-11 and

adociaquinol likely resembles that of AS-4 but, due to the absence of the 6,7-epoxide,

would continue on to the terminal olefin. Deprotonation at the C9 methyl would

introduce the exocyclic alkene. AS-12 likely undergoes hydride and methyl shifts prior to

deprotonation, allowing the more stable trisubstituted alkene to form. The structure of

AS-8 can be reached with a single proton-initiated cascade with extensive

rearrangements, such that the end result appears structurally distinct from the

adociasulfate core (**Figure 1.8-B**).[15] In total, five hydride shifts and four methyl shifts

would need to occur before an attack by water at the bridgehead carbon C7. Cyclization

of the initial bicyclic ring system of toxiusol likely involves both of the atypical pro-(*S*)-

directed attacks of C23 on C18, as predicted for AS-11 and adociaquinol, and C19 on

C14, as predicted for the group III compounds (**Figure 1.8-C**).[18, 25] Two hydride shifts

and a methyl transfer would occur before deprotonation ends the first cyclization. The

second proton initiated cascade of toxiusol would follow that of AS-12 but would include

a second hydride transfer prior to deprotonation, placing the trisubstituted alkene on the

opposite ring. The final sponge merotriterpenoid, akaterpin, likely follows a similar

cyclization scheme as toxiusol but would involve an alkyl shift during the first

cyclization that moves the remaining linear isoprene chain from C14 to the bridgehead

carbon, C19 (**Figure 1.8-D**).[28]

After critical consideration of the origin of adociasulfates, it should be clear that

all sponge merotriterpenoids of the hydrobenzoquinone family are related

biosynthetically. In each adociasulfate discovery reported, mixtures of compounds from

multiple groups were identified, suggesting a common synthetic route that is independent

of the epoxidation state of the substrate.[13-14, 16-18, 25] Of this class of compounds, all but

one member has been isolated from sponges within the family *Chalinidae*. The exception

is akaterpin, which was reportedly discovered from *Callyspongia sp.*[28] Though

Callyspongia is a member of the same order as *Chalinidae* (order *Haplosclerida*),

Callyspongia is far enough removed in this case to be considered unrelated (Mary Kay

Harper, personal communication). Thus, these compounds can be used as taxonomic

identifiers, potentially due to a shared biosynthetic pathway.

1.5 Considerations of the enzymatic origin of sponge merotriterpenoids

Only a few key biosynthetic steps are required for all four groups of

merotriterpenoids described above: aromatic prenylation, proton-initiated cyclization, and

sulfation. Epoxidation also occurs for the majority of these compounds, with the

exception of those in group IV. The potential enzyme families responsible for these key steps of adociasulfate construction are discussed in this section. The source of the terpene and benzoquinone precursors is also considered, as these metabolites can be derived from multiple routes and the enzymes involved in their synthesis may be components of an adociasulfate biosynthetic gene cluster. In addition to the enzymatic origins of sponge merotriterpenoids, the identity of the producing organism is taken into account, as this will dramatically affect the genetic organization of the pathway (or lack thereof).

1.5.1 Origin of adociasulfate precursors. The majority of the adociasulfate structure is constructed of five carbon isoprene units. There are two known biosynthetic pathways for isoprene production: the mevalonate (MEV) pathway, which provides the precursors for steroid anabolism in eukaryotes but is also present in some bacteria, and the 1-deoxy-D-xylulose-5 phosphate (MEP/non-mevalonate) pathway unique to plants, bacteria, and some parasites (**Figure 1.9**). Both of these are considered primary metabolic pathways. It is possible that the adociasulfate pathway draws IPP directly from an endogenous metabolite pool and lacks any dedicated genes for IPP/DMAPP synthesis. However, the producing organism's native isoprene source does not necessarily imply that pathway's involvement in secondary metabolism. Bacteria normally lacking the MEV pathway, including the prolific natural product producing actinobacteria, are known to horizontally acquire MEV pathway genes and incorporate them into meroterpenoid biosynthetic clusters as a pathway-specific source of IPP/DMAPP.[29-38] Some MEP pathway bacteria contain duplications of MEP genes in secondary metabolite clusters.[39-40] Gene duplication and divergence of endogenous biosynthetic proteins is one way in which enzymes with alternative substrate preferences or novel enzymatic functions have

evolved, especially in the context of secondary metabolism.[41-44] The role of these seemingly redundant genes may be to enhance production of precursor metabolites, to impose distinct regulations on gene expression, or to provide a unique function. Both bacteria and eukaryotes are capable of this phenomenon. Thus, copies of genes with primary metabolic functions might be responsible for meroterpenoid production, as opposed to endogenous genes taking on dual primary and secondary metabolic roles. With several possible means of obtaining IPP/DMAPP for polyprenyl diphosphate synthesis and no guarantee that MEV- or MEP-pathway genes would be present in a hypothetical adociasulfate gene cluster, pathway identification should focus on the biosynthetic steps fundamental to adociasulfate production, and only consider the presence of isoprene pathway elements as a secondary indication of a terpene pathway.

The adociasulfate prenyl donor, consisting of six isoprene units, is almost certainly a product of a *trans* isoprenyl diphosphate synthase (IPP synthase). IPP synthases are soluble, Mg^{2+}-dependent prenyltransferases (PTases) mechanistically related to aromatic UbiA-like PTases discussed later on.[45] These enzymes are responsible for producing prenyl diphosphates of different lengths for various biological functions, including polyprenyl diphosphates of 30-50 carbons in length used in the biosynthesis of ubiquinone, and the FPP used to make squalene in steroid biosynthesis. IPP synthases are sometimes components of meroterpenoid gene clusters.[29-34, 37, 39] Their inclusion in secondary metabolite pathways may reflect a selection mechanism for a particular length polyprenyl substrate by establishing a distinct substrate pool for meroterpenoid biosynthesis separate from the endogenous isoprenyl diphosphate pool. In other

pathways, however, the lack of any IPP synthase gene suggests that native IPP synthases provide the prenyl substrate for secondary metabolism.

While only two known pathways are responsible for isoprene production, the aromatic prenyl acceptor can be derived from a number of distinct pathways. All meroterpenoid pathways discussed in this review contain genes responsible for providing or modifying existing aromatic precursors. Understanding the function of such genes could aid in the recognition of a hydroquinone meroterpenoid pathway. Like prenyl diphosphates, quinones are also derived from primary metabolic pathways. According to the Kyoto Encyclopedia of Genes and Genomes (KEGG), hydroquinone and 4-hydroxyphenylacetate (4HPA), a potential prehydroxylation precursor of the 5'-glycolic acid substituted adociasulfates, are both components of phenylalanine/tyrosine metabolism (**Figure 1.10-A**). In the model put forth by KEGG, 4HPA can be derived from 4-hydroxyphenylpyruvate (4HPP) and funnels into either the 4HPA degradation pathway of *E. coli* or the homogentisate pathway used by both prokaryotes and eukaryotes to metabolize phenylalanine and tyrosine, while hydroquinone can be obtained from homogentisate in a few enzymatic steps.[46-47] However, poorly characterized, broad-substrate enzymes like aldehyde dehydrogenases and aldehyde oxidases are implicated in 4HPA and hydroquinone synthesis.[48-49] The placement of these metabolites in the KEGG model is based on feeding studies with aromatic carbon sources in which aromatic substrates were shown to be metabolized via the 4HPA pathway, homogentisate pathway, or by a variation of the 3-oxoadipate pathway.[46-47, 50] They are not endogenous metabolites and thus lack the potential to be used as prenyl acceptors as it may appear in the model. However, the enzymes of tyrosine metabolism may represent

functions required by the adociasulfate pathway. Oxidative decarboxylation, such as that catalyzed by 4HPP dioxygenase, an Fe^{2+}-dependent internal ketoacid dioxygenase, could be used to generate hydroxy-4HPA from 4HPP.[51] In a less direct route, 4HPA could potentially be obtained from 4HPP via 4HPA decarboxylase, which normally produces *p*-cresol from 4HPA in *Clostridium difficile* and is a member of the glycyl radical enzymes (GRE) of the radical-SAM superfamily.[52] Alternatively, decarboxylation of 4HPP to the aldehyde with subsequent oxidation to 4HPA by either an aldehyde dehydrogenase (ALDH), as indicated in the KEGG model, or an aldehyde oxidase (AOX), as shown by KEGG for the oxidation of gentisaldehyde to gentisate, could be possible. Both the $NAD(P)^+$-dependent ALDHs and flavin-dependent molybdenum/tungsten AOXs are described as broad-substrate and are largely uncharacterized.[48-49] Subsequent hydroxylation of the 4HPA acyl side-chain could be carried out by an α-ketoglutarate-dependent Fe^{2+} enzyme or a cytochrome P450 (P450).[53-54] Hydroquinone could be derived from gentisate by decarboxylation, potentially requiring a nonoxidative decarboxylase like 5-carboxyvanillate or γ-resorcylate decarboxylase, both members of the ACMSD decarboxylase family.[55-57] Oxidative decarboxylation of aromatic substrates can also be carried out by flavin monooxygenases.[58] Though it is unclear what role tyrosine metabolism might play in providing the aromatic prenyl acceptor for meroterpenoid biosynthesis, the adociasulfate pathway may incorporate elements of these pathways in order to supply hydroquinone.

Aromatic prenylation substrates are not solely derived from tyrosine or shikimate and could be generated *de novo* for adociasulfate biosynthesis. Hydroquinone prenyl acceptors of known meroterpenoid pathways are derived primarily from polyketides,[30-31,]

[33, 37, 59-63] but can also be derived from tyrosine[40, 64] and, quite remarkably, from the carbohydrate *sedo*-heptulose-7-phosphate (**Figure 1.10-B**).[32] Another possibility is that the prenyl acceptor is extensively modified after the initial prenylation event, as is the case in ubiquinone synthesis (**Figure 1.10-C**). 4-hydroxybenzoate (4HB) and homogentisate, similar in structure to 4HPA and hydroquinone, are known prenyl acceptors in the ubiquinone and plastoquinone/tocopherol pathways, respectively.[65-66] Thus, there is a precedent for metabolites of tyrosine catabolism to act as prenyl donors. Prenyl-4HB/homogentisate could be decarboxylated and then hydroxylated to generate the precursor of adociasulfate cyclization. From the examples described here of the potential origin of the aromatic prenyl acceptor, it is unlikely that the adociasulfate pathway can derive its prenyl acceptor solely from primary metabolism. Thus, the pathway should include genes involved in the production of an aromatic substrate. The presence and interpretation of these genes can be used to facilitate meroterpenoid pathway identification.

1.5.2 Prenylation. Prenyltransferase is the first true step in adociasulfate biosynthesis. A variety of aromatic prenyltransferases (PTases) are known to generate prenylated aromatic products resembling the linear adociasulfate precursors shown in **Figure 1.4-A**. The earliest to be characterized of these enzymes is 4HB-PTase, which is involved in the biosynthesis of ubiquinone.[65, 67] 4HB-PTases are present in all forms of life, as ubiquinone is an essential component of biological redox reactions like the electron transport chain. The mechanism of prenyl transfer by UbiA, the 4HB-PTase of *E. coli*, involves activation of the isoprene diphosphate to form a carbocation, initiating the electrophilic addition of 4HB.[68-69] Both carbocation formation and electrophilic

addition require an active site constructed from two separate Asp-rich regions that are highly conserved amongst membrane-associated aromatic PTases.[45, 68] Only recently have crystal structures become available for this family of PTases, though only structures of archaeal proteins that have not been biochemically characterized have been solved.[69-70] However, these and earlier modeling studies agree that coordination of the isoprenyl diphosphate with an active site Mg^{2+}-ion induces formation of an isoprenyl carbocation that then condenses with 4HB, resulting in a Friedel-Crafts type alkylation.[68-69] UbiA and related PTases are broadly substrate selective *in vitro*, especially with regard to the length of isoprenes that can be incorporated into their product.[65, 71-72] Ubiquinone prenyl groups vary greatly in length within and between species.[72] The protein structures of two archaeal UbiA homologs reveal a lateral opening into the hydrophobic interior of the lipid membrane that may provide access of long-chain isoprenes to the active site.[69-70] The prenyl chain presumably extends into the bilayer rather than be contained within the enzyme, allowing UbiA extreme substrate flexibility. UbiA also exhibits broad substrate specificity for prenyl acceptors, provided that these substrates are *para*-alcohol- or amino-substituted benzoates.[73] In fact, membrane-associated aromatic PTases utilize a wide variety of aromatic prenyl acceptors in the biosynthesis of plastoquinones/tocopherols, menaquinone, and even secondary metabolites; a testament to their vast biosynthetic potential.[74] It is likely, owing in particular to their accommodation of variable isoprene chain lengths, that membrane aromatic PTases are involved in sponge merotriterpenoid biosynthesis. The same rationale can be extended to meroterpenoids of other isoprene classes.

Prenylation of aromatic compounds is not unique to the UbiA family of PTases. The ABBA-family of aromatic PTases, so named for their alternating, antiparallel α-ß-ß-α folds (dubbed the PT-fold or PT-barrel), are a more recently discovered family of soluble aromatic prenyltransferases involved in secondary metabolism of bacterial and fungal natural products.[75-76] Since the characterization of the first ABBA PTase CloQ, responsible for condensation of DMAPP with 4HPP in the biosynthesis of the antibiotic clorobiocin, several additional members of the ABBA PTase family have been described.[77] Some of these members utilize naphthalene-like prenyl acceptors: NphB and Fnq26 geranylate flaviolin in the biosynthesis of the antioxidant naphterin and furanonaphthoquinone I, respectively; HypSc adds DMAPP to 1,6-dihydroxynapthalene *in vitro*; SCO7190 was shown to prenylate hydroxynaphthalene, coumarin, and 1,3-hydroxybenzene derivatives, including resveratrol; the fungal ABBA PTases Ptf$_{At}$, Ptf$_{Ss}$, and Ptf$_{Bf}$ condense 2,7-dihydroxynapthalene and DMAPP *in vitro*, though their biological substrates are unknown.[30, 33, 78-79] Other ABBA PTases prenylate amino acids or peptides: the cyanobacterial AmbP1, WelP1, and FidP1 geranylate a tryptophan derivative, *cis*-indolyl vinyl isonitrile, in the biosynthesis of ambiguine, welwitindolinone, and fischerindole, respectively; LynF O-prenylates Tyr residues of various cyanobactins, which encompass a broad class of cyanobacteria-derived macrocylic peptides.[39, 80-81] Still other ABBA PTases use more diverse aromatic prenyl acceptors: PpzP adds DMAPP to 5,10-dihydrophenazine-1-carboxylate in endophenazine A biosynthesis; DzmP is the only known ABBA PTase to use FPP to farnesylate its aromatic substrate, diabenzodiazepinone.[29, 36] Fungal dimethylallyltryptophan synthases (DMATs) are structurally related to the CloQ-type and prenylate indole rings.[82-83] DMATs are even

more limited than ABBA PTases in that they almost exclusively employ DMAPP.

Additionally, the specificity of DMATs for aromatic prenyl acceptors other than indoles

is either limited or has not been explored.[84-85] In summary, ABBA PTases, though

broadly substrate selective with regard to the aromatic prenyl acceptor, are restricted in

the length of the prenyl donor to two or fewer isoprene units. While it is possible that

ABBA prenyltransferases are responsible for merosesquiterpenoid biosynthesis, only one

ABBA PTase is known to accept FPP as a prenyl donor.[29] Despite the significant role of

ABBA PTases in secondary metabolism, the comparison between PTase families better

supports the idea that a membrane-associated PTase is involved in sponge meroterpenoid

biosynthesis.

　　　1.5.3 Cyclization. Cyclization of linear terpenes is an electrophilic reaction

catalyzed by terpene cyclases (also called terpene synthases). Triterpene cyclases of the

class II squalene-hopane cyclase (SHC) and oxidosqualene-lanosterol cyclase (OSC)

families are known for their broad substrate selectivity and their extreme product

diversity *in vitro*.[86-89] This latter property results from the potential for deprotonation

after each cyclization step and at different positions, and from potential rearrangement.

Terpene cyclases in general have proven to be highly engineerable; single amino acid

changes can significantly alter product specificity.[90-91] OSCs are eukaryotic in origin and

are limited to formation of four-ring products from 2,3-oxidosqualene, while bacterial

SHCs are responsible for production of the five-ring hopenes from squalene (**Figure 1.3**).

SHCs accept a variety of terpene substrates *in vitro*, including 2,3-oxidosqualene,

indicative of their greater substrate flexibility relative to OSCs.[86, 92] Both SHCs and OSCs

follow the same scheme with regard to product formation: substrate-binding in product-

like conformation, initiation by protonation of the terminal isopropylidene alkene or oxirane, chaperoning cyclization by stabilization of carbocation intermediates, and termination by deprotonation or quenching with water. The initiation step imparts some control of substrate specificity for OSCs, as demonstrated by mutational studies of SHC wherein the ability of an SHC to cyclize squalene was abolished when the conserved SHC motif DDTAVV was replaced with the conserved OSC motif, DCTAEA.[91] At the same time, improved *in vitro* cyclization of 2,3-oxidosqualene by the mutant SHC was observed. Product diversity is in part controlled by aromatic residues that line the active site of the enzyme and stabilize carbocation intermediates via π-cation interactions, and also by residues positioned at the end of the cyclization cascade that are responsible for the deprotonation. Terpene cyclization is a highly stereospecific reaction, resulting in characteristic configurations about the chiral bridgehead and methyl-substituted carbons. The shape of the active site cavity likely plays a large role in determining the arrangement of the rings in the final product, and whether deprotonation terminates cyclization or the addition of water. Many adociasulfates display sterol-like stereochemistry within rings A-C, indicative of the "prechair" conformation assumed by group I adociasulfates prior to cyclization that is characteristic of both sterol and hopene cyclizations.[86] The position of the methyl at C21 in the adociasulfate core is shifted by one carbon relative to hopene and lanosterol due to the use of a head-to-tail versus head-to-head terpene precursor (**Figure 1.3**). Some adociasulfates exhibit early termination, resulting in only a two-ring system. This phenomenon has also been observed for SHC/OSCs *in vitro*.[86, 89] While flexibility of SHCs and OSCs has been shown *in vitro*, this does not necessarily translate to the same flexibility *in vivo*. Still, a class II terpene

cyclase is a likely candidate for an adociasulfate synthase given the nature of SHC and OSC substrates.

SHCs and OSCs have been shown to accept linear hydroquinone meroterpenoids as substrates *in vitro*.[88, 93-95] In these examples, SHCs are able to cyclize the prenyl side chain of the linear meroterpenoid substrate, but their products lack fusion of the aromatic moiety with the terpene ring system.[93, 95] An exception is SHC from *Zymomonas* mobilis, (ZmoSHC1) which fused a phenol ring with C_{10} and C_{15} isoprenes, but not C_{20}.[88] The OSC lupeol synthase (LUP1) from *Arabidopsis thaliana* is capable of forming *in vitro* petromindole, a fungal indole meroditerpenoid, from linear 3-(ω-oxidogeranylgeranyl)indole, successfully fusing the aromatic indole ring to the prenyl side chain.[94] This may reflect enzyme-specific preferences of OSC for the epoxide substrate, or a mechanistic difference in how LUP1 and ZmoSHC1 deal with termination of cyclization. The linear adociasulfate substrate is likely epoxidized at the 6,7 or 10,11 positions rather than at the 2,3 position. Some OSCs are able to cyclize 2,3-22,23-dioxidosqualene, but cyclization of internal epoxides is not known.[96-97] Proton initiation at the internal epoxides of adociasulfate precursors is potentially responsible for allowing some adociasulfates to escape cyclization of the 2,3 alkene. Additionally, adociasulfate cyclization events sometimes involve heterocycle formation via hydroxyls produced by putative epoxide ring openings. SHCs are capable of heterocycle formation in this way.[98] Despite the presence of epoxides, some adociasulfate cyclizations do not involve epoxide openings, even amongst those with epoxide substrates like AS-3/4 (**Figure 1.7-A**). Thus, it is more likely that an SHC adapted to initiation at epoxides is responsible for adociasulfate cyclization. Sponge meroterpenoids of smaller prenyl lengths do not exhibit

epoxidation of substrates and would not likely be cyclized by OSCs. Alternatively, class II diterpene cyclases exist that could act on C_{20} and possibly C_{15} substrates by proton-initiated cyclization.[99-101] In addition to the classical terpene cyclases, other meroterpenoid pathways may contain terpene cyclases specialized for linear aromatic meroterpenoid precursors.[32, 102] These have the potential to be pathway-specific and mechanistically novel, but also difficult to recognize as terpene cyclases. For example, P450s have been show to catalyze terpene cyclizations.[103-104]

Epoxidation of squalene in eukaryotes is carried out by squalene monooxygenase (SM), a membrane-bound flavin-dependent protein that requires molecular oxygen and reduced NADPH, as well as a P450 reductase partner.[105] The requirement for a P450 reductase is unique amongst flavin-dependent monooxygenases, as there is no structural relationship between SM and P450s, but several groups of flavin monooxygenases are known to require other flavin reductase partners.[58] There is evidence that a second, non-P450 type flavin reductase may be also be able to supply reduced NADPH to SM.[106] There is a precedent for SM in secondary metabolism, as the diterpene phenalinolactone, produced by a *Streptomyces* strain, includes an SM homolog in its biosynthetic gene cluster.[107] This SM homolog is believed to introduce an epoxide at the terminal olefin of the C20 geranylgeranyl diphosphate substrate. However, SM produces a single isomer of oxidosqualene, introducing an oxirane ring at the 2,3 position in the (*S*) configuration. It is possible that a related enzyme is responsible for the formation of the internal 6,7- and 10,11 epoxides of the adociasulfate precursor, but due to the rigid specificity of SM for terminal olefins it is likely that an unrelated monooxygenase is involved. P450 monooxygenases are involved in oxidative tailoring reactions in numerous natural

product pathways and are capable of performing a wide variety of chemical modifications on diverse substrates, including epoxidation. P450s are also important in the metabolism of xenobiotics and drugs in animals.[108] All P450s obtain reduced flavin via a P450 reductase partner, similar to SM.[109] Owing to their incredible diversity in both function and substrate specificity, this class of enzymes is a likely candidate for epoxidation in the adociasulfate pathway.

1.5.4 Sulfation. The final step in the synthesis of adociasulfates is sulfation of the hydroquinone moiety. In eukaryotes, sulfation is carried out by sulfotransferases (SULTs) that utilize 3'-phosphoadenosine 5'-phosphosulfate (PAPS) as a sulfonate (SO_3^-) donor. Sulfation substrates include glycosaminoglycans, proteins, lipids, hormones, and various drugs.[110] The role of sulfation varies, but is involved in cell signaling, molecular recognition, and detoxification.[111] Though SULTs are less prevalent in bacteria, a diversity of sulfated molecules exists, for many of which sulfation has a profound impact on biological function. Some of these functions involve interactions with eukaryotes. Commensal gut bacteria have been implicated in the sulfation of antibiotics in mammals.[112] Sulfated glycolipids are involved in establishing stable infections of symbionts within their eukaryotic hosts. and also play a role in bacterial pathogenesis.[113-114] Sulfation has been incorporated into secondary metabolism as well, where SULT domains have been identified within polyketide synthases to generate sulfated products, or, in one case, sulfation activates a substrate for decarboxylation.[115] Sulfated glycopeptide antibiotics exhibit reduced activation of resistance in antibiotic-resistant bacteria relative to their nonsulfated counterparts, potentially by preventing dimerization of antibiotics.[116] The role of sulfation in adociasulfate activity can only be guessed, as the

native biological function of adociasulfates is not known. However, with regard to kinesin, the sulfates only prevent membrane penetration and do not affect inhibition.[6, 13] Sulfation could be a mechanism for elimination to avoid toxicity associated with kinesin inhibition, or it could enhance secretion to facilitate exposure to predators. Not all sponge merotriterpenoids are sulfated, however, but these compounds have not been tested for kinesin inhibition.[24, 26] It has been suggested that an AS-14 analog containing an esterified glycolic acid moiety and lacking sulfation might be membrane permeable and still inhibit kinesin, making it a good anticancer lead.[13] Haliclotriol A and haliclotriol triacetate closely resemble this hypothetical analog and should be screened for kinesin inhibition.[26] Still, sulfation is not essential for adociasulfate biosynthesis, and the genes involved need not reside in the same gene cluster or even the same genome as the rest of the pathway. While either host or symbiont may produce adociasulfates, the other symbiotic partner may be responsible for their sulfation.

In addition to SULTs, a distinct clade of sulfotransferases, the arylsulfate sulfotransferases (ASSTs), exists that is PAPS-independent, using instead aryl sulfates, often phenolic, as sulfonate donors.[117] This small group of sulfotransferases is restricted to bacteria, including commensal and pathogenic bacteria.[112, 118] Native aryl sulfonate donors remain unknown for each of the few identified ASSTs, except for Cpz4 of the biosynthetic pathway for the liponucleoside antibiotic, caprazamycin.[119] Encoded within the *cpz* gene cluster is the ASST *cpz4* in addition to a SULT and a type III PKS (*cpz8* and *cpz6*, respectively). The Cpz6 product acts as the sulfonate donor after it itself is sulfated by Cpz8. Additionally, Cpz4 transfers its sulfonate group to a non-aryl substrate, suggesting there is potential for broader substrate diversity amongst ASSTs. Thus,

SULTs are not the only sulfotransferase capable of adociasulfate sulfation; bacterial ASSTs may be responsible.

1.6 Concluding remarks

Sponge meroterpenoids exhibit broad biological activity, but some display unique mechanisms of action. For these compounds, a sustainable and economic source is highly desirable. Chemical synthesis of adociasulfates has proved difficult and isolation from nature is not environmentally feasible.[120-122] A biosynthetic strategy would require knowledge of the native pathway. This review is aimed at dissecting the adociasulfate structure, highlighting key structural features that provide mechanistic clues toward the identification of the enzymes involved. The pattern of adociasulfate cyclization suggests of a squalene-hopene cyclase (**Figure 1.11**). The positions of the epoxides in the linear adociasulfate precursor suggest that squalene monooxygenase is unlikely to be involved; instead, a P450 may be responsible. The use of a head-to-tail linear triterpene precursor instead of squalene supports the idea that a polyprenyl synthase supplies the precursor to cyclization. These observations, which encompass the more distinct features of the adociasulfate structure, imply a bacterial origin. No biosynthetic pathways for sponge meroterpenoids have ever been identified, so there is no reference to base this conclusion on. In only one case has a producing organism been claimed to have been identified – for the production of avarol, a merosesquiterpenoid, by the sponge *Dysidea avara*. In these studies, avarol was traced to a specific sponge cell type and production was later observed from an axenic primary sponge culture.[123-124] No publications have followed these studies in nearly 17 years. In an effort to identify the biosynthetic genes involved in ilimaquinone production from *Hippospongia sp.*, I could not identify any biosynthetic

gene clusters by metagenome mining consistent with the biosynthetic features proposed for adociasulfate biosynthesis. This could reflect a lack of clustering and/or the absence of bacterial production, or a biosynthetic pathway entirely unrelated to that proposed for the adociasulfates. Thus, the possibility exists that adociasulfates and related meroterpenoids are sponge-derived, though merosesquiterpenoid biosynthesis may differ substantially with that of merotriterpenoids. The most consistent sponge-based biosynthetic origin would involve an off-shoot of ubiquinone synthesis, with cyclization catalyzed by an oxidosqualene cyclase.

Though the structure of adociasulfates favors symbiont-derived production over sponge-derived, no clear verdict can be obtained without experimental investigation. Clues as to what types of enzymes are responsible have been described here. Targeted searches of genes with these functions could help to identify the adociasulfate pathway. Metagenomic approaches may complicate data interpretation in that several to hundreds of gene homologs may be identified within a single metagenome, especially for those genes related to primary pathways, such as *ubiA*. A comparative metagenomics approach may resolve these issues, in which the metagenomes of nonproducing *Chalinidae* sponges are sequenced alongside adociasulfate-producing specimens to enable comparative analysis. While metagenome sequencing was attempted for *Cladocroce aculeata*, from which AS-8, -13, and -14 were characterized by Chris Ireland's lab, poor quality DNA was isolated from frozen samples.[13] Care must be taken to collect and prepare samples for analysis of by chemical and DNA sequencing approaches. Following the guidelines for pathway identification laid out in this review may result in successful

recognition of a meroterpenoid pathway, opening the door for biosynthetic approaches to the natural products supply problem that surrounds these valuable compounds.

rose bengal lactone   Dawson structure
(RBL)   (NSC 622124)
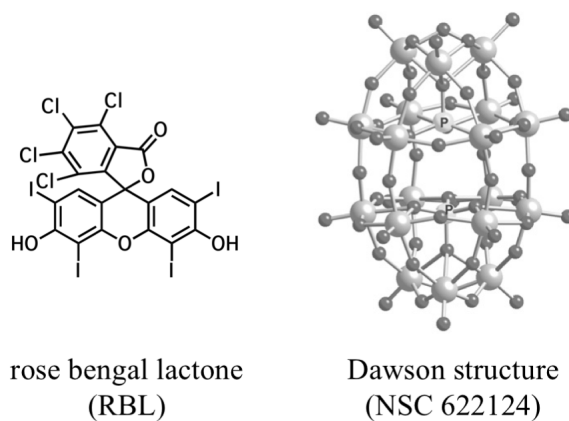
Figure 1.1. Structures of microtubule-competitive kinesin inhibitors. The polyoxometalate NSC 622124 $[K_6Mo_{18}O_{62}P_2]^{6-}$ assumes a Dawson structure. Dawson structure image reprinted from *Chemistry & Biology,* 15, R. Prudent *et al.*, "Identification of Polyoxometalates as Nanomolar Noncompetitive Inhibitors of Protein Kinase CK2", 683–692, Copyright 2008, with permission from Elsevier.[7]
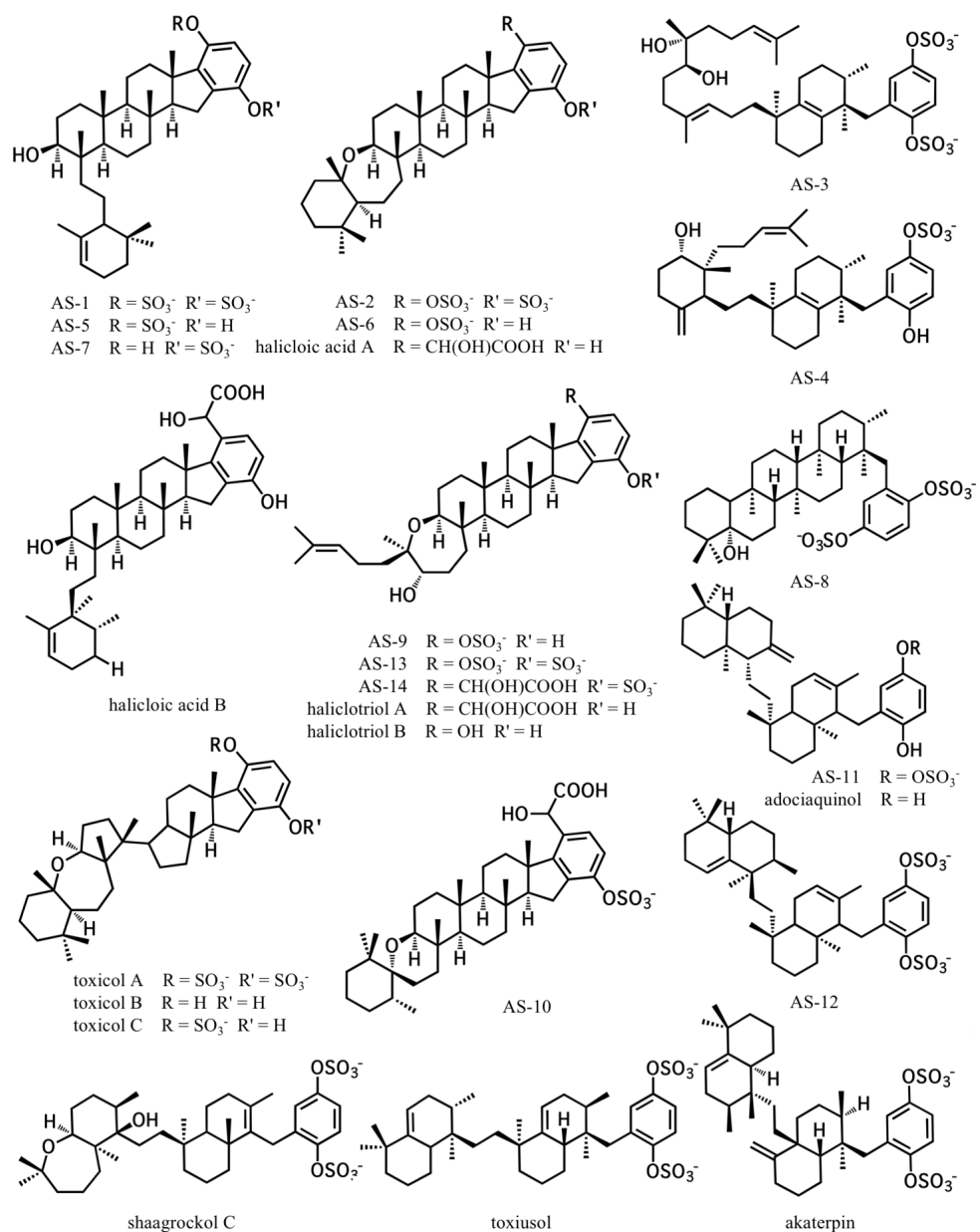
Figure 1.2. Chemical structures of related sponge merotriterpenoids. The adociasulfates and related compounds are derived from sponges of the family *Chalinidae*, with the exception of akaterpin, isolated from *Callyspongia sp*.
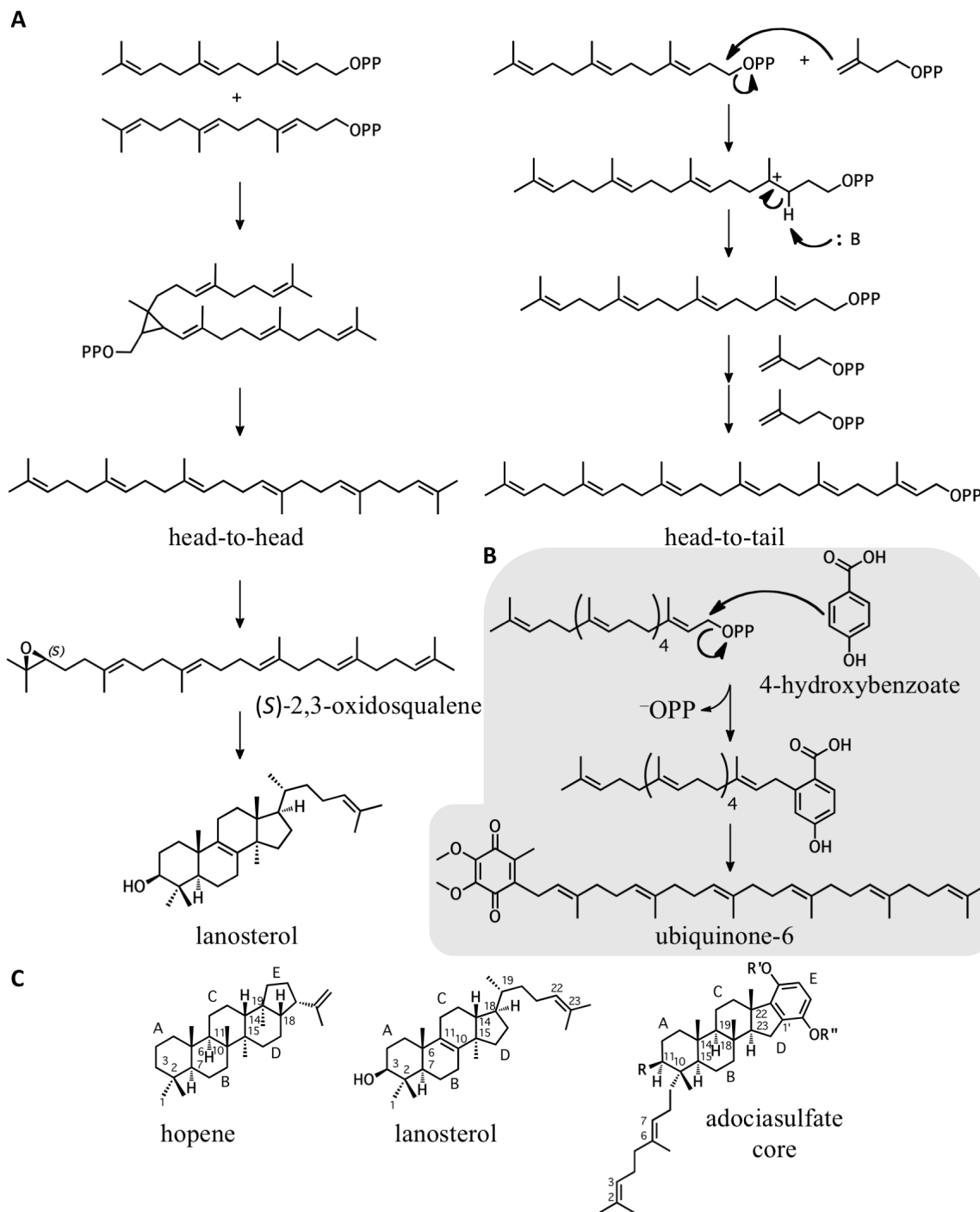
Figure 1.3. A comparison of the biosynthesis of squalene (left column) and polyprenyl diphosphate (right column). A) The condensation of two FPP molecules in head-to-head orientation yields squalene, while iterative head-to-tail prenylation results in hexaprenyl diphosphate. 2,3-oxidosqualene is the precursor for cyclization in the synthesis of hopenes and sterols. B) Polyprenyl diphosphates are substrates for aromatic prenylation in the synthesis of ubiquinone. C) The adociasulfate core present in many sponge merotriterpenoids resembles rings A-C of hopene and rings A-D of lanosterol in all configurations.
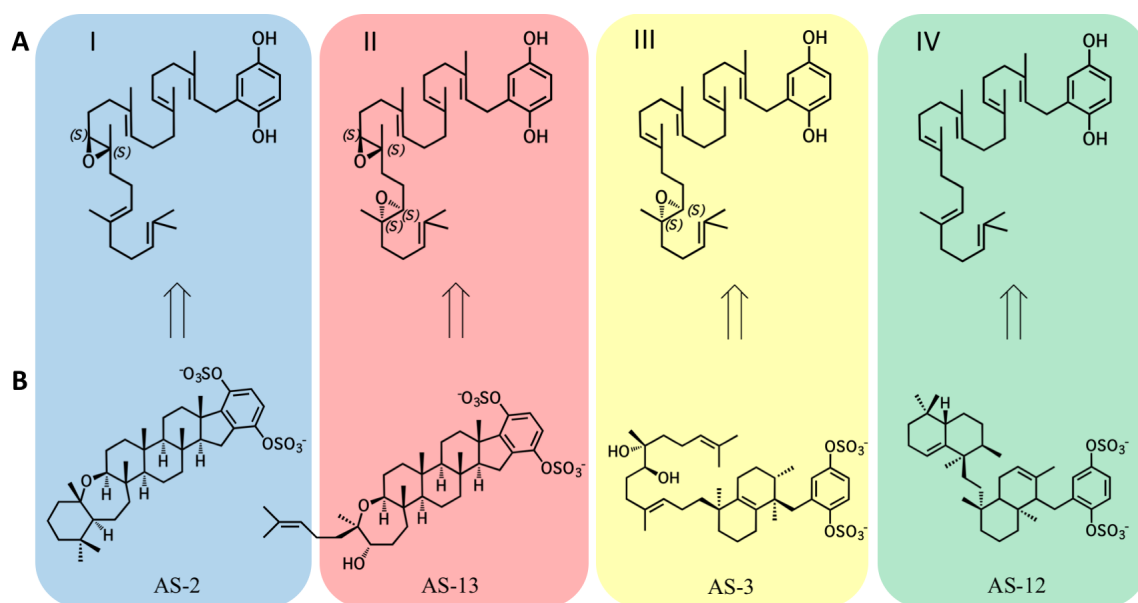
Figure 1.4. The four putative adociasulfate biosynthetic classes. A) Sponge merotriterpenoids can be divided into four groups by the number and position of epoxidations of the linear precursor. B) Representative adociasulfates derived from each major group.

Figure 1.5. Putative cyclization routes of group I sponge merotriterpenoids. Group I compounds are derived from 10,11-epoxyhexaprenyl diphosphate. A) Cyclization of the adociasulfate core of AS-1, -2, -5, -6, -7, and halicloic acid A occurs with an "all chair" arrangement similar to hopene/sterol cyclization. A second cyclization involving the remaining terminal olefins event involves either direction deprotonation (i) or attack by the C11 hydroxyl (ii). B) The initial cyclization event of AS-10 is identical to those in Figure 1.5-A. The second cyclization involves a proton transfer prior to attack by the C11 hydroxyl. C) The initial cyclization event of halicloic acid B is identical to those in Figure 1.5-A. The second cyclization follows that of AS-10 but the proton shift is followed by a methyl transfer. Deprotonation occurs instead of hydroxyl attack. D) The initial cyclization event of the toxicols may follow that in Figure 1.5-A but would include a rearrangement that collapses the six-membered A-ring into a five-membered ring (i). Alternatively, the initial cyclization may yield the five-membered ring directly (ii). The second cyclization follows that of AS-2, -6, and halicloic acid A.

A

i)

ii)

AS-1/5/7

AS-2/6/halicloic acid A

AS-10

Figure 1.5 Continued

C

halicloic acid B
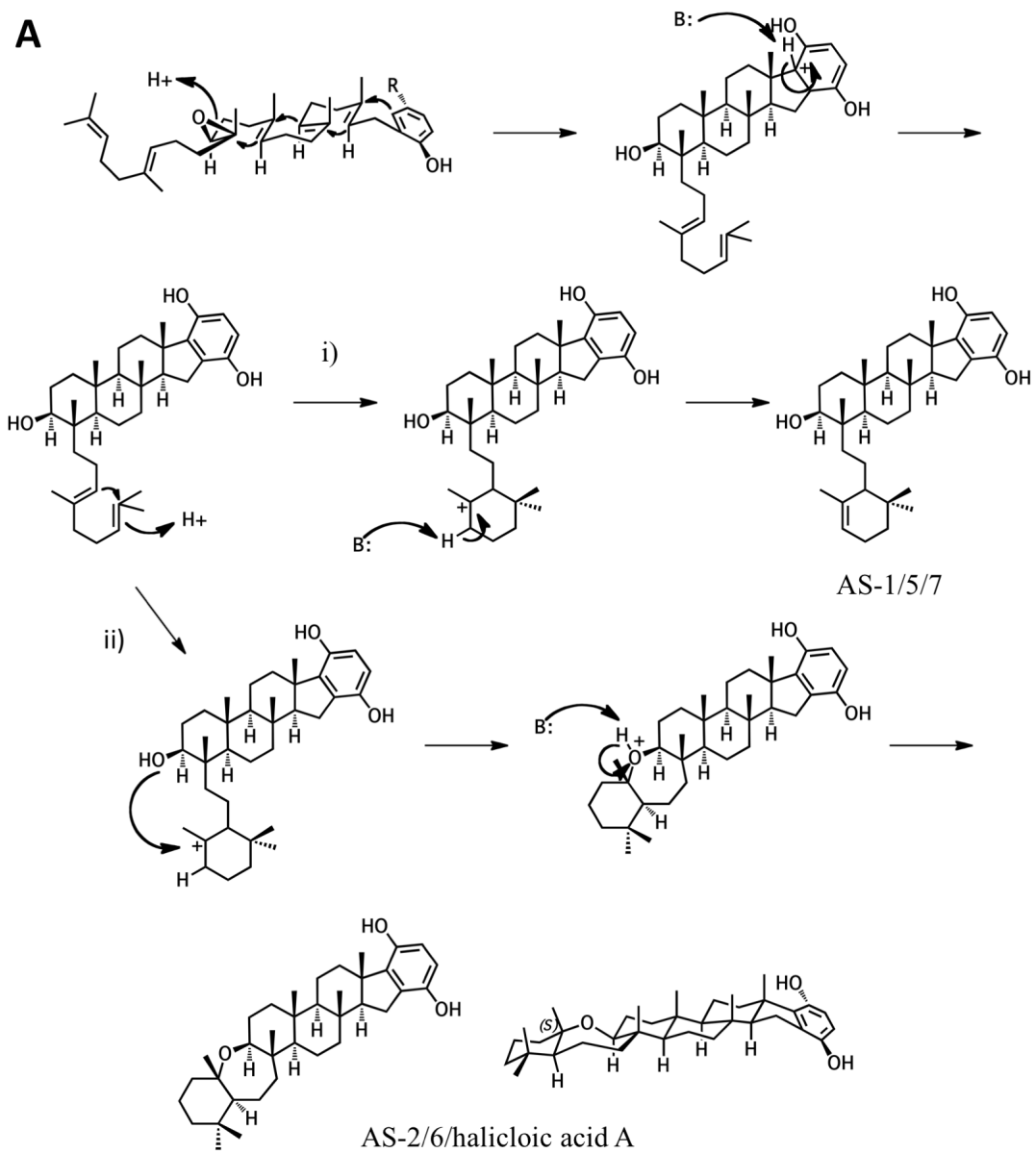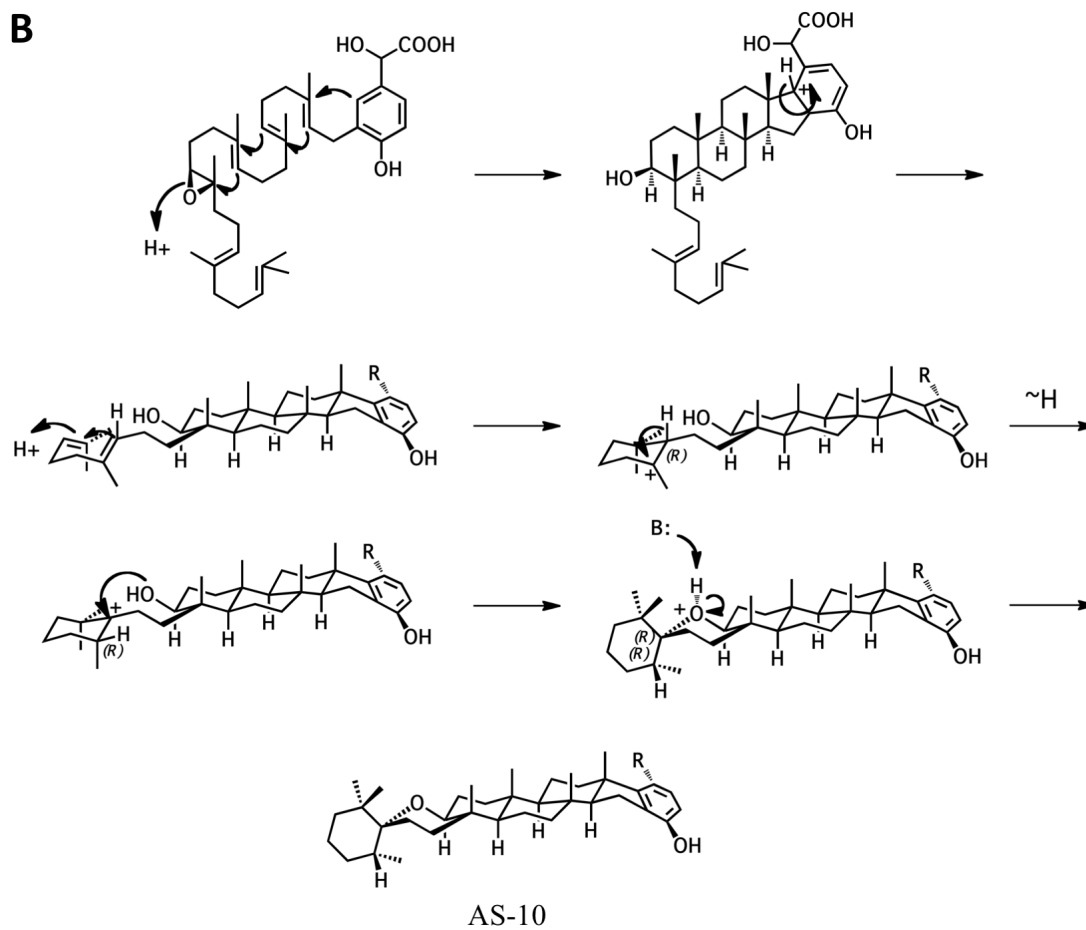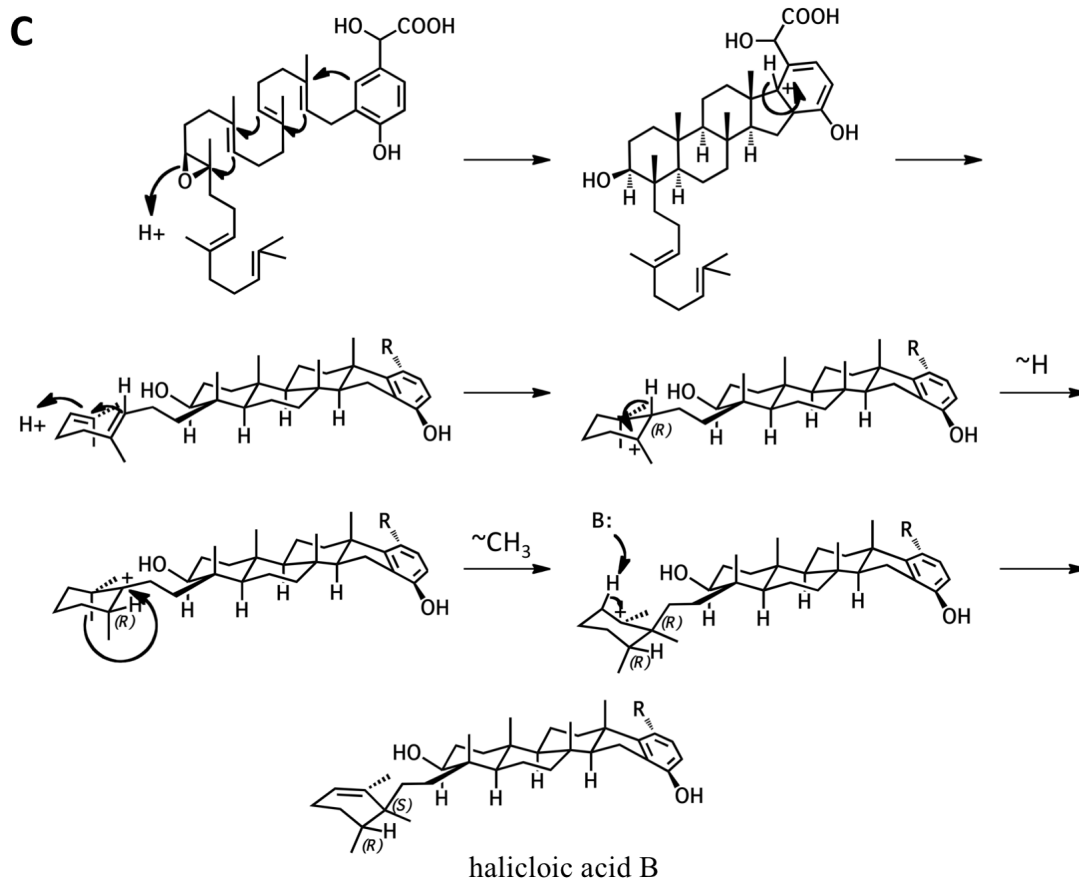
Figure 1.5 Continued

Figure 1.5 Continued

AS-9/13/14/
haliclotriols A-B
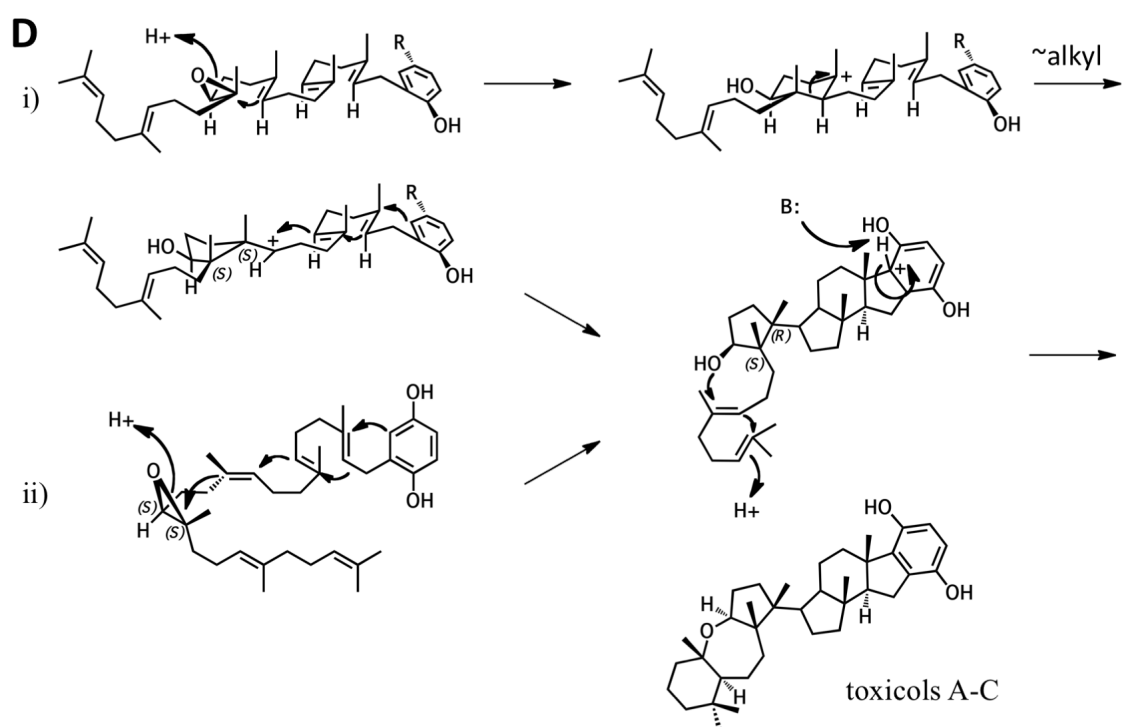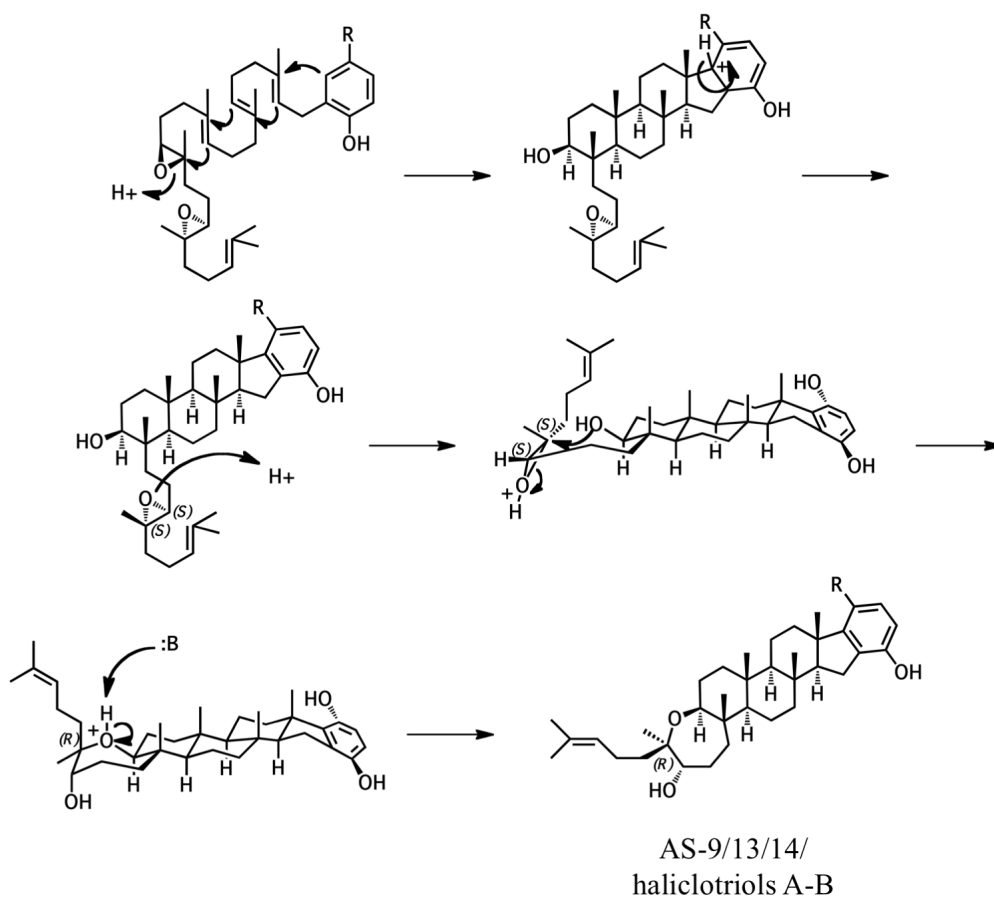
Figure 1.6. Putative cyclization routes of group II sponge merotriterpenoids. Group II compounds are derived from 6,7-10,11-diepoxyhexaprenyl diphosphate. The initial cyclization event is identical to that shown in Figure 1.5-A, resulting in the adociasulfate core ring system. The second cyclization involves epoxide opening of the 6,7-oxirane ring by the C11 hydroxyl, leaving a free terminal olefin.

Figure 1.7. Putative cyclization routes of group III sponge merotriterpenoids. Group III compounds are derived from 6,7-epoxyhexaprenyl diphosphate. A) The initial cyclization event of AS-3 and -4 yields a drimane-like two-ring system with a free hydroquinone moiety. The second cyclization involves epoxide opening of the 6,7-oxirane ring either by attack the 10,11 alkene for AS-4, producing a second bicyclic ring system (i), or attack by water for AS-3, yielding a diol (ii). B) The initial cyclization of shaagrockol C is similar to that of AS-3 and -4 but involves direct deprotonation without rearrangement. The second cyclization involves epoxide opening of the 6,7-oxirane ring by attack the 10,11 alkene followed by attack by water, then a third cyclization between the terminal alkene and the C7 hydroxyl.

AS-4

AS-3

shaagrockol C
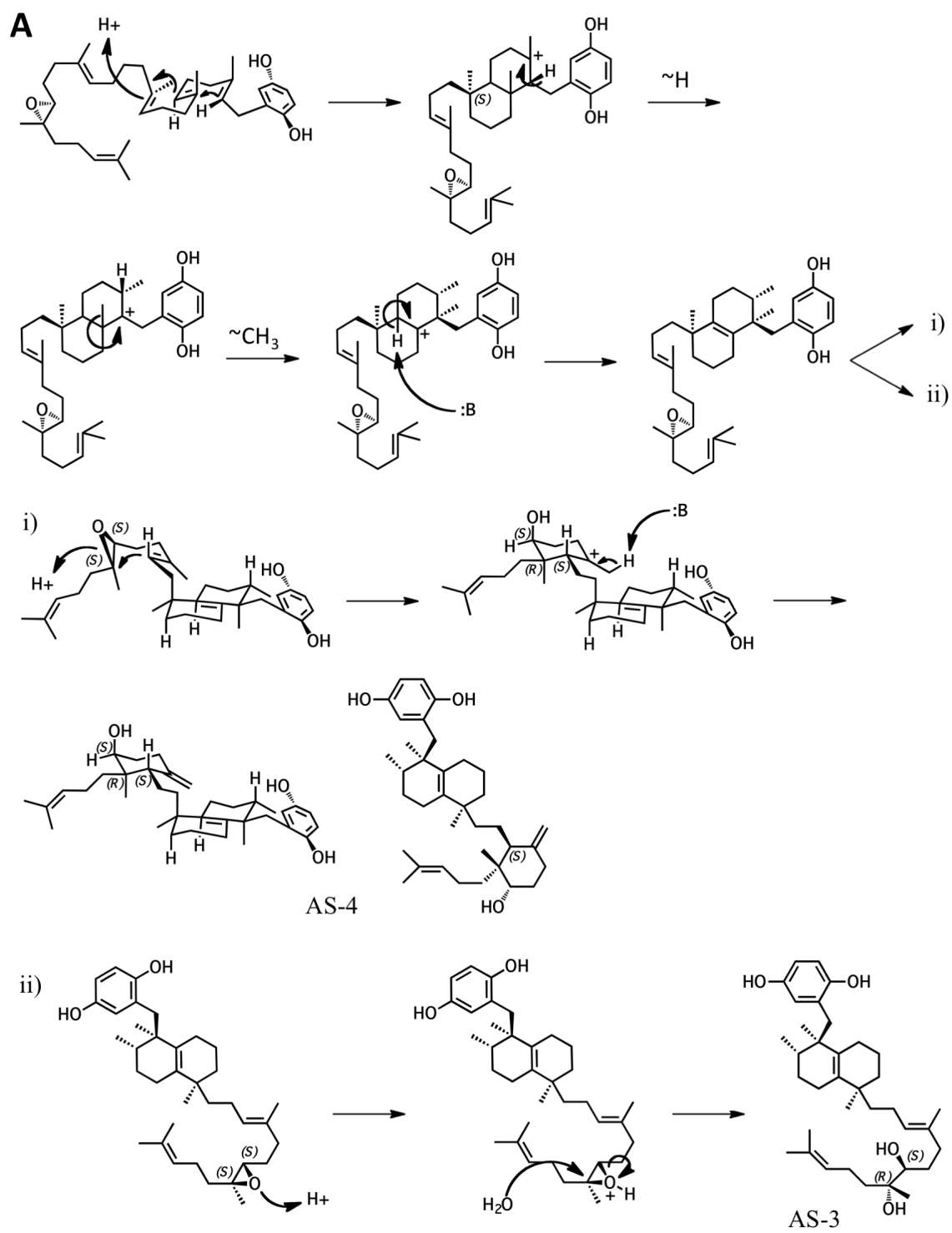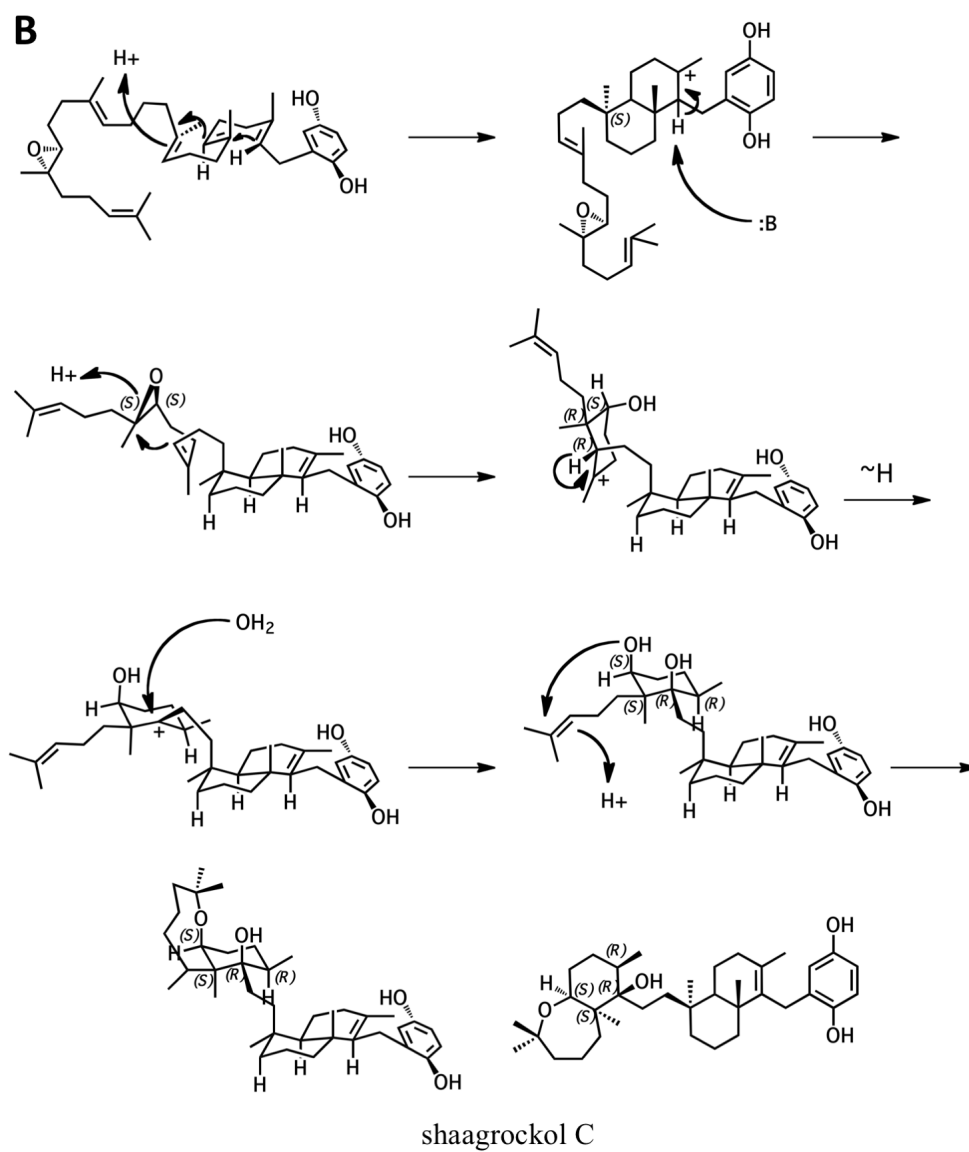
Figure 1.7 Continued

Figure 1.8. Putative cyclization routes of group IV sponge merotriterpenoids. Group IV compounds are derived from hexaprenyl diphosphate. A) The initial cyclization of AS-11, -12, and adociaquinol is similar to that of AS-3 and -4 but the attack of C19 on C14 occurs on a different face of the 14,15 alkene. The second cyclization results in a second bicyclic ring system, with either direct deprotonation to produce AS-11 and adociaquinol, or rearrangement followed by deprotonation to generate a more stable tertiary alkene in AS-12. B) The cyclization of AS-8 involves all six alkenes, but not hydroquinone, resulting in a five-membered ring system. Extensive rearrangement ensues, totaling five proton shifts and four methyl transfers. A final attack by water quenches the carbocation. C) The initial cyclization of toxiusol is similar to that of AS-3 and -4 in Figure 1.7-A. The second cyclization event follows that of AS-12 but includes an additional proton transfer across the bicyclic ring system. D) The initial cyclization of toxiusol is similar to that of AS-3 and -4 in Figure 1.7-A. The second cyclization event follows that of AS-12 but includes an additional proton transfer across the bicyclic ring system.

A

i)

AS-11/adociaquinol

ii)

AS-12

AS-8

Figure 1.8 Continued

C

toxiusol
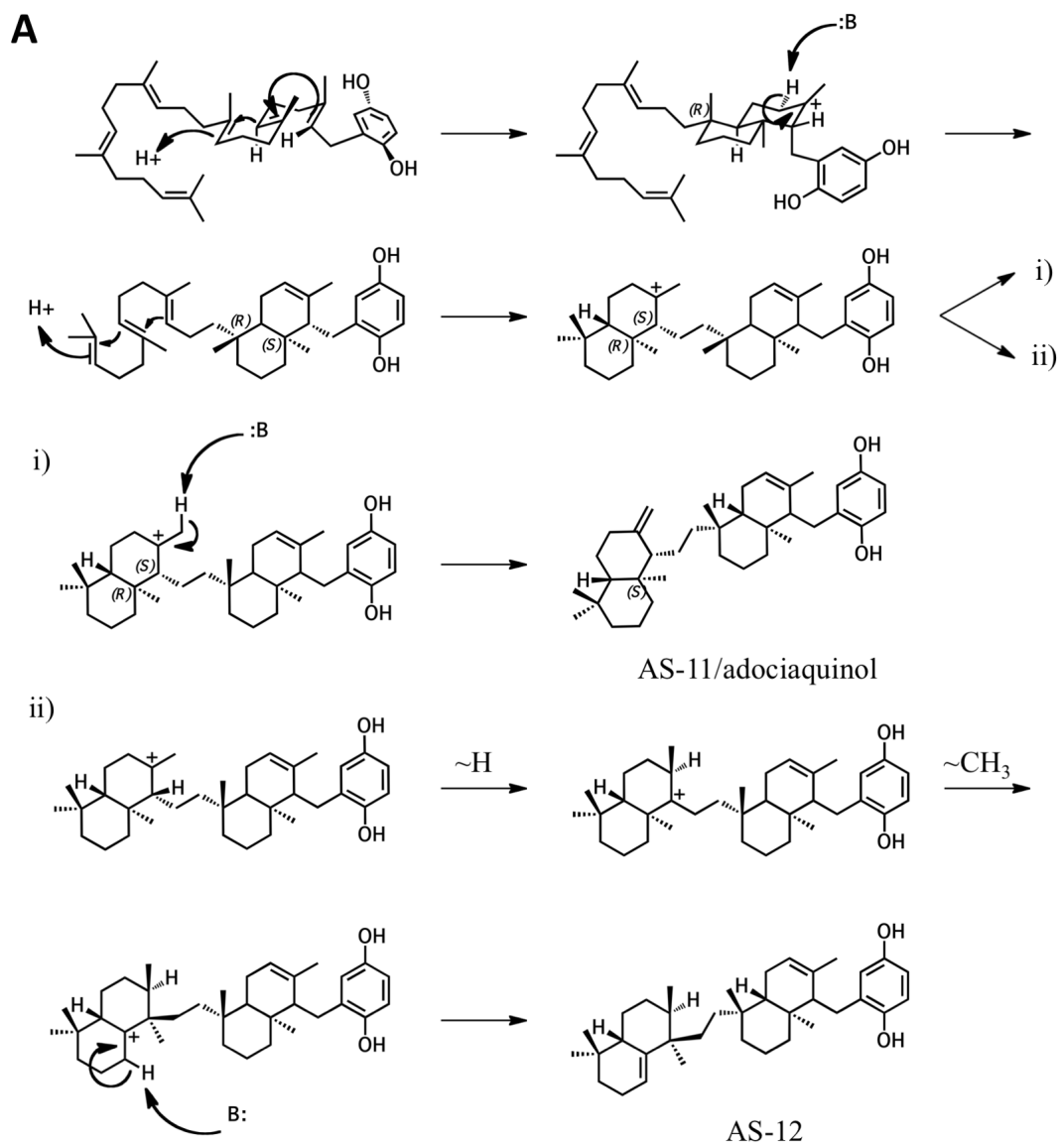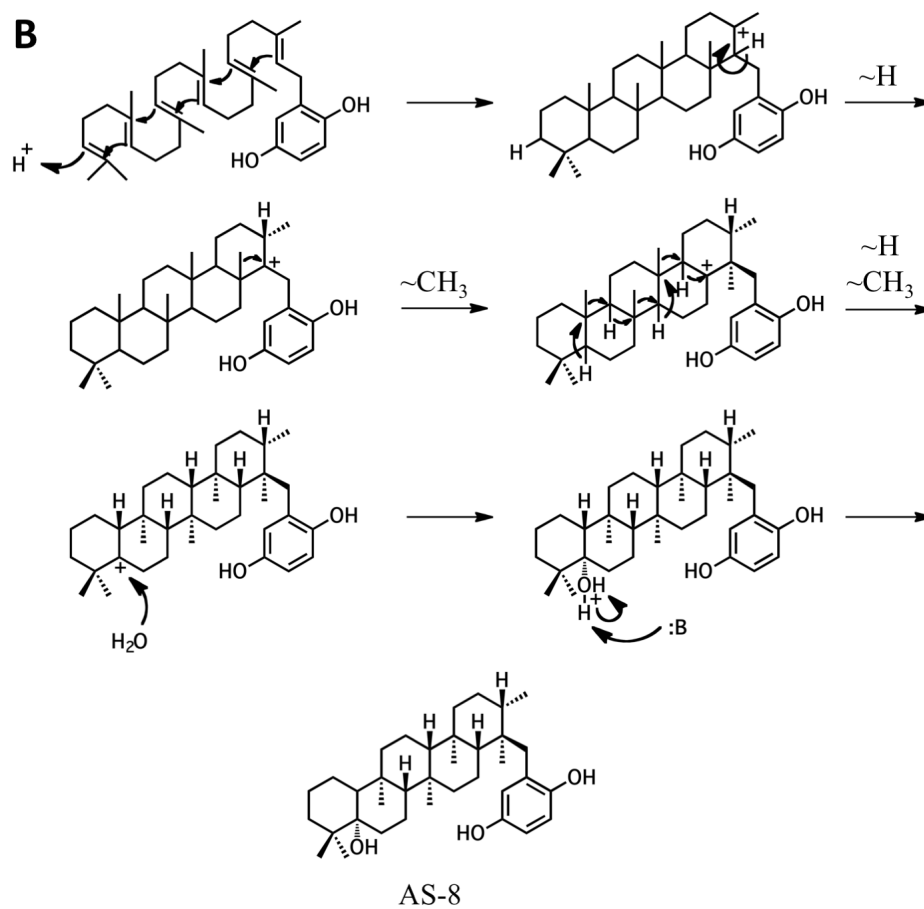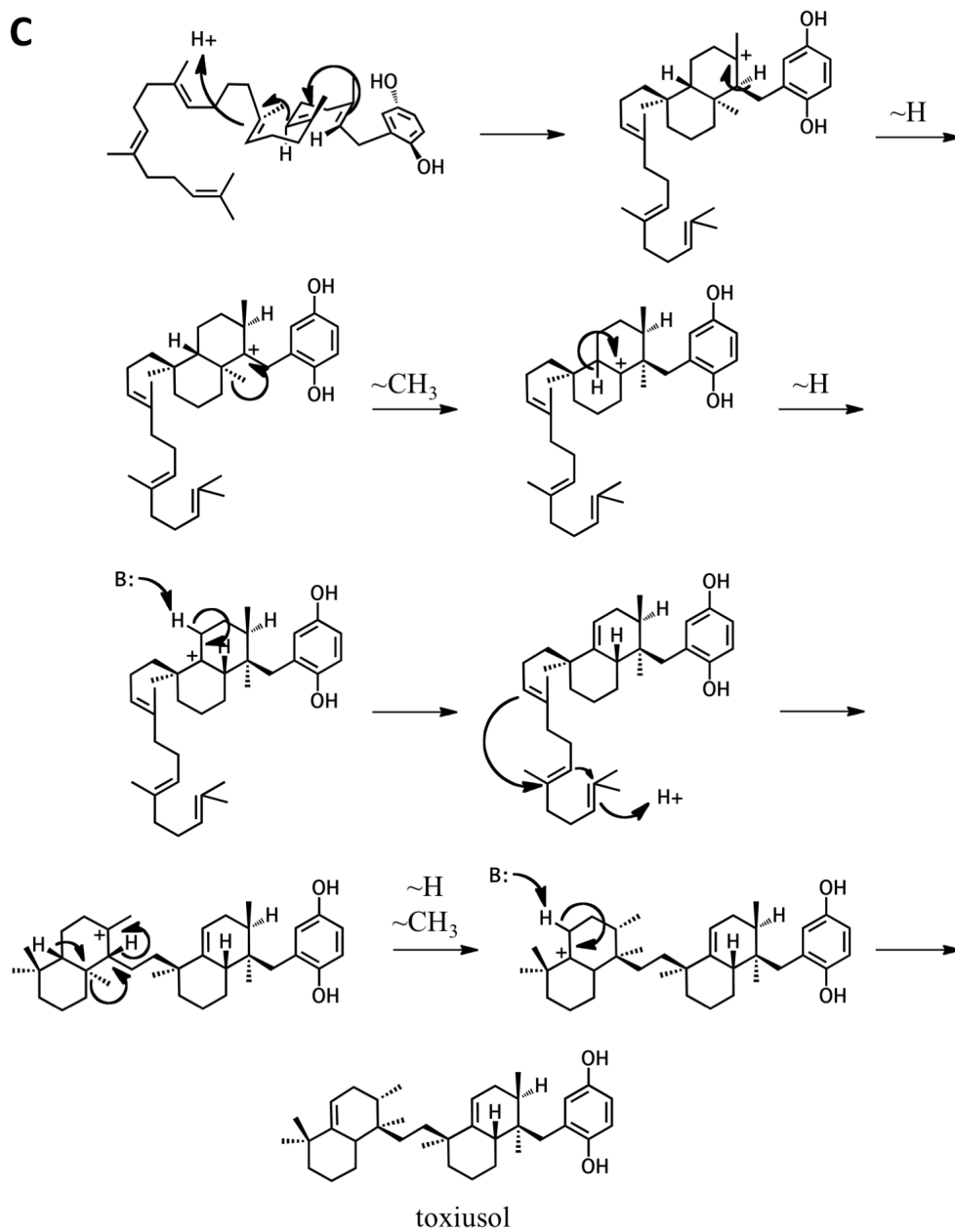
Figure 1.8 Continued
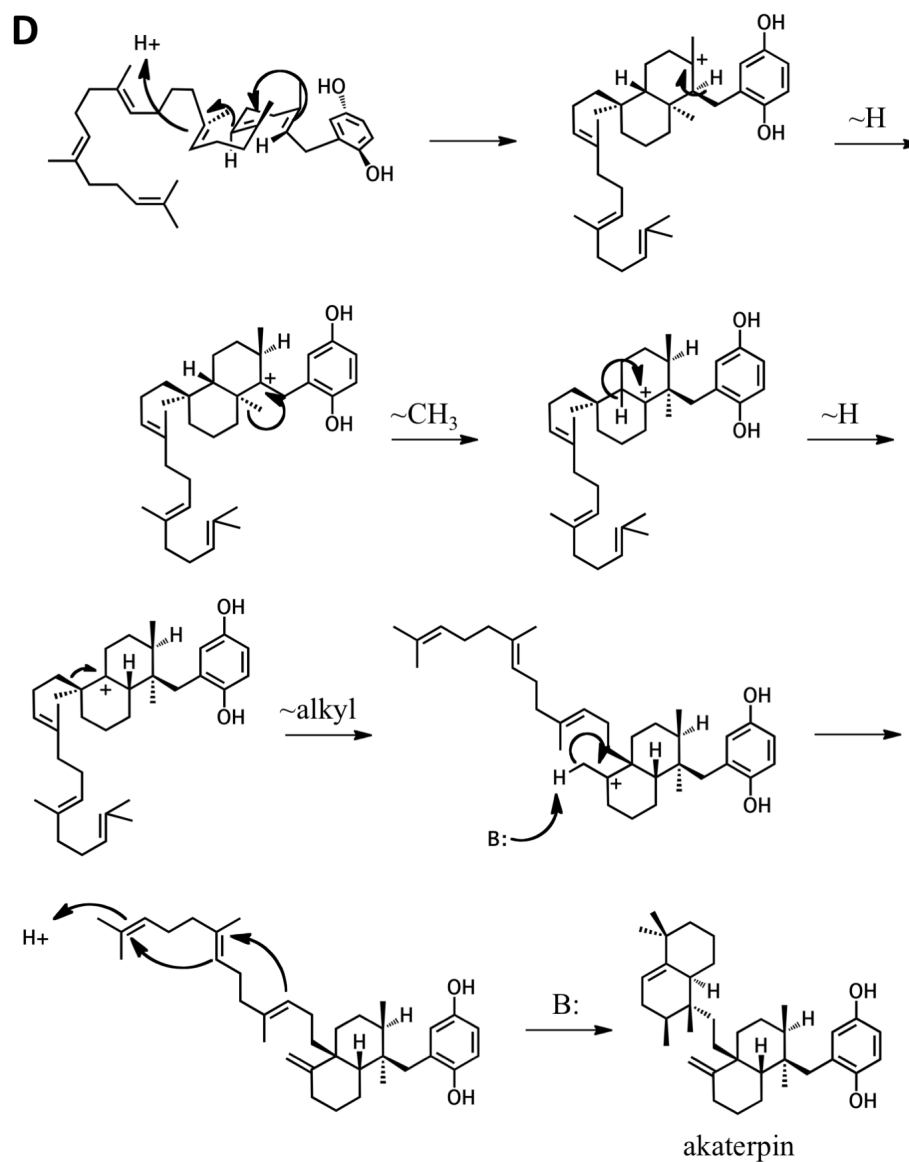
Figure 1.8 Continued
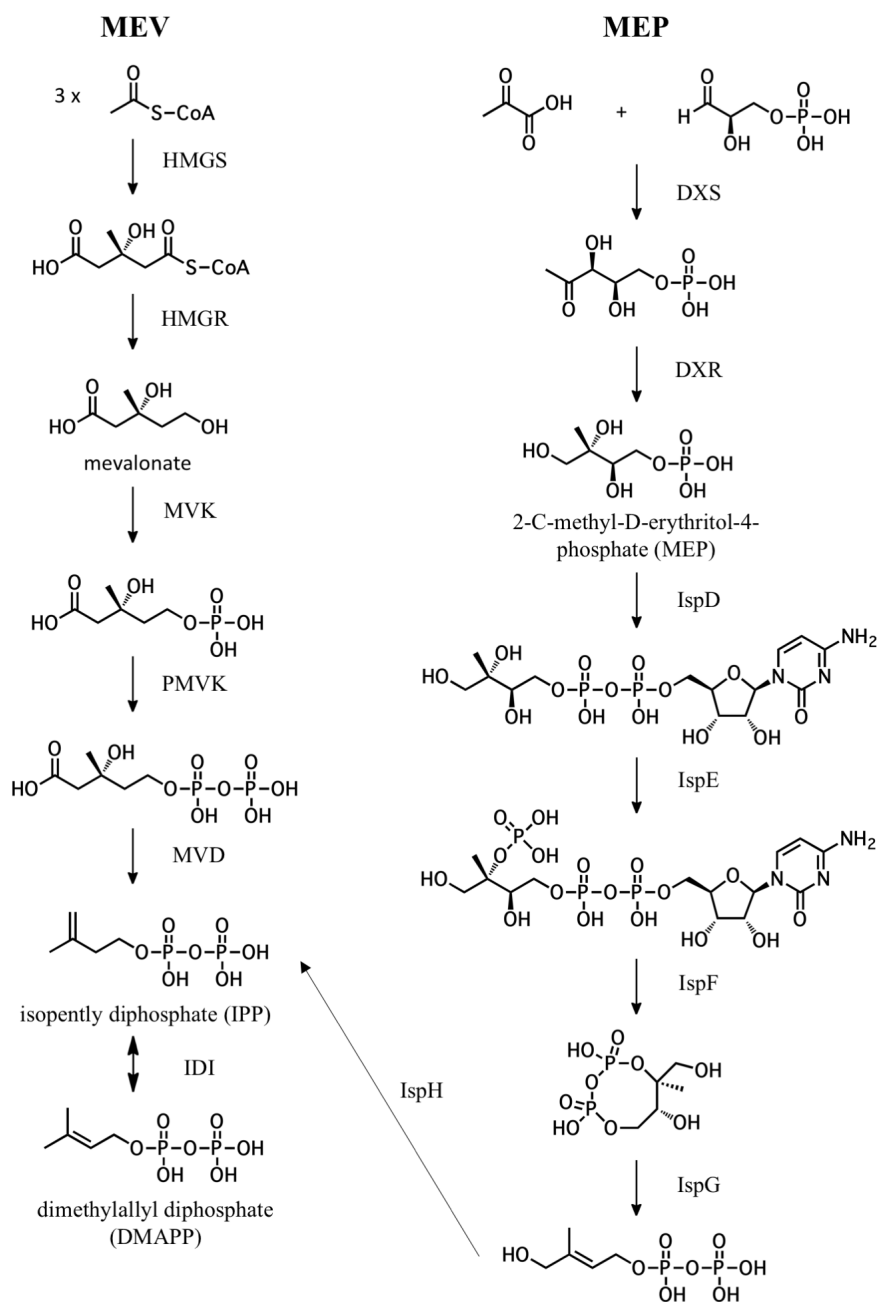
Figure 1.9. Isoprene biosynthesis by the mevalonate (MEV) and non-mevalonate (MEP) pathways. The MEV and MEP pathways are named based on their intermediate metabolites mevalonate and 2-C-methyl-D-erythritol-4-phosphate, respectively. Genes encoding portions of these pathways have been found in bacterial meroterpenoid pathways.
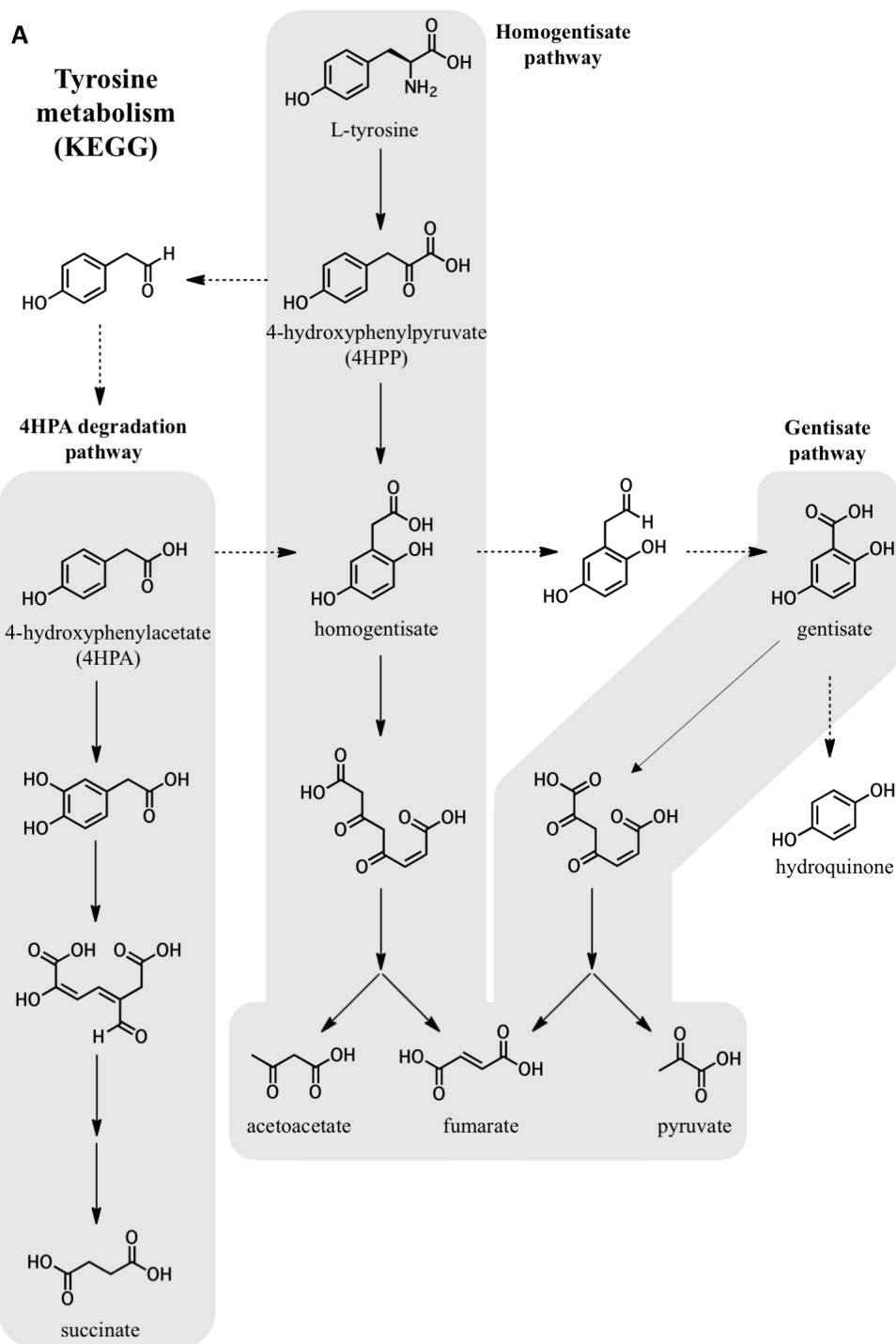
Figure 1.10. Potential biosynthetic origins of the aromatic prenyl acceptor. A) Tyrosine metabolism pathways, as mapped by KEGG. Dashed arrows indicate improbable enzymatic transformations. B) Examples of unique metabolic origins of the aromatic moiety in quinone meroterpenoid natural product pathways.[32, 59] C) The aromatic prenyl acceptor of ubiquinone undergoes several enzymatic transformations after the initial prenylation event.

B

polyketide

PKS   :B   PKS

3,5-dimethylorsellinic acid

carbohydrate

sedo-heptulose-7-phosphate

2-*epi*-3-*epi*-
valinone synthase

alcohol
dehydrogenase,
oxidoreductase

C

UbiA   UbiD/UbiX

UbiG   UbiH

UbiE
UbiF
UbiG

ubiquinone-6

Figure 1.10 Continued

A   **hypothetical bacterial adociasulfate gene cluster**

gene

function
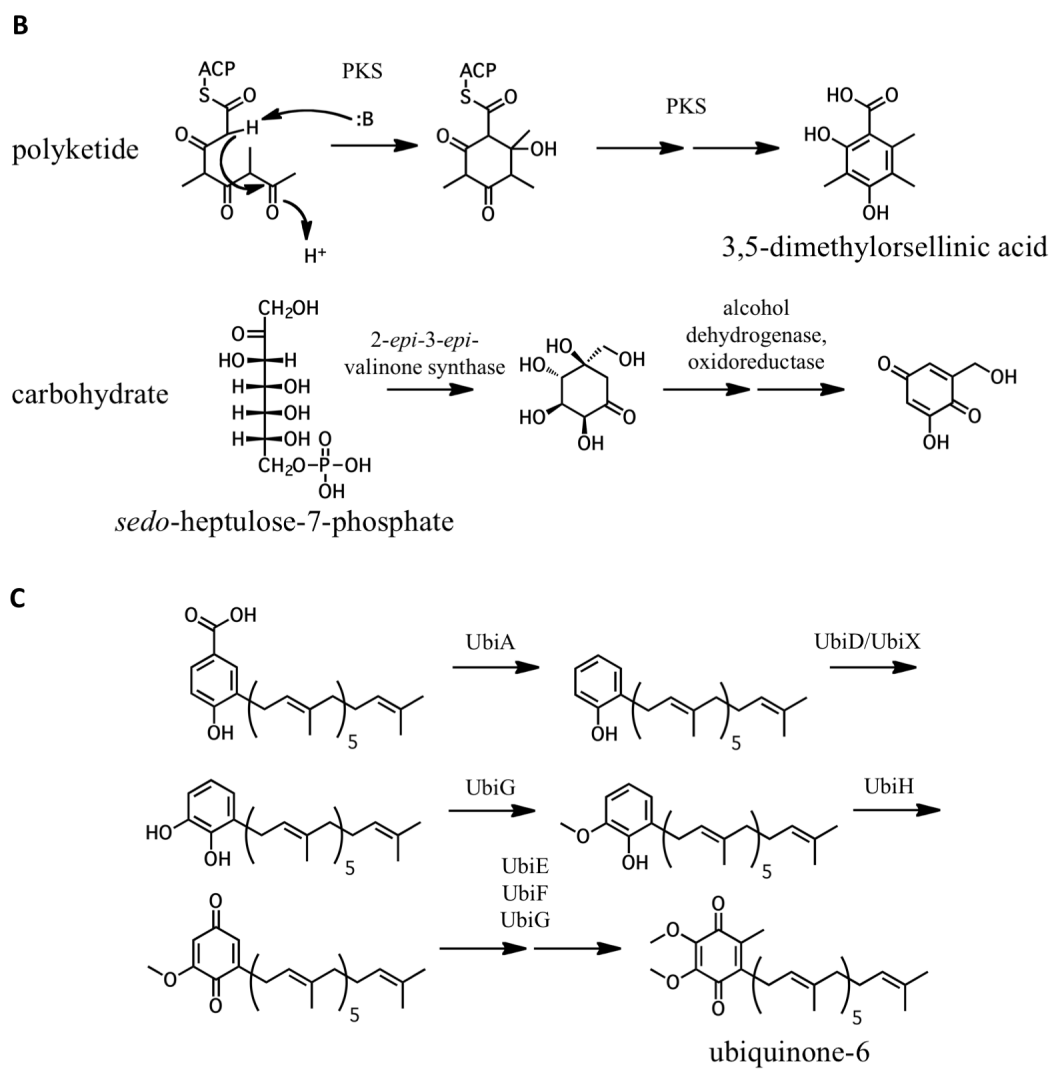
B   **biogenetic hypothesis**
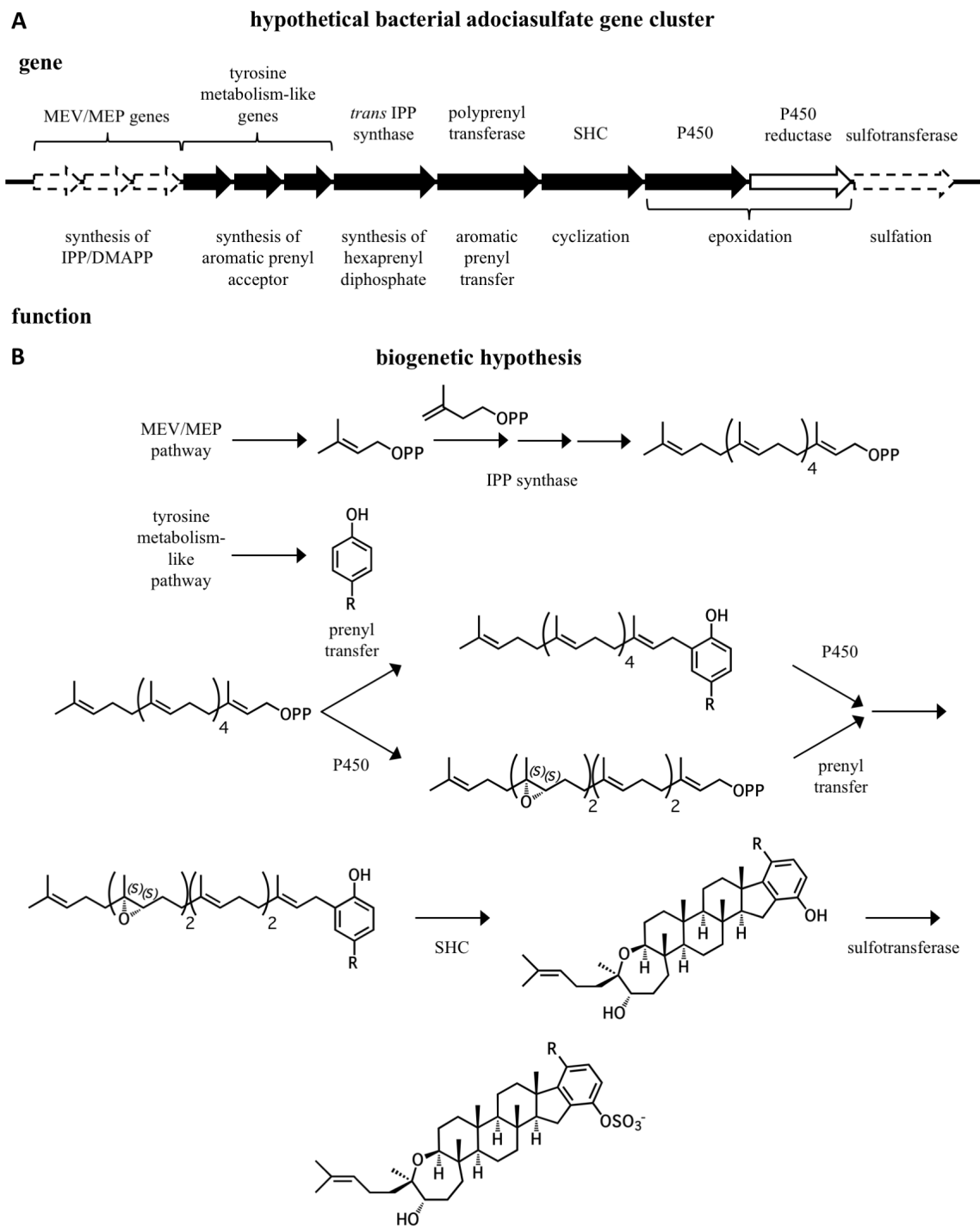
Figure 1.11. A biogenetic hypothesis for the adociasulfates. A) A hypothetical adociasulfate gene cluster was constructed based on the most likely biosynthetic origin, as addressed in this review. In this scenario, the pathway is assumed to be part of a bacterial genome. Black genes represent those directly involved in biosynthesis, white genes are those indirectly involved in biosynthesis, and those bordered with a dashed line have the potential to be entirely absent from the cluster. B) A biosynthetic scheme summarizing the proposed biogenetic hypothesis for the origin of adociasulfates.

1.7 References

1. Hirokawa, N.; Noda, Y.; Tanaka, Y.; Niwa, S. *Nat. Rev. Mol. Cell. Bio.* **2009**, *10*, 682–696.

2. Kapitein, L. C.; Peterman, E. J. G.; Kwok, B. H.; Kim, J. H.; Kapoor, T. M.; Schmidt, C. F. *Nature* **2005**, *435*, 114–118.

3. Cross, R. A. *Trends. Biochem. Sci.* **2004**, *29*, 301–309.

4. Hopkins, S. C.; Vale, R. D.; Kuntz, I. D. *Biochemistry* **2000**, *39*, 2805–2814.

5. Learman, S. S.; Kim, C. D.; Stevens, N. S.; Kim, S.; Wojcik, E. J.; Walker, R. A. *Biochemistry* **2009**, *48*, 1754–1762.

6. Sakowicz, R.; Berdelis, M. S.; Ray, K.; Blackburn, C. L.; Hopmann, C.; Faulkner, D. J.; Goldstein, L. S. B. *Science* **1998**, *280*, 292–295.

7. Prudent, R.; Moucadel, V.; Laudet, B.; Barette, C.; Lafanechere, L.; Hasenknopf, B.; Li, J.; Bareyt, S.; Lacote, E.; Thorimbert, S.; Malacria, M.; Gouzerh, P.; Cochet, C. *Chem. Biol.* **2008**, *15*, 683–692.

8. McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. *J. Med. Chem.* **2002**, *45*, 1712–1722.

9. Drake, D. M.; Pack, D. W. *J. Pharm. Sci.* **2008**, *97*, 1399–1413.

10. Lee, C. M. *Biochim. Biophys. Acta.* **2014**, *1843*, 2027–2036.

11. Narasimhan, K.; Pillay, S.; Bin Ahmad, N. R.; Bikadi, Z.; Hazai, E.; Yan, L.; Kolatkar, P. R.; Pervushin, K.; Jauch, R. *ACS Chemical Biology* **2011**, *6*, 573–581.

12. Reddie, K. G.; Roberts, D. R.; Dore, T. M. *J. Med. Chem.* **2006**, *49*, 4857–4860.

13. Smith, T. E.; Hong, W.; Zachariah, M. M.; Harper, M. K.; Matainaho, T. K.; Van Wagoner, R. M.; Ireland, C. M.; Vershinin, M. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 18880–18885.

14. Blackburn, C. L.; Hopmann, C.; Sakowicz, R.; Berdelis, M. S.; Goldstein, L. S. B.; Faulkner, D. J. *J. Org. Chem.* **1999**, *64*, 5565–5570.

15. Kalaitzis, J. A.; Leone, P.; Harris, L.; Butler, M. S.; Ngo, A.; Hooper, J. N. A.; Quinn, R. J. *J. Org. Chem.* **1999**, *64*, 5571–5574.

16.   Kalaitzis, J. A.; Quinn, R. J. *J. Nat. Prod.* **1999**, *62*, 1682–1684.

17.   Blackburn, C. L.; Faulkner, D. J. *Tetrahedron* **2000**, *56*, 8429–8432.

18.   West, L. M.; Faulkner, D. J. *J. Nat. Prod.* **2006**, *69*, 1001–1004.

19.   Menna, M.; Imperatore, C.; D'Aniello, F.; Aiello, A. *Mar. Drugs* **2013**, *11*, 1602–1643.

20.   Braekman, J. C.; Daloze, D.; Hulot, G.; Tursch, B.; Declercq, J. P.; Germain, G.; van Meerssche, M. *Bull. Soc. Chim. Belg.* **1978**, *87*, 917–926.

21.   Cimino, G.; De Luca, P.; De Stefano, S.; Minale, L. *Tetrahedron* **1975**, *31*, 271–275.

22.   Takizawa, P. A.; Yucel, J. K.; Veit, B.; Faulkner, D. J.; Deermick, T.; Soto, G.; Ellisman, M.; Malhotra, V. *Cell* **1993**, *73*, 1079–1090.

23.   Cichewicz, R. H.; Kenyon, V. A.; Whitman, S.; Morales, N. M.; Arguello, J. F.; Holman, T. R.; Crews, P. *J. Am. Chem. Soc.* **2004**, *126*, 14910–14920.

24.   Williams, D. E.; Steino, A.; de Voogd, N. J.; Mauk, A. G.; Andersen, R. J. *J. Nat. Prod.* **2012**, *75*, 1451–1458.

25.   Isaacs, S.; Hizi, A.; Kashman, Y. *Tetrahedron* **1993**, *49*, 4275–4282.

26.   Crews, P.; Harrison, B. *Tetrahedron* **2000**, *56*, 9039–9046.

27.   Isaacs, S.; Kashman, Y. *Tetrahedron Lett.* **1992**, *33*, 2227–2230.

28.   Fukami, A.; Ikeda, Y.; Kondo, S.; Naganawa, H.; Takeuchi, T.; Furuya, S.; Hirabayashi, Y.; Shimoike, K.; Hosaka, S.; Watanabe, Y.; Umezawa, K. *Tetrahedron Lett.* **1997**, *38*, 1201–1202.

29.   Bonitz, T.; Zubeil, F.; Grond, S.; Heide, L. *PLoS ONE* **2013**, *8*, e85707.

30.   Haagen, Y.; Gluck, K.; Fay, K.; Kammerer, B.; Gust, B.; Heide, L. *Chembiochem* **2006**, *7*, 2016–2027.

31.   Kawasaki, T.; Hayashi, Y.; Kuzuyama, T.; Furihata, K.; Itoh, N.; Seto, H.; Dairi, T. *J. Bacteriol.* **2006**, *188*, 1236–1244.

32.   Kawasaki, T.; Kuzuyama, T.; Furihata, K.; Itoh, N.; Seto, H.; Dairi, T. *J. Antibiot.* **2003**, *56*, 957–966.

33.   Kuzuyama, T.; Noel, J. P.; Richard, S. B. *Nature* **2005**, *435*, 983–987.

34.	McAlpine, J., B.; Banskota, A. H.; Charan, R. D.; Schlingmann, G.; Zazopoulos, E.; Piraee, M.; Janso, J.; Bernan, V. S.; Aouidate, M.; Farnet, C. M.; Feng, X.; Zhao, Z.; Carter, G. T. *J. Nat. Prod.* **2008**, *71*, 1585–1590.

35.	Saikia, S.; Nicholson, M. J.; Young, C.; Parker, E. J.; Scott, B. *Mycol. Res.* **2008**, *112*, 184–199.

36.	Saleh, O.; Gust, B.; Boll, B.; Fiedler, H. P.; Heide, L. *J. Biol. Chem.* **2009**, *284*, 14439–14447.

37.	Winter, J. M.; Moffitt, M. C.; Zazopoulos, E.; McAlpine, J. B.; Dorrestein, P. C.; Moore, B. S. *J. Biol. Chem.* **2007**, *282*, 16362–16368.

38.	Zeyhle, P.; Bauer, J. S.; Kalinowski, J.; Shin-ya, K.; Gross, H.; Heide, L. *PLoS ONE* **2014**, *9*, e99122.

39.	Hillwig, M. L.; Zhu, Q.; Liu, X. *ACS Chemical Biology* **2014**, *9*, 372–377.

40.	Steffensky, M.; Mühlenweg, A.; Wange, Z.-X.; Li, S.-M.; Heide, L. *Antimicrob. Agents Ch.* **2000**, *44*, 1214–1222.

41.	Gevers, D.; Vandepoele, K.; Simillion, C.; Van de Peer, Y. *Trends Microbiol.* **2004**, *12*, 148–154.

42.	Keeling, C. I.; Weisshaar, S.; Lin, R. P. C.; Bohlmann, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1085–1090.

43.	Kliebenstein, D. J. *PLoS ONE* **2008**, *3*, e1838.

44.	Lin, Z.; Torres, J. P.; Tianero, M. D.; Kwan, J. C.; Schmidt, E. W. *Appl. Environ. Microbiol.* **2016**, *82*, 3450–3460.

45.	Heide, L. *Curr. Opin. Chem. Biol.* **2009**, *13*, 171–179.

46.	Arias-Barrau, E.; Olivera, E. R.; Luengo, J. M.; Fernandez, C.; Galan, B.; Garcia, J. L.; Diaz, E.; Minambres, B. *J. Bacteriol.* **2004**, *186*, 5062–5077.

47.	Prieto, M. A.; Díaz, E.; García, J. L. *J. Bacteriol.* **1996**, *178*, 111-120.

48.	Garratini, E.; Fratelli, M.; Terao, M. *Human Genomics* **2009**, *4*, 119–130.

49.	Yoshida, A.; Rzhetsky, A.; Hsu, L. C.; Chang, C. *Eur. J. Biochem.* **1998**, *251*, 549–557.

50.	Holesova, Z.; Jakubkova, M.; Zavadiakova, I.; Zeman, I.; Tomaska, L.; Nosek, J. *Microbiology* **2011**, *157*, 2152–2163.

51. Crouch, N. P.; Adlington, R. M.; Baldwin, J. E.; Lee, M.-H.; MacKinnon, C. H. *Tetrahedron* **1997**, *53*, 6993–7010.

52. Selmer, T.; Andrei, P. I. *Eur. J. Biochem.* **2001**, *268*, 1363–1372.

53. Prescott, A. G.; Lloyd, M. D. *Nat. Prod. Rep.* **2000**, *17*, 367–383.

54. Urlacher, V. B.; Girhard, M. *Trends. Biotechnol.* **2012**, *30*, 26–36.

55. Liu, A.; Zhang, H. *Biochemistry* **2006**, *45*, 10408–10411.

56. Peng, X.; Masai, E.; Kitayama, H.; Harada, K.; Katayama, Y.; Fukuda, M. *Appl. Environ. Microbiol.* **2002**, *68*, 4407–4415.

57. Yoshida, M.; Fukuhara, N.; Oikawa, T. *J. Bacteriol.* **2004**, *186*, 6855–6863.

58. Huijbers, M. M.; Montersino, S.; Westphal, A. H.; Tischler, D.; van Berkel, W. J. *Arch. Biochem. Biophys.* **2014**, *544*, 2–17.

59. Itoh, T.; Tokunaga, K.; Matsuda, Y.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T. *Nat. Chem.* **2010**, *2*, 858–864.

60. Itoh, T.; Tokunaga, K.; Radhakrishnan, E. K.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T. *Chembiochem* **2012**, *13*, 1132–1135.

61. Lo, H. C.; Entwistle, R.; Guo, C. J.; Ahuja, M.; Szewczyk, E.; Hung, J. H.; Chiang, Y. M.; Oakley, B. R.; Wang, C. C. *J. Am. Chem. Soc.* **2012**, *134*, 4709–4720.

62. Matsuda, Y.; Awakawa, T.; Abe, I. *Tetrahedron* **2013**, *69*, 8199–8204.

63. Matsuda, Y.; Wakimoto, T.; Mori, T.; Awakawa, T.; Abe, I. *J. Am. Chem. Soc.* **2014**, *136*, 15326–15336.

64. Pojer, F.; Li, S.-M.; Heide, L. *Microbiology* **2002**, *148*, 3901–3911.

65. Melzer, M.; Heide, L. *Biochim. Biophys. Acta* **1994**, *1212*, 93–102.

66. Norris, S. R.; Barrette, T. R.; DellaPenna, D. *The Plant Cell* **1995**, *7*, 2139–2149.

67. Young, L. G.; Leppik, R. A.; Hamilton, J. A.; Gibson, F. *J. Bacteriol.* **1972**, *110*, 18–25.

68. Bräuer, L.; Brandt, W.; Wessjohann, L. A. *J. Mol. Model.* **2004**, *10*, 317–327.

69.	Huang, H.; Levin, E. J.; Liu, S.; Bai, Y.; Lockless, S. W.; Zhou, M. *PLoS Biol.* **2014**, *12*, e1001911.

70.	Cheng, W.; Weikai, L. *Science* **2014**, *343*, 878–881.

71.	El Hachimi, Z.; Samuel, O.; Azerad, R. ***Biochimie*** 1974, *56*, 1239–1247.

72.	Okada, K.; Suzuki, K.; Kamiya, Y.; Zhu, X.; Fujisaki, S.; Nishimura, Y.; Nishino, T.; Nakagawa, T.; Kawamukai, M.; Matsuda, H. *Biochim. Biophys. Acta* **1996**, *1302*, 217–223.

73.	Wessjohann, L.; Sontag, B. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1697–1699.

74.	Li, W. *Trends. Biochem. Sci.* **2016**, *41*, 356–370.

75.	Saleh, O.; Haagen, Y.; Seeger, K.; Heide, L. *Phytochemistry* **2009**, *70*, 1728–1738.

76.	Tello, M.; Kuzuyama, T.; Heide, L.; Noel, J. P.; Richard, S. B. *Cell. Mol. Life Sci.* **2008**, *65*, 1459–1463.

77.	Pojer, F.; Wemakor, E.; Kammerer, B.; Chen, H.; Walsh, C. T.; Li, S.-M.; Heide, L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2316–2321.

78.	Haug-Schifferdecker, E.; Arican, D.; Bruckner, R.; Heide, L. *J. Biol. Chem.* **2010**, *285*, 16487–16494.

79.	Kumano, T.; Richard, S. B.; Noel, J. P.; Nishiyama, M.; Kuzuyama, T. *Bioorg. Med. Chem.* **2008**, *16*, 8117–8126.

80.	Liu, X.; Hillwig, M. L.; Koharudin, L. M.; Gronenborn, A. M. *Chemical Communications* **2016**, *52*, 1737–1740.

81.	McIntosh, J. A.; Donia, M. S.; Nair, S. K.; Schmidt, E. W. *Journal of the American Chemical Society* **2011**, *133*, 13698–13705.

82.	Bonitz, T.; Alva, V.; Saleh, O.; Lupas, A. N.; Heide, L. *PLoS ONE* **2011**, *6*, e27336.

83.	Steffan, N.; Grundmann, A.; Yin, W.-B.; Kremer, A.; Li, S.-M. *Curr. Med. Chem.* **2009**, *16*, 218–231.

84.	Yu, X.; ZXie, X.; Li, S.-M. *Appl. Microbiol. Biotech.* **2011**, *92*, 737–748.

85.	Pockrandt, D.; Sack, C.; Kosiol, T.; Li, S.-M. *Appl. Microbiol. Biotech.* **2014**, *98*, 4987–4994.

86.   Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev.* **1993**, *93*, 2189–2208.

87.   Abe, I.; Tanaka, H.; Noguchi, H. *J. Am. Chem. Soc.* **2002**, *124*, 14514–14515.

88.   Hammer, S. C.; Dominicus, J. M.; Syrén, P.-O.; Nestl, B. M.; Hauer, B. *Tetrahedron* **2012**, *68*, 7624–7629.

89.   Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Kolesnikova, M. D.; Onak, C. S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2007**, *129*, 11213–11222.

90.   Criswell, J.; Potter, K.; Shephard, F.; Beale, M. H.; Peters, R. J. *Organic Letters* **2012**, *14*, 5828–5831.

91.   Dang, T.; Prestwich, G. D. *Chem. Biol.* **2000**, *7*, 643–649.

92.   Abe, I.; Rohmer, M. *J. Chem. Soc. Perk. T 1* **1994**, *7*, 783–791.

93.   Tanaka, H.; Noguchi, H.; Abe, I. *Organic Letters* **2005**, *7*, 5873–5876.

94.   Xiong, Q.; Zhu, X.; Wilson, W. K.; Ganesan, A.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2003**, *125*, 9002–9003.

95.   Yonemura, Y.; Ohyama, T.; Hoshino, T. *Org. Biomol. Chem.* **2012**, *10*, 440–446.

96.   Boutaud, O.; Dolis, D.; Schuber, F. *Biochem. Bioph. Res. Co.* **1992**, *188*, 898–904.

97.   Shan, H.; Segura, M. J. R.; Wilson, W. K.; Lodeiro, S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 18008–18009.

98.   Abe, T.; Hoshino, T. *Org. Biomol. Chem.* **2005**, *3*, 3127–3139.

99.   Dairi, T.; Hamano, Y.; Kuzuyama, T.; Itoh, N.; Furihata, K.; Seto, H. *J. Bacteriol.* **2001**, *183*, 6085–6094.

100.   Kwon, M.; Cochrane, S. A.; Vederas, J. C.; Ro, D. K. *FEBS Lett.* **2014**, *588*, 4597–4603.

101.   Morrone, D.; Chambers, J.; Lowry, L.; Kim, G.; Anterola, A.; Bender, K.; Peters, R. J. *FEBS Lett.* **2009**, *583*, 475–480.

102.   Simpson, T. J.; Ahmed, S. A.; McIntyre, C. R.; Scott, F. E.; Sadler, I. H. *Tetrahedron* **1997**, *53*, 4013–4034.

103.   Chooi, Y. H.; Hong, Y. J.; Cacho, R. A.; Tantillo, D. J.; Tang, Y. *J. Am. Chem. Soc.* **2013**, *135*, 16805–16808.

104. Zhao, B.; Lei, L.; Vassylyev, D. G.; Lin, X.; Cane, D. E.; Kelly, S. L.; Yuan, H.; Lamb, D. C.; Waterman, M. R. *J. Biol. Chem.* **2009**, *284*, 36711–36719.

105. Laden, B. P.; Tang, Y.; Porter, T. D. *Arch. Biochem. Biophys.* **2000**, *374*, 381–388.

106. Li, L.; Porter, T. D. *Arch. Biochem. Biophys.* **2007**, *461*, 76–84.

107. Dürr, C.; Schnell, H. J.; Luzhetskyy, A.; Murillo, R.; Weber, M.; Welzel, K.; Vente, A.; Bechthold, A. *Chem. Biol.* **2006**, *13*, 365–377.

108. Ingelman-Sundberg, M. *Naunyn Schmiedeberg's Arch. Pharmacol.* **2004**, *369*, 89–104.

109. Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R. *Biochim. Biophys. Acta* **2007**, *1770*, 330–344.

110. Gamage, N.; Barnett, A.; Hempel, N.; Duggleby, R. G.; Windmill, K. F.; Martin, J. L.; McManus, M. E. *Toxicol. Sci.* **2006,** *90*, 5–22.

111. Chapman, E.; Best, M. D.; Hanson, S. R.; Wong, C. H. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 3526–3548.

112. Kim, D.-H.; Kobashi, K. *Biochem. Pharmacol.* **1986**, *35*, 3507–3510.

113. Hanin, M.; Jabbouri, S.; Quesada-Vincens, D.; Freiberg, C.; Perret, X.; Promé, J.-C.; Broughton, W. J.; Fellay, R. *Mol. Microbiol.* **1997**, *24*, 1119–1129.

114. Mougous, J. D.; Green, R. E.; Williams, S. J.; Brenner, S. E.; Bertozzi, C. R. *Chem. Biol.* **2002**, *9*, 767–776.

115. McCarthy, J. G.; Eisman, E. B.; Kulkarni, S.; Gerwick, L.; Gerwick, W. H.; Wipf, P.; Sherman, D. H.; Smith, J. L. *ACS Chemical Biology* **2012**, *7*, 1994–2003.

116. Kalan, L.; Perry, J.; Koteva, K.; Thaker, M.; Wright, G. *J. Bacteriol.* **2013**, *195*, 167–171.

117. Malojcic, G.; Owen, R. L.; Grimshaw, J. P. A.; Brozzo, M. S.; Dreher-Teo, H.; CGlockshuber, R. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19217–19222.

118. Grimshaw, J. P.; Stirnimann, C. U.; Brozzo, M. S.; Malojcic, G.; Grutter, M. G.; Capitani, G.; Glockshuber, R. *J. Mol. Biol.* **2008**, *380*, 667–680.

119. Tang, X.; Eitel, K.; Kaysser, L.; Kulik, A.; Grond, S.; Gust, B. *Nat. Chem. Biol.* **2013**, *9*, 610–605.

120.   Bogenstätter, M.; Limberg, A.; Overman, L. E.; Tomasi, A. L. *J. Am. Chem. Soc.* **1999**, *121*, 12206–12207.

121.   Darne, C. P., M.S. Thesis, The University of Georgia, May 2005.

122.   Erben, F.; Specowius, V.; Wölfling, J.; Schneider, G.; Langer, P. *Helv. Chim. Acta* **2013**, *96*, 924–930.

123.   Müller, W. E. G.; Böhm, M.; Batel, R.; De Rosa, S.; Tommonaro, G.; Müller, I. M.; Schröder, H. C. *J. Nat. Prod.* **2000**, *63*, 1077–1081.

124.   Uriz, M. J.; Turon, X.; FGalera, J.; Tur, J. M. *Cell. Tissue Res.* **1996**, *285*, 519–527.

CHAPTER 2

SINGLE-MOLECULE INHIBITION OF HUMAN KINESIN BY

ADOCIASULFATE-13 AND -14 FROM THE SPONGE *Cladocroce aculeata*

Single-molecule inhibition of human kinesin by adociasulfate-13 and -14 from the sponge *Cladocroce aculeata*

Thomas E. Smith[a,1], Weili Hong[b,1], Malcolm M. Zachariah[a], Mary Kay Harper[a], Teatulohi K. Matainaho[c], Ryan M. Van Wagoner[a], Chris M. Ireland[a,2], and Michael Vershinin[d,2]

Departments of [a]Medicinal Chemistry and [b]Physics and Astronomy and [d]Center for Cell and Genome Science, University of Utah, Salt Lake City, UT 84112; and [c]Discipline of Pharmacology, School of Medicine and Health Sciences, University of Papua New Guinea, National Capital District 111, Papua New Guinea

Two merotriterpenoid hydroquinone sulfates designated adociasulfate-13 (1) and adociasulfate-14 (2) were purified from *Cladocroce aculeata* (Chalinidae) along with adociasulfate-8. All three compounds were found to inhibit microtubule-stimulated ATPase activity of kinesin at 15 μM by blocking both the binding of microtubules and the processive motion of kinesin along microtubules. These findings directly show that substitution of the 5′-sulfate in 1 for a glycolic acid moiety in 2 maintains kinesin inhibition. Nomarski imaging and bead diffusion assays in the presence of adociasulfates showed no signs of either free-floating or bead-bound adociasulfate aggregates. Single-molecule biophysical experiments also suggest that inhibition of kinesin activity does not involve adociasulfate aggregation. Furthermore, both mitotic and nonmitotic kinesins are inhibited by adociasulfates to a significantly different extent. We also report evidence that microtubule binding of nonkinesin microtubule binding domains may be affected by adociasulfates.

single-molecule biophysics | natural products | mechanism of action | microtubule-based motors | terpenes

Kinesin motor proteins are implicated in several vital eukaryotic cellular processes, including vesicle transport (1) and mitosis (2). They are composed of three distinct domains: a motor domain that both hydrolyzes ATP and steps along microtubules (MTs), a linker domain involved in dimerization, and a cargo binding tail domain. Inhibitors of these enzymes can provide key information concerning the mechanism of coupling ATP hydrolysis to the characteristic stepping motion of kinesins. In addition, kinesins control cellular functions that are often implicated in disease. Hence, kinesin inhibitors are highly sought. Of the known kinesin inhibitors, monastrol (3), terpendole E (4), HR22C16 (5), CK0106023 (6), S-trityl-L-cysteine (7), and the dihydropyrazoles (8), including ispinesib (9), inhibit ATPase activity of Eg5 allosterically, allowing ATP binding but preventing ADP release. Rose bengal lactone (RBL) inhibits MT-stimulated ATPase activity of kinesin (10). Thiazole inhibitors compete with ATP directly to inhibit Eg5 (11), whereas biaryl compounds GSK-1 and GSK-2 block interactions with nucleotides through an allosteric binding site (12). Adociasulfates are unique in that they are the only kinesin inhibitors with mechanisms of action that involve competition for binding to MTs (13, 14). Thus, they have the potential to be used as probes for kinesin functions unaffected by other inhibitors or drugs that target these functions specifically.

Adociasulfates are a subfamily of sulfated triterpenoid hydroquinone compounds derived from marine sponges of the family Chalinidae. They have received attention for their inhibitory effect on kinesin family motors and H⁺-ATPase proton pump enzymes, where their activity has been linked to the presence of at least one sulfate group (15). Much of what is known about adociasulfate activity comes from studies of adociasulfate-2 [AS-2 (**4**)]. The compound is known to bind to kinesin and interfere with MT binding with minor effects on nucleotide interactions (13, 14). The idea that adociasulfates are 1:1 kinesin inhibitors

has been questioned in a recent study suggesting that **4** forms extended aggregates that mimic the negatively charged microtubule surface and thereby, inhibit kinesin activity (16). The mode of binding to kinesin is a critical question for future drug development. Specific inhibition would make these compounds more suitable to target a small class of enzymes, whereas aggregation would make practical in vivo applications problematic. At the same time, aggregations of small molecules could be potentially interesting in nanobiological engineering, where artificial microtubule tracks are highly desirable.

We report the discovery of two previously undescribed adociasulfates, which we designate adociasulfate-13 (**1**) and adociasulfate-14 (**2**), isolated from the marine sponge *Cladocroce aculeata* Pulitzer-Finali, 1982 (Fig. 1). Adociasulfate-8 (**3**) was also isolated from the same organism. We have used single-molecule biophysical measurements to assess the activity of these compounds against kinesin-1 and -5 family motor proteins. We also assessed whether the inhibitory activity of these compounds is connected to the formation of extended adociasulfate aggregates. Our results suggest that compounds **1**, **2**, and **3** inhibit both binding of kinesin to MTs and processive motion. The replacement of the 5′-sulfate with a glycolic acid moiety has little effect on kinesin inhibition, indicating that alternative functional groups may be capable of maintaining activity. We note that this observation significantly

---

Significance

Kinesin motor proteins are central to cellular processes and considered good drug targets, but very few reported kinesin inhibitors exhibit potential as drugs. Adociasulfates uniquely inhibit kinesins by competing with microtubules for binding. A declining interest in these compounds resulted from a report of large aggregates of adociasulfate-2 responsible for kinesin inhibition, poor cell permeability, and broad kinesin inhibition, limiting potential therapeutic applications. In this study, we show that kinesin inhibition is likely a 1:1 interaction and does not involve aggregates. We suggest a means by which cell permeability may be improved and show that adociasulfates inhibit kinesin-1 and -5 families of motors to a significantly different extent. These results collectively bring adociasulfates back to the foreground of chemical biology.

---

enhances the value of adociasulfates as a target for future drug development. We corroborate and extend the previously proposed theory that adociasulfates act as MT mimics, but we find no evidence that large aggregates of individual adociasulfate molecules could be responsible for the observed activity. We also provide evidence that adociasulfates inhibit kinesins at the single-molecule level.

Results

Structure Elucidation of AS-13 and AS-14. High-resolution mass measurement of the [M-H]$^-$ ion of AS-13 (**1**) resulted in a predicted chemical formula of $C_{36}H_{54}O_{10}S_2$ with 10 double-bond equivalents. A loss of 80 Da in the negative ion mode suggested the presence of a sulfate group (Fig. S1A). The carbon skeleton of AS-13 (**1**) is identical to AS-9 (**5**), which was determined from ^1H, ^{13}C, and heteronuclear multiple bond correlation (HMBC) NMR data (Figs. S2 and S3D, summarized in Table S1) (17). Rotating-frame nuclear Overhauser effect correlation spectroscopy (ROESY) correlations confirmed that the configurations of all methyls, ring junction methines, and hydroxymethines match those configurations previously reported (Fig. S3F and Table S1). A quaternary carbon presumed to be C9 was not observed by ^{13}C or HMBC NMR experiments. A ^{13}C spectrum recorded in DMSO-d_6 revealed all 36 carbons (Fig. S2C). Taken together, these data are consistent with **1** being the disulfated version of **5**.

Adociasulfate-14 (**2**) has a molecular formula of $C_{38}H_{56}O_9S$ with 11 double-bond equivalents as determined from the high-resolution mass of the [M-H]$^-$ ion, indicating the addition of two carbons relative to **1** and **3** (Fig. S1B). The ^1H NMR spectrum recorded in DMSO-d_6 identified an additional oxymethine relative to **1** that was significantly more deshielded than the other two (δ_H = 4.78, 3.58, and 3.44 ppm for **2**; δ_H = 3.76 and 3.51 for **1**; Fig. S4A). A carbonyl carbon (δ_C = 175.7 ppm) and an additional heteroatom-bound carbon were also identified by ^{13}C NMR (Fig.

S3B). The carbon skeleton and stereocenter configurations were shown to be similar to **1** by HMBC and ROESY correlations (Fig. S5 D and F, summarized in Table S2). The configuration of H-6 could not be confidently assigned because of the presence of ROESY correlations from methyl H$_3$-26 to both H-6 and 6-OH. The splitting of H-6 is broad in DMSO-d_6 (Fig. S4A), but it is a triplet with J = 6.9 Hz in MeOH-d_4 (Fig. S4B), suggesting that multiple conformations may exist in which H-6 is either pseudoaxial or -equatorial. The position of the glycolic acid group on the aromatic ring relative to the sulfate group was confirmed from ROESY correlations between H-7′ and both H-20$_{ax}$ and H-20$_{eq}$. The configuration of the stereocenter at position 7′ was not determined. Two methylene protons at positions 15 and 19 were not observed throughout all NMR experiments, likely because of overlapping chemical shifts. Collectively, these data suggest that **2** is identical to **1** with one sulfate replaced by a glycolic acid moiety. Adociasulfate-10 (**6**) is the only other adociasulfate reported with such a feature (15).

Adociasulfate-8 (**3**) was identified from the high-resolution mass of the [M-H]$^-$ ion (m/z = 695.3300; Δ = −1.0 ppm) predicting a chemical formula of $C_{36}H_{55}S_2O_9$ (Fig. S1C) and comparison of NMR data with reported chemical shift values and HMBC/ROESY correlations in DMSO-d_6 (18). A ^1H NMR spectrum is provided in the supplementary material to indicate sample purity (Fig. S6).

Adociasulfates Are Potent Inhibitors of Kinesin. The adociasulfates reported here share significant structural similarity with AS-2 (**4**), AS-10 (**6**), and AS-6 (**7**), which are known to be strong inhibitors of MT-stimulated activity of kinesin ATPases. We used a bead motility assay to test whether AS-13 (**1**), AS-14 (**2**), and AS-8 (**3**) also showed this inhibitory activity. Beads preincubated with increasing amounts of kinesin heavy chain isoform 5A (KIF5A) motors showed increased binding to MTs (Fig.
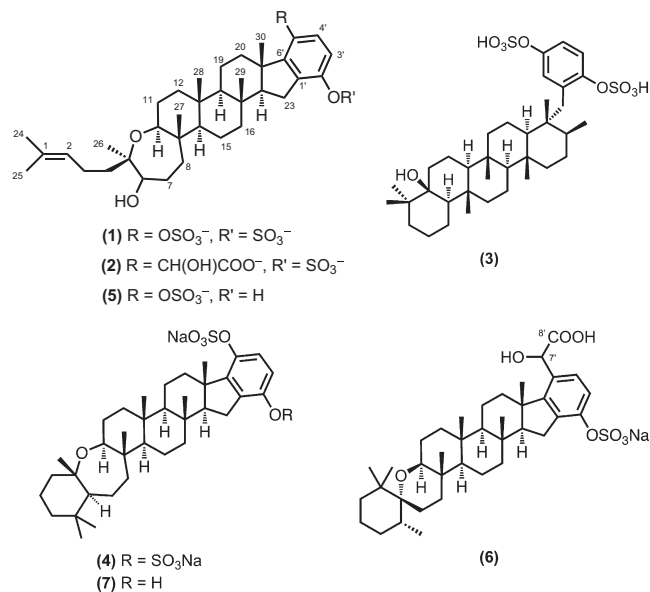
(1) R = OSO$_3^-$, R' = SO$_3^-$
(2) R = CH(OH)COO$^-$, R' = SO$_3^-$
(5) R = OSO$_3^-$, R' = H

(3)

(4) R = SO$_3$Na
(7) R = H

(6)

Fig. 1. Chemical structures of adociasulfates.

Fig. 2. The effect of AS-8, -13, and -14 on kinesin–MT binding. (*A*) Beads preincubated with different amounts of kinesin-I motors were tested for MT binding activity. All binding fraction data sets were well fit by single-molecule Poisson distribution $P(n) = 1 - \exp(-n/b)$ (solid lines; n, relative motor concentration; b, effective binding affinity). Estimated binding affinities were $3.5e-6 \pm 0.3e-6$, $4.3e-6 \pm 0.3e-6$, $5.1e-5 \pm 0.9e-5$, $5.5e-5 \pm 0.7e-5$, and $1.1e-4 \pm 0.2e-4$ for KIF5A alone, KIF5A + cholesterol, KIF5A + AS-8 (**3**), KIF5A + AS-13 (**1**), and KIF5A + AS-14 (**2**), respectively. Error bars: 95% confidence interval for binomial distribution. (*B*) Mean squared displacement curves for floating KIF5A beads preincubated with **1**, without **1**, and with short MTs in the absence of ATP (*n* = 9, 10, and 17, respectively). Computed mean squared displacement yielded excellent linear fits (solid lines; all adjusted R square > 0.9995; error bars = SEM) indicative of pure diffusion. The MT assay mean squared displacement curve indicates lower bead diffusion than for the other two assays.

2). The addition of 15 µM adociasulfates dramatically reduced bead–MT binding. Beads that bound nearly universally to MTs in the absence of adociasulfates showed negligible binding in the presence of **1**, **2**, or **3**. MT binding could be rescued by incubating the beads with higher concentrations of kinesin. Compound **2** showed a pattern of kinesin–MT binding inhibition similar to the observations for **3**, whereas **1** showed the strongest inhibitory activity (Fig. 2 and Fig. S7). The half maximal inhibitory concentration of **4**, **6**, and **7** varies somewhat with tubulin concentration but has been reported to be within the 2–6 µM range, with near complete inhibition of ATPase activity at 15 µM (14, 15, 19). Therefore, **1**, **2**, and **3** seem to have comparable inhibition with **4**, **6**, and **7**. No significant effect on kinesin–MT binding was

seen in the presence of 15 µM cholesterol (Fig. 2), which shares many structural features with **1** and **2** but lacks the sulfated aromatic ring.

The three adociasulfates examined were also observed to severely inhibit processive motion of kinesin on MTs. In assays where **1**, **2**, or **3** was present and the kinesin–MT bead binding fraction was below 30% (corresponding to a regime where only a single motor is likely to be active), kinesin stepping along MTs was below our detection: less than 40–50 nm (five to six steps) if at all present. Note that bound beads underwent constrained Brownian motion that made it difficult to observe extremely short and fast processive motion events. Single-kinesin processive motion was readily observable in the absence of adociasulfates.

Large Aggregates of Adociasulfates Are Not Required for Kinesin Inhibition. It is presently unclear whether adociasulfates are 1:1 inhibitors of kinesin or whether the binding to MTs is inhibited by extended adociasulfate aggregates. Large aggregates of **4** rather than individual molecules have been previously reported to be responsible for kinesin inhibition (16). We attempted to detect such aggregates in our assays and elucidate the mode by which **1**, **2**, and **3** inhibit kinesin. Aggregates extending 200 nm across, which were described by Reddie et al. (16), are well within the resolution of our imaging, which allows us to clearly resolve individual microtubules (~25-nm diameter). We observed no evidence of adociasulfates forming extended or globular structures in our assays (Fig. S8*A*) and found no evidence of adociasulfate aggregation in turbidity assays (Fig. S8*C*). Additionally, bead motion displayed comparable Brownian character between adociasulfate and control assays, suggesting that no extended objects were attached to the bead-bound motors (Fig. 2*B*). Finally, we observed no elevated bead-to-bead binding in adociasulfate assays. Such cross-linking would be expected if beads bound by multiple motors were interacting with extended structures in solution. Reddie et al. (16) reported that activity of **4** changed with longer adociasulfate incubation times. Our results were not altered when preincubation times with adociasulfates were extended to 10 min.

Adociasulfates Show Broad Microtubule Mimic Activity. We assessed whether adociasulfates inhibit MT binding for nonkinesin MT binding domains by examining the MT binding of two mitotic kinesins: Eg5 and BimC (Fig. 3). Both proteins belong to the kinesin-5 family of motor proteins. BimC possesses a unique N-terminal domain rich in positive residues and vaguely homologous to the MT binding region of microtubule-associated protein 2 (MAP2) (20). This domain is known to dramatically increase BimC binding to MTs relative to motor proteins lacking such a domain (20). Note that MT binding domains of kinesins have been directly implicated as the sites of adociasulfate binding and that Eg5 and BimC possess highly homologous MT binding regions within the motor domains (Fig. 3*A*) (13, 14). Overall, inhibition is significantly less pronounced for mitotic kinesins versus kinesin-1 motors. This observation is consistent with a previous report that IC$_{50}$ values of AS-2 (**4**) for kinesin-1 and centromere protein E (CENP-E) motors differ by as much as a factor of five (14). Therefore, although adociasulfates inhibit kinesins across many motor families (13, 14), they may be useful at low micromolar concentrations as specific inhibitors of some nonmitotic kinesins. Additional work is needed to investigate how different classes of kinesins are affected by adociasulfates.

Discussion

The increasing diversity of known adociasulfate inhibitors of kinesin allows us to speculate about key structural features required for binding to kinesin. Knowledge of the structure activity relationship (SAR) of adociasulfates is limited by incomplete information on the kinesin activity of all known family members.
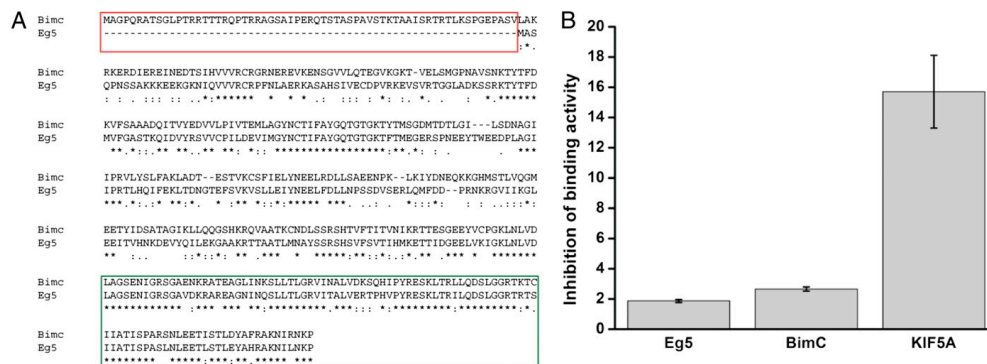
Fig. 3. Eg5 and BimC inhibition by AS-13. (*A*) The motor domains of BimC and Eg5 are highly homologous, particularly the region containing their MT binding sites (green box). BimC has a unique N-terminal domain that has independent MT binding activity (red box) (20). Clustal Omega (European Molecular Biology Laboratory-European Bioinformatics Institute) was used for alignment. Clustal Omega consensus symbols are shown to indicate degree of residue similarity (28). (*B*) Binding of Eg5, BimC, and KIF5A is inhibited by AS-13 (**1**). The ratio of binding affinities was directly estimated in parallel assays with and without **1** (1.87 ± 0.10 for Eg5 and 2.66 ± 0.14 for BimC; three independent measurements for Eg5 and BimC). The relative KIF5A–MT affinities (as well as corresponding error bars) were computed as described in Fig. 2. Higher ratio indicates stronger binding inhibition by **1**.

It is clear, however, that the 2′ sulfate does not significantly contribute to activity (19). In addition, substitution of the 5′ sulfate for a glycolic acid moiety in **6** does not significantly alter kinesin inhibition compared with **4** (15). The similarity between the activities of **1** and **2** is in agreement with this observation and is more convincing because these two compounds are otherwise identical. This result suggests that an α-hydroxyacid is a sufficient bioisostere for the 5′ sulfate, perhaps because of the negative charge shared by the two functional groups, raising the possibility that functional groups other than sulfates may confer bioactivity. If so, these findings may aid in the discovery of more cell-permeable kinesin inhibitors. For example, the methyl esters of glycolic acid-substituted adociasulfates could potentially act as prodrugs, penetrating the cell membrane and inhibiting kinesin upon hydrolysis by intracellular esterases. The structure of positions 1–10 in **1** seems to have little impact on activity. For example, **1** resembles a linear version of **4**. Despite dramatic steric consequences, **4** and **1** exhibit comparable activities. The ring system of **3** differs from the ring systems of most other adociasulfates in the position of its axial methyl groups but still exhibits *trans* geometries across all ring junctions. The absence of the pentacyclic ring linking the aromatic system to the adociasulfate core does not result in reduced activity, suggesting that some degree of conformational freedom of the aromatic ring is tolerated. Thus, our finding that **3** is also a strong kinesin inhibitor reinforces the idea that the broader meroterpenoid hydroquinone class of compounds may affect kinesin activity. Compounds from the extended family [e.g., disidein (21), haliclotriols (22), and flabellinol (23)] should be tested for their effect on kinesin activity in the future.

We note that a modified steroid core coupled to a benzene ring is a common feature of most of the strongest known adociasulfate inhibitors of kinesin (including **1** and **2**). Because the most variable structural features of adociasulfates do not seem to affect kinesin inhibition, we reason that the benzylated sterol core is responsible for the observed activity. Compounds **1**, **4**, and **7** are strong kinesin inhibitors and share this motif with **5**. Thus, it is likely that **5** is also a kinesin inhibitor and should be tested for such bioactivity. The steroid nucleus is a common chemical skeleton and not likely responsible for kinesin inhibition by itself. Indeed, our measurements show no notable kinesin inhibitory activity for cholesterol—a sterol lacking a benzene ring.

There is, however, a lack of true SAR data surrounding the core, and additional exploration of the adociasulfate motif is necessary.

We find no evidence to support the presence of extended inhibitory aggregates of the adociasulfates in kinesin assays. We did not observe free-floating extended aggregates as described in the work by Reddie et al. (16) through differential interference contrast (DIC) imaging. We examined whether smaller aggregates might be present on beads using bead diffusion assays with functional kinesin-1 preadsorbed to the beads. Kinesin was fully capable of MT binding and therefore, also capable of interacting with any aggregates acting as MT mimics. The random attachment of extended structures to spherical beads typically increases the effective Stokes radius of the beads and hence, lowers the bead diffusion coefficient. We observed no difference in bead diffusion between assays with and without 15 μM **1** (Fig. 2*B*). We did observe lower diffusion for beads preincubated with short microtubules (Fig. 2*B*), indicating that our measurements were sensitive enough to detect micrometer-long aggregates on beads. Finally, our single-molecule biophysics experiments provide evidence that even nanoscale aggregates are not needed for kinesin inhibition. In single-motor assays (bead binding fraction below 30%), bead binding to MTs resulted in kinesin stepping in the absence of adociasulfates but failed to produce subsequent motility on incubation with **1**.

Although MT binding is inhibited by adociasulfates, binding events still occur to a small extent, implying that at least one head of the kinesin-1 dimer is available to bind MTs. However, the stepping motion after the initial MT binding event is notably inhibited. Because adociasulfate activity is thought to result from the compound binding to the kinesin head, we conclude that the unattached head is bound by **1**, halting the processive motion of kinesin. It is tempting to speculate that the mode of inhibition of stepping occurs by either blocking the binding of the adociasulfate-bound head to the MT or sterically preventing the unattached head from passing the MT-bound head, thus blocking the powerstroke-induced isomerization of the kinesin dimer. Regardless of the exact mechanism, we conclude that, under our assay conditions, adociasulfates can bind a single head of a kinesin dimer without inhibiting the second head of the dimer (either sterically or allosterically).

The failure of adociasulfates to cross-link kinesin heads is not limited to the two heads of the same dimer. Beads precoated with kinesin-1 motors are not cross-linked in the presence of 15 μM of any adociasulfate that we have tested. However, similarly prepared beads are, indeed, efficiently cross-linked by microtubules in solution in the absence of ATP, consistent with previous observations of multiple kinesins binding or even densely covering the same MT (Fig. S8*B*) (24, 25). Based on the above evidence, we propose that aggregate formation is not a factor in adociasulfate inhibitory activity and that kinesin inhibition by adociasulfates needs to be considered on the scale of an individual MT binding domain.

We examined whether MT binding domains other than the motor head domain of kinesin are affected by adociasulfates. We compared MT binding of Eg5 and BimC, the latter of which has an N-terminal nonkinesin MT binding domain. We found that MT binding of BimC is inhibited by 15 μM **1** to a greater extent than Eg5. Thus, in the presence of **1**, the N-terminal domain of BimC does not enhance its binding to MTs. This result suggests that MT affinity of both MT binding domains of BimC is inhibited by adociasulfates. The interaction of the N-terminal domain with MTs is largely electrostatic and could be efficiently screened by the negatively charged adociasulfates, resulting in greater relative inhibition of binding for BimC. The exact mechanism by which adociasulfates affect the activity of the N-terminal domain of BimC requires additional investigation, but the result directly shows that MT binding activity of domains other than the one found in kinesin family motors may be affected by adociasulfates. Therefore, future drug discovery efforts based on adociasulfates and related structures need to screen broad effects on MT-associated proteins rather than confine their focus to kinesins.

In conclusion, our work has shown that AS-8 (**3**), AS-13 (**1**), and AS-14 (**2**) inhibit kinesin MT binding as well as the processive motion of kinesin. The ability of **2** to inhibit kinesin indicates that glycolic acid may be substituted for a sulfate without a significant loss in activity. Thus, the possibility exists that other functional groups can be used to effectively replace sulfates and enhance the drug-like properties of adociasulfates. The absence of visible small molecule aggregations and the lack of detectable cross-linking between kinesin-bound beads suggest that single molecules of adociasulfates bind individual kinesin motors, enhancing their potential as subjects of future drug development. Moreover, both mitotic and nonmitotic kinesins are inhibited, expanding the range of possible therapeutic targets (13, 14). Work by Brier et al. (13) provided evidence that adociasulfates are useful as molecular probes to understand kinesin activity. The observation that mitotic and nonmitotic kinesins tend to be inhibited to a different extent not only has therapeutic implications but also makes adociasulfates intriguing tools for future in vitro experiments. The broad range of potential adociasulfate targets is in contrast to the mechanism of monastrol, which is specific for the mitotic kinesin Eg5, further enhancing the versatility of adociasulfates as molecular probes (3). Finally, widespread kinesin inhibition by this family of compounds, despite structural variability, implies that additional related compounds may be active. In particular, AS-9 (**5**) and AS-10 (**6**) are likely kinesin inhibitors. We base this prediction on the presence of the benzylated sterol-like core in many of the most active adociasulfates. Investigation of the SAR of this motif would require synthesis and derivitization of the core or use of different or novel merotriterpenoids in kinesin assays.

Methods

General Experimental Procedures. Optical rotations were measured using a PerkinElmer Model 343 Polarimeter set at 589 nm and 20 °C with a 5.0-s integration time and a 1-dm path length cell. UV spectra were obtained using a Hitachi U-4100 spectrophotometer. NMR spectra were recorded on a Varian INOVA 600 NMR spectrometer (^{1}H 600 MHz, ^{13}C 150 MHz) equipped with a 5-mm Nalorac IDTG probe with a z-axis gradient or a Varian INOVA 500 NMR spectrometer (^{1}H 500 MHz, ^{13}C 125 MHz) using a 3-mm Nalorac MDBG probe with a z-axis gradient. Residual solvent signals were used for referencing (δ_H = 2.50 ppm, δ_C = 39.52 ppm for DMSO-d_6 and δ_H = 4.87 and 3.31 ppm, δ_C 49.15 ppm for CD$_3$OD). Liquid chromatography–MS analyses, including high-resolution mass spectra, were recorded using a Waters Micromass Q-TOF Micro mass spectrometer in negative ion mode with ion source and desolvation temperatures of 100 °C and 400 °C, respectively, and desolvation with nitrogen gas at a 400-L/h flow rate. A Beckman System Gold 126 solvent module with a 168 PDA detector was used for analytical and semipreparative HPLC. All reagents were purchased and used without additional purification.

Biological Material. Sponge material identified as *C. aculeata* Pulitzer-Finali, 1982 (order Haplosclerida, family Chalinidae) was collected by scuba from Eastern Fields, Papua New Guinea (S 10°00.395′; E 145°43.924′) and immediately frozen. A voucher specimen is maintained at the University of Utah under accession number PNG11-17–127.

Extraction and Isolation. Frozen sponge (190 g wet mass) was extracted three times with MeOH. Pooled extracts were fractionated on Diaion HP20SS resin using isopropanol (IPA)/H$_2$O mixtures in 25% (vol/vol) increments followed by a 100% MeOH wash to generate five fractions with a cumulative mass of 1,314.6 mg. A portion of the F1 fraction (25/75 IPA/H$_2$O; 203.9 mg) was subjected to C$_{18}$ flash chromatography eluting with MeOH/H$_2$O in 15% steps from 40% to 100% MeOH to generate 24 fractions. Fractions C–H were combined, yielding 56.7 mg material, and further fractionated by HPLC using a Luna 5μ C$_{18}$ column (100 Å, 250 × 10 mm) and a 1%/min linear gradient from 45% acetonitrile (ACN) to 70% ACN with 0.2 M NaCl. Salt was removed by loading samples onto preequilibrated Waters SEP-PAK C18 cartridges and flushing with three column volumes of 10% MeOH followed by elution of desired compounds with 100% MeOH. A broad peak eluting from 10 to 12.5 min (21.5 mg) was subjected to a final HPLC fractionation using a Luna 5μ pentafluorophenyl column (100 Å, 250 × 4.6 mm) and an isocratic solvent method at 76% MeOH/24% 0.2 M NaCl yielding compounds **1** (t_R = 10.0 min, 5.5 mg), **3** (t_R = 12.0 min, 6.9 mg), and **2** (t_R = 15.0 min, 2.3 mg).

Adociasulfate-13 (1). White solid; $[\alpha]^{20}_D$ −21.7 (c 0.55, MeOH); UV (MeOH) λ_{max} (log ε) 220 nm (3.93), 265 (2.47) nm; ^{1}H and ^{13}C NMR (Table S1); high-resolution electrospray ionization mass spectrometry (HRESIMS) *m/z* 709.3102 [M–H]$^-$ (calculated for C$_{36}$H$_{53}$O$_{10}$S$_2$, 709.3086; Δ −2.3 ppm).

Adociasulfate-14 (2). White solid; $[\alpha]^{20}_D$ −42.9 (c 0.13, MeOH); UV (MeOH) λ_{max} (log ε) 220 nm (3.86), 266 (2.60) nm; ^{1}H NMR and ^{13}C NMR (Table S2); HRESIMS *m/z* 687.3553 [M–H]$^-$ (calculated for C$_{38}$H$_{55}$O$_9$S, 687.3572; Δ 2.8 ppm).

Protein Material. BimC from *Aspergillus nidulans* and human Eg5 was purchased from Cytoskeleton. Human kinesin-1 (KIF5A heavy chain) with C-terminal 6xHis and FLAG tags and an N-terminal maltose-binding protein (MBP) tag was bacterially expressed in BL21DE3. The strategy of using an MBP solubility tag on the N terminus follows the work by Wong and Rice (26) and allowed us to successfully express high quantities of nonaggregated full-length KHC. Lysis was accomplished by sonication for 30 min at 4 °C. Lysis buffer: 50 mM sodium phosphate, pH 7.8, 300 mM NaCl, 10% glycerol, 20 mM imidazole, 2 mM MgCl$_2$, 0.25 mM ATP, 2 mM bME, and 0.2 μM PMSF with EDTA-free Roche mixture inhibitors. Cell lysis was followed by immobilized metal ion affinity chromatography purification (two washes and elution). Wash buffer 1: 50 mM sodium phosphate, pH 7.8, 500 mM NaCl, 10% glycerol, 40 mM imidazole, 0.02% Triton X-100, 2 mM MgCl$_2$, 0.25 mM ATP, and 2 mM bME. Wash buffer 2: 50 mM sodium phosphate, pH 7.8, 300 mM NaCl, 10% glycerol, 100 mM imidazole, 2 mM MgCl$_2$, 0.25 mM ATP, 3 mM glutathione (GSH), and 0.3 mM glutathione disulfide (GSSG). Elution buffer: 50 mM sodium phosphate, pH 7.5, 300 mM NaCl, 5% glycerol, 500 mM imidazole, 2 mM MgCl$_2$, 0.25 mM ATP, 3 mM GSH, and 0.3 mM GSSG. Material was then loaded on an anti-FLAG column. Equilibration buffer: 50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 15% glycerol, 0.005% Triton X-100, 2 mM MgCl$_2$, 0.25 mM ATP, 3 mM GSH, and 0.3 mM GSSG. Cleavage of the N-terminal MBP tag was performed using TEV protease on the same column. Elution buffer: 50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 15% glycerol, 0.005% Triton X-100, 2 mM MgCl$_2$, 0.25 mM ATP, 3 mM GSH, 0.3 mM GSSG, and 0.1 mg/mL 3× FLAG Peptide. Gene synthesis of KIF5A and purification were performed by Bionexus, Inc.

In Vitro Motility Assay. In vitro motility assays were prepared at room temperature. Flow cells were prepared by attaching a clean poly-L-lysine–coated

coverslip to a glass slide using double-sided tape. Assay buffer was PMEE (35 mM Pipes, 5 mM MgCl2, 1mM EGTA, 0.5 mM EDTA), pH 7.2. Taxol-stabilized MTs were first diluted in flow buffer (assay buffer + 20 μM taxol + 1 mM GTP) and then rapidly flown into the flow cell so that they were typically flow-aligned and attached to the polylysinated surface parallel to each other. After a brief incubation, mobile MTs were washed away, and the surface was blocked with buffer containing 20 mg/mL casein (Sigma-Aldrich).

Kinesin-1 at a selected concentration was first incubated for 10 min at room temperature with carboxylated polystyrene beads (1-μm diameter; Polysciences, Inc.) in the presence of saturating ATP (1 mM ATP). The selected adociasulfate stock solution (1.5 mM adociasulfate in DMSO) was first diluted 10-fold in assay buffer and then added to the kinesin bead mixture for a final adociasulfate concentration of 15 μM. The mixture was then incubated at room temperature. Incubation times did not affect the observed results. Typical incubation lasted 15 min. On incubation, the mixture was admitted into the flow cells.

Measurement of Binding. Optical trapping experiments were performed using a previously described measurement setup (27). Briefly, beads and MTs were viewed using an inverted DIC microscope (Nikon Eclipse T*i*-U) equipped with a high-magnification, high-NA objective (Nikon Plan Apo VC 100× oil, 1.40 NA). A high-resolution camera (Andor iXon+ DU897) and Nikon NIS elements AR 3.1 software were used to record experiments. Custom-built image processing and tracking software written in Matlab (Mathworks) was used for video tracking. Additional analysis and fitting were performed using Origin software (Orginlabs).

To measure the fraction of beads bound by kinesin, a bead was optically trapped (~4 mW laser power at the sample) and moved close to MTs. After a testing period of at least 40 s, beads that exhibited no MT binding were scored as such.

1. Sheetz MP (1996) Microtubule motor complexes moving membranous organelles. *Cell Struct Funct* 21(5):369–373.
2. Sharp DJ, Rogers GC, Scholey JM (2000) Microtubule motors in mitosis. *Nature* 407(6800):41–47.
3. Mayer TU, et al. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286(5441):971–974.
4. Nakazawa J, et al. (2003) A novel action of terpendole E on the motor activity of mitotic Kinesin Eg5. *Chem Biol* 10(2):131–137.
5. Hotha S, et al. (2003) HR22C16: A potent small-molecule probe for the dynamics of cell division. *Angew Chem Int Ed Engl* 42(21):2379–2382.
6. Sakowicz R, et al. (2004) Antitumor activity of a kinesin inhibitor. *Cancer Res* 64(9):3276–3280.
7. DeBonis S, et al. (2004) In vitro screening for inhibitors of the human mitotic kinesin Eg5 with antimitotic and antitumor activities. *Mol Cancer Ther* 3(9):1079–1090.
8. Cox CD, et al. (2005) Kinesin spindle protein (KSP) inhibitors. Part 1: The discovery of 3,5-diaryl-4,5-dihydropyrazoles as potent and selective inhibitors of the mitotic kinesin KSP. *Bioorg Med Chem Lett* 15(8):2041–2045.
9. Sorbera LA, Bolós J, Serradell N, Bayés M (2006) Ispinesib mesilate. *Drugs Future* 31(9):778–787.
10. Hopkins SC, Vale RD, Kuntz ID (2000) Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* 39(10):2805–2814.
11. Rickert KW, et al. (2008) Discovery and biochemical characterization of selective ATP competitive inhibitors of the human mitotic kinesin KSP. *Arch Biochem Biophys* 469(2):220–231.
12. Luo L, et al. (2007) ATP-competitive inhibitors of the mitotic kinesin KSP that function via an allosteric mechanism. *Nat Chem Biol* 3(11):722–726.
13. Brier S, et al. (2006) The marine natural product adociasulfate-2 as a tool to identify the MT-binding region of kinesins. *Biochemistry* 45(51):15644–15653.
14. Sakowicz R, et al. (1998) A marine natural product inhibitor of kinesin motors. *Science* 280(5361):292–295.
15. Blackburn CL, Faulkner JD (2000) Adociasulfate 10, a new merohexaprenoid sulfate from the sponge *Haliclona* (aka *Adocia*) sp. *Tetrahedron* 56(43):8429–8432.
16. Reddie KG, Roberts DR, Dore TM (2006) Inhibition of kinesin motor proteins by adociasulfate-2. *J Med Chem* 49(16):4857–4860.
17. Kalaitzis JA, Quinn RJ (1999) Adociasulfate-9, a new hexaprenoid hydroquinone from the Great Barrier Reef sponge *Adocia aculeata*. *J Nat Prod* 62(12):1682–1684.
18. Kalaitzis JA, et al. (1999) Adociasulfates 1, 7, and 8: New bioactive hexaprenoid hydroquinones from the marine sponge *Adocia* sp. *J Org Chem* 64(15):5571–5574.
19. Blackburn CL, et al. (1999) Adociasulfates 1–6, inhibitors of kinesin motor proteins from the sponge *Haliclona* (aka *Adocia*) sp. *J Org Chem* 64(15):5565–5570.
20. Stock MF, Chu J, Hackney DD (2003) The kinesin family member BimC contains a second microtubule binding region attached to the N terminus of the motor domain. *J Biol Chem* 278(52):52315–52322.
21. Cimino G, De Luca P, De Stefano S, Minale L (1975) Disidein, a pentacyclic sesterterpene condensed with an hydroxyhydroquinone moiety, from the sponge *Disidea pallescens*. *Tetrahedron* 31(3):271–275.
22. Crews P, Harrison B (2000) New triterpene-ketides (merotriterpenes), haliclotriol A and B, from an Indo-Pacific *Haliclona* sponge. *Tetrahedron* 56(46):9039–9046.
23. Sabry OMM, et al. (2005) Neurotoxic meroditerpenoids from the tropical marine brown alga *Stypopodium flabelliforme*. *J Nat Prod* 68(7):1022–1030.
24. Marx A, Müller J, Mandelkow E-M, Hoenger A, Mandelkow E (2006) Interaction of kinesin motors, microtubules, and MAPs. *J Muscle Res Cell Motil* 27(2):125–137.
25. Vale RD, Schnapp BJ, Reese TS, Sheetz MP (1985) Organelle, bead, and microtubule translocations promoted by soluble factors from the squid giant axon. *Cell* 40(3):559–569.
26. Wong YL, Rice SE (2010) Kinesin's light chains inhibit the head- and microtubule-binding activity of its tail. *Proc Natl Acad Sci USA* 107(26):11781–11786.
27. Butterfield J, Hong W, Mershon L, Vershinin M (2013) Construction of a high resolution microscope with conventional and holographic optical trapping capabilities. *J Vis Exp*, 10.3791/50481.
28. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7. Available at www.ebi.ac.uk/Tools/msa/clustalo/help/faq.html. Accessed October 22 2013.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

CHEMISTRY

CHAPTER 3

CHEMICAL DIVERSIFICATION ENABLES SYMBIOTIC MICROBIOTA TO

AFFORD FUNCTIONALLY DISTINCT PEPTIDES

3.1 Abstract

The microbiota is thought to provide host animals with bioactive metabolites,

such as antibiotics and other forms of chemical defense. Here, we show how chemical

diversification in the microbiota of marine tunicates leads to functionally distinct

secondary metabolites, resembling a "call-and-response" type adaptation. Small colonial

tunicates of the species *Didemnum molle* contained a family of structurally novel

peptides, some of which exhibited anti-HIV effects. By marrying structure elucidation,

whole microbiota sequencing methods, and synthetic biology, we uncovered the

divamide family of natural products. The anti-HIV compound divamide A and

derivatives were synthesized recombinantly using *Escherichia coli*, eliminating the need

to harvest large amounts of animals from coral reefs for chemical and pharmacological

evaluation. The ensuing structure-activity studies revealed a sequence-specific

requirement for the potent anti-HIV activity observed within the divamide series. These

results reveal how small animals, enabled by their microbiota, may achieve new or

modified functions through simple changes in protein sequence. In this way, chemical

diversification may reflect a weapon in the evolutionary arms race between the tunicate symbiosis and its environment.

3.2 Introduction

Chemistry guides the interactions between living organisms. Natural chemical diversity directs the specificity of these interactions and structures communities. An individual species or strain is commonly thought to display a distinctive chemical profile that reflects its particular role or function within the biological community. However, in certain situations organisms utilize biosynthetic pathways that inherently produce diverse chemicals, establishing populations of similar individuals with variable chemistry. This phenomenon is repeated throughout nature, indicating an evolutionary advantage of displaying chemical variability. A potential role of chemical diversity is to generate functional diversity. Chemical diversity enables adaptation to a changing environment, allowing either existing functions, perpetrated by chemicals, to be modified or new functions to be obtained. This may play a role in competition between species. Such properties would be useful from a medicinal chemistry perspective for the design of new and effective drugs in an era of increasing resistance.

Here, we show that the flexibility of a single biochemical pathway leads to useful functional differences that can be applied to treating human disease. From a remote reef in the Coral Sea, we found small colonies of tunicates containing related lanthipeptides for which minor structural variations result in major functional differences. We used a new approach integrating organic chemistry, metagenome sequencing, and synthetic biology to discover, produce, and characterize the biological action of a series of novel anti-HIV compounds: the divamides. We show that these compounds are produced by

uncultivated symbiotic bacteria that generate diverse chemicals using a highly conserved biosynthetic pathway and use this knowledge to produce two natural analogs using a single set of enzymes.

<center>3.3 Results</center>

We identified the organic extract from a small marine tunicate, *Didemnum molle* (E11-036) from Papua New Guinea, as a promising lead in an anti-HIV assay (**Figure 3.1-A, left**). The active compound, divamide A, was purified (<1 mg) by bioassay-guided fractionation and determined to be the active component (**Figure S3.1**). Using spectroscopic methods, divamide A was determined to be a novel peptide containing the modified amino acid lanthionine (Lan), and its partial sequence was elucidated. However, because of the limited available material, the structure could not be completed. Therefore, we turned to metagenome sequencing as a novel method for chemical discovery.

DNA was extracted from macerated tissues of the whole animal, sequenced, and assembled. The putative tetramer "GTTR" was used to probe the assembled metagenome, yielding a hit to a short coding sequence (CDS) for a precursor peptide (*divA*) from a ribosomally and posttranslationally modified peptide (RiPP) biosynthetic pathway (**Table S3.5**).[1] The 20 C-terminal residues of DivA encompassed the divamide A amino acids predicted by NMR as well as the sequence "GTTK," suggesting the Arg residue of the anticipated sequence was misinterpreted from NMR data. Immediately adjacent to *divA* was a gene (*divM*) bearing homology to the type II lanthionine synthetases,[2] confirming the lanthipeptide association. In addition, a series of other genes was predictive of the divamide A structure. Divamide A incorporates three methyllanthionine residues (MeLan), lysinoalanine (Lal), ß-hydroxy aspartic acid (Hya), and N-terminal

trimethylation, a rare posttranslational modification in nature (**Figure 3.1-B**).[3-5] Thus, a traditional bioassay-guided natural products approach was integrated with metagenomics-based methods to easily elucidate the structure of an otherwise highly challenging compound. Finally, with the planar structure in hand, stereochemical analysis was performed using chemical degradation.

A second, morphologically distinct *D. molle* specimen (E11-037) was collected in the same location as E11-036 and its extract exhibited no anti-HIV properties (**Figure 3.1-A, right; Figure S3.1**). Several divamide-like peptides were identified in the E11-037 extract including divamides B and C, but only divamide B could be fully characterized by the same spectroscopy-metagenome methodology used for divamide A (**Figures S3.4, S3.5**). All of the same posttranslational modifications adorn these compounds, which differ from divamide A only in amino acid sequence (**Figure 3.1-C**). Metagenomic analysis revealed a nearly identical biosynthetic gene cluster to that of divamide A, although only a partial sequence was recovered due to the challenging nature of the sample (**Figure 3.1-D, Table S3.5**).

The lack of any anti-HIV activity in divamide B or C starkly highlights how modest changes in structure govern the pharmacology. We sought to determine the molecular basis for this disparity and to provide material to determine the mechanism of action of divamide A. Given the scarcity of isolated material, we first aimed to produce divamides by recombinant expression. Such an *in vivo* total synthesis could also be used as a structural confirmation. Finally, it would represent the first example of a new strategy for drug discovery, marrying the best of traditional natural products chemistry with tools of metagenomics and synthetic biology.

The *div* clusters were found embedded not in the genomes of the host animal, but in those of symbiotic cyanobacteria, *Prochloron didemni*.[6] *P. didemni* often produces bioactive, presumably defensive chemicals in the interior of tunicates,[7] but lanthipeptides were previously not known. The *div* cluster is comprised of genes encoding five biosynthetic proteins, an export pump, and two genes of unknown function. In addition to *divA* and *divM*, these include an α-ketoglutarate-dependent Fe(II) monooxygenase ß-Asp hydroxylase (*divX*), a homolog of the uncharacterized *cinorf7* thought to be involved in lysinoalanine formation (*divLA*),[8] a SAM-dependent methyltransferase (*divMT*), and an ABC-type transporter (*divT*). Our expression strategy aimed to use these biosynthetic genes to reconstitute synthesis in *Escherichia coli*. However, the source organism is highly problematic for cloning experiments, being rich in mucopolysaccharides, and compounding the problem *P. didemni* was relatively rare in the samples and intractably embedded in animal mucus.

Therefore, we synthesized the expression vector by codon optimizing the biosynthetic genes and using native coding sequences in the intergenic regions, then placing the whole artificially organized operon under control of *lac* (**Figure 3.2-A**). This initial vector yielded no detectable products. Through a series of trials, we found that replacement of codon-optimized *divM* with the native gene sequence (shown in black), amplified from metagenomic DNA (**Figure S3.6**), resulted in production of divamide intermediates with three MeLan residues and dehydroalanine (Dha) but lacking Lal and methylation (**Figure 3.2-A, step 1**). While the 20-mer peptide was detected as a major product of the pathway, peptides of one, two, and three residues longer were also observed, reflecting cleavage at four different positions within the putative recognition

site "DIAA" (**Figure S3.8**). This implies the existence of an *E. coli* protease capable of performing a similar function as that of a symbiotic cyanobacterial protease. As the *div* and other lanthipeptide pathways lack a designated protease,[2] this observation is particularly interesting in the context of heterologous expression.

To complete the synthesis of divamide A from the desmethyl deslysinoalanine intermediate, Lal was introduced *in vitro*, forming spontaneously under basic conditions as described for cinnamycin (**Figure 3.2-B, step 2**).[8] DivMT, expressed and purified from *E. coli* BL21(DE3) (**Figure S3.7**), was unable to accept deslysinoalanine intermediates as a substrate, catalyzing N-terminal trimethylation only after introduction of Lal and only on the 20-mer substrate (**Figure 3.2-B, step 3**). The stringency of DivMT substrate specificity suggests that nonenzymatic Lal formation is stereoselective, likely due to structural constraints imposed by the prior installation of three MeLan residues. From our synthesis of divamide A, a functional role could be assigned to all biosynthetic genes in the pathway except *divLA* (**Figure 3.2-C**).

Scale up of this initial process involved several improvements, the most notable of which was adding cysteine to the expression media. As previously found in work with cyanobactin RiPP products, cysteine significantly improved the *in vivo* production of divamide A.[9] Currently, the best method produces 13.7 µg deslysinoalanine derivative per L of *E. coli* culture broth, after purification to homogeneity. From about 90 L of culture, 157 µg of pure divamide A were obtained. This represents a 12.7% yield from the delysinoalanine divamide A. Accurate measurement of submilligram quantities required the use quantitative NMR or LC/MS (**Figures S3.9, S3.10; Tables S3.6, S3.7**).

NMR analysis confirmed that the synthesized divamide A was identical to the natural product, confirming the elucidated structure (**Figures 3.2-D, S3.11**).

With material now in hand, we evaluated the biological activities of the divamides and their biosynthetic intermediates alongside cinnamycin, a closely related lanthipeptide produced by the soil-actinobacterium *Streptomyces cinnamoneous*.[10] Human CEM 1a2 tat/rev++ T-cells were incubated with compound and infected with vIIIB$^{\Delta Tat/Rev}$ HIV pseudovirus. Cell survival was determined four days after infection using a colorimetric MTT assay.[11] Potent cytoprotective activity was observed for both original and synthetic divamide A, as well as desmethyl divamide A, the desmethyl extra-proteolytic divamide A species, and an uncharacterized methyl ester of divamide A (**Figure 3.3, Figure S3.12**). This activity was absent, however, for deslysinoalanine divamide A, divamide B and cinnamycin. Additionally, significant cytotoxicity was observed for all but one divamide compound at concentrations just 10-fold higher than cytoprotection IC$_{50}$ values, indicating a relatively narrow therapeutic window. Cinnamycin displayed the most potent cytotoxicity, consistent with previous reports of its hemolytic and pro-apoptotic properties.[12-13] Interestingly, divamide C, an uncharacterized lanthipeptide isolated from E11-037 extracts, exhibited neither cytoprotection nor cytotoxicity.

These results reveal several key structure-activity relationships. N-methylation is not required for anti-HIV activity. Extension of the amino terminus with additional amino acid residues did not alter the overall phenotypic effects of divamide, though it appeared to slightly widen the therapeutic window of divamide A. Introduction of a methyl ester, likely at either the Glu or Hya side-chain carboxylic acid, also improved this aspect of divamide activity. However, the shape of the molecule conferred by MeLan and Lal

residues was essential for cytoprotection by divamide A, as is indicated by the loss of

activity for the deslysinoalanine intermediate. Thus, posttranslational modifications

outside of the chemical scaffold defined by MeLan/Lal have only minor impacts on the

overall activity of divamide A. On the other hand, the amino acid sequence has a

dramatic impact on activity. Five amino acid substitutions between divamide A and B

lead to the loss of the cytoprotective phenotype for divamide B, highlighting a

hypervariable region between residues 8 and 14 that is responsible for the cytoprotection

phenotype. Divamide C displays more sequence substitutions than divamide B while

maintaining all of the same posttranslational modifications (**Figure S3.12**), yet it lacks

any activity at all in the context of this assay. Similarly, the structural differences

between divamide A and cinnamycin, which mostly reflect changes in sequence, also

lead to a loss of HIV cytoprotection and enhanced cytotoxicity in cinnamycin.

Collectively, these observations suggest that while posttranslational modification

establishes a bioactive scaffold, it is the variation between amino acid sequences that

determines the biological activities of the divamides.

Divamides A-C do not provide a complete picture of the chemical diversity within

the tunicates described. While only two peptides were isolated in purity from E11-037, a

number of other divamide-like ions were observed by LC/MS to display Hoffman

elimination of trimethylamine,[14] a characteristic of divamides possessing N-terminal

trimethylation (**Table S3.8, Figure S3.13**). Still other species were observed resembling

divamide but lacking N-trimethylation, likely representing partially or nonmethylated

intermediates. This is distinct from E11-036, from which a single divamide was observed

and whose metagenome contains a single *div* pathway and precursor gene. Only one

precursor gene encoding divamide B was found for E11-037 due to the challenging nature of the sample, but the presence of additional divamides with alternative amino acid sequences implies the presence of multiple gene clusters or precursors. Lanthipeptide pathways from other organisms were identified by BLAST search that exhibit multiple precursor genes per cluster, or multiple clusters per genome (**Figure S3.14**). In addition, some of these pathways display the potential for novel posttranslational modifications not found in related but characterized pathways. An alignment of precursor protein cores indicates a signature chemical scaffold consisting of Lan, Lal, Hya, and Gly residues, while most residues in between these conserved positions are highly variable (**Figure S3.15**). Thus, there exists convincing evidence to suggest that the *div* pathway and its extended lanthipeptide family encompass a broader family of peptides that is chemically diverse within a conserved scaffold.

3.4 Discussion

This work represents the first example of a direct discovery-supply strategy of a symbiotic natural product. Prior knowledge of the compound's biological activity and structure obtained via a bioassay-guided approach greatly facilitated pathway identification from a complex metagenome, enabling access to novel chemistry. A compound of such limited sample size would likely be abandoned in the context of a traditional natural products approach, while the organism would not have been targeted for metagenome mining without significant cause, such as promising biological activity or unique chemistry. The discovery process also provided us with a platform to produce divamides biosynthetically. We have already synthesized more divamide than was originally isolated from the natural source.

Nature uses chemical diversity to obtain functional diversity. This ability is highly prized by medicinal chemists seeking to discover new drugs. Nature is particularly adept at creating complex, biologically relevant chemical scaffolds, but also excels at incorporating diversity into these scaffolds using simple biochemical concepts. This is especially true for natural products. The *div* pathway exhibits elements of diversity-generating RiPPs, including broad-substrate enzymes and precursor gene divergence,[15] but lacks the conformational and scaffold diversity displayed by the prochlorosins or cyanobactins.[16-17] Nonetheless, internal amino acid changes have dramatic effects on divamide activity. The divamide scaffold may be particularly well suited for biological interactions, or this may reflect fine-tuning of a conserved function, perhaps in response to similar adaptability in the native target. Such a "call-and-response" is reminiscent of elements of the innate immune response and may be well suited to deal with viruses.[18]

3.5 Supplementary results

Divamide A was isolated from E11-036 as a white powder with *m/z* 1010.96248 $[M+H]^{2+}$, determined by Fourier transform mass spectrometry (calculated for $C_{87}H_{139}N_{21}O_{28}S_3^{2+}$, 1010.96299; Δ -0.504 ppm). A ^1H NMR spectrum recorded in D_2O revealed amino acid-like chemical shifts (**Figure S3.2-A**). Homonuclear (gCOSY, zTOCSY, ROESY) and heteronuclear (gHSQC, gHMBC, gHSQCTOXY) experiments conducted in the same solvent led to the identification of individual amino acid spin systems, including Gly, Ala, Val, Ile, Phe, Arg, Thr, and Ser (**Figure S3.2-B–F**).

The presence of several unusual amino acid spin systems was also noted. H_β- H_β or H_N-H_α NOESY correlations occur between three pairs of Cys-like and Thr-like spin systems, suggesting the presence of MeLan (**Table S3.2**). Additionally, an HMBC

correlation was observed from a Cys-H_β to a Thr-C_β that was later determined to correspond to Cys15 and Thr5. A residue resembling ß-hydroxy Asp (Hya) displayed a strongly deshielded ß-methine (δ_H 4.75 ppm, δ_C 73.1 ppm). A prominent heteronuclear correlation between δ_H 3.28 ppm and δ_C 53.5 ppm appearing in both HSQC and HMBC spectra indicated a quaternary trimethylamine functional group, which was associated with a Glu residue via NOE correlation between the N-methyl protons and Glu-H_α, and via HMBC correlation between N-methyl protons and Glu-C_α. This interpretation was supported by the failure of amino acid sequencing by Edman degradation and by observation of a characteristic 63 Da mass differential between positive and negative ion mode LC/MS spectra resulting from Hoffman elimination of a trimethylamine (**Figure S3.13**).

Partial amino acid connectivity was determined from either H_N-H_α or H_N-H_N NOEs using 2D NMR data acquired in 90% H_2O/10% D_2O to enable detection of exchangeable amide backbone protons. Full deuterium/proton exchange was achieved by incubation with 0.1% ammonium hydroxide at 37˚C overnight prior to collection of NMR data using protonated solvents. Interpretation of these data resulted in the sequences "GTTR" and "AST" (**Figure S3.2-H–I**). The former was used to search the assembled E11-036 metagenome for a RiPP precursor gene using tBLASTn. A short CDS was found containing the sequence "ECASTCSFGIVTIVCDGTTK" (*divA*). The gene was clustered with a class II lanthionine synthetase (*divM*) and other CDSs with functional annotations that complied with the modifications predicted. Re-evaluation of NMR data showed complete agreement with the sequence, indicating an incorrect assignment of the Lys spin system to Arg and establishing three MeLan linkages between

Cys2-Thr19, Thr5-Cys15, and Cys6-Thr12. The sequence and Lan connectivities greatly

resemble those of cinnamycin, which also incorporates the unusual amino acids Hya and

Lal. Our data also support the presence of these residues. The presence of Lys instead of

Arg made obvious an NOE between Lys20-H_ε and Ser7-H_β indicative of Lal. A ^{13}C

spectrum obtained using 500 MHz instrument with a carbon-optimized cryogenic probe

included 87 carbons (**Figure S3.2-K**), though not all could be assigned due to the absence

of HMBC signals for those carbons. Additional 2D experiments, collected in 70%

CD_3OH/ 30% H_2O/ 0.1% TFA using a 900 MHz spectrometer to improve the separation

between amide proton signals and enhance signal-to-noise, agreed with previous data

(**Figure S3.2-J–M**). Thus, the planar structure of divamide A was elucidated partly by

spectroscopy and partly using metagenomics.

Stereochemical characterization of divamide was accomplished by degradative

analysis. Mixtures of amino acid standards were subjected to acid hydrolysis, esterified

with methanol, and then amidated using pentafluoropropionyl anhydride. These were

subjected to chiral GC/MS alongside similarly treated divamide A for comparative

analysis of retention times and fragmentation patterns (**Table S3.3**). All conventional

amino acids were detected in L configurations, while atypical amino acids L-*erythro*-

Hya, D-L MeLan, and L-L Lal exhibited identical chirality to cinnamcyin (**Figure S3.3**).

Lal standards (LL and DL) were run on an achiral GC/MS column. In the course of

divamide A hydrolysis, some scrambling of the Lal configuration was observed. The

residual excess of the LL enantiomer indicated the initial Lal configuration was LL. The

formation of Lan and MeLan results in inversion of the relative C_α configuration of the

recipient Ser or Thr residue. This is distinct from Lal, in which no change in

stereochemistry occurs. The configuration of N-trimethylglutamate is most likely to be L based on overlapping NOEs with biosynthetic divamide A (**Figure S3.11-G**).

Divamide B was isolated from E11-037 as a white powder with *m/z* 965.9194 $[M+H]^{2+}$ by high-resolution electrospray ionization mass spectrometry (HRESIMS) (calculated for $C_{79}H_{129}N_{21}O_{29}S_3^{2+}$, 965.921325; Δ 1.99 ppm). Hoffman elimination, observed by LC/MS, suggested similar posttranslational modifications as found in divamide A (**Figure S3.13**). NMR data recorded in either 70% CD_3OD/ 30% D_2O/ 0.1% TFA or 70% CD_3OH/ 30% H_2O/ 0.1% TFA revealed many overlapping spin systems with divamide A as well as some significant differences (**Table S3.1, Figure S3.4**). While no Val was detected, Ala and Pro spin systems were recognized. Two Ile spin systems were identified that differed significantly from those of divamide A, suggesting alternative positions. Two sets of correlations were observed in homonuclear spectra for most divamide B spin systems. This can be explained by the potential for conformational isomerization introduced by Pro. The reported chemical shifts are similar to those of divamide A and represent the major divamide B rotamer. A precursor protein containing the sequence "ECASTCSSGPITAICDGTTK" was identified from a partial *div* cluster found in the assembled E11-037 metagenome. The predicted lanthipeptide of this pathway, containing modifications identical to divamide A, fit well with the observed mass and NMR data. Some controversy exists regarding the presence of aromatic chemical shifts despite the apparent replacement of Phe8 with Ser. However, expression of pDiv2 in *E. coli* using the divamide B core with the divamide A biosynthetic enzymes resulted in production of *m/z* 944.9, the predicted mass of desmethyl divamide B, as well as the same extra-proteolytic species that were observed with divamide A expression

(**Figure S3.8-B**). These data suggest that the precursor protein sequence from E11-037 is the correct divamide B sequence.

Divamide C was isolated from E11-037 as a white powder with *m/z* 1039.4535 $[M+H]^{2+}$ by HRESIMS. Similar to divamide B, Hoffman elimination in negative ion mode was observed, suggesting a relationship to divamide A (**Figure S3.13**). 2D NMR data were difficult to interpret due to a low signal-to-noise ratio, but conserved residues of divamides A and B were observed (**Figure S3.5-B–E, G–I**). No additional precursor genes or *div* clusters were identified from E11-037, preventing us from obtaining an amino acid sequence to aid in structure elucidation. In an attempt to predict the sequence, we used the online tool Mass Analysis Peptide Sequence Prediction (MAPSP) to generate a list of all possible peptide masses that could be obtained given a number of sequence restrictions (http://mapsp.ifg.uni-muenster.de).[19] From an alignment of all related divamide precursor proteins, eight variable positions become readily apparent (**Figure S3.15**). We limited the number of responses by allowing only those residues known to occur in these positions. Even with these restrictions, we obtained 722 possible sequence masses that occur within 5 ppm of the accurate mass obtained for divamide C. However, NMR spectra suggest that divamide C resembles divamide B in the presence of Pro10- and Ile14-like spin systems, but incorporates a Val11-like residue akin to that in divamide A (**Figure S3.5-C**). Imposing these additional restrictions reduces the number of possible sequences to 12. The sequence ECQSTCSYGPVTVICDGTTK best fits all available data (calculated for $C_{88}H_{138}N_{22}O_{30}S_3^{2+}$, 2078.91105; Δ 1.95 ppm), but could not be confirmed using spectroscopic or genomic techniques.

A putative degradation product of divamide A that was 14 Da higher in mass accumulated in the original E11-036-derived divamide A material over a period of more than one year, such that the +14 Da derivative contributed significantly to the overall mass ratio (**Table S3.7**). This peptide mixture displayed a wider therapeutic window than pure divamide A by HIV cytoprotection assay, suggesting the derivative may exhibit improved pharmacological properties (**Figure 3.3**). The +14 Da species was identified by HRESIMS as the methylation product of divamide A (calculated for $C_{88}H_{141}N_{21}O_{28}S_3^{2+}$, 1017.9708; Δ 1.38 ppm). Though the exact methylation position was not determined, it is possible that either of the two acidic residues Glu or Hya or the C-terminus was esterified with methanol during deuterium back-exchange in 0.1% ammonium hydroxide.

3.6 Methods

3.6.1 General. The majority of NMR spectra were recorded on a Varian INOVA 600 NMR spectrometer (^{1}H 600 MHz, ^{13}C 150 MHz) equipped with a 5 mm Nalorac IDTG cryogenic probe or a Varian INOVA 500 NMR (^{1}H 500 MHz, ^{13}C 125 MHz) with 5 mm Nalorac IDTG probe, using residual solvent signals for reference (δH = 4.80 ppm for D_2O and δH = 4.87 ppm, δC = 49.15 ppm for D_2O/CD_3OD mixtures). For the initial structural characterization of divamide A, a Bruker Advance III HD 900 MHZ instrument (^{1}H 900 MHz, ^{13}C 225 MHz) with 5 mm ^{1}H(^{13}C/^{15}N) cryogenic probe and a Bruker Advance III 500 MHz (^{1}H 500 MHz, ^{13}C 125 MHz) with a 5 mm ^{13}C/^{15}N(^{1}H) cryogenic probe were also used. Liquid chromatography–MS spectra were recorded using a Micromass Q-ToF Micro mass spectrometer (Waters). High-resolution mass spectra were recorded on a LTQ-FTMS (ThermoElectron) or a maXis-II ETD Q-ToF (Bruker). Analytical HPLC was performed on a Beckman System Gold 126 solvent module with a

168 PDA detector or a Hitachi Primaide 1110 Pump with 1430 Diode Array Detector. An

HP 6890 Series GC System equipped with an HP 7683 Series Injector and an HP 5973

Mass Selective Detector was used for all GC/MS experiments. A Fisher Multiscan FC

plate reader was employed for biological assays. Amino acid standards included standard

L and D amino acids (Cyclo Chemical or Sigma-Aldrich), DL-threo-ß-hydroxyaspartic

acid (Sigma-Aldrich), L-erythro-ß-hydroxyaspartic acid (Wako Chemicals), and DL-

lysinoalanine (Bachem). Methyllanthionine GC/MS standards were generously provided

by John Vederas (University of Alberta). Cinnamycin was purified from spent culture

broth of *Streptomyces cinnamoneus* ATCC 11874 grown in ISP-2 (1L) at 30˚C in

Fernbach flasks (2.8L) shaking at 150 rpm for three days.

 3.6.2 Animal collection. Tunicates identified as *Didemnum molle* (Herdman,

1886) were collected by scuba from the Eastern Fields of Papua New Guinea (S 10˚

16.217'; E 145˚ 38.679'). Chemistry samples were immediately frozen, while samples for

metagenome sequencing were first processed by slicing into or placing small samples in

Rnalater, then leaving them at 4˚C overnight with occasional inversion of tube, then

freezing. Vouchers are maintained at the University of Utah under accession numbers

E11-036 and E11-037, or PNG11-9-099 and PNG11-9-100, respectively.

 3.6.3 Extraction and isolation. Frozen tunicate (200 g) was extracted three times

in MeOH. Pooled extract was filtered by gravity and fractionated by Diaion HP20SS

resin using isopropanol (IPA)/H_2O mixtures of 25% increments of increasing IPA to

generate five fractions with a cumulative weight (excluding 100% H_2O fractions) of

581.8 mg for E11-036 and 685.2 mg for E11-037. E11-036 F1 fractions (eluted in 25%

IPA; 211.3 mg) were subjected to C_{18} flash chromatography starting at 10% MeOH and

increasing stepwise by 30% up to 100% MeOH, yielding eight fractions. Fraction E (70% MeOH; 19.1 mg) was further fractionated by HPLC. Using a Luna 5µ C_{18} column (100 Å, 250 x 4.60 mm) at 1 ml/min, the solvent 30% ACN/70% 0.2 M NaCl was held for 15 min then increased to 100% ACN over 5 min. Two peaks were collected and desalted using Waters SEPPAK C18 cartridges, resulting in divamide A (t_R = 8.87 min, 315.1 µg,) and oxidized divamide A (t_R = 8.09 min, not quantified). A degradation product, later identified as a methylated divamide A derivative, accumulated in the divamide A sample over time. The resulting mixture was fractionated by C_{18} using a method similar to that described later for lysinoalanine formation, yielding 77.8 µg of pure divamide A and 237.3 µg of a mixture of the two peptides.

C_{18} flash chromatography E11-037 F1 fractions (25% IPA; 172.9 mg) followed a similar procedure to generate 12 fractions. Fraction F (40% MeOH, 10.1 mg) was subjected to HPLC using the same Luna C_{18} column at 31% ACN/ 69% 0.2 M NaCl held for 11 min then increased to 60% ACN over 4 min and held at 60% for 8 min with a 1ml/min flow rate, yielding divamide B (t_R = 8.78 min). All elutions excluding the pure peptide were recombined (7.6 mg) and further purified by HPLC on a Luna 5µ Phenyl-Hexyl column (100 Å, 250 x 4.60 mm) at 1 ml/min starting at 20% ACN/80% 0.1% TFA for 4 min, then increasing to 24% ACN over 4 min and holding for 6 min, increasing to 30% ACN over 6 min and holding for 3 min, then finally to 60% ACN over 15 min. This yielded additional divamide B (t_R = 11.45 min, added to divamide B described above for a combined weight of 100 µg) as well as divamide C (t_R = 20.12 min, 31.7 µg). Reported yields were determined by NMR and LC/MS quantification as described below.

3.6.4 Quantitative NMR. Quantitative NMR experiments followed previous protocols using external standards.[20] Samples were suspended in deuterated solvents and dried multiple times to ensure full deuterium exchange. A 1:2 dilution series of an L-Trp stock solution was prepared in D_2O encompassing a range of concentrations from 2.7 to 0.093 μg/ml to generate a standard curve. These concentrations were obtained by measuring UV absorbance using a Nanodrop 2000 ($\varepsilon_{280, Trp}$ = 5170.9 nm^{-1}, $\varepsilon_{276, Tyr}$ = 1322.2 nm^{-1}). Similarly, an L-Tyr standard was also prepared to assess the accuracy of the curve. L-Trp and L-Tyr samples were also fully exchanged in D_2O prior to data acquisition. All samples were then suspended in D_2O (120 μl) and transferred to new, dry Norell S-3-500-7 NMR tubes. The relaxation time t1 was determined for all proton signals intended for quantification and d1 was set to 10 times t1 of the highest observed d1. This value, as well as all other parameters, was maintained for all subsequent experiments. A 1D ^1H spectrum was recorded for each sample and care was given to optimize shims between each experiment. Signals of interest were chosen based on isolation from other peaks and known proton assignment. Spectra were phased, baseline corrected, and integrated, and the absolute integral values were used to produce a standard concentration curve (**Figure S3.9**). Concentrations of samples were calculated using the resulting linear equation. To reduce any signal-specific effects, three standard curves were generated from three separate L-Trp proton signals and the concentration of analyte was calculated from each of them. Similarly, at least two proton signals were chosen from each analyte spectrum and the concentrations determined from each L-Trp standard curve were averaged. **Table S3.6** summarizes the results of quantitative NMR. The concentration of L-Tyr was estimated 0.351 ± 0.0095 mg/ml while the UV-

determined concentration was 0.326 mg/ml. A 7.6% difference was calculated, suggesting that analyte concentrations may be slightly overestimated.

3.6.5 Quantitative LC/MS. A 1:2 dilution series of desmethyl divamide A, suspended at the same initial concentration as was determined by NMR (1.935 mM) and diluted 100-fold, was prepared to generate a standard curve by LC/MS (**Figure S3.10**). Each sample (100 µl) was spiked with an internal cyanobactin standard (10 µl; *m/z* 883.5) to assess variability in injection volume between each run. Each desmethyl divamide A concentration and each analyte were run in triplicate and in random order. The area under the curve (AUC) was determined for doubly charged ions in selected ion mode (SIM) for all peaks of interest within a sample. Concentrations were calculated from the average of the three AUC values. The internal standard AUC fluctuated between runs of like-samples within two standard deviations (SD = 35.7) of the average AUC (AUC = 132.3), or ± ~50%, indicating a moderate degree of run-to-run variability. This was deemed acceptable variation given the difficulty in obtaining an accurate mass of small amounts of material by weighing or by NMR. The internal standard AUC showed an up to 3-fold difference between runs of different samples, depending on the co-injected analyte. Biosynthetic divamide A, also quantified by NMR (0.664 mM), was used to assess the agreeability between NMR and LC/MS quantification curves. A concentration of 0.624 ± 0.0363 mM for *E. coli*-derived divamide A was calculated – a difference of 5.92% less than the NMR concentration. The results of LC/MS quantification are summarized in **Table S3.7**.

3.6.6 Chiral GC/MS. Amino acid standards included Gly and L and D configurations of all canonical amino acids as well as L- and D-*allo*-Ile and Thr.

Noncanonical amino acid standards included: L-*erythro*-Hya (2S,3S), L-*threo*-Hya (2S,3R), D-*threo*-Hya (2R,3S), LL Lan (2R,6R), *meso*-DL Lan (2S,6R), LL MeLan (2R,3R,6R), DL MeLan (2S,3S,6R), L-*allo*-MeLan (2R,3S,6R), D-*allo*-MeLan (2S,3R,6R), LL Lal (2S, 9S), and DL Lal (2R, 9S). Peptides (less than 0.1 mg) and standards were hydrolyzed in 6 M HCl or 1 M NaOH at 100°C overnight then dried completely under a stream of nitrogen gas followed by high vacuum. Peptide hydrolysate and amino acid standards were then esterified with MeOH (200 µl) and acetyl chloride (50 µl) in a sealed reaction vial at 100°C for 1 hour, dried to completion, and finally treated with 100 µl each of DCM and PFP-anhydride at 100°C for 15 min. For the CP-Chirasil L-Valine column (25 m x 0.25 mm, 0.12 µm), an initial temperature of 80°C was held for 3 min before increasing linearly to 200°C over 30 min, then holding at 200°C for 12 min. For the Zebron ZB-5MSi Guardian column (30 m x 0.25 mm ID, 0.25 µm), a starting temperature of 60°C was held for 1.5 min before increasing linearly to 250°C over 31.7 min, then holding at 330°C for 13 min. All samples were prepared at 0.1 mg/ml in ethyl acetate and 5 µl injected per run using a split ratio of 2 and helium mobile phase gas.

3.6.7 DNA Extraction. Extraction of DNA from *D. molle* samples E11-036 and E11-037 was complicated by the presence of copious amounts of polysaccharide-rich mucus. We utilized a protocol previously adapted from Sokolov for the precipitation of polysaccharides found in marine invertebrate slime.[21-22] Briefly, tissue was sliced thinly and squashed between two aluminum foil sheets, then treated with proteinase K (Qiagen). Polysaccharides and proteins were then precipitated with saturated KCl. Additional KCl was required beyond what was called for in the original protocol. Samples were

centrifuged and an equal volume of isopropanol added to the supernatant. White precipitate formed upon incubation at room temperature. The mixture was centrifuged again and pellet washed with 70% ethanol. Excess solvent was removed by air drying and the pellet resuspended in QBT buffer (QIAGEN). From this point forward, DNA extraction followed the standard protocol for the QIAGEN Genomic-tip kit. DNA from sample E11-036 was sequenced via Pacific Biosciences (PacBio) single molecule real time (SMRT) technology using 5 SMRT cells. Also, an Illumina library was prepared using the E11-036 DNA with ~700 bp insert sizes and then sequenced on an Illumina Miseq sequencer using a 250 bp paired end run. For *D. molle* sample E11-037, a 350 bp insert size library was prepared and sequenced (100 bp paired end) on an Illumina HiSeq 2500.

3.6.8 Assembly of the div *cluster*. To elucidate the full lanthipeptide cluster, E11-036 raw Illumina reads were filtered for quality and length (phred score >30 for greater than 40 bp) using the windowed adaptive trimming tool Sickle.[23] The filtered reads were then assembled with the Velvet assembler v. 1.2.09.[24] The assembly (kmer size 61) was analyzed and produced a contig containing the precursor peptide gene and genes homologous to those of the cinnamycin cluster of *S. cinnamoneus*. In addition, analysis of this contig showed unique CDSs, such as a putative methyltransferase. Based on the minimum genes required for *in vitro* cinnamycin biosynthesis by Ökesli et al.,[8] a second contig with a gene sharing 44% identity to *cinorf7,* a protein believed to be involved in lysinoalanine formation, was included in the divamide cluster. Extensions of the contigs using matching reads were then optimized to produce a single contig 17 kbp in length containing the entire cluster (12 kbp from precursor gene to the *cinorf7* homolog, *divLA*).

This sequence was deposited at GenBank under the accession number KY115608. E11-037 raw sequencing reads were assembled similarly. Two contigs were identified containing genes homologs to those observed in the *div* gene cluster. One contig 1152 bp in length contained the full *divA* gene and a partial *divM* sequence, consisting of 111 bp of the 3' end. The second contig, 1655 bp in length, contained additional partial genes, consisting of 604 bp of the 5' end of *divX* and 603 bp of the 3' end of *divMT*, while also containing a complete but truncated version of *divY*, resembling only the 5' end of E11-036 *divY*. By comparison with the divamide A gene cluster, we estimate that a 3144 bp gap exists between these two contigs that would account for the remainder of the *divM* and *divX* genes. The E11-037 partial *div* cluster was deposited at GenBank under the accession number KY115609.

3.6.9 Construction of pDiv and pRSFDuet-DivMT. Four DNA pieces containing portions of the *div* pathway with putative biosynthetic genes codon-optimized were synthesized by Integrated DNA Technologies (IDT) and preassembled into two larger pieces by overlapping PCR (**Table S3.9**). Pieces 1a and 2 were joined via an initial reaction with 15 cycles of denaturing (98˚C, 15 s), annealing (62˚C, 30 s), and elongation (72˚C, 1 min) using Phusion® DNA polymerase (New England BioLabs). This was followed by another 30 cycles of denaturing (98˚C, 15 s), annealing (62˚C, 15 s), and elongation (72˚C, 2 min) with primers div1a-fwd and div2-rvs, using the first reaction directly as the template (**Table S3.10**). Similarly, DNA pieces 3 and 1b were joined using primers div3-fwd and div1b-rev. The final products were gel purified using Ultrafree-DA Centrifugal Filters (EMD Millipore). The vector backbone was obtained by digestion of pPat with SacII and KpnI and gel purification of the resulting 4450 bp fragment.[9]

The backbone was transformed into the uracil auxotroph *Saccharomyces cerevisiae* BY4741 along with the PCR-joined 1a-2 and 3-1b DNA pieces for assembly by yeast recombination. The resulting colonies, grown on uracil-deficient SD agar, were combined and grown overnight in liquid uracil-deficient SD. Plasmid DNA was purified from the overnight culture using a QIAprep Spin Miniprep kit (QIAGEN) with protocol modifications to enhance cell lysis: the cells were resuspended in buffer P1 with 0.5 mm glass beads and mixed by vortex on high for 10 min. Yeast plasmid was transformed into DH10ß *E. coli* and the resulting colonies screened for the presence of *div* genes by sequencing. Additional plasmid was generated from one of the *div*-positive clones and fully sequenced using primers 5 through 15, confirming the desired sequence of pDiv.

Expression of pDiv in *E. coli* did not yield any divamide A-like products, so modifications were made to the vector to induce production. We had previously observed expression of soluble, recombinant DivM only with the native gene sequence (data not shown), and reasoned that the codon-optimization of *divM* might have resulted in poor expression or incorrect protein folding, rendering the initial biosynthetic step catalyzed by DivM nonfunctional in pDiv.[25] The native sequence *divM* gene was cloned by PCR from E11-036 metagenomic DNA using primers intergenic-divM-native-fwd and intergenic-divM-native-rvs with Platinum® *Taq* DNA Polymerase High Fidelity (ThermoFisher Scientific) and 35 cycles of denaturation (94˚C, 30 s), annealing (72˚C, 30 s), and elongation (68˚C, 3 min 30 s) to yield a faint band (**Figure S3.6-A**). The product was diluted 100-fold and amplified by PCR with Phusion® with the same primers by 30 cycles of denaturation (98˚C, 15 s), annealing (65/72˚C, 30 s), and elongation (72˚C, 3 min 30 s) and purified using a QIAquick PCR Purification kit (QIAGEN) (**Figure S3.6-**

B). pDiv was digested with MluI-HF and AflII to remove the majority of the codon-optimized *divM* gene and the resulting 8355 bp band was gel purified using a QIAEX II Gel Extraction kit (QIAGEN). The native *divM* gene and linearized pDiv were assembled by yeast homologous recombination as previously described. Yeast lysis was further enhanced by incubation with buffer P2 at 55˚C for 10 min after pulverization with glass beads by vortex. Plasmids were extracted from individual *E. coli* colonies transformed with yeast DNA and digested with SpeI and NdeI to confirm replacement of the codon-optimized *divM* with the native gene, then sequenced with primers 17–22 to verify the desired sequence of *divM* in pDiv2.

The pDiv3 divamide B expression vector was constructed in two separate phases, with replacement of *divA* in pDiv with a divamide-B encoding *divA* sequence, followed by replacement of codon-optimized *divM* with the native *divM* gene. First, pDiv was digested with ClaI and NheI restriction enzymes. The product was then gel purified using a centrifugal filter as described previously. The linearized vector was recombined with synthetic DNA divA-divamideB (IDT; **Table S3.9**) using yeast as described above. Plasmids obtained from pDiv-containing *E. coli* colonies transformed with extracted yeast DNA were sequenced with primer 1, confirming the altered sequence of *divA* with a single noncoding C to G mutation, the position of which is underlined in the sequence shown in **Table S3.9**. The native *divM* gene was then introduced as described for the construction of pDiv2, using the divamide B-encoding pDiv as a template for recombination. During the restriction digest of pDiv2, XhoI was used in addition to AflII and MluI-HF to cut the codon-optimized *divM* gene and prevent reinsertion.

To construct the methyltransferase expression plasmid pRSFDuet-DivMT, the codon-optimized *divMT* gene was PCR amplified using *pDiv* as a template with primers divMT-MCS1-fwd and divMT-MCS1-rvs, Platinum® *Taq* Hi-Fi, and 2% DMSO by 30 cycles of denaturation (94°C, 30 s), annealing (45.5°C, 30 s), and elongation (68°C, 2 min). The PCR product was purified as described above. Both purified PCR product and the empty pRSFDuet vector were then digested with SacI and NotI restriction enzymes. Reaction mixtures were heat inactivated and the linearized pRSFDuet was dephosphorylated using Antarctic Phosphatase (New England BioLabs). The two pieces were then ligated together using T4 DNA ligase (New England BioLabs). The ligation reaction was transformed into DH10ß *E. coli* and the transformed cells plated on LB kanamycin selection media. Individual colonies were picked and grown overnight in liquid culture. Cells were harvested and plasmids purified using the QIAprep Spin Miniprep kit. Plasmids were triple digested with ScaI, MluI, and PstI to verify the expected band sizes, and positive plasmids were sequenced for confirmation.

3.6.10 Expression and purification of the pDiv2-E. coli product. pDiv2 was transformed into DH10ß *E. coli* and six colonies used to inoculate individual wells of a 24-well plate, each containing LB (6 ml) with ampicillin (50 µg/ml), which was shaken overnight at 150 rpm and 30°C. Wells were combined and 3 ml used to seed each of eight 2xYT cultures (1L), each containing ampicillin (50 µg/ml) and cysteine hydrochloride monohydrate (5 mM) from a freshly prepared, filter-sterilized stock solution (1.66 M), in Fernbach flasks (2.8L). These were grown at 30°C for 48 hours shaking at 150 rpm before harvesting cells by centrifugation at 4000 rpm and 4°C for 30 min and decanting the supernatant. Cell pellets were frozen and lyophilized before extracting with DCM and

sonicating for 30 min. The DCM extract was removed by centrifugation at 4000 rpm, 4˚C

for 30 min. Residual DCM was removed by drying the pellets under N_2 gas. A second

extraction with 1:1 acetone:EtOAc followed the same procedure of sonication,

centrifugation, and drying. The final extraction with 1:1 ACN:H_2O was performed

similarly and in duplicate, with both extractions combined. The organic solvent was

removed by rotovap and remaining aqueous extract frozen and lyophilized. Typically,

about 1.2 g of dry extract was obtained per 24 L of culture. Extractions of more than 24 L

of combined cultures resulted in poor compound recovery due to impaired

chromatographic resolution.

Extract was either resuspended in water after weighing or dried directly onto 230-

400 mesh grade 60 silica gel and added atop a silica flash column containing a 10-fold

excess of silica gel relative to the estimated dry extract weight (i.e. 12 g of silica per

extract from 24 L of culture), pre-equilibrated with $CHCl_3$. The column was then eluted

with three column volumes each of $CHCl_3$, EtOAc, 1:1 EtOAc:MeOH, MeOH, and 1:5

H_2O:MeOH. For fractions containing MeOH and H_2O, a piece of filter paper was folded

and held tightly in place at the mouth of the column to filter out any silica particulate co-

eluting with the more polar solvents. The MeOH and 1:5 H_2O:MeOH elutions were

pooled and concentrated to dryness by rotavap and/or lyophilization. Dried silica column

elutions were resuspended in 30% MeOH in H_2O and loaded onto a second column of 4 g

Bakerbond 40 μm C_{18} resin per 24 L pooled culture, pre-equilibrated with MeOH, 50%

MeOH:H_2O, and 10% MeOH:H_2O. The C_{18} column was then eluted with five column

volumes each of 10%, 30%, 50%, and 100% MeOH.

The latter two C_{18} flash column elutions were pooled and dried, then resuspended in no more than 5 ml total volume. HPLC was performed using a Luna 5μ Phenyl-Hexyl 250 x 10 mm column (Phenomenex) running a 30% ACN/ 70% 0.1% TFA in H_2O isochratic method at 4 ml/min, with peak A eluting at ~14.3 min and peak B eluting at ~16.1 min. Retention times tended to increase with repeated injections, and peak shapes broadened with injections of higher volume. Organic solvent was removed from HPLC fractions by rotovap and the remaining acidic water removed by lyophilization. Typical yields by weight for 16 L of culture were ~0.6 mg of peak A and ~1.5 mg of peak B with some variability from batch to batch. Both peaks consisted of desmethyl, deslysinoalanine divamide, with peak A representing the 20-mer peptide sequence and peak B consisting of a mixture of three different extra-proteolytic species.

3.6.11 Chemical formation of lysinoalanine. HPLC-purified divamide intermediates were suspended in 0.1 M Tris buffer pH 10.8 and incubated at 37°C overnight. To remove Tris buffer and salt, the suspension was dried, resuspended in 10% MeOH in H_2O, and loaded onto a small plug of Fluka C_{18} resin (90 Å pore size) in a Pasteur pipette column pre-equilibrated as described above for the large-scale C_{18} flash column. The plug was eluted with 10% MeOH (2-5 ml), 50% MeOH (1.2 ml), and 100% (1.2 ml) MeOH. The 50% and 100% MeOH elutions were pooled.

The initial confirmation of lysinoalanine formation was carried by desulfurization of MeLan residues using *in situ* Ni_2B, prepared from $NiCl_2$ and $NaBH_4$ as described for other lanthipeptides.[26-30] In our hands, the conditions that worked best were as follows: peptide (150 μg, relative to amount estimated by weight) was suspended in of 7:1 MeOH:H_2O (500 μl) with $NiCl_2$ (1-2 mg) pre-dissolved in a 1.5 ml screw-cap reaction

vial with stir bar. About a five-fold excess of dry $NaBH_4$ was added directly to the vial, which was immediately capped and place in an oil bath at 50˚C for 30 min. The liquid, containing black Ni_2B precipitate, was transferred to a 1.7 ml microcentrifuge tube and centrifuged to remove the majority of precipitate. The supernatant was dried *in vacuo* and desalted on a C_{18} Pasteur pipette column as described above, washing with 10% MeOH in H_2O and eluting in 50% and 100% MeOH. The latter two elutions were combined and suspended in 75% MeOH and subjected to LC/MS and LC/MS/MS analyses.

3.6.12 Expression and purification of DivMT. The expression plasmid pRSFDuet-DivMT was transformed into BL21(DE3) *E. coli*, and individual colonies were picked to inoculate eight wells of a 24-well plate, each containing LB (6 ml) with kanamycin (50 µg/ml), which was shaken at 150 rpm and 30˚C overnight. The wells were pooled and used to inoculate cultures of 2xYT (1 L) media with kanamycin (50 µg/ml) in Fernbach flasks (2.8 L), with 5 ml seed culture per flask. Cultures were shaken at 150 rpm and 30˚C until an OD_{600} of 0.7 was reached, at which time cells were induced with IPTG (1 mM). The temperature was then reduced to 18˚C and the cultures allowed to incubate overnight. Cells were harvested by centrifugation at 4000 rpm and 4˚C and cell pellets of 2 L combined culture were frozen. Pellets were thawed on ice and suspended in lysis buffer (35 ml; 0.5 M NaCl, 25 mM imidazole, 10% glycerol) with lysozyme (0.4 mg/ml), PMSF (1 mM), $MgCl_2$ (10 mM) and DNaseI (<1 mg) freshly added, using a glass stir rod. The mixture was sonicated on ice for three cycles of 2 min pulses at 40% amplitude (4 s pulse, 10 s delay), then centrifuged at 13 krpm and 4˚C for 30 min. The supernatant was filtered using a 0.45 µm syringe filter onto Ni-NTA resin (2 ml; Qiagen) pre-equilibrated with lysis buffer and allowed to rock at 4˚C for 30 min to equilibrate. The

column was rinsed with lysis buffer (100 ml) followed by transition buffer (50 ml; 0.1 M

NaCl, 25 mM MOPS, 25 mM imidazole, 10% glycerol, pH 8.0) and equilibrated another

10 min, then eluted with 10 ml each of the following imidazole concentrations, all

prepared from transition buffer: 50 mM, 100 mM, 200 mM, and 500 mM imidazole. The

resin was given a 10 min equilibration between each buffer change. SDS-PAGE was used

to assess the elution time and purity of the 32.6 kDa protein (**Figure S3.7**). The 200 mM

elution was then dialyzed in dialysis buffer (1 L; 0.1 M NaCl, 25 mM MOPS, 10%

glycerol, pH 8.0) at 4˚C. The buffer was exchanged four times within 24 hours, adding

BME at a final concentration of 5 mM to the first two buffer exchanges only. The

concentration of the protein prep as determined by UV absorbance at 280 nm using a

Nanodrop 2000 was 3.77 µM/0.123 mg/ml (MW = 32566.8 Da, $\varepsilon_{280\ (reduced)}$ = 46215).

    3.6.13 Enzymatic N-trimethylation. Desmethyl divamide was suspended to an

estimated concentration of 2 mg/ml in nanopure water. An analytical scale methylation

assay was prepared by combining the desmethyl divamide solution (1 µl) with 1 µl of

each of the following in 20 µl total volume: 0.5 M MOPS (pH 7.5), 50 mM DTT, 100

mM SAM, and 3.77 µM DivMT. Reactions were mixed and incubated at 37˚C for 1-2

days using a thermocycler with a heated lid. MeOH was added to 20% of total volume

and the entire sample subjected to LC/MS analysis. For large-scale DivMT reactions, the

entire desmethyl divamide solution was used (typically around 300 µl at 2 mg/ml was

obtained per 24 L batch) and the remaining reagents scaled accordingly. The reaction

solution was mixed and aliquoted into PCR tube strips, 80 µl per tube, and was incubated

at 37˚C in a thermocycler for 2 days. All reactions were then pooled and dried *in vacuo*,

then desalted using a C$_{18}$ Pasteur pipette column, eluting this time with 10% MeOH (5-10

ml) followed by of 30%, 50%, and 100% MeOH (1.2 ml each). The latter two elutions

were pooled and subjected to HPLC using a Gemini C6-Phenyl column (150 x 4.60 mm,

110Å, 3 µ) running an isochratic method of 27% ACN/ 68% 0.1 M NaCl/ 0.1% TFA

aqueous buffer, yielding divamide A at t_R = 18.7 min. Residual extra-proteolytic

divamide intermediates eluted earlier, at t_R = 16.7 min. Organic solvent was removed

from the collected peak prior to lyophilization. Dried material was suspended in 10%

MeOH and desalted using a C_{18} Pasteur pipette column as described before.

3.6.14 Bioassays. T-cell lymphoid CEM 1a2 tat/rev++ cells are a clone of

CEM^{Tart} tat/rev++ cells that apoptose upon infection with HIV.[31] These cells were

employed for all HIV cytoprotection assays with divamide compounds. Cell lines were

maintained in RPMI-1640 media supplemented with sodium bicarbonate (2 g/L), fetal

bovine serum (20%), and antibiotic-antimycotic solution (100 X, GIBCO, Grand Island,

NY, USA), and were cultivated in 5% CO_2 at 37˚C. Assays were run in 96-well round-

bottom plates. Cells (200 µl) were seeded at a concentration of 50,000 cells/ml (12,500

cells/well) and sample compounds/controls applied in 1 µl DMSO. Azidothymidine

(AZT) was used as a positive control at a stock concentration of 0.1 and 0.05 mg/ml in

DMSO. HIV tat/rev- was then applied at an appropriate titer. After a four-day incubation,

yellow MTT (3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide; 10 µl, 5

mg/ml) was added and the plates incubated for 1-2 hours. Plates were then centrifuged,

the supernatant removed, and DMSO (100 µl) added to solubilize the MTT, now reduced

and blue in color due to the metabolic activity of live cells. The plate was analyzed by

plate reader at 590 nm and the results tabulated.

To best compare sample responses between HIV infected and uninfected states,

data were first normalized to the uninfected control, adjusted by the normalized HIV

infected control response, and then scaled for ease of comparison. In this way, data could

be interpreted as the percent cell survival relative to HIV infected cells. To obtain the

normalized sample response, S_N, equation 3.1 was used:

$$S_N = \frac{S_{avg} - B_{avg}}{U_{avg}} \tag{3.1}$$

Variables represent average assay response values from three similarly treated wells of

sample (S), blank (B), and uninfected control (U). Infected cells with and without AZT

treatment were also normalized to obtain A_N and I_N, respectively. The highest AZT

response of the two concentrations employed was used. Normalized data were then

scaled the fit a range of 0-100, with the 0 and 100 % thresholds representing the infected

and uninfected states, respectively. The scaling factor, s, was calculated using equation

3.2, while sample and AZT control data were scaled using equation 3.3:

$$s = \frac{100}{U_N - I_N} \tag{3.2}$$

$$S_S = (S_N - I_N)s \tag{3.3}$$

Under these conditions, the scaled uninfected control response U_S was equal to 100 %

relative survival.

Dose-response curves were generated and fit using KaleidaGraph:

$$y = \left(D_1 + \frac{A_1 - D_1}{1 + 10^{(x - \log C_1)B_1}}\right) + \left(D_2 + \frac{A_2 - D_2}{1 + 10^{(x - \log C_2)B_2}}\right) \tag{3.4}$$

For a typical sigmoidal curve, A and D represent y-min and -max values, respectively,

while x = log concentration, C is the IC_{50}, and B the Hill coefficient. Because equation

3.4 combines two unrelated dose-dependent effects, A and D values do not accurately

reflect y-min and -max values. *B* and *C*, however, are independent of *y*, the percent relative survival. C_1 and C_2 provide the IC_{50} and CC_{50} values, while B_1 and B_2 represent the steepness of their respective curve. Some sample curves showed a dip in the cytoprotection curve and could not be fit using the above model (**Figure 3.12**). For these compounds, IC_{50} and CC_{50} values were estimated by eye (**Figure 3.3-A**).
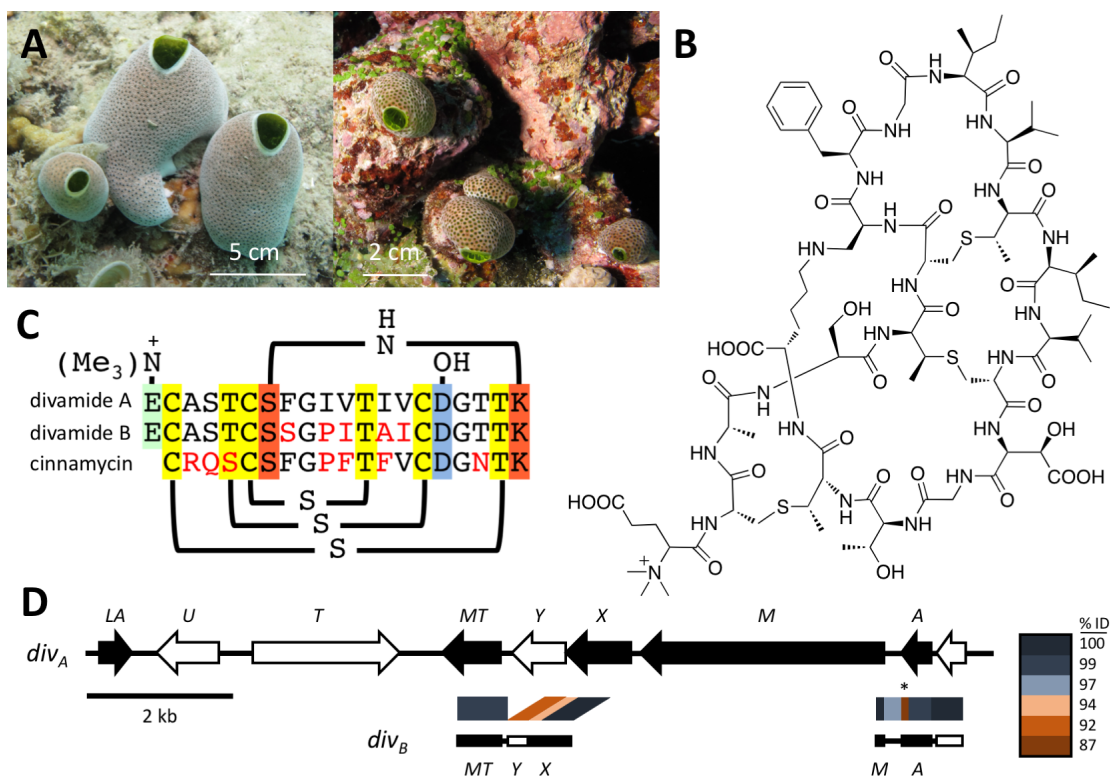
Figure 3.1. Discovery of the divamides from *Didemnum molle.* A) The divamide-producing *Didemnum molle* tunicates E11-036 (left) and E11-037 (right) were found living side-by-side in the Eastern Fields of Papua New Guinea, approximately 5-10 m underwater. The two tunicates exhibit distinct morphologies that may represent genetically distinct species.[32] B) The chemical structure of the anti-HIV lanthipeptide divamide A, produced by E11-036, elucidated by NMR (**Figure S3.2; Tables S3.1, S3.2**) and chiral GC/MS (**Figure S3.3; Table S3.3**). C) Divamides A-C display an identical scaffold to cinnamcyin, incorporating lanthionine (yellow), lysinoalanine (red), and ß-hydroxyaspartic acid (blue) in the same sequential positions, but include an additional N-terminal N-trimethylated residue (green). Differences in sequence from divamide A are shown in red text. D) The E11-036 *div* biosynthetic gene cluster (*div_A*; top) exhibits nearly 100% identity with the partial E11-037 *div* gene cluster (*div_B*; bottom), with the greatest disparity occurring within the hypervariable core of *divA* (*) (**Tables S3.4, S3.5**).
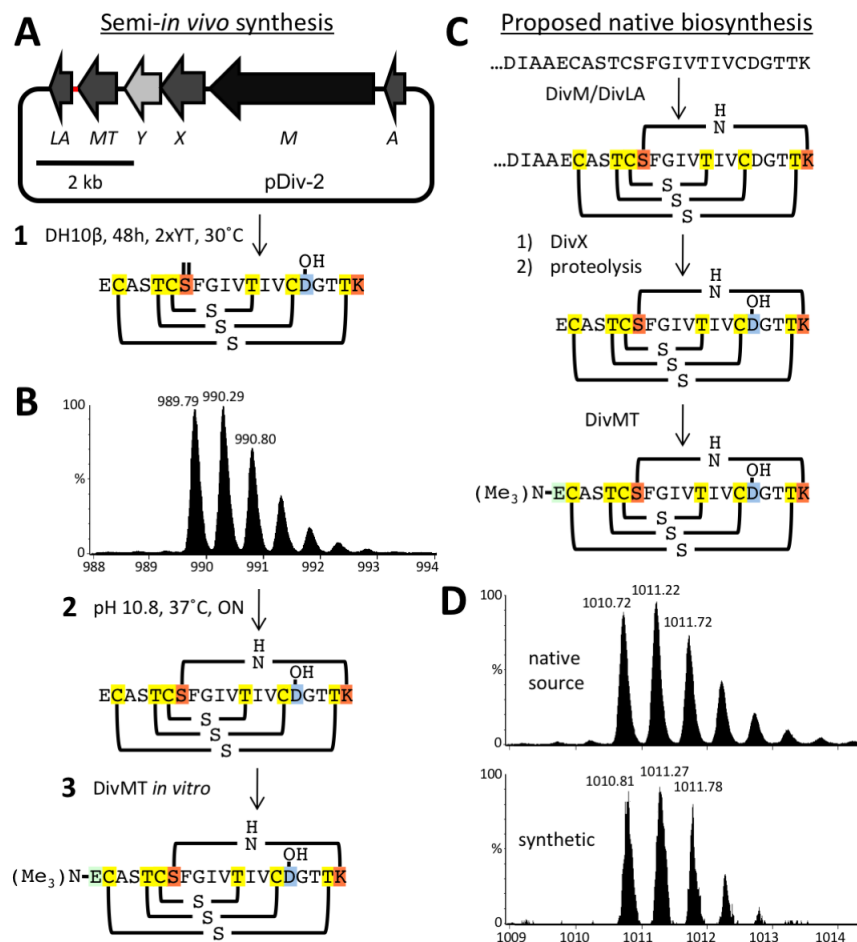
Figure 3.2. Synthesis of divamide A. The semi-*in vivo* synthesis of divamide A (left column) involves *in vivo*, chemical, and enzymatic steps. A) The divamide A expression plasmid, pDiv2, enables *in vivo* expression of a divamide intermediate that incorporates three MeLan residues, Hya, and dehydroalanine (Dha) by DH10ß *E. coli* (Step 1). B) The desmethyl, deslysinoalanine intermediate, purified from *E. coli* ($m/z = 989.8$), is incubated in base to induce formation of Lal (Step 2), the desmethyl divamide A product of which is the substrate of SAM-dependent N-trimethyltransferase DivMT (Step 3). C) The putative native biosynthetic route of divamide A requires DivM, DivX, proteolysis by an endogenous protease, and DivMT. DivLA may be involved in Lal formation in the cyanobacterial tunicate endosymbiont *Prochloron didemni*.[8] D) Comparison of tunicate-derived (top) and synthetic divamide A (bottom) by mass spectrometry ($m/z = 1010.8$) and NMR (**Figure S3.11**) confirm the two are chemically identical.
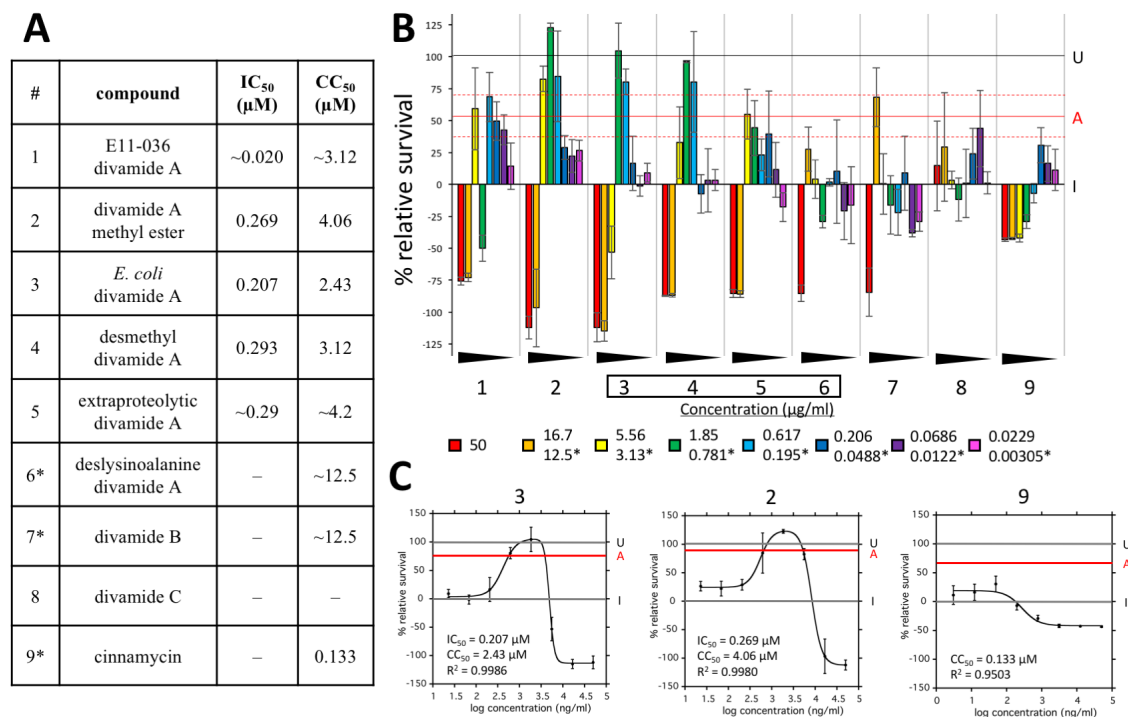
Figure 3.3. Biological activity of the divamides. Dose-response curves were generated for all divamide species generated alongside cinnamycin using the same cytoprotection assay used in the initial screening of tunicate extracts. A) Nine compounds were assayed in total. IC_{50} values correspond to anti-HIV activity, while CC_{50} values reflect cytotoxicity. For some compounds, values were estimated by eye (~) due the difficulties associated with fitting the resulting atypical curves (samples 5-7). A dash indicates the absence of that particular biological response. B) The dose-response curve of each compound assayed is shown as a bar graph. I = response level of infected cells, U = response level of uninfected cells, A= response level of AZT. Dashed red lines represent the AZT response ± one standard deviation. Boxed compound numbers represent synthetic species. The concentrations of each sample follow the top values shown in the key, corresponding to a 1:3 dilution series, while those specifically indicated in the table with an * follow the bottom values, corresponding to a 1:4 dilution series. C) Individual dose-response curves were fit with the sum of two four-parameter IC_{50} equations to simultaneously determine the IC_{50} and CC_{50} values. A wider therapeutic window was observed for the divamide A methyl ester (2) than for divamide A (3), while cinnamycin exhibited no cytoprotection.
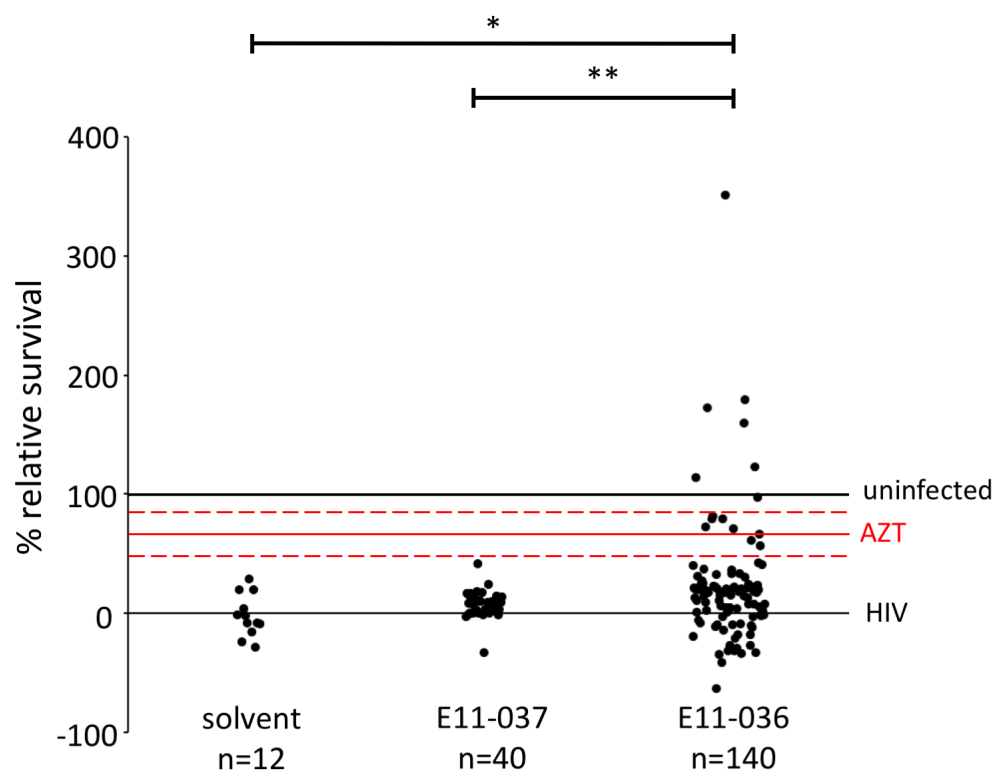
Figure S3.1. Anti-HIV screen of tunicate extracts. All fractions generated as described in the supplementary methods were screened in an HIV cytoprotection assay to identify active anti-HIV agents. The number of cells living four days after infection with HIV is represented by 0% relative survival, while 100% relative survival indicates uninfected cells. The average assay response of the positive control, AZT, is shown as a red line, with dashed lines representing the average response ± the standard deviation. Each point represents an individual fraction at a specific concentration. Fractions included both crude extracts and HPLC-purified compounds. A portion of fractions derived from E11-036 showed cytoprotection comparable to the effects elicited by AZT, while fractions from E11-037 and solvent controls did not display anti-HIV activity. Taking into account the assay results of all fractions from a given source, a student's t-test showed statistical significance between E11-036 and both control (*; p-value = 0.00362) and E11-037 (**; p-value = 0.0177) fractions. The large number of E11-036 data points represent multiple rounds of fractionation and repeats of assays, inflating the population size relative to the other two groups. This inflation should skew the t-test in favor of rejecting statistical significance, yet the data show significance nonetheless, indicating authentic anti-HIV activity from E11-036.

Figure S3.2. Natural divamide A NMR data. 600 MHz NMR data collected in D_2O include: A) ^1H spectrum, B) gradient correlation spectroscopy (gCOSY) spectrum, C) z-filter total correlation spectroscopy (zTOCSY) spectrum, D) nuclear Overhauser effect spectroscopy (NOESY) spectrum, E) gradient heteronuclear single quantum coherence (gHSQC) spectrum, and F) gradient heteronuclear multiple bond correlation (gHMBC) spectrum. 600 MHz data collected in 90% H_2O/ 10% D_2O include: G) jump and return (1-1) decoupling in the presence of scalar interactions (DIPSI) spectrum and H) 1-1 NOESY spectrum. NMR data collected in 70% CD_3OH/ 30% H_2O/ 0.1% TFA at 36˚C include: I) 500 MHz ^{13}C spectrum, J) 900 MHz 1-1 ^1H spectrum, K) 900 MHz 1-1 NOESY spectrum, L) 900 MHz 1-1 DIPSI spectrum, M) 900 MHz gradient carbon HSQC (gCHSQC) spectrum.

A
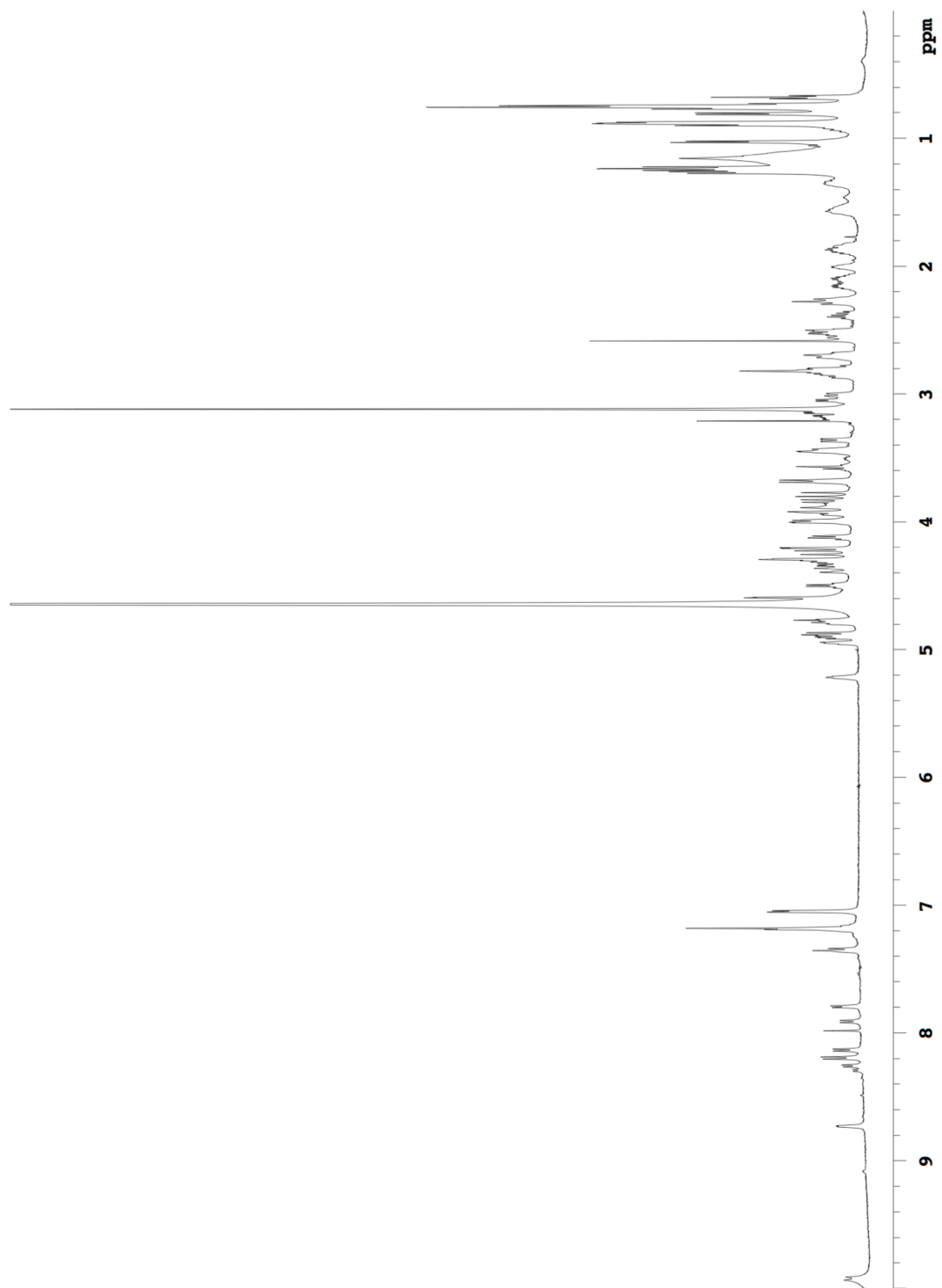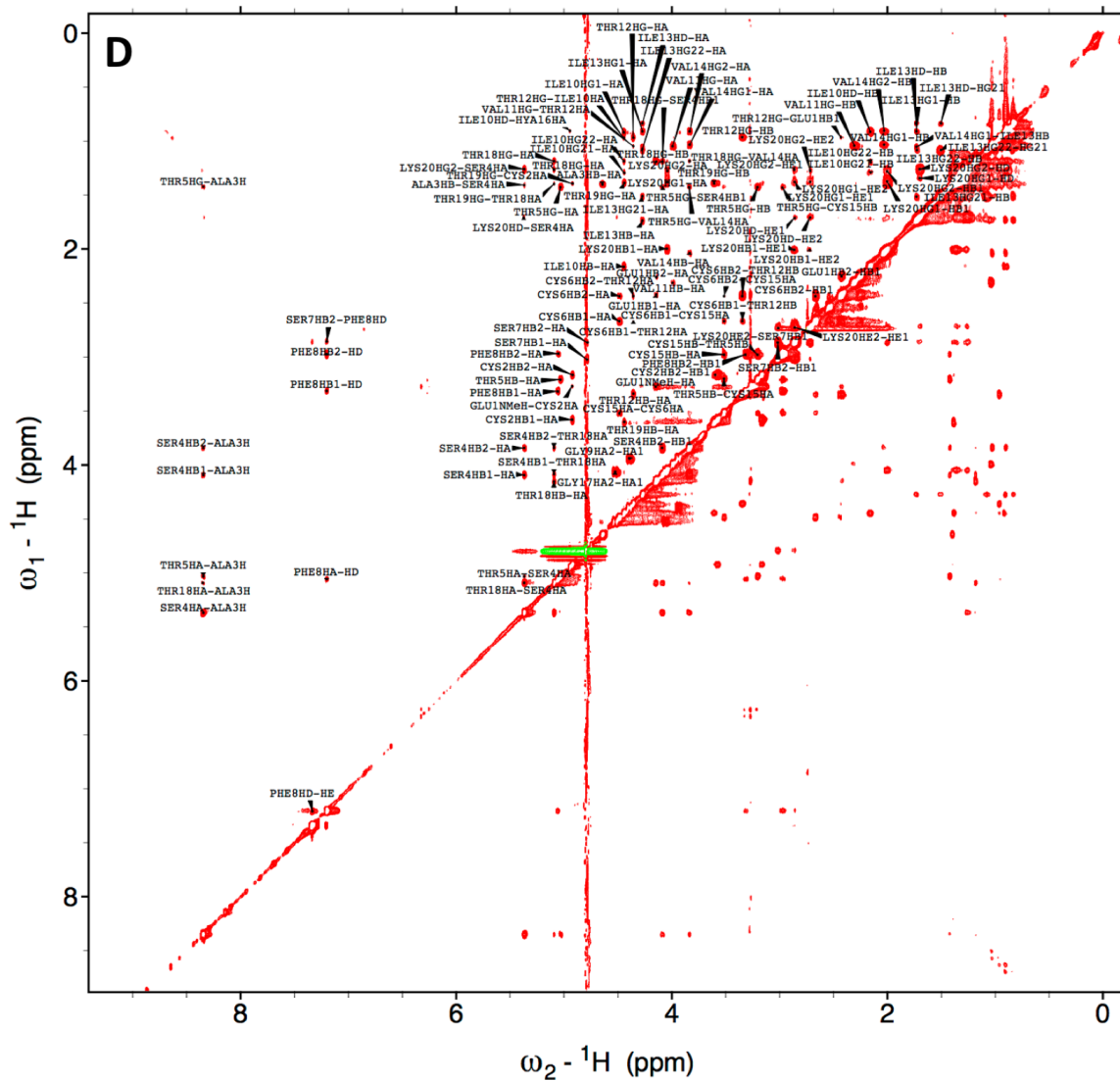
Figure S3.2 Continued
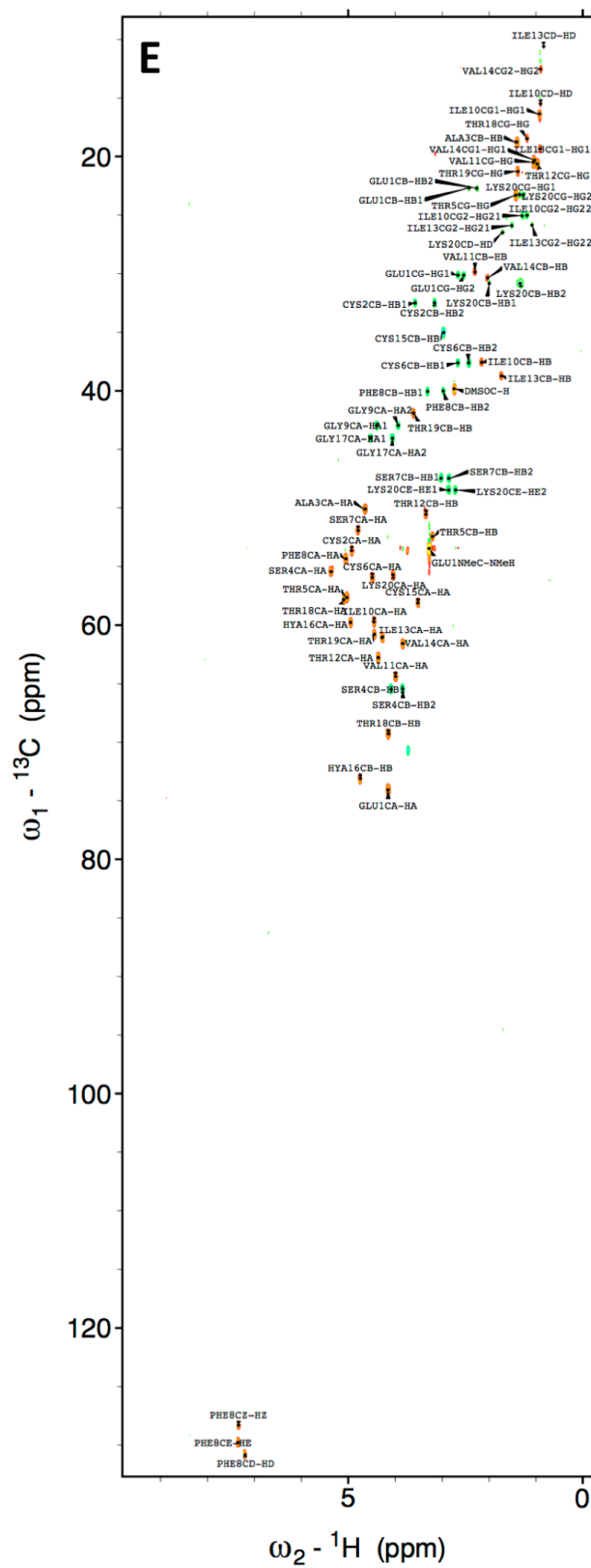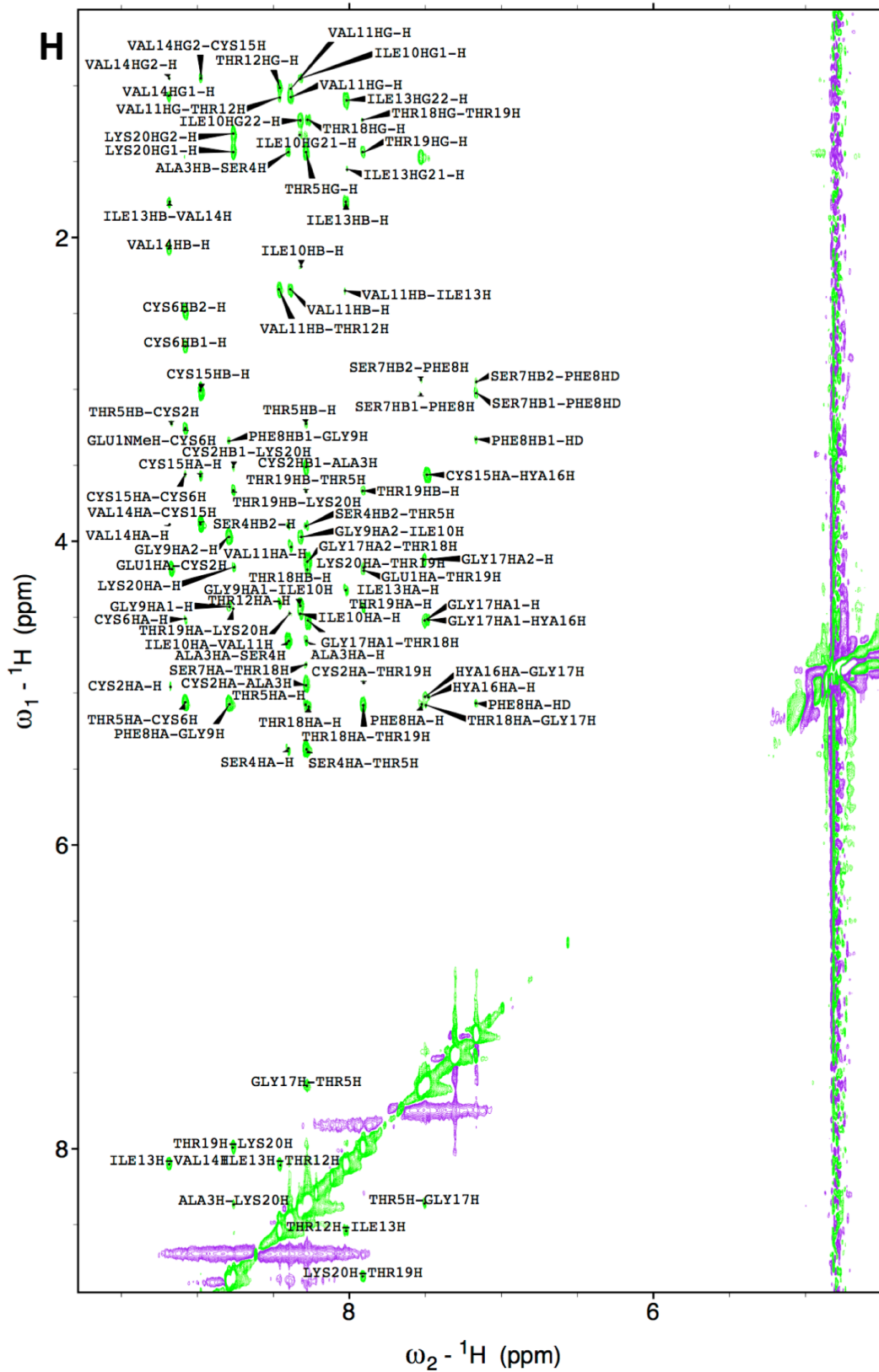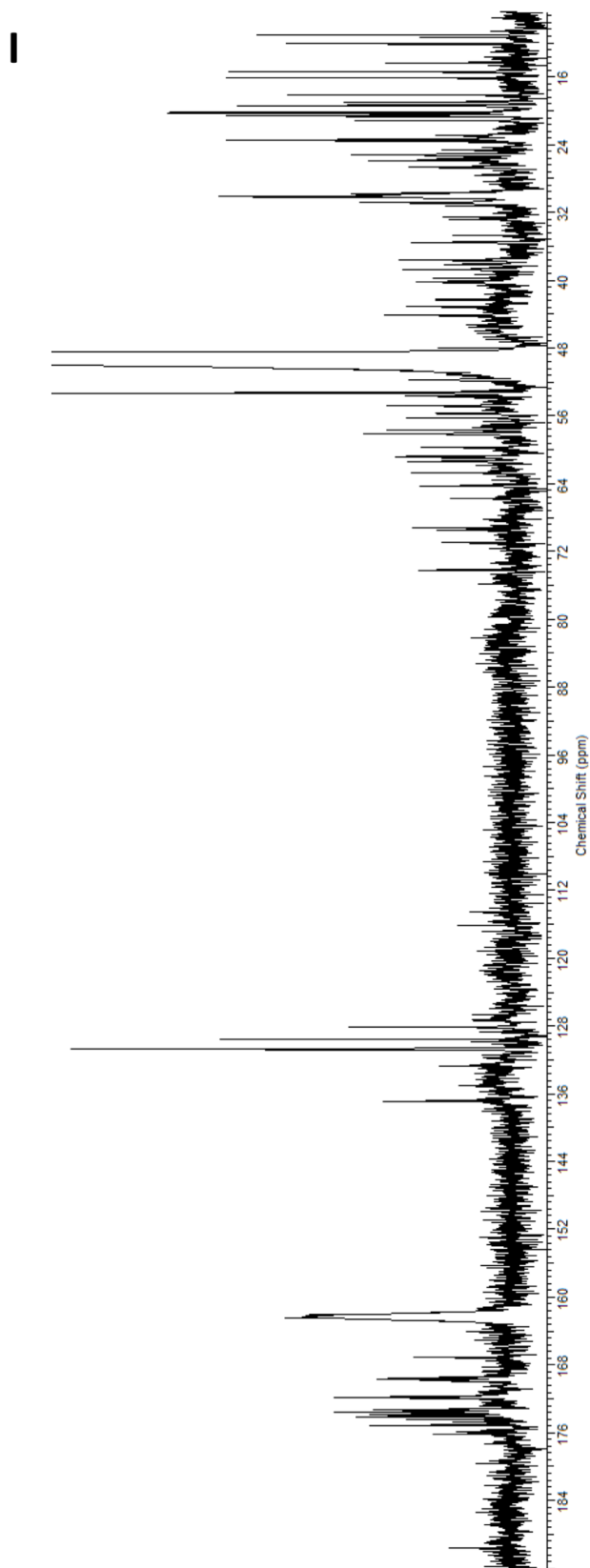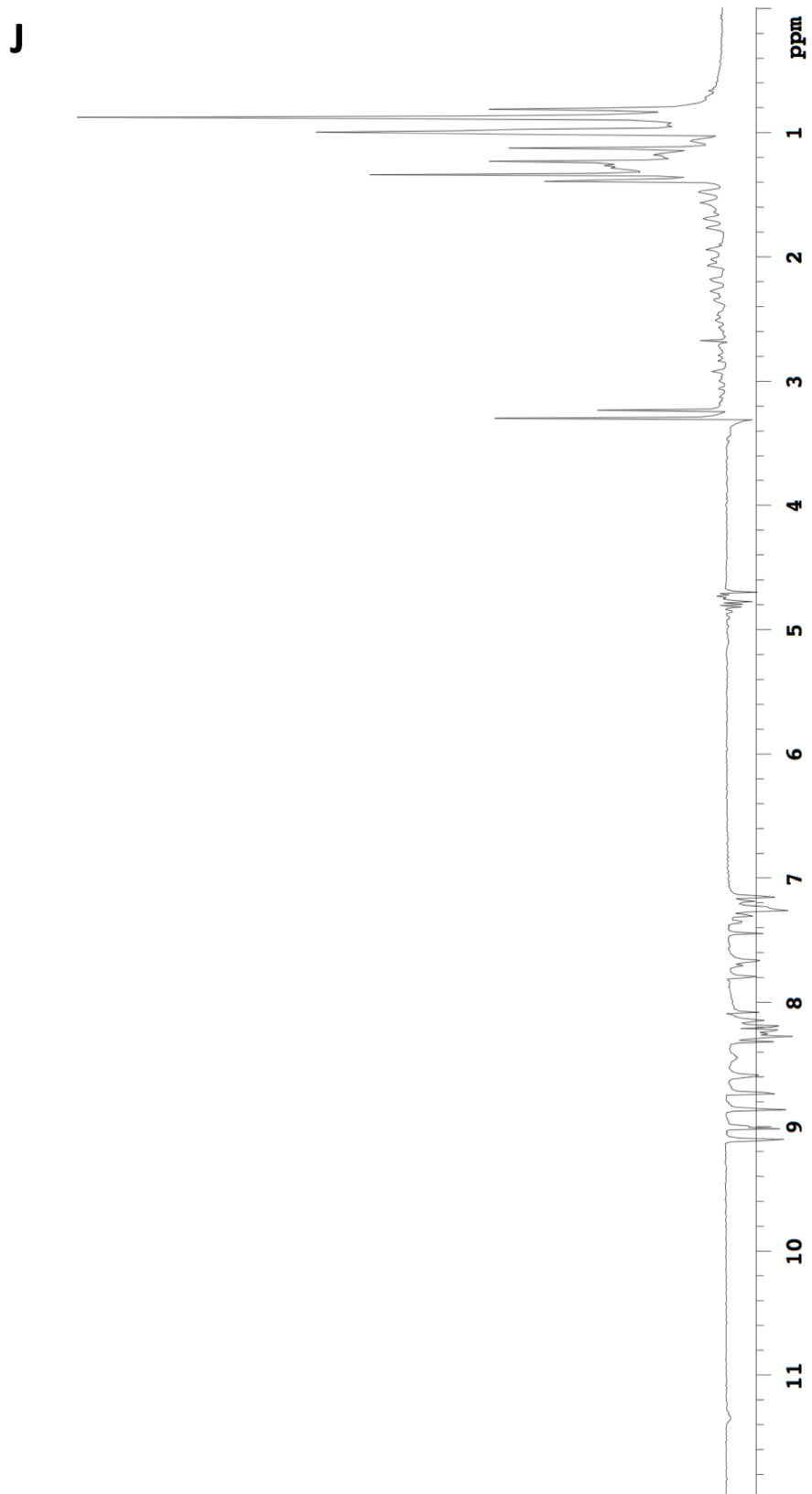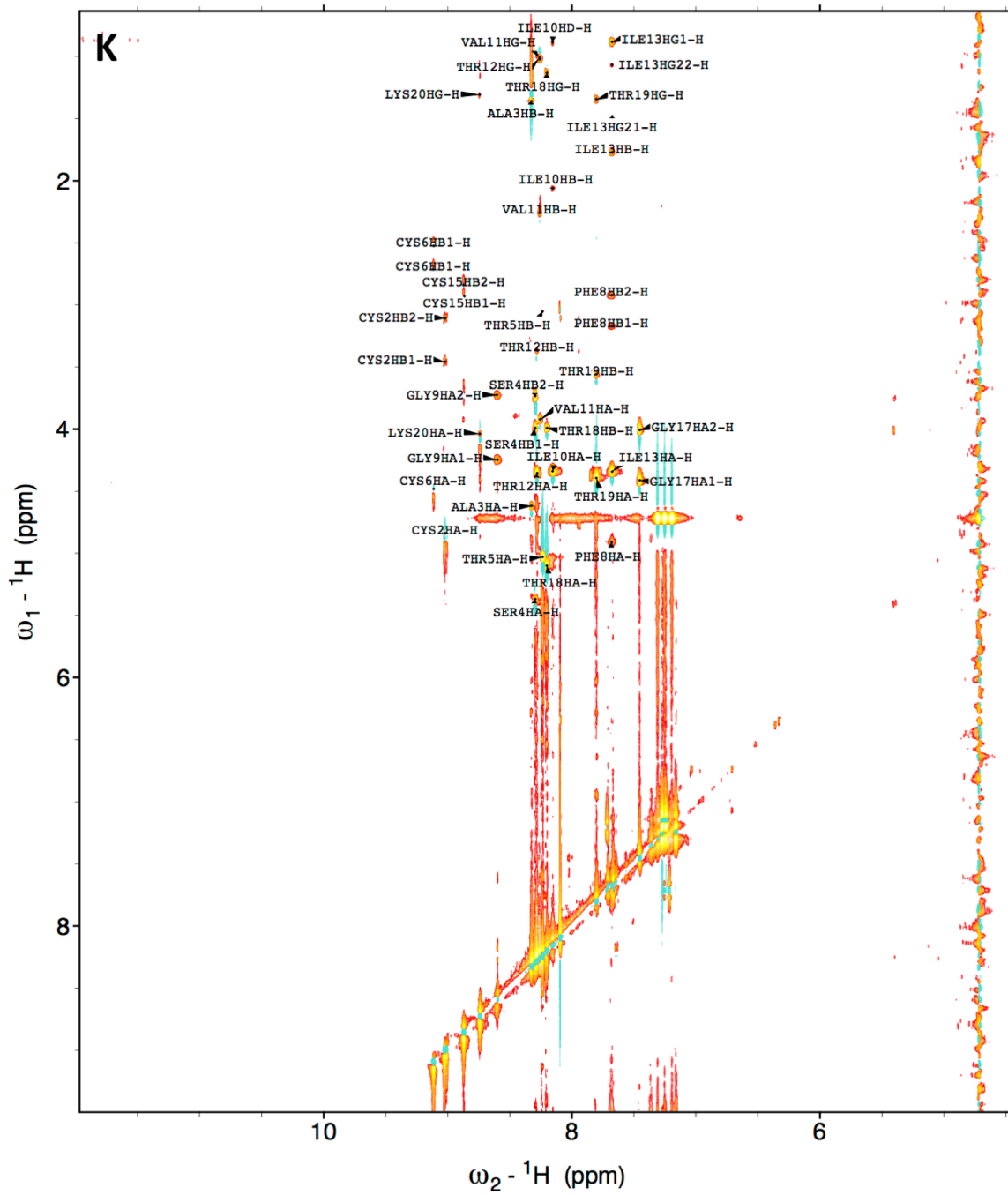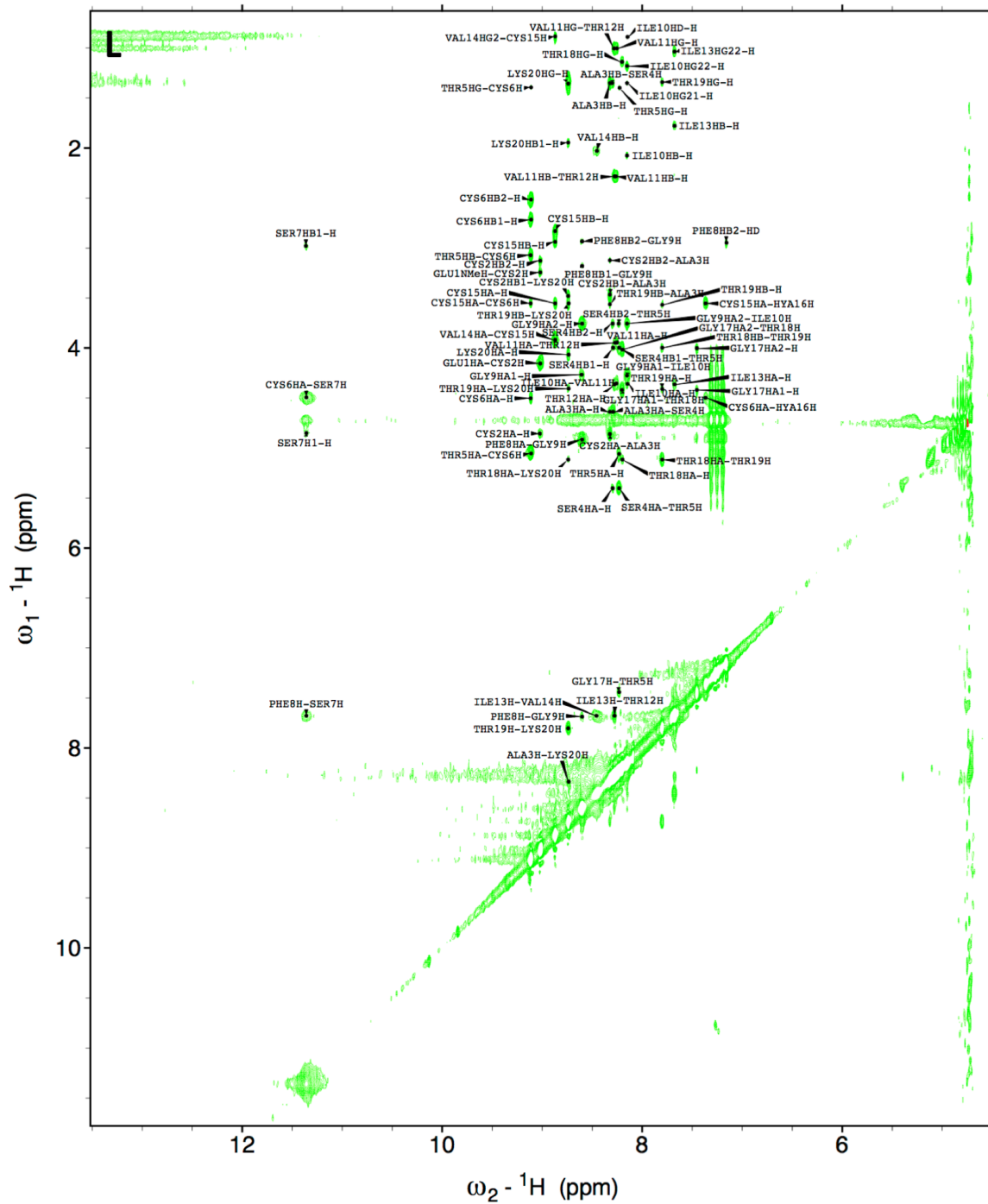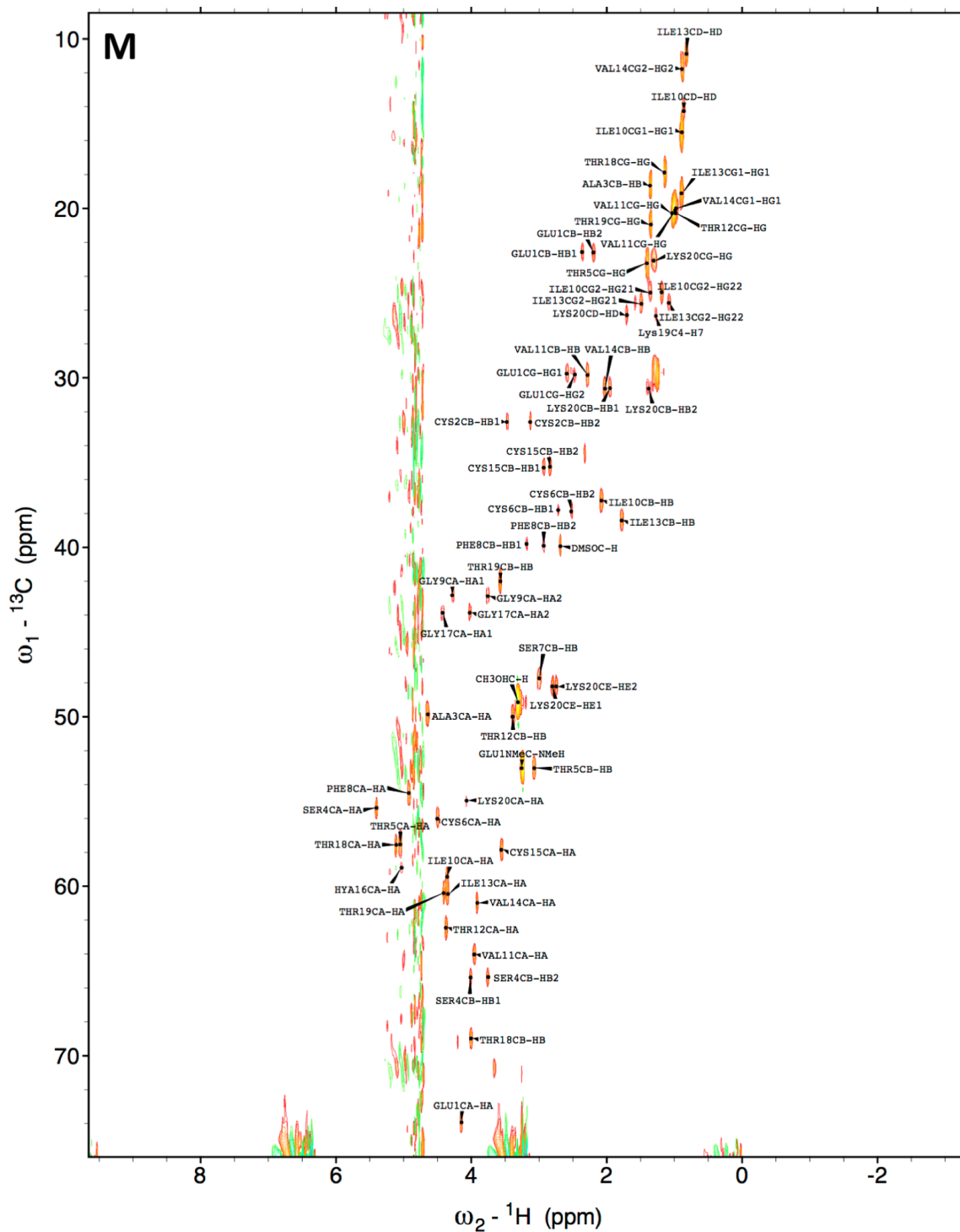
Figure S3.2 Continued

Figure S3.2 Continued

E

Figure S3.2 Continued

Figure S3.2 Continued

G

PHE8HB2-H
PHE8HB1-H
THR19HB-H
SER4HB2-H
GLY9HA2-H
VAL11HA-H
SER4HB1-H   THR18HB-H   GLY17HA2-H
GLY9HA1-H   THR12HA-H   ILE13HA-H
                              THR19HA-H
CYS6HA-H   ILE10HA-H   GLY17HA1-H
                              HYA16HB-H
CYS2HA-H
                              HYA16HA-H
THR5HA-H   PHE8HA-H
       THR18HA-H
          SER4HA-H

$\omega_1 - {}^1H$ (ppm)

$\omega_2 - {}^1H$ (ppm)

Figure S3.2 Continued

Figure S3.2 Continued

Figure S3.2 Continued

J

Figure S3.2 Continued

Figure S3.2 Continued

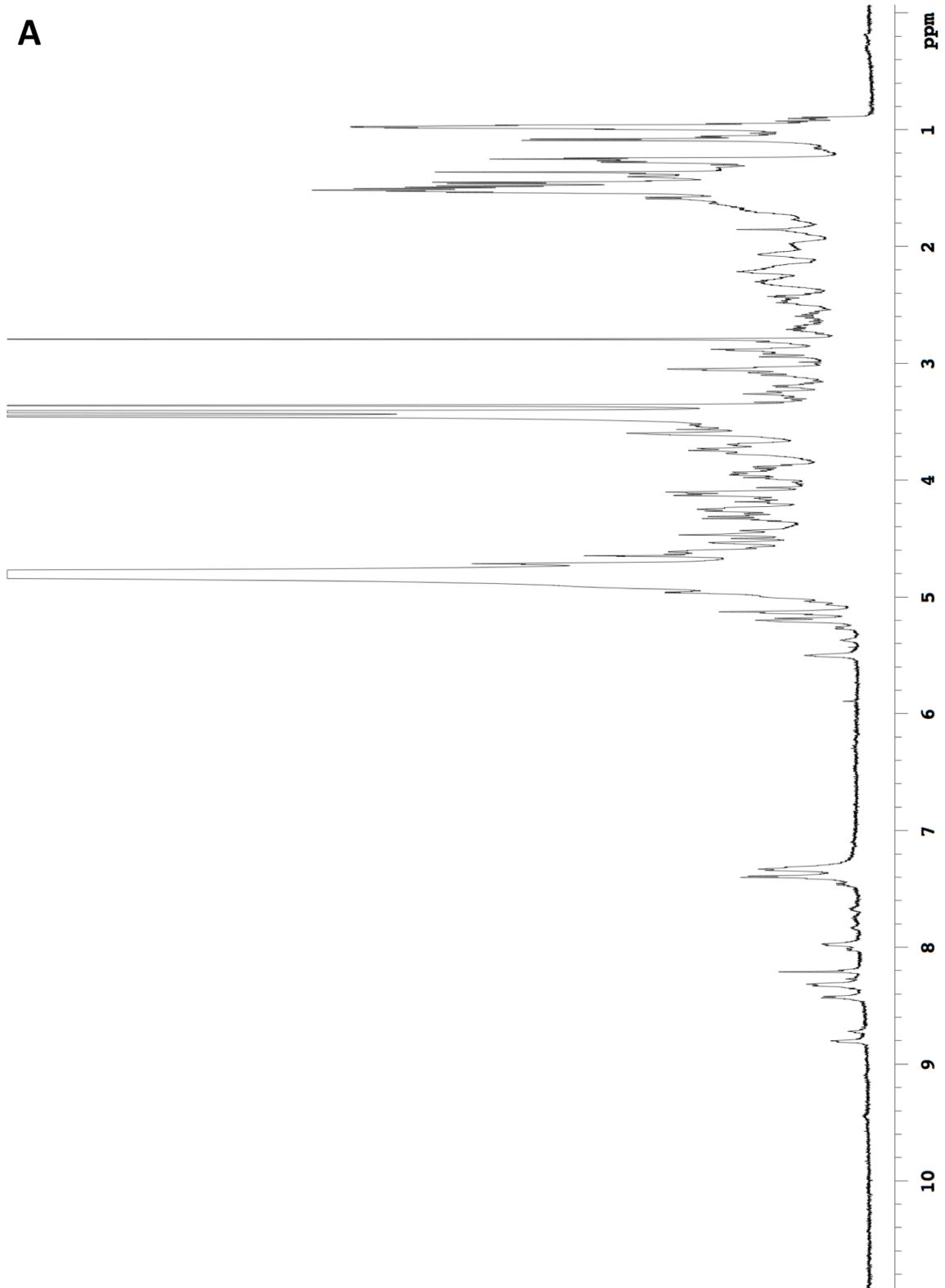**Figure S3.2 Continued**

Figure S3.2 Continued
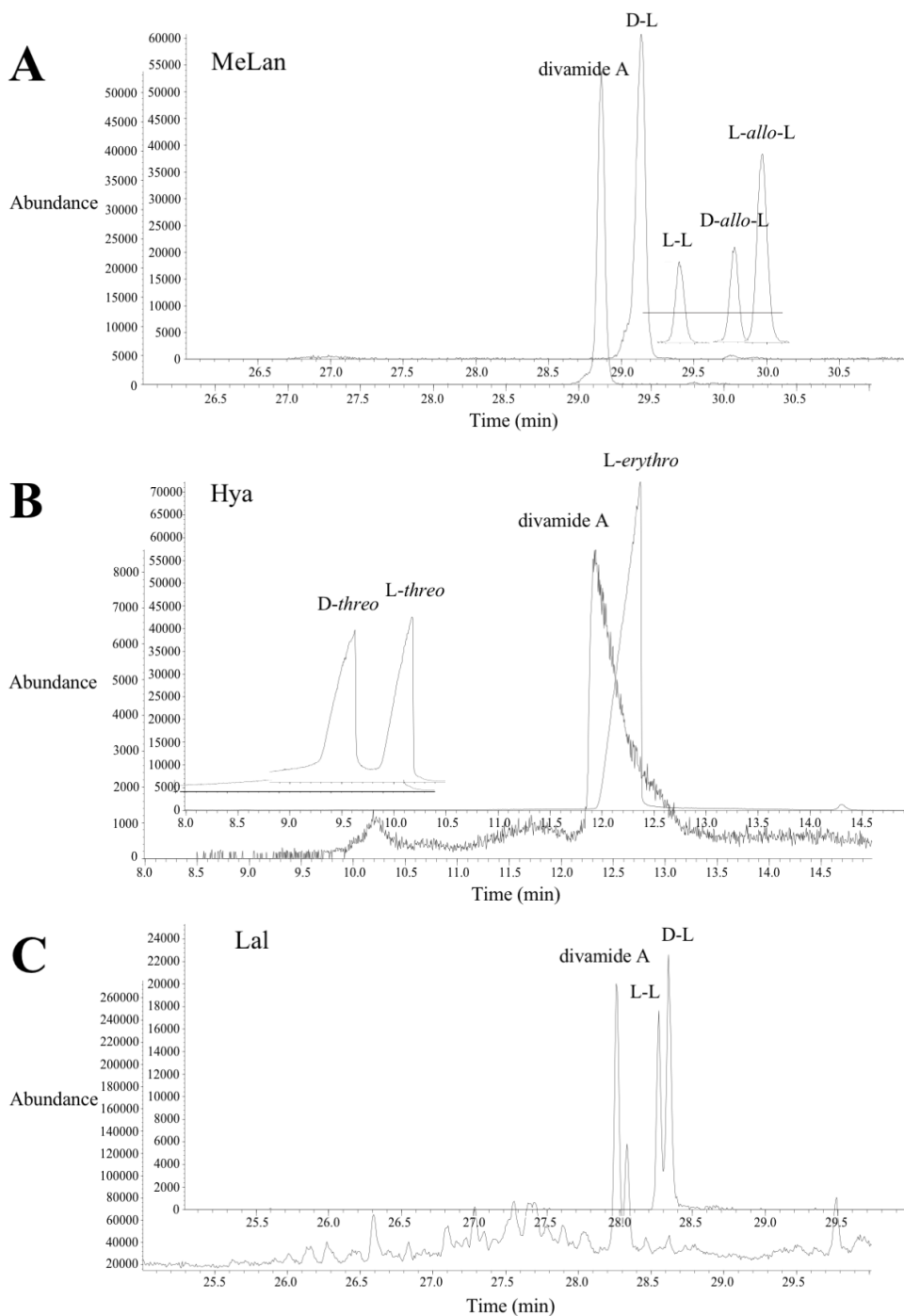
Figure S3.3. Chiral GC/MS chromatograms. Divamide A stereochemistry was determined by comparing GC/MS fragmentation patterns and retention times of acid hydrolyzed and chemically derivatized peptide with amino acid standards using a chiral stationary phase. A) Selected Ion Mode (SIM) chromatogram of MeLan showing alignment of the MeLan peak in divamide A with the DL-MeLan standard. B) SIM chromatogram of Hya showing alignment of the divamide A Hya peak with L-*erythro*-Hya. C) SIM chromatogram of Lal showing alignment of the divamide A Lal peak with LL-Lal using an achiral stationary phase.

Figure S3.4. Natural divamide B NMR data. 600 MHz NMR spectra collected in 70% CD$_3$OD/ 30% D$_2$O/ 0.1% TFA at 36˚C include: A) ^{1}H spectrum, B) gCOSY spectrum, C) zTOCSY spectrum, D) NOESY spectrum, E) gHSQC spectrum, and F) gHMBC spectrum. Data collected in 70% CD$_3$OH/ 30% H$_2$O/ 0.1% TFA include: G) 1-1 ^{1}H spectrum, H) 11-DIPSI spectrum, I) 1-1 NOESY spectrum, J) gCHSQC spectrum.
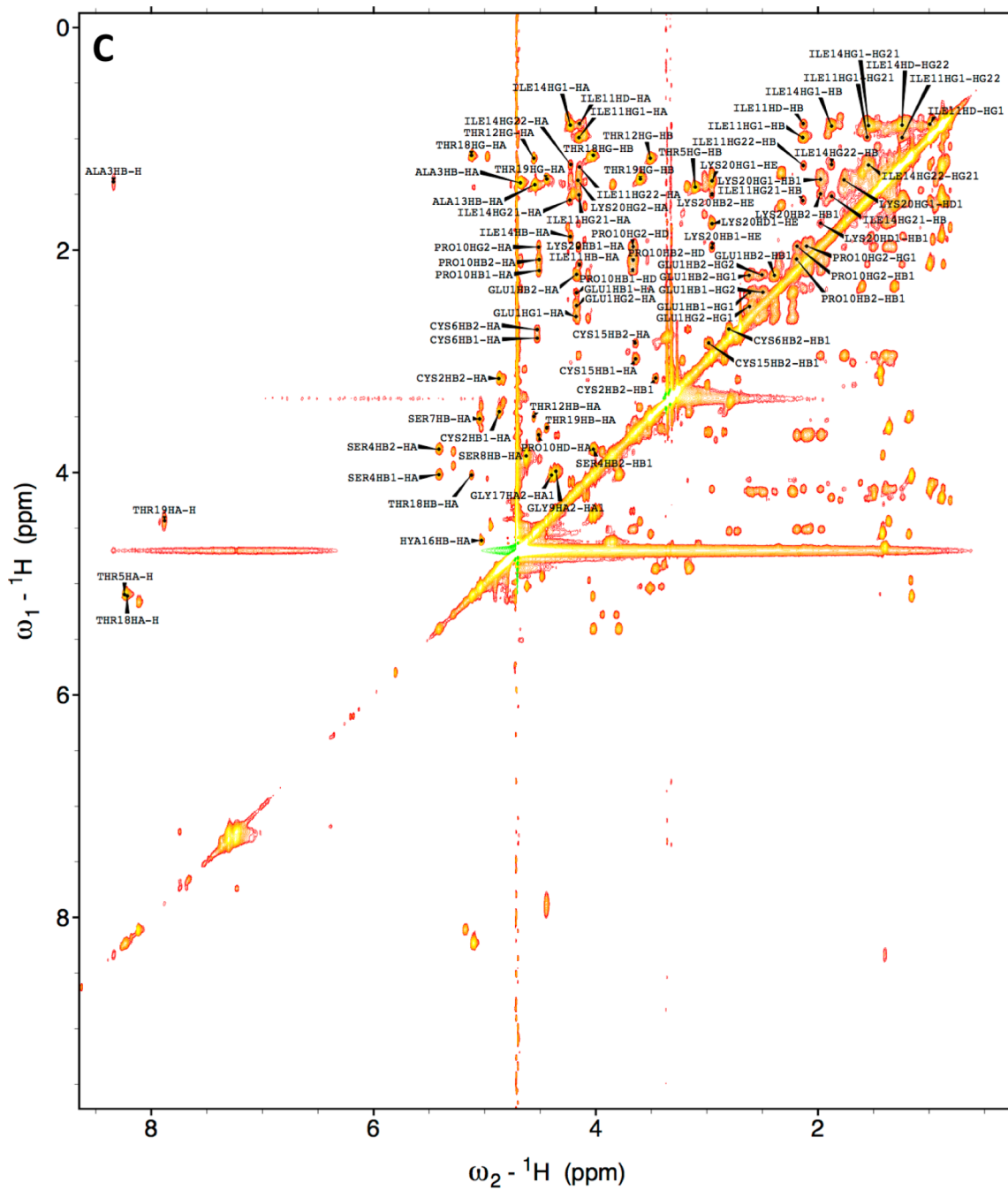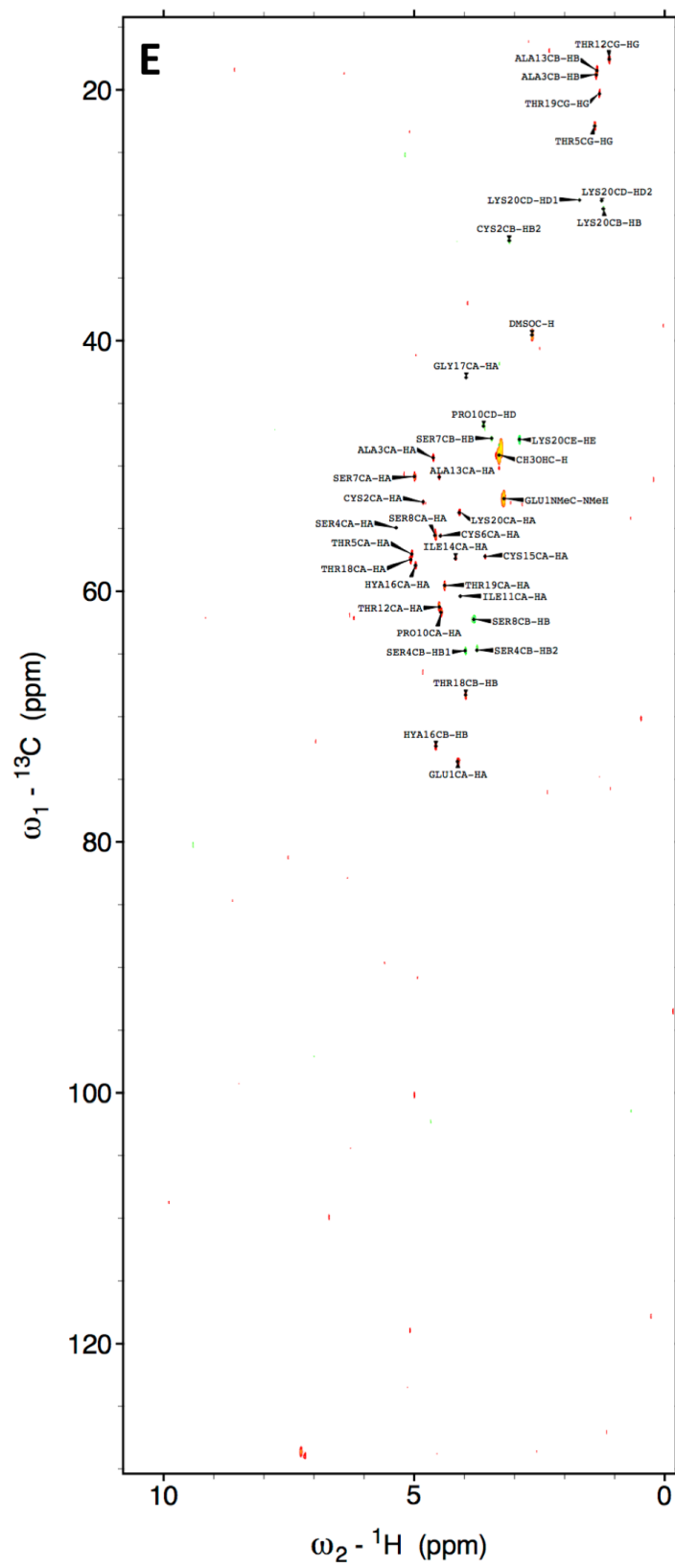
A

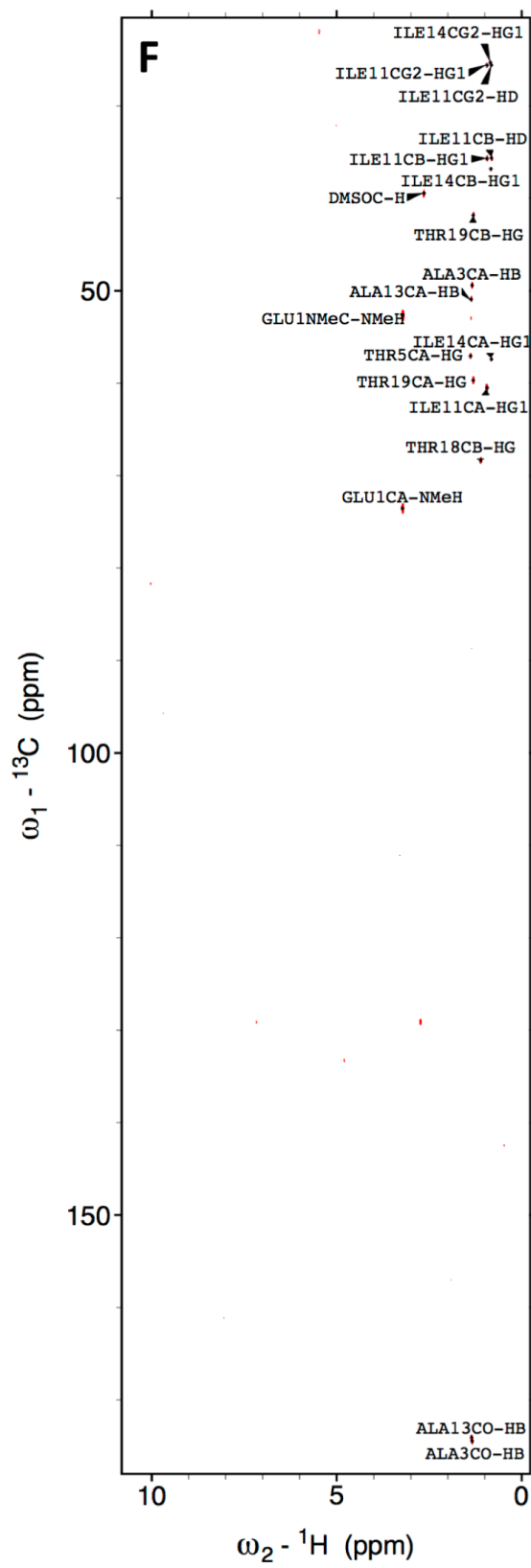**Figure S3.4 Continued**

Figure S3.4 Continued

Figure S3.4 Continued

Figure S3.4 Continued

Figure S3.4 Continued

G

ppm

Figure S3.4 Continued

Figure S3.4 Continued

Figure S3.4 Continued
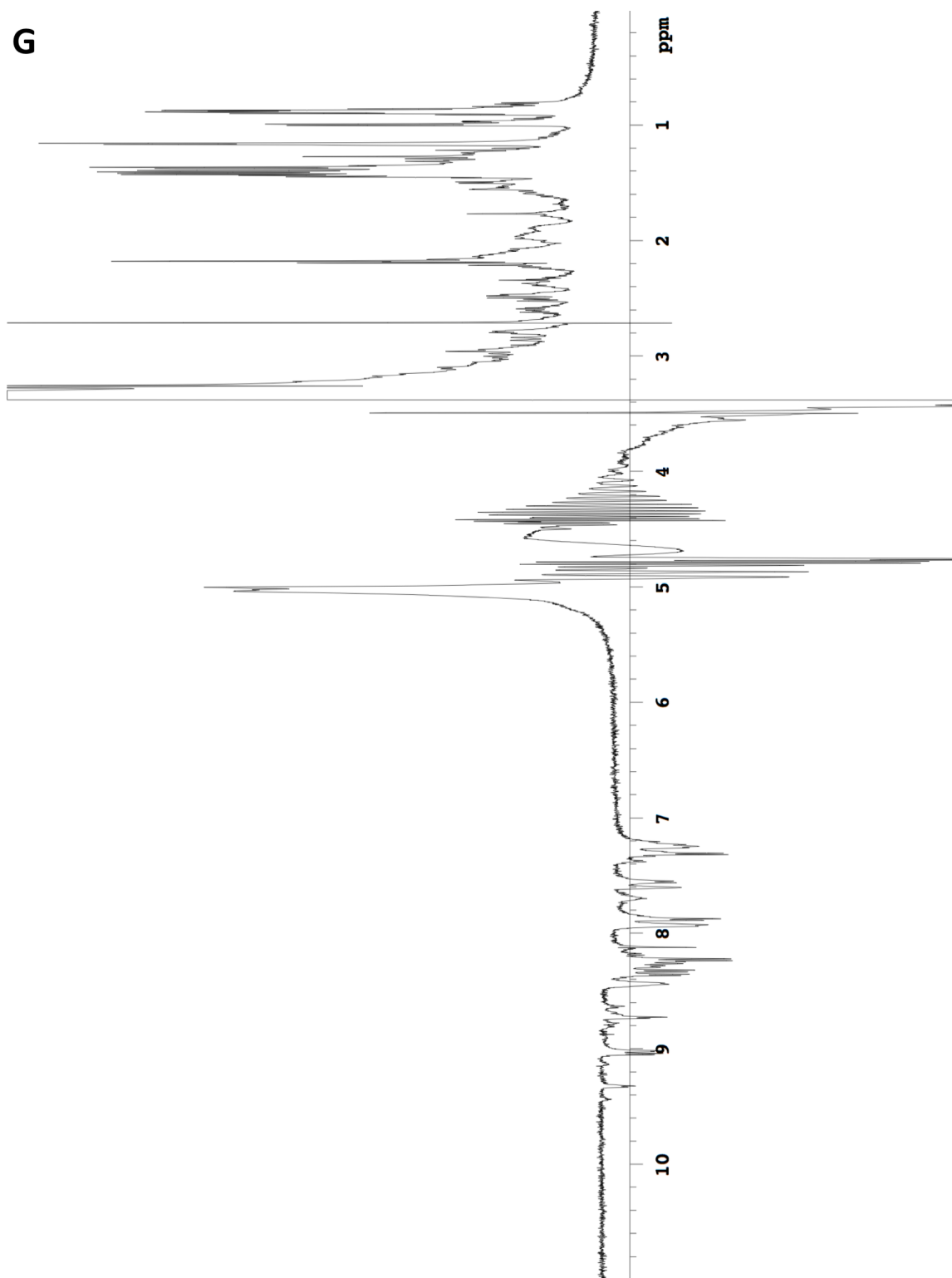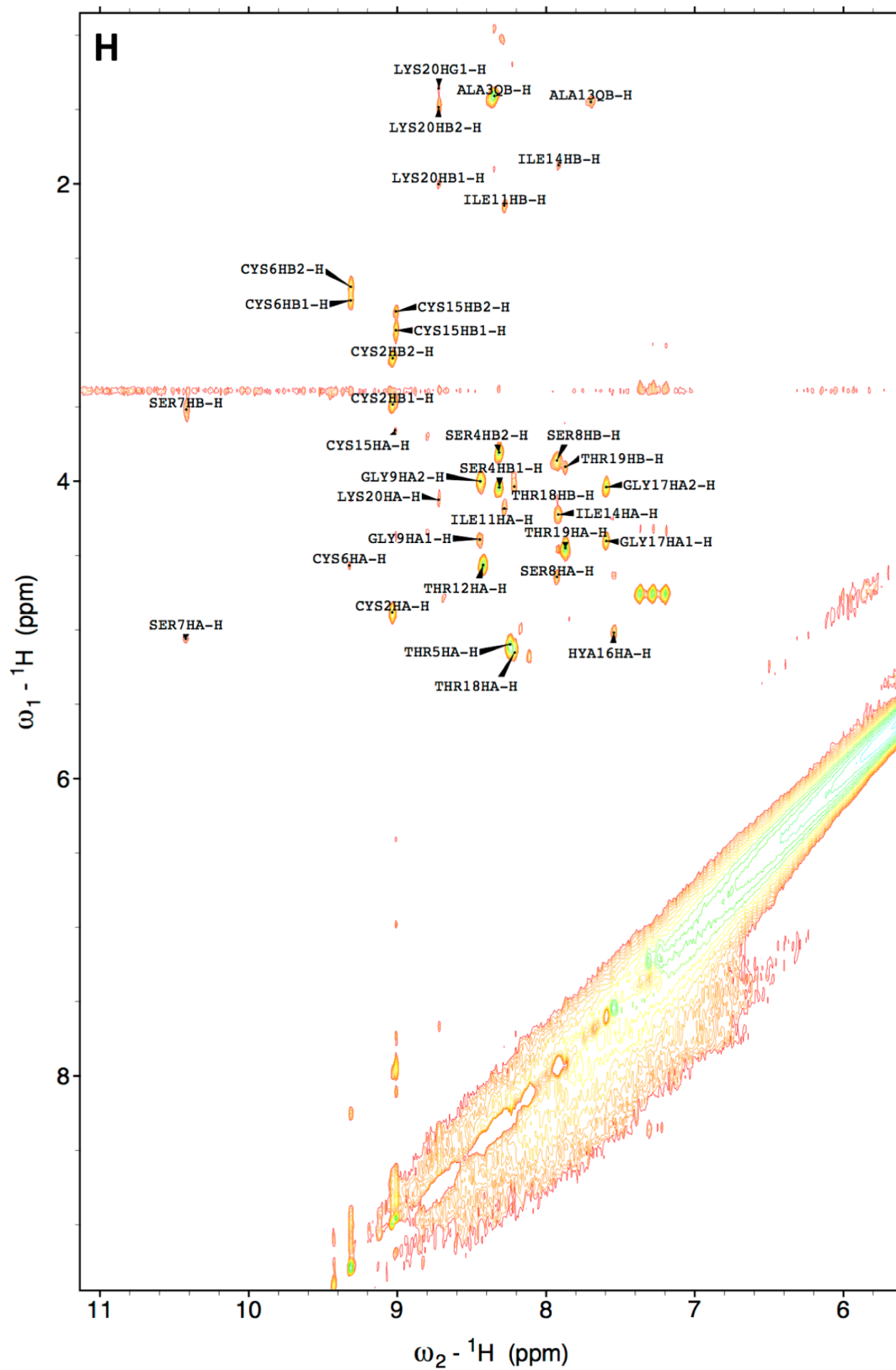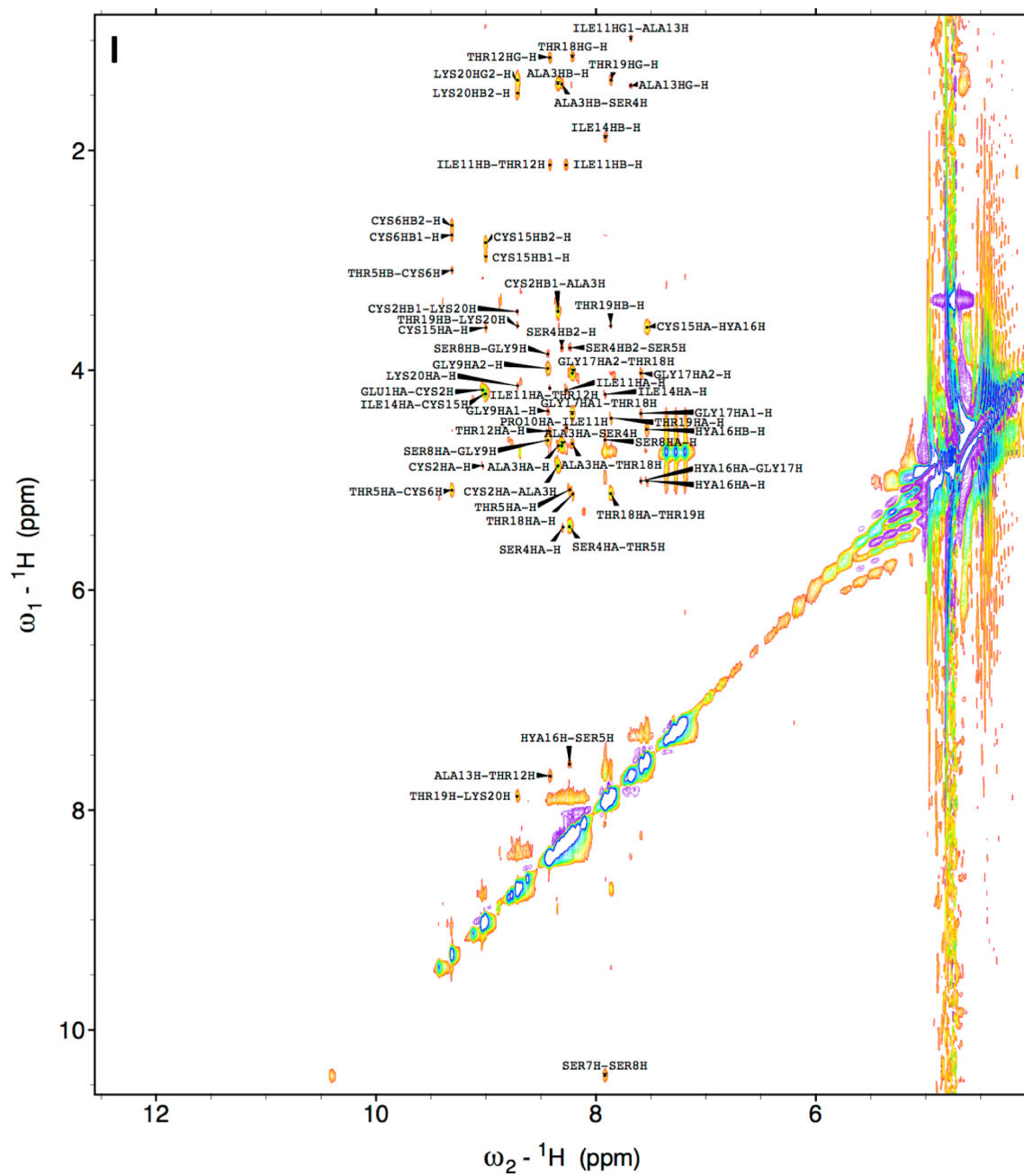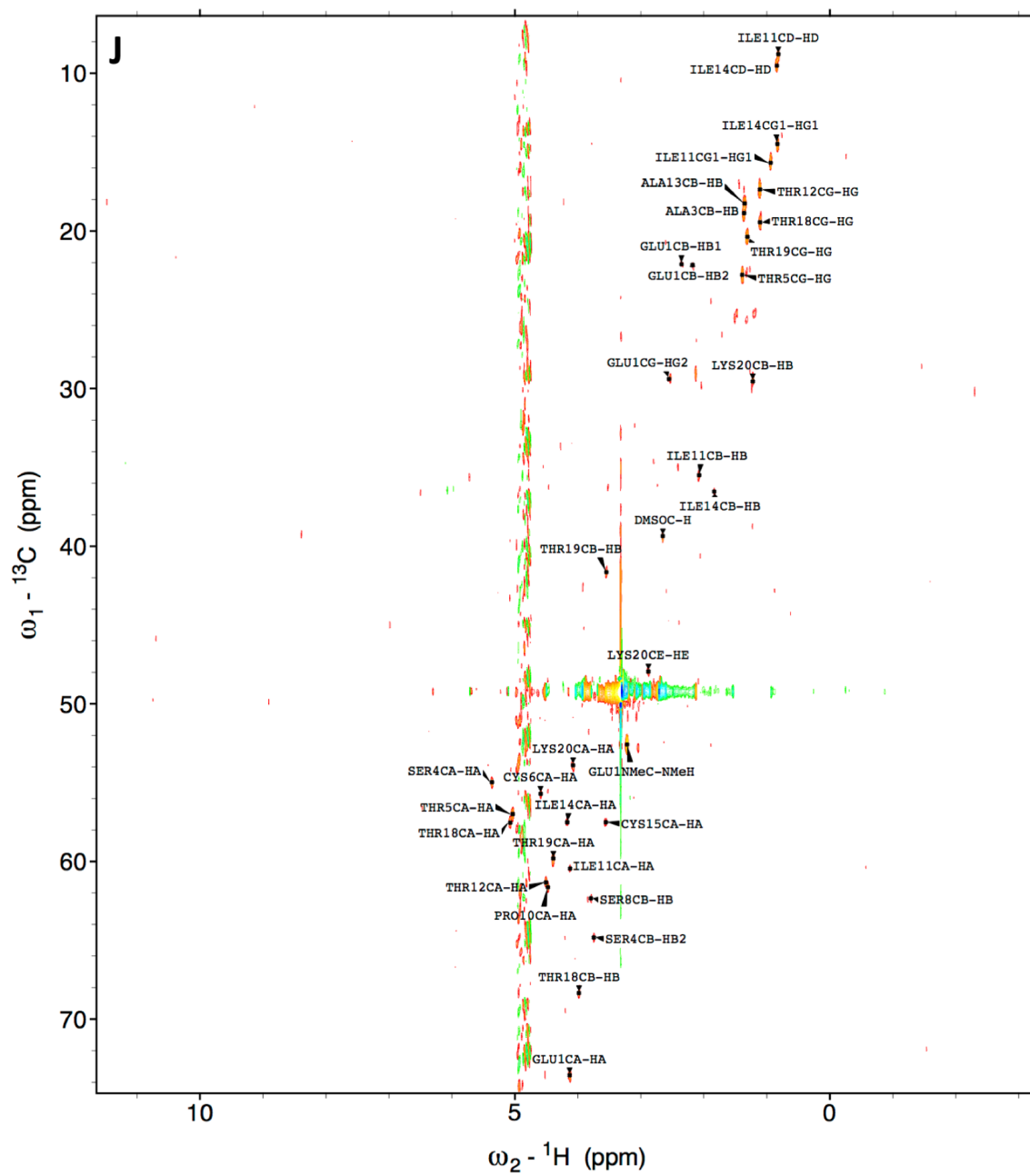
Figure S3.4 Continued

Figure S3.5. Natural divamide C NMR data. 600 MHz NMR spectra collected in 70% CD$_3$OD/ 30% D$_2$O/ 0.1% TFA at 36˚C include: A) ^{1}H spectrum, B) gCOSY spectrum, C) zTOCSY spectrum, D) NOESY spectrum, and E) gHSQC spectrum. F) Data collected in 70% CD$_3$OH/ 30% H$_2$O/ 0.1% TFA include: 1-1 ^{1}H spectrum, G) 11-DIPSI spectrum, H) 1-1 NOESY spectrum, and I) gCHSQC spectrum.
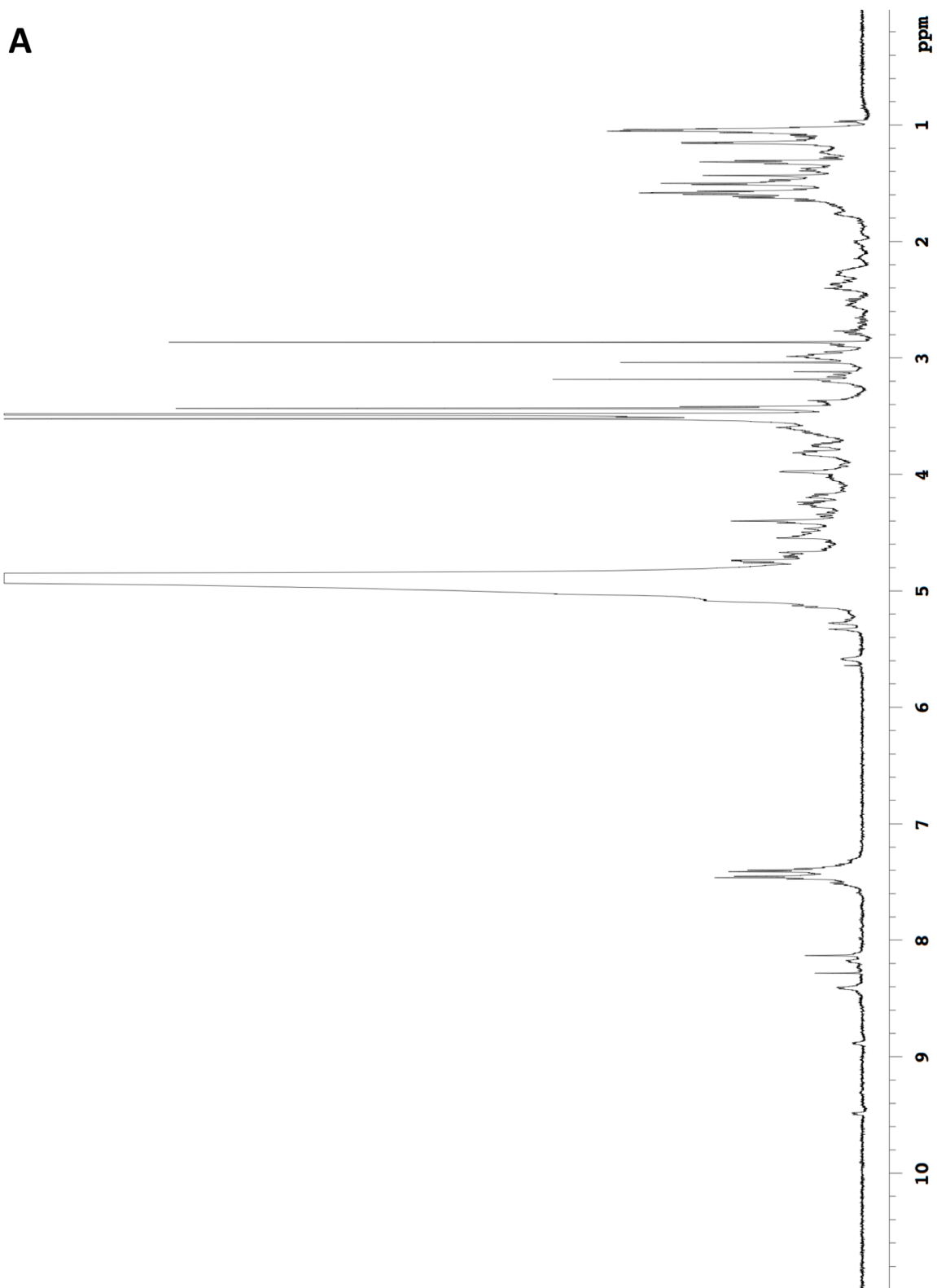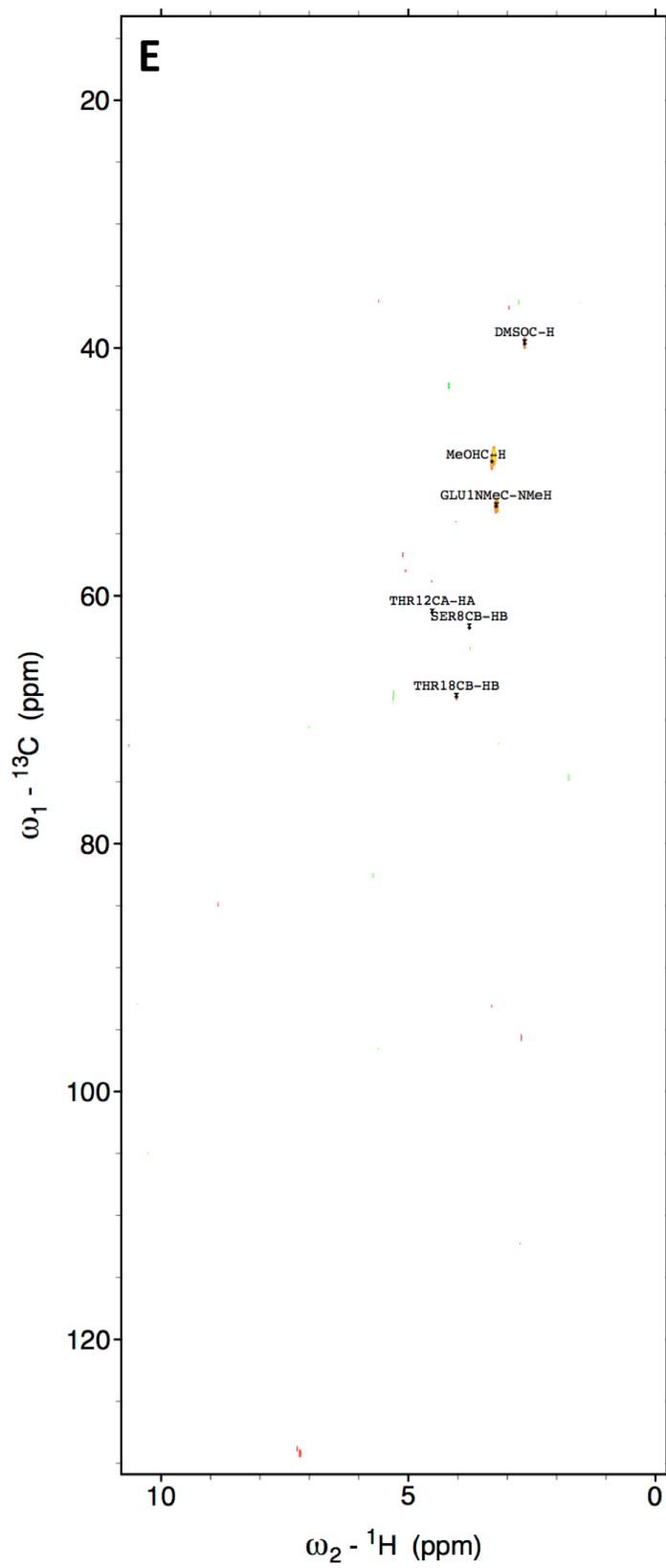
A

Figure S3.5 Continued

Figure S3.5 Continued

Figure S3.5 Continued

Figure S3.5 Continued

F
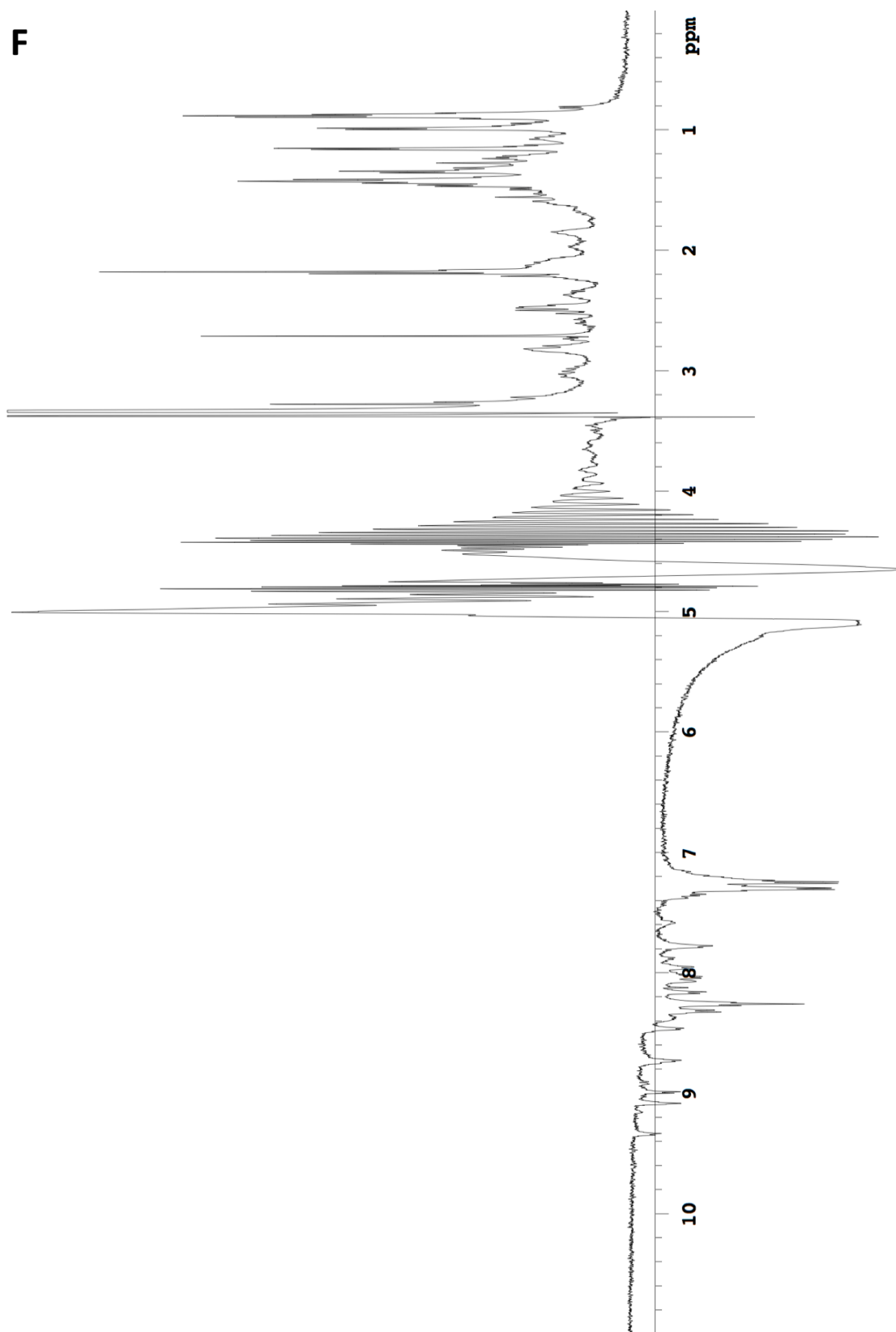
Figure S3.5 Continued
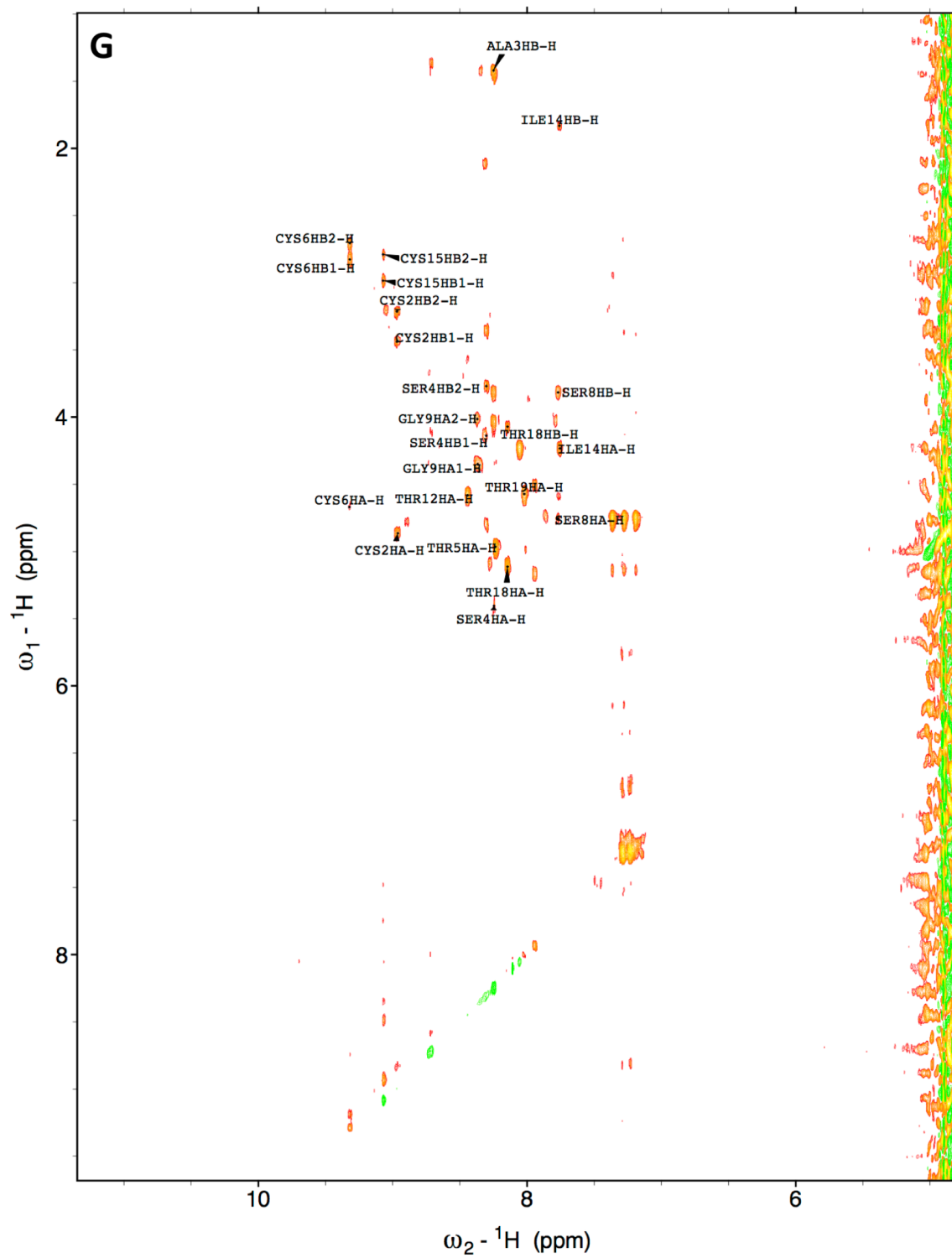
Figure S3.5 Continued
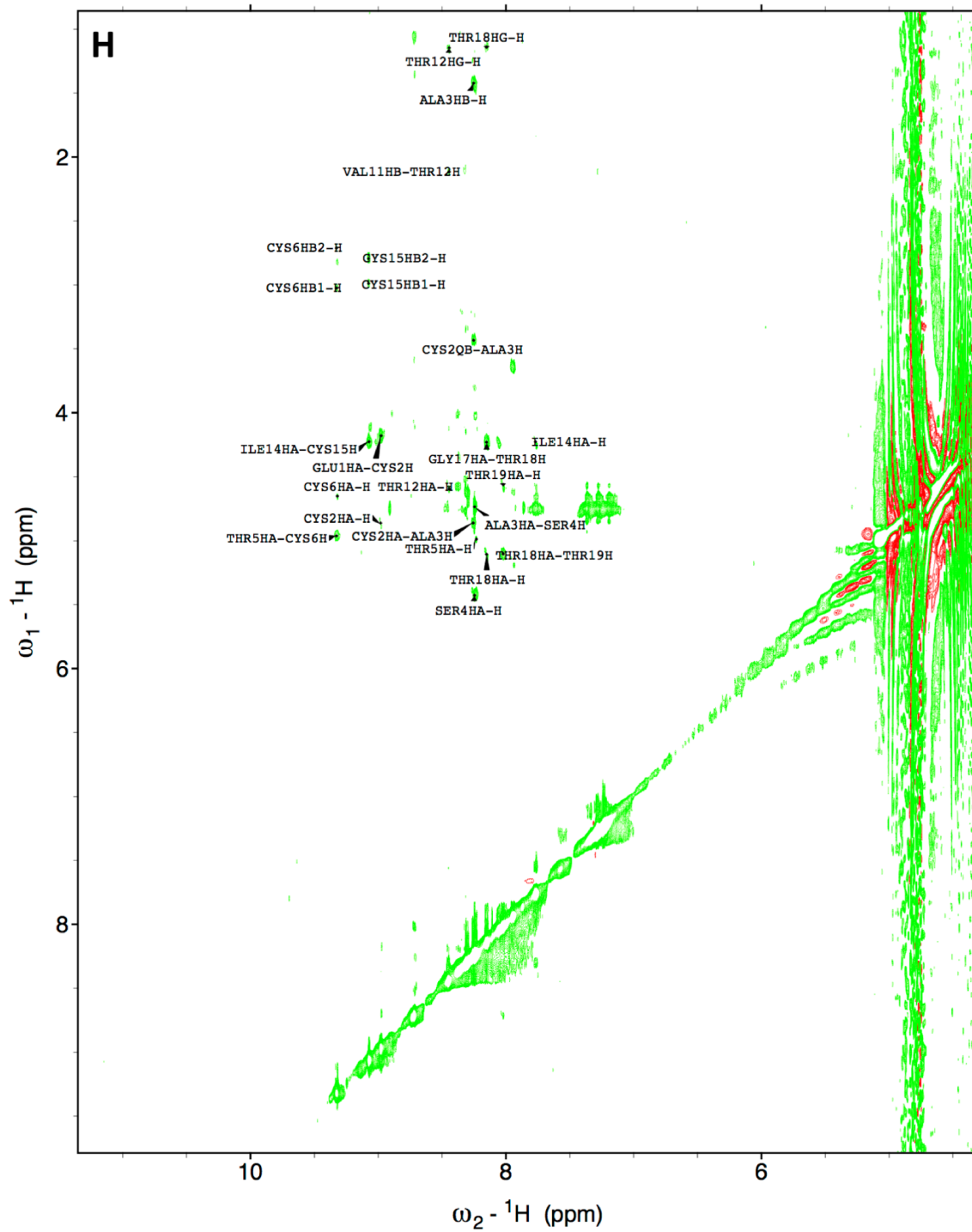
Figure S3.5 Continued
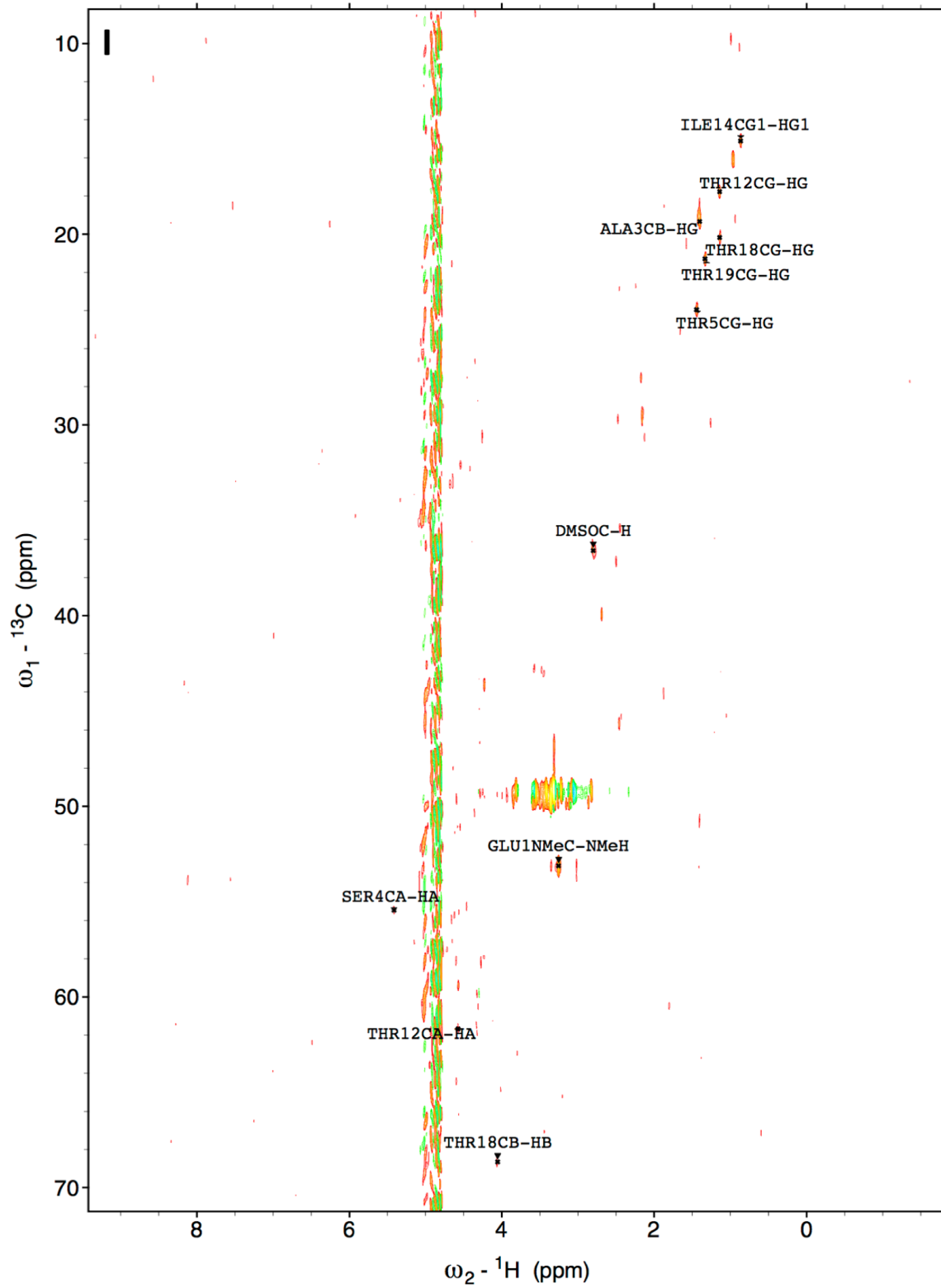
Figure S3.5 Continued
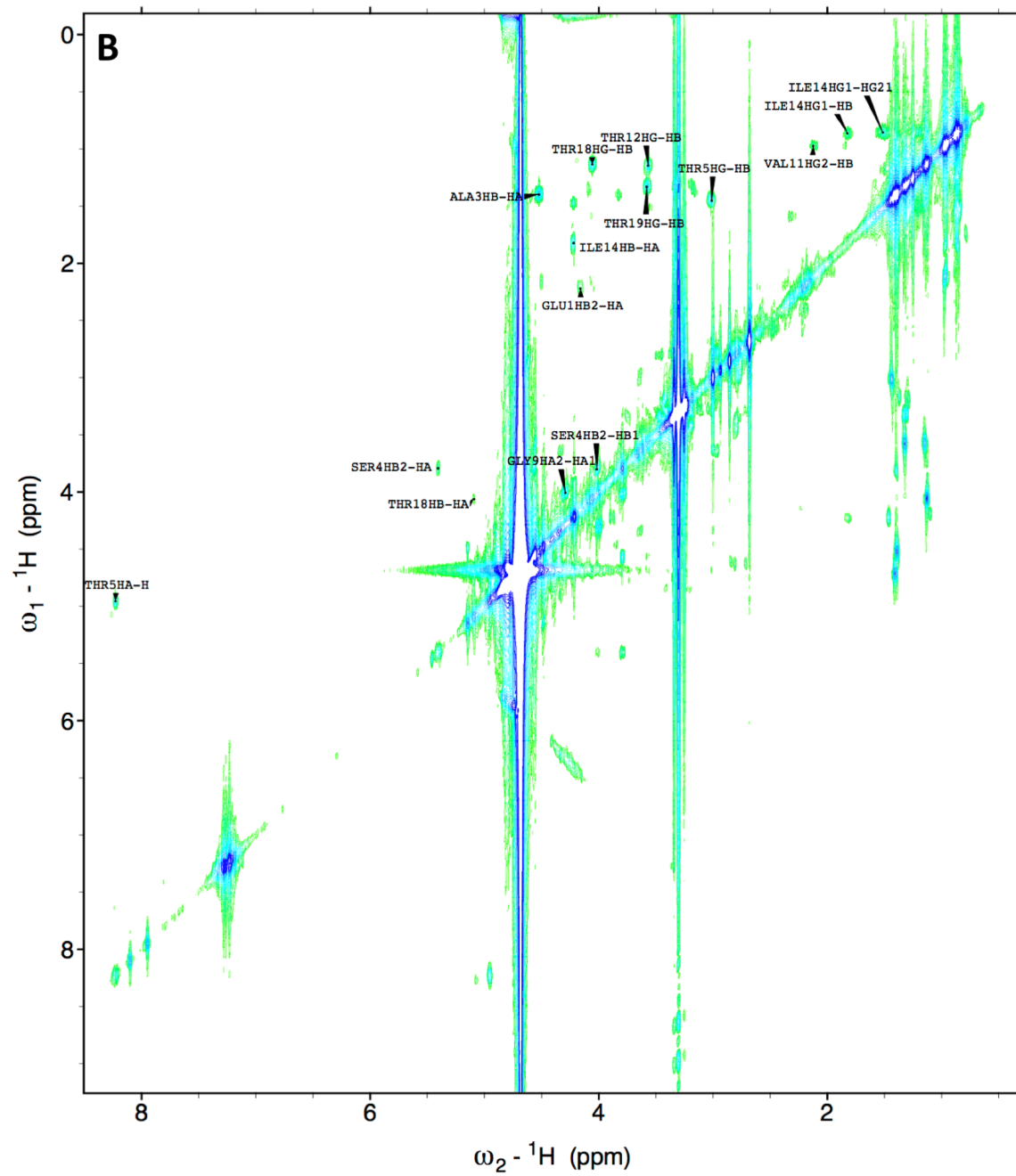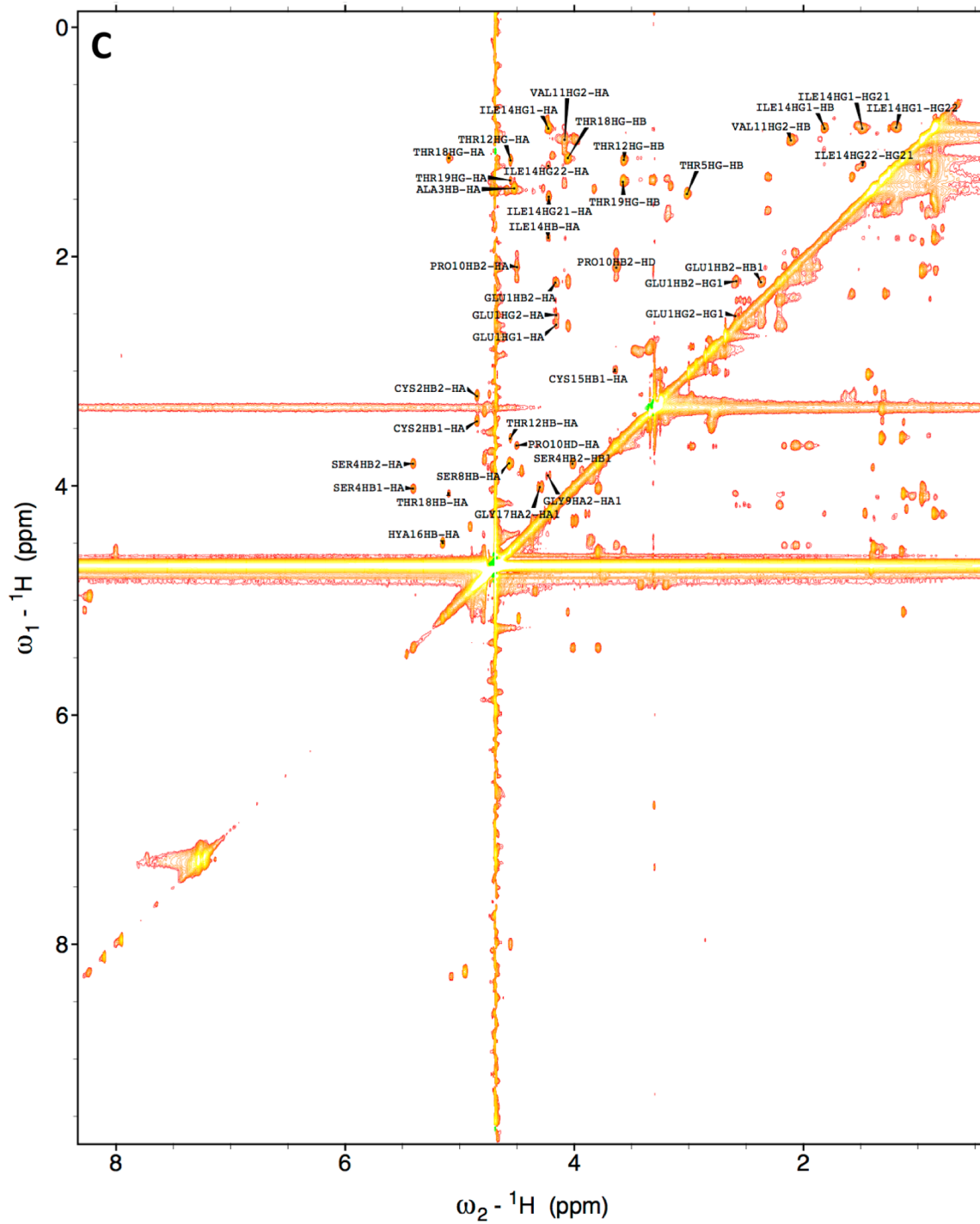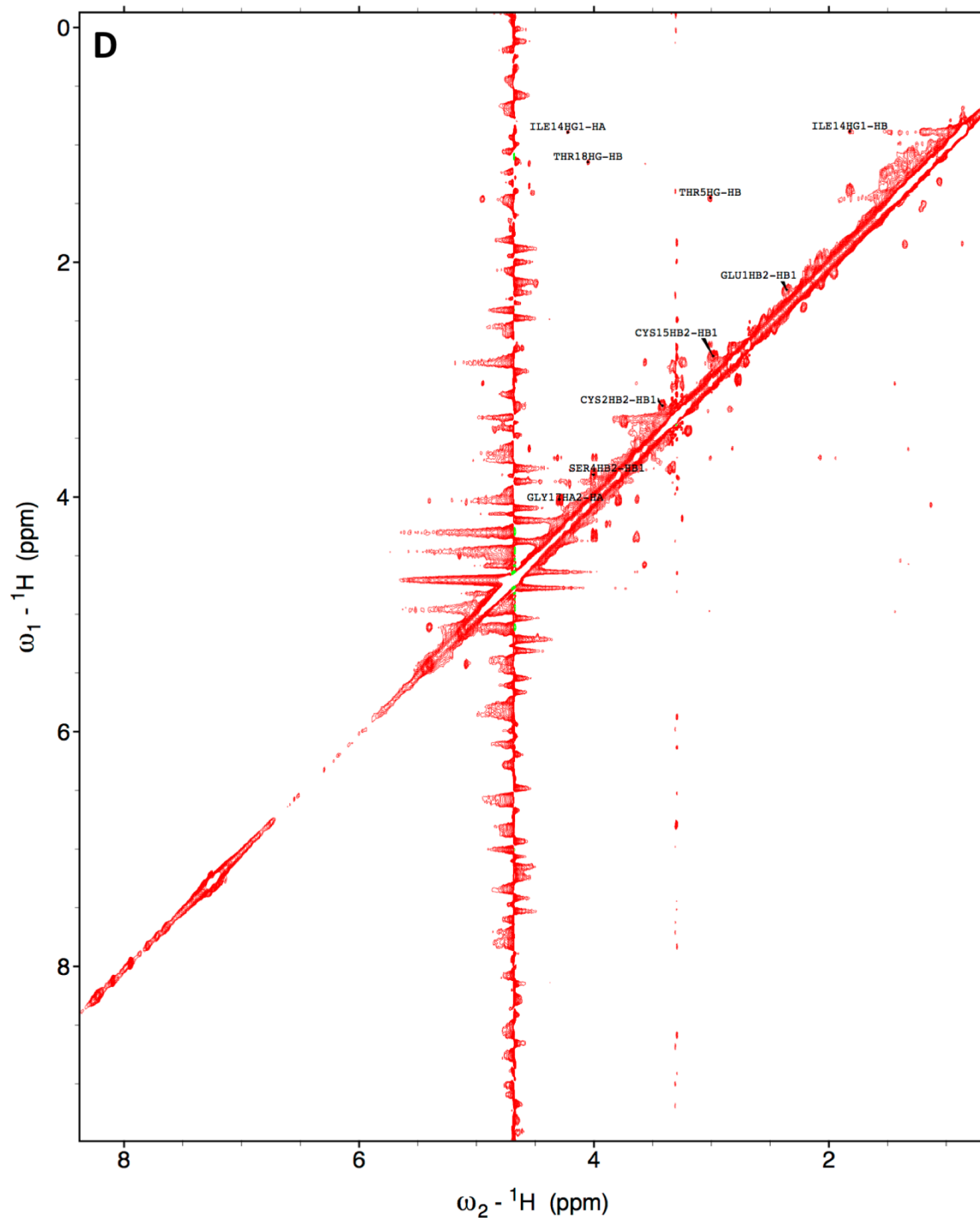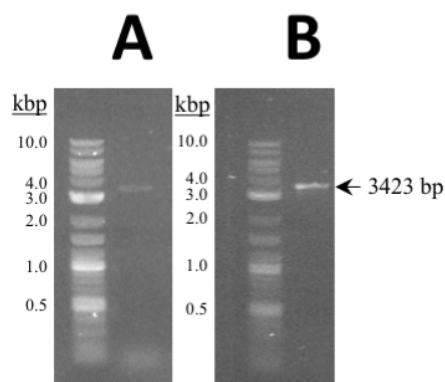
Figure S3.6. Native divM PCR amplification. A) Amplification of the native *divM* gene from E11-036 metagenomic DNA by PCR resulted in a faint band of the anticipated length. B) The product of the first PCR amplification was used as a template for amplification cycle to generate additional *divM* for the replacement of the codon-optimized *divM* gene in pDiv.
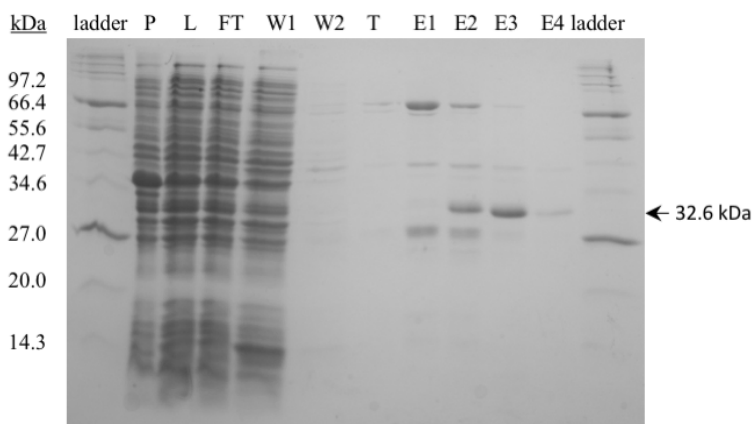
Figure S3.7. Expression and purification of DivMT. DivMT was purified on Ni-NTA fractions from the cell lysate of pRSFDuet-DivMT-expressing *E. coli* BL21(DE3). The Ni-NTA fractions were analyzed by SDS-PAGE. P = pellet, L = filtered cell lysate (supernatant), FT = Ni-NTA column flowthrough, W = column wash with lysis buffer, T = column wash with transition buffer, E = elutions of increasing imidazole concentration in transition buffer (1 = 50 mM, 2 = 100 mM, 3 = 200 mM, 4 = 500 mM imidazole). A band of the expected molecular weight for His_6-DivMT was observed in E3, and this fraction was used in the *in vitro* N-trimethylation of desmethyl divamide A.
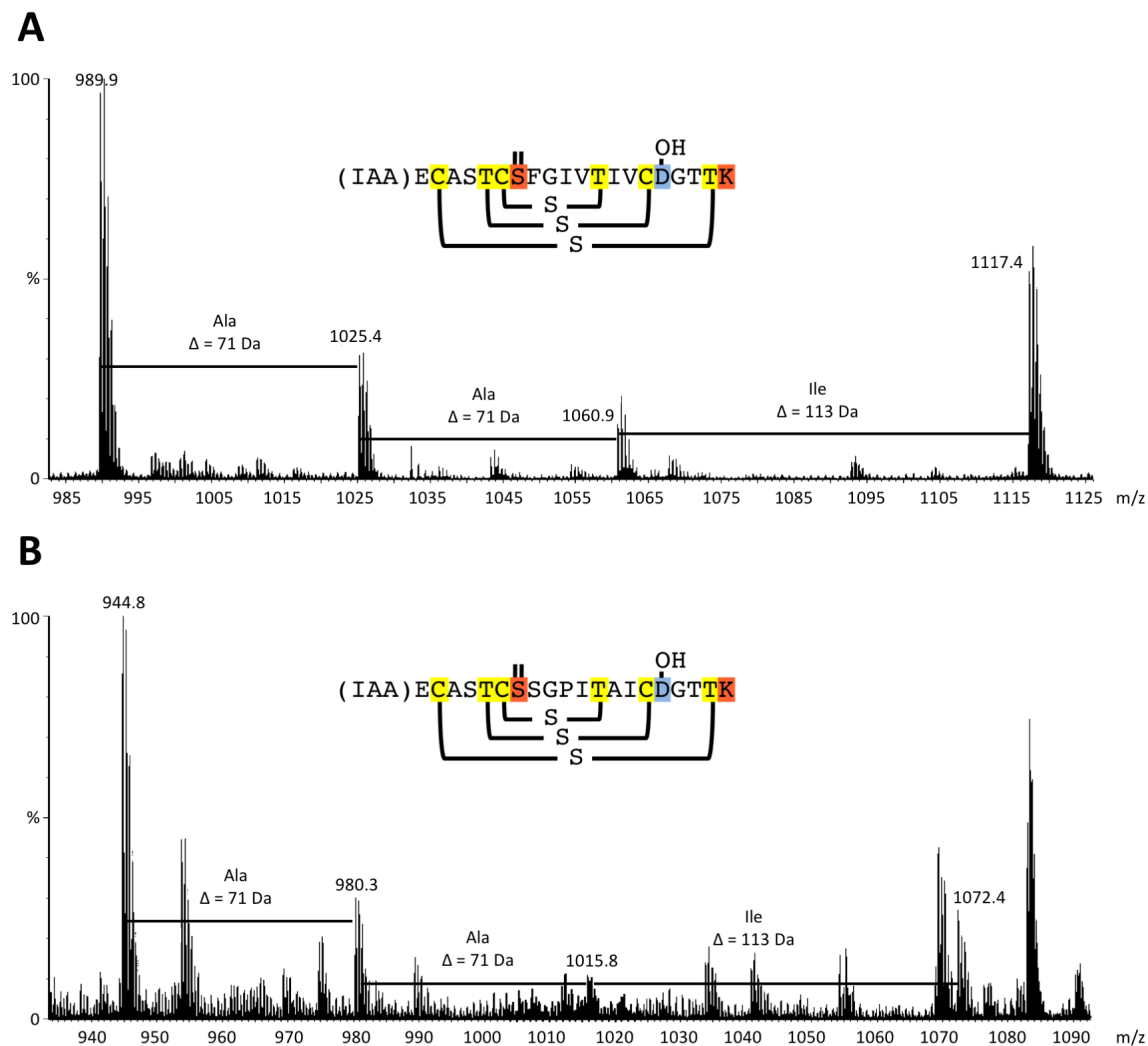
Figure S3.8. Mass spectra of divamide expression in *E. coli.* Organic extracts of *E. coli* expressing pDiv2 (A) and pDiv3 (B) were analyzed by LC/MS. In each spectrum, the primary expression product observed is the desLal divamide derivative (*m/z* 989.9 and 944.8 for desLal divamide A and B, respectively), but extra-proteolytic species were also identified that represent cleavage products at various points within the putative protease site "DIAA."
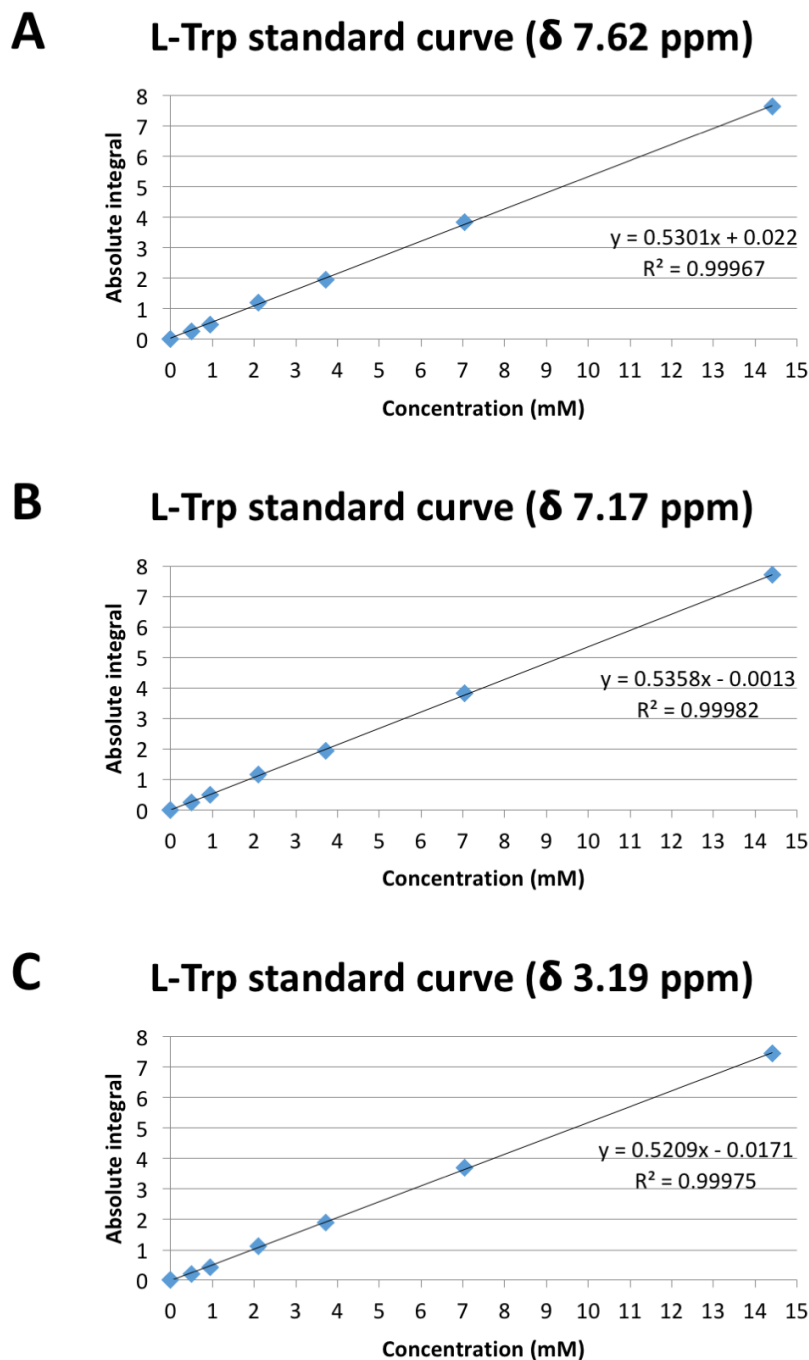
Figure S3.9. Quantitative NMR standard curves. Standard curves for the quantification of low abundance materials were generated by NMR using a dilution series of L-Trp in H_2O, the concentration of which was determined by UV-absorbance ($\varepsilon_{280,\ Trp} = 5170.9$ nm^{-1}). The absolute integral was plotted as a function of concentration using three different proton signals aromatic signals: H_{E3} (A), H_{Z3} (B), and methylene H_{B2} (C). The linear equations generated were used to determine yields of divamide analytes of unknown concentration (**Table S3.6**).
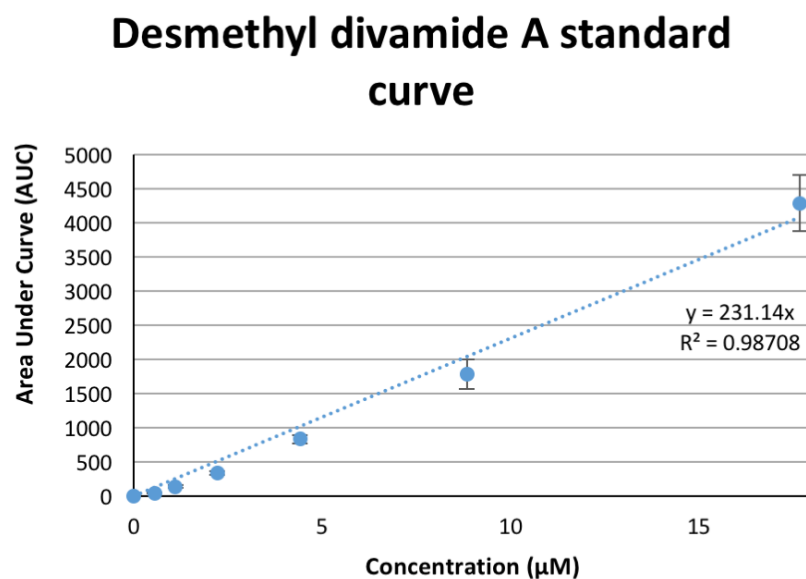
Figure S3.10. Quantitative LC/MS standard curve. A standard curve for the quantification of divamide intermediates was generated using a dilution series of synthetic desmethyl divamide A, previously quantified by NMR. The resulting linear equation was used to determine concentrations of divamide analytes and estimate relative ratios of compound mixtures, such as those obtained from expression in *E. coli*. Quantification data are summarized in **Table S3.7**.

Figure S3.11. Synthetic divamide A NMR data. NMR spectra of *E. coli*-derived divamide A intermediates obtained in 70% CD_3OH/ 30% H_2O/ 0.1% TFA include the following: A) 600 MHz 1-1 ^1H spectrum, B) 11-NOESY spectrum, C-E) gCHSQC spectra of the aliphatic (C), alpha carbon (D), and aromatic (E) regions of desmethyl deslysinoalanine divamide A. NMR spectra of synthetic divamide A obtained in 70% CD_3OH/ 30% H_2O/ 0.1% TFA include: F) 500 MHz 1-1 ^1H spectrum, G) 11-NOESY spectrum, and H) gCHSQC spectrum of the aliphatic region.
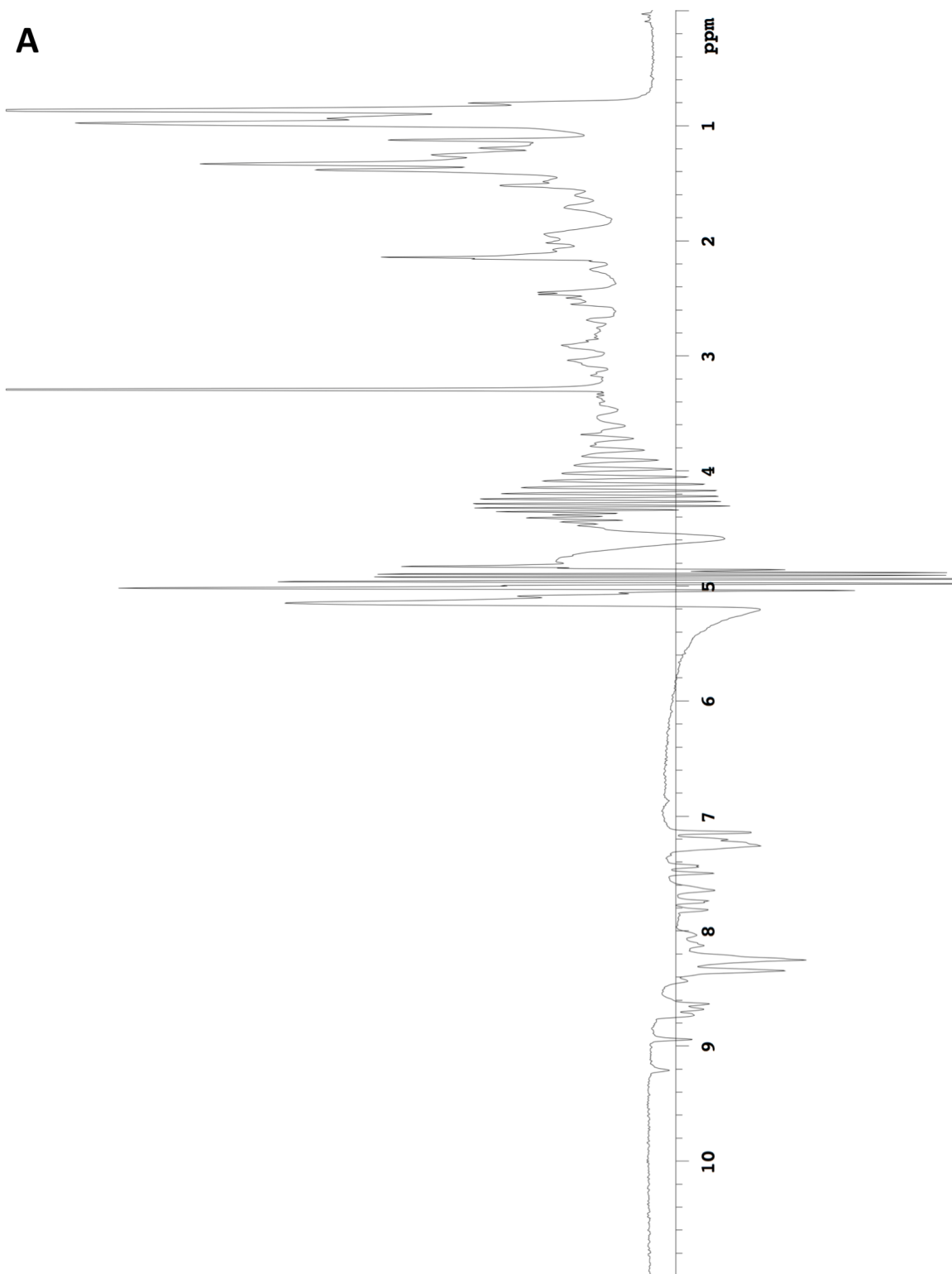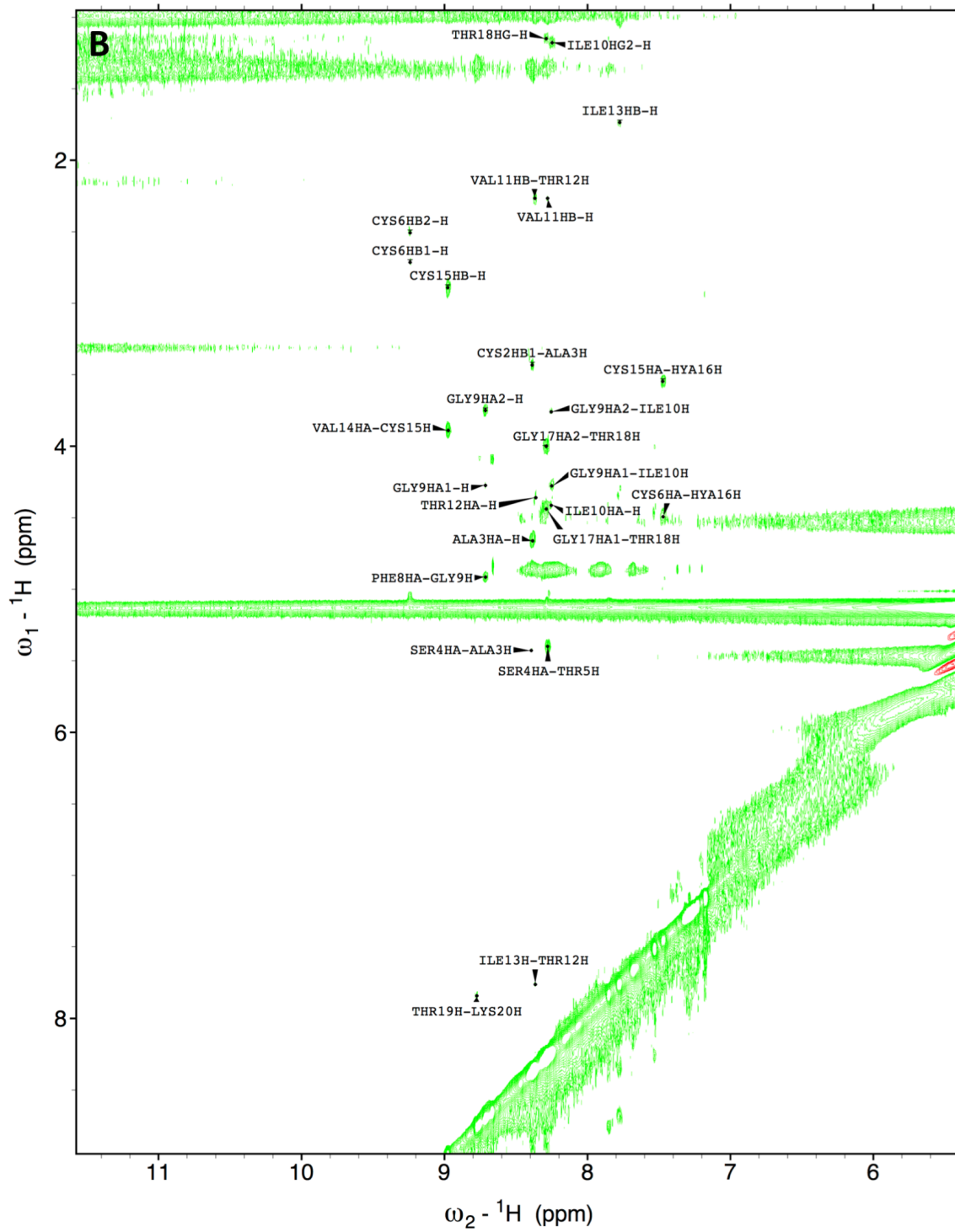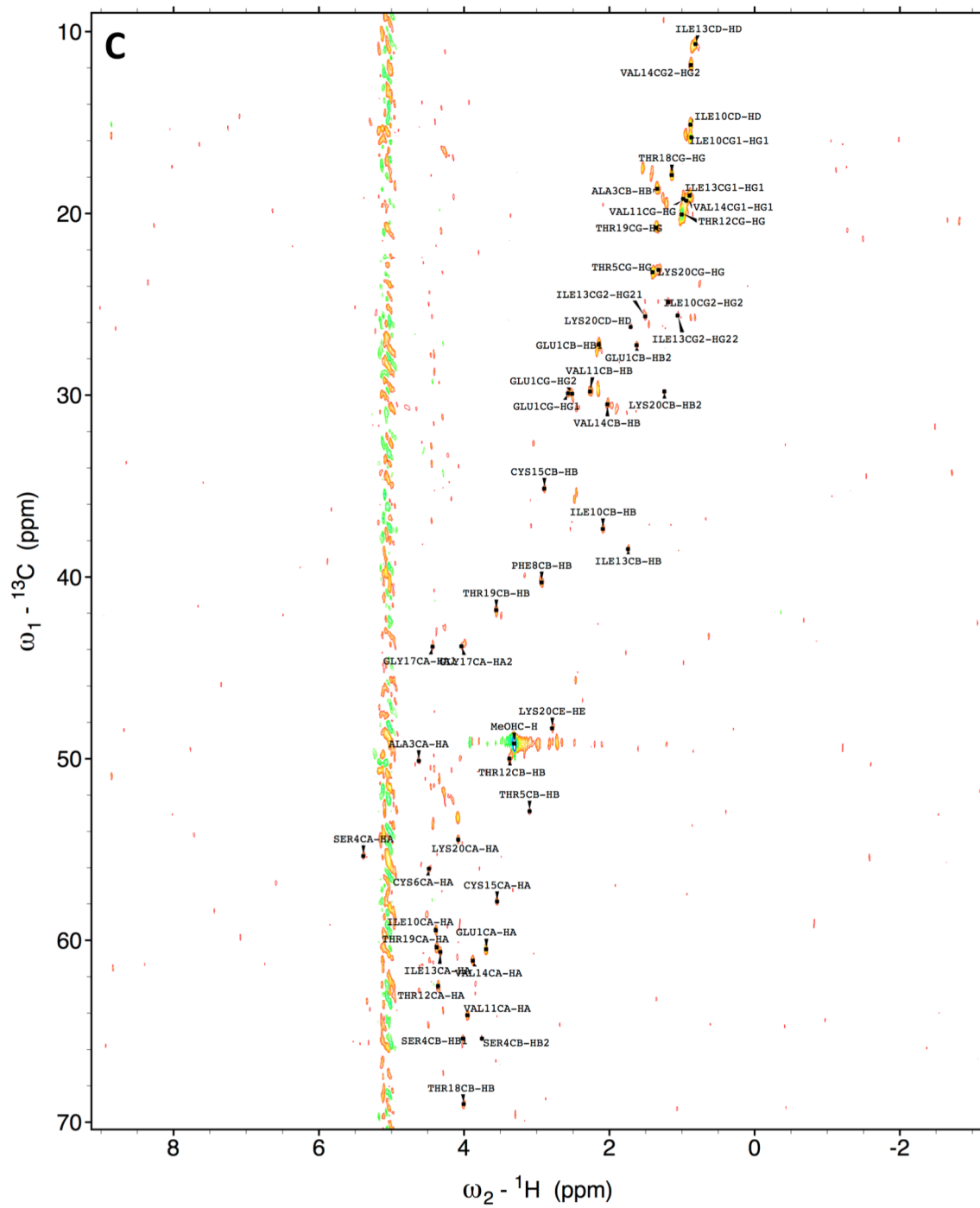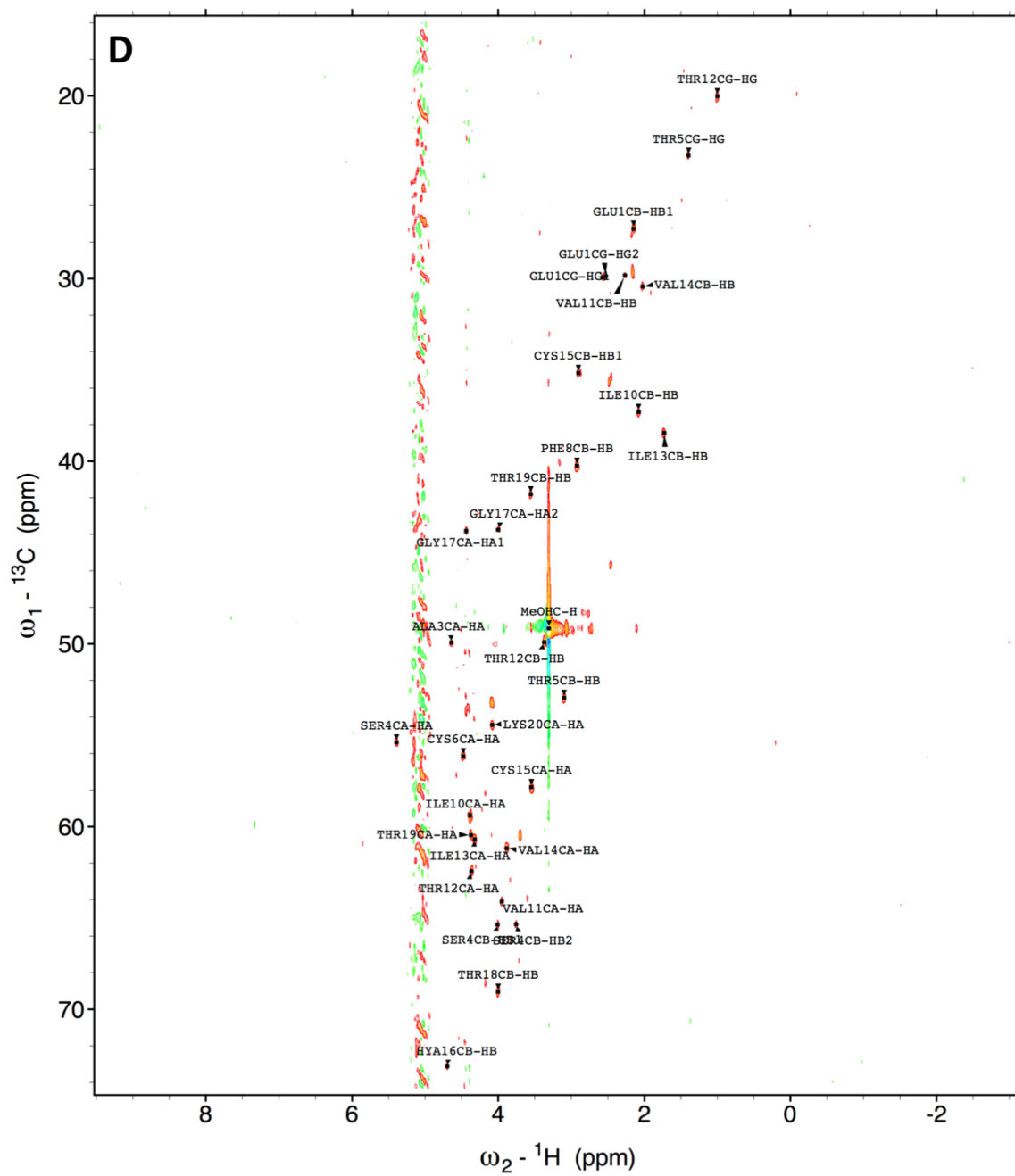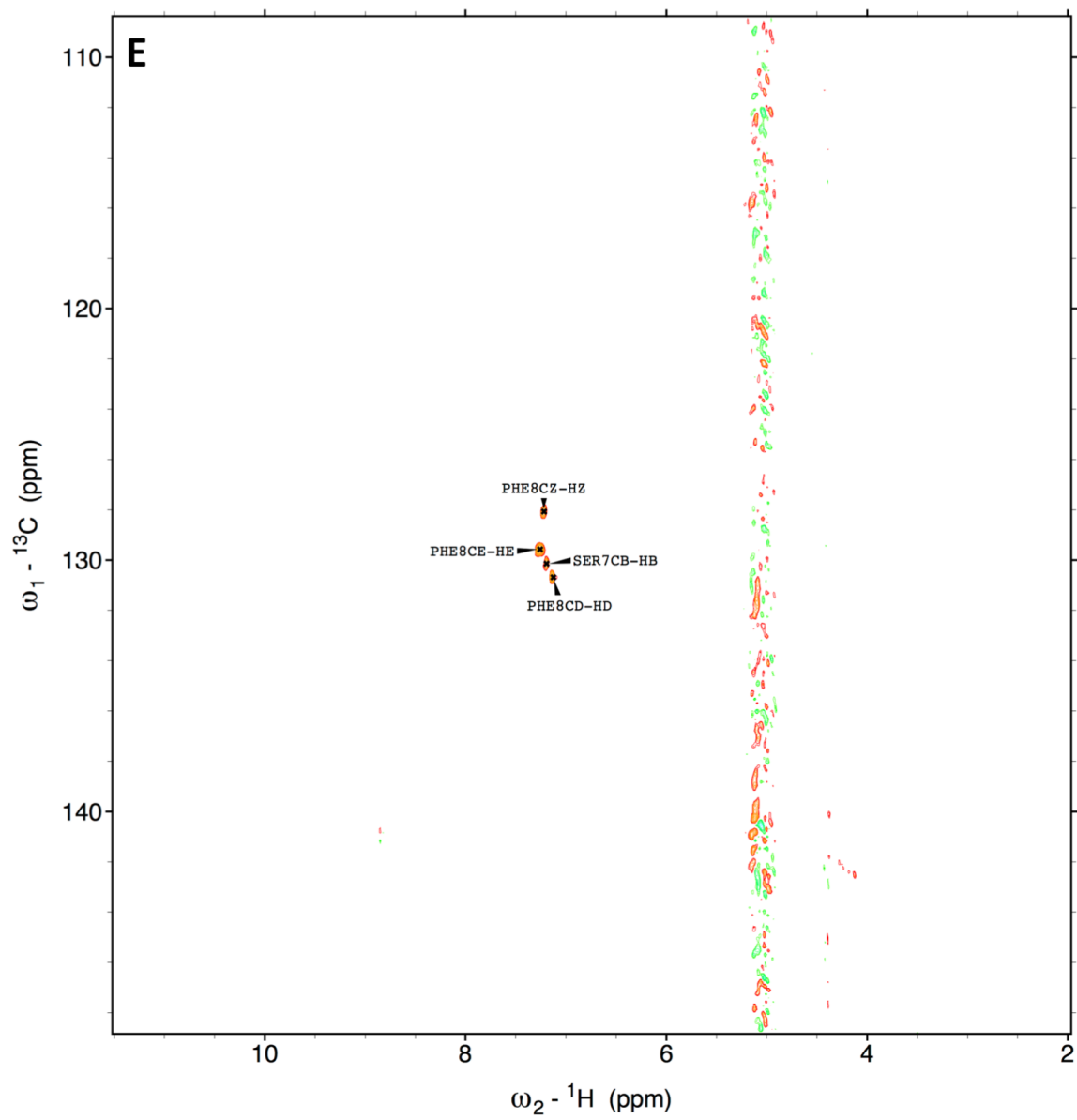
A

Figure S3.11 Continued

Figure S3.11 Continued

Figure S3.11 Continued

Figure S3.11 Continued

F
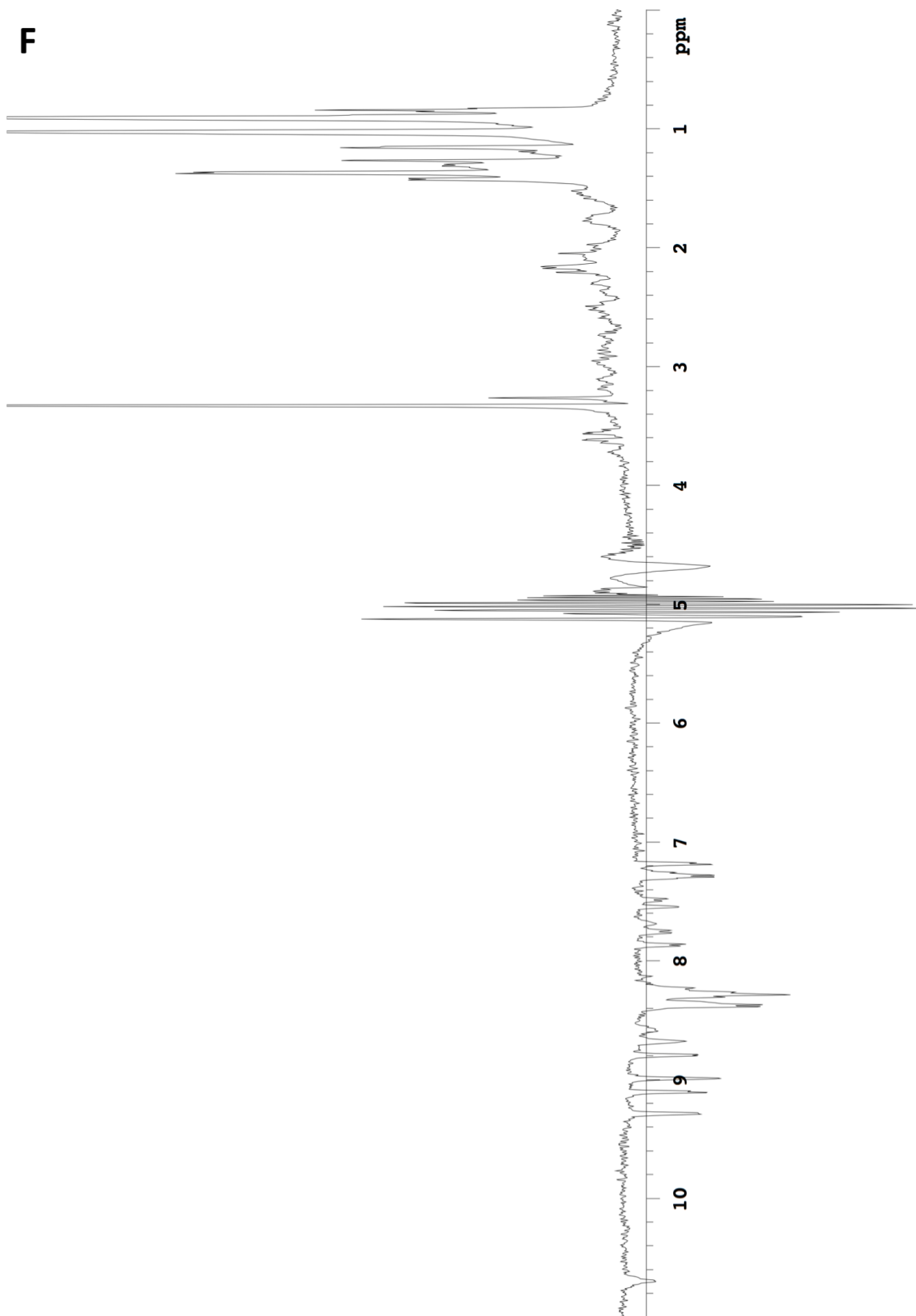
Figure S3.11 Continued
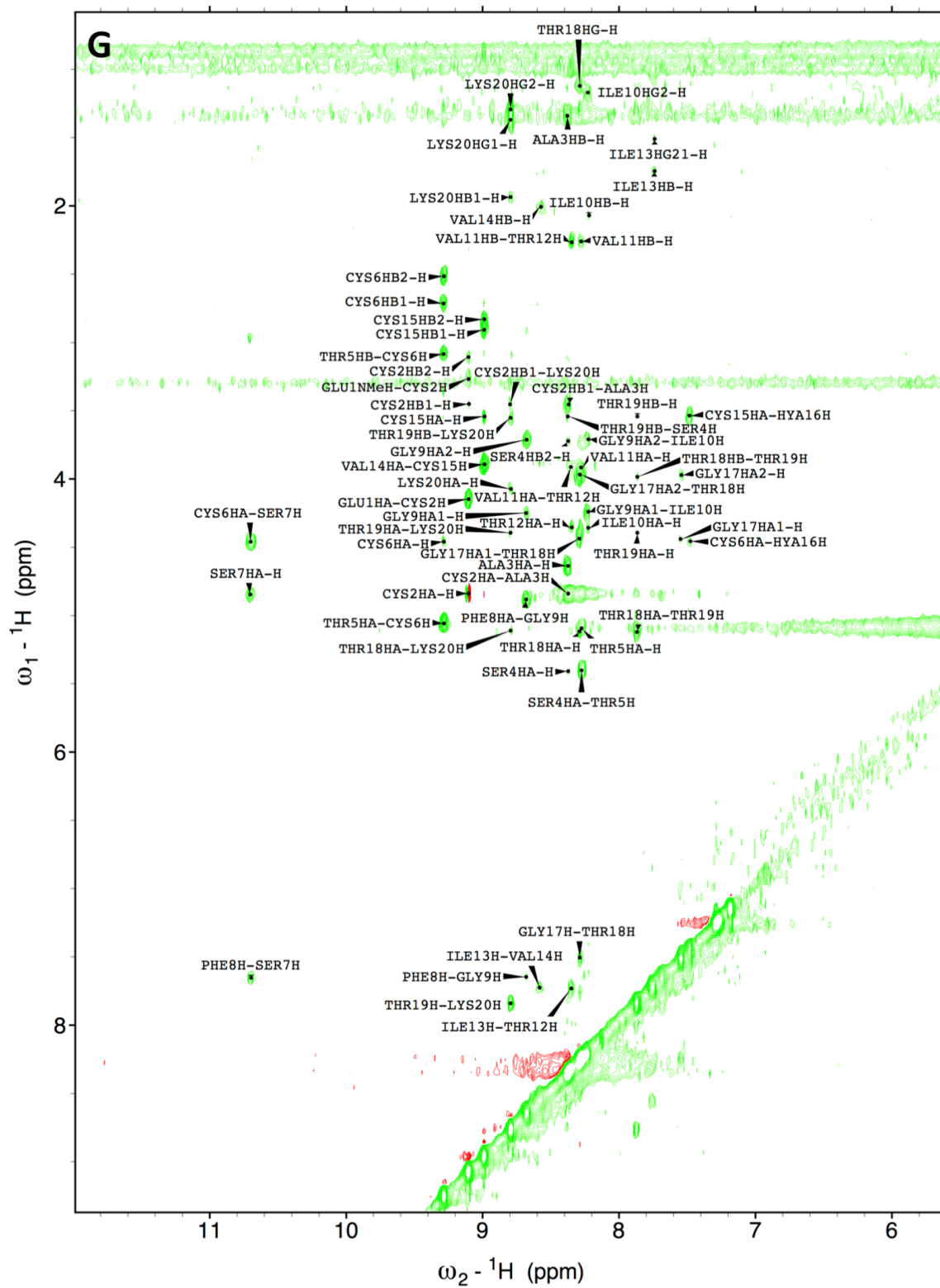
Figure S3.11 Continued
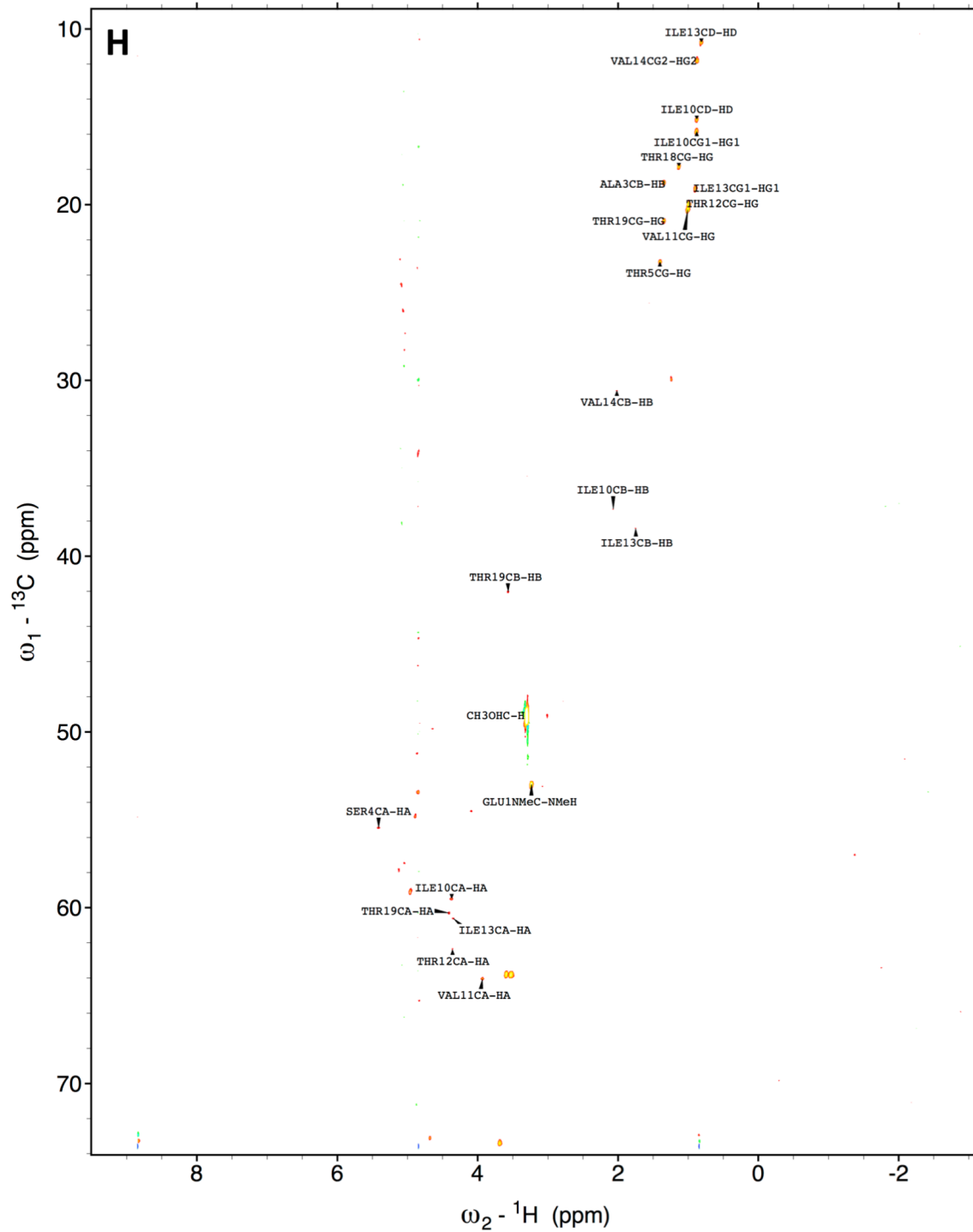
Figure S3.11 Continued
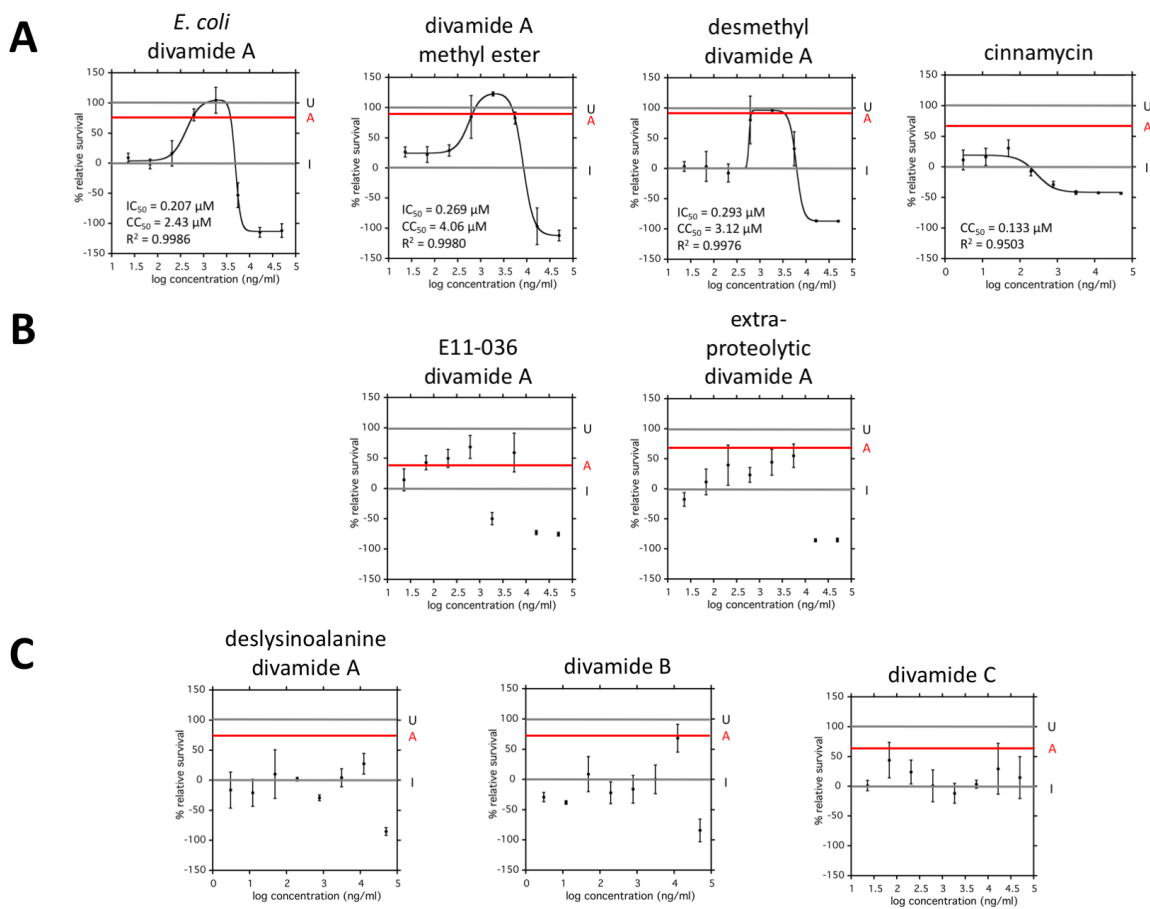
Figure S3.12. HIV cytoprotection dose-response curves. All divamide species available were investigated. U = uninfected, I = infected, and A = AZT response. A) Compounds for which curves could be fit with equation 3.4 or a single four-parameter IC_{50} equation. B) HIV cytoprotective compounds for which the curve could not be fit due to an abnormal shape characterized by a dip in response at middle concentrations. C) HIV inactive compounds, two of which still display cytotoxicity (deslysinoalanine divamide A and divamide B), for which the curve could not be fit due to a similar abnormal shape.
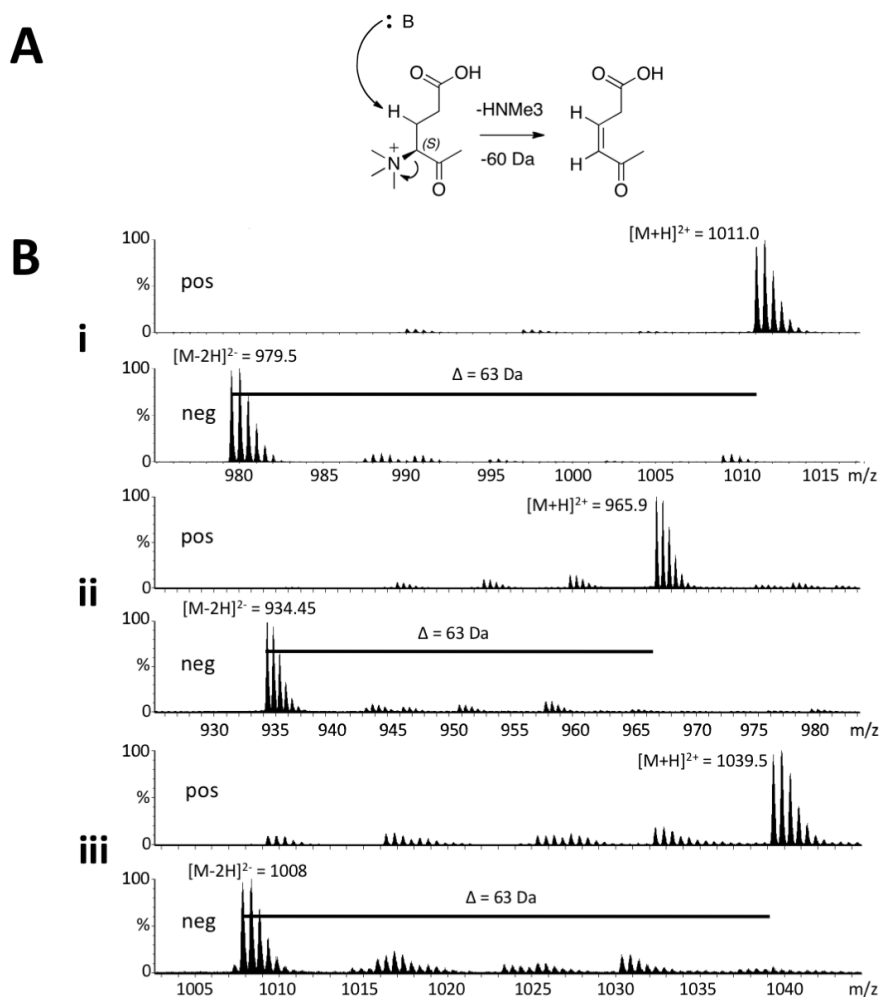
Figure S3.13. Detection of N-trimethylation by Hoffmann elimination. A) The Hoffman elimination of trimethylamine from a quaternary N-trimethylamine results in a net loss of 60 Da and the elimination of a positive charge. B) This characteristic was observed for N-trimethylglutamate-containing lanthipeptides divamide A (i), B (ii), and C (iii) by LC/MS in negative ionization mode.
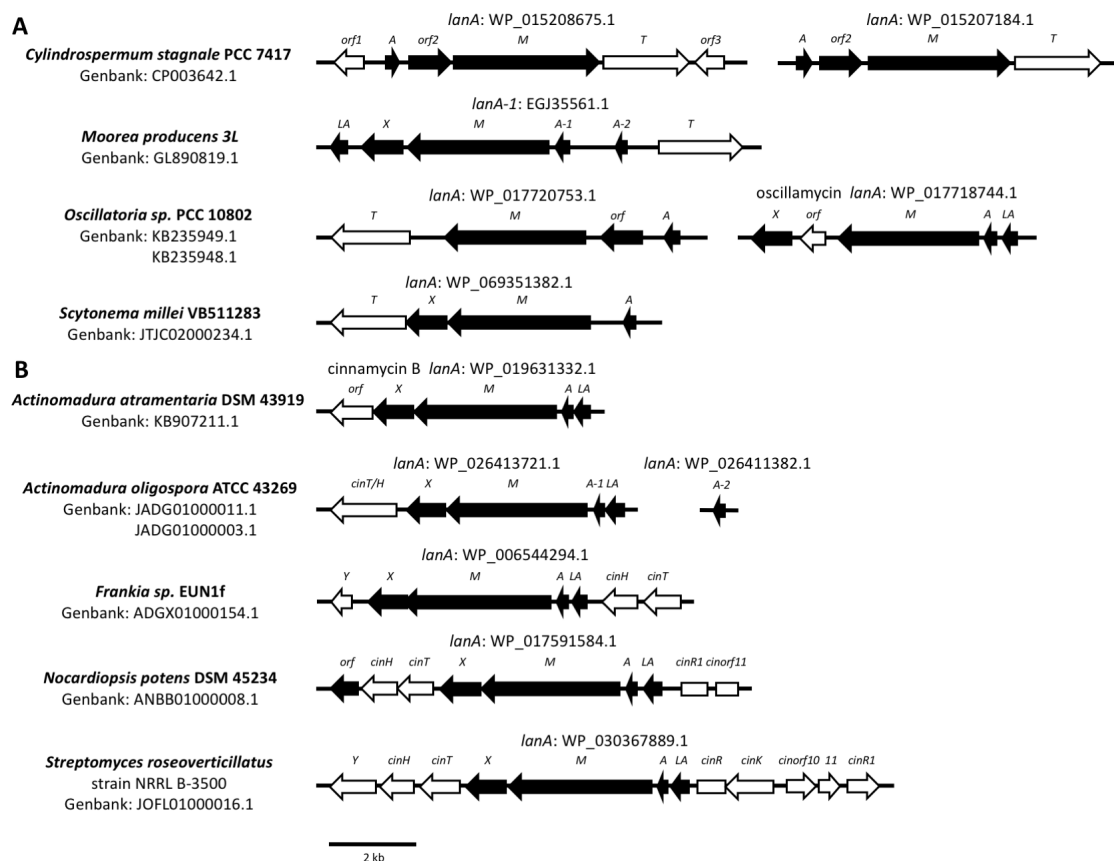
Figure S3.14. Divamide-like lanthipeptide gene clusters. All divamide-like lanthipeptide gene clusters that could be identified by BLAST search of *div* pathway genes are shown. They occur only in cyanobacteria (A) and actinobactera (B). Cyanobacteria may contain multiple lanthipeptide clusters within a single genome or multiple *lanA* precursor proteins in a single cluster, suggesting multiple products are produced from a single strain. Each pathway maintains the genes responsible for introducing conserved modifications, including Lan/MeLan and Hya. Most also contain *divLA* homologs, but this gene does not appear to be 100% conserved across all pathways. Some cyanobacterial pathways have incorporated unique modifying enzymes. A methyltransferase bearing no resemblance to DivMT can be found in the first pathway from *Cylindrospermum stagnale* PCC7417 ("*orf1*"). The recently reported oscillamycin is the product of the second cluster from *Oscillatoria sp.* PCC 10802 and contains hydroxyproline, presumably introduced by the putative proline hydroxylase located in the first cluster ("*orf*").[33] Many of the actinobacterial cluster strongly resemble that of cinnamycin, containing cinnamycin regulatory gene homologs, while cinnamycin B, recently reported from *Actinomadura atramentaria* NRBC 14695 (the cluster described here was identified by BLAST search from *Actinomadura atramentaria* DSM 43919),[34] includes only the essential divamide-like biosynthetic genes. Thus, the *div* family lanthipeptides display several key features of a diversity-generating pathway.[15]

Figure S3.15. Alignment of divamide extended family LanA protein sequences. The alignment was generated using ClustalW2 and then manually adjusted. The occurrence of multiple cassettes, as is seen for putative proteins 3-5, is a rare feature among RiPPs but common within the diversity-generating cyanobactins.[35] The alignment reveals conservation of Lan/MeLan, Lal, Gly, and Hya positions within the core as well as hypervariable positions (*).

Table S3.1. Summary of NMR data for divamides A and B. Divamide A data were collected in either D_2O and 90% H_2O/ 10% D_2O at 600 MHz (a), or 70 % CD_3OD/ 30% D_2O/ 0.1% TFA and 70 % CD_3OH/ 30% H_2O/ 0.1% TFA at 36˚C and 900 MHz (b). Divamide B data was collected in 70 % CD_3OH/ 30% H_2O/ 0.1% TFA at 36˚C and 600 MHz (c). Unassigned carbonyl chemical shifts observed by ^{13}C NMR spectroscopy in 70 % CD_3OH/ 30% H_2O/ 0.1% TFA are included (**Figure S3.2-I**). 22 total carbonyl signals were detected.

| divamide A | | divamide A [a] | | divamide A [b] | | divamide B [c] | | divamide B | |
Residue	Position	δ_C	δ_H	δ_C	δ_H	δ_C	δ_H	Position	Residue
Glu1	NMe	53.5	3.27	53.0	3.25	52.6	3.23	NMe	Glu1
	A	74.1	4.16	74.0	4.15	73.6	4.17	A	
	B	22.7	2.43	22.6	2.37	22.1	2.38	B	
			2.26		2.20		2.22		
	G	30.1	2.67	29.8	2.59	29.4	2.61	G	
			2.54		2.47		2.51		
	D	176.8		–		–		D	
	CO	–		–		–		CO	
Cys2	N		9.16		9.02		9.03	N	Cys2
	A	53.6	4.93	–	4.85	52.9	4.86	A	
	B	32.5	3.57	32.6	3.47	32.0	3.46	B	
			3.17		3.12		3.15		
	CO	–		–		–		CO	
Ala3	N		8.33		8.33		8.34	N	Ala3
	A	50.1	4.65	49.9	4.64	49.4	4.67	A	
	B	18.8	1.41	18.6	1.35	18.8	1.38	B	
	CO	175.0		176.3		174.4		CO	
Ser4	N		8.40		8.30		8.31	N	Ser4
	A	55.4	5.37	55.4	5.39	54.9	5.38	A	
	B	65.5	4.09	65.4	4.00	64.7	4.02	B	
			3.85		3.75		3.79		
	CO	–		–		–		CO	
Thr5	N		8.32		8.23		8.24	N	Thr5
	A	57.6	5.04	57.5	5.05	57.0	5.08	A	
	B	52.4	3.21	53.0	3.07	–	3.10	B	
	G	23.4	1.43	23.2	1.40	22.8	1.42	G	
	CO	–		–		–		CO	
Cys6	N		9.07		9.11		9.31	N	Cys6
	A	55.9	4.49	56.0	4.50	55.6	4.54	A	
	B	37.6	2.67	37.8	2.67	–	2.82	B	
			2.44		2.52		2.73		
	CO	–		–		–		CO	
Ser7	N		–		11.36		10.4	N	Ser7
	A	51.9	4.79	–	4.85	50.9	5.03	A	
	B	47.5	3.02	47.7	2.99	47.8	3.50	B	
			2.88						
	CO	–		–		–		CO	
Phe8	N		7.54		7.68		7.92	N	Ser8
	A	54.3	5.06	54.5	4.92	55.5	4.62	A	
	B	40.0	3.32	39.8	3.18	62.3	3.84	B	
			2.97		2.93				
	G	136.5		–					
	D	130.9	7.20	130.7	7.15				
	E	129.8	7.35	129.6	7.24				
	Z	128.3	7.33	128.1	7.22				
	CO	173.8		–		–		CO	
Gly9	N		8.81				8.44	N	Gly9
	A	43.0	4.40	42.9	8.60	–	4.37	A	
			3.95		4.27		4.00		
	CO	172.5		–	3.75	–		CO	
Ile10	N		8.33		8.15			N	Pro10
	A	59.7	4.45	59.5	4.35	61.7	4.51	A	
	B	37.6	2.16	37.3	2.07	–	2.19	B	
	G1	16.4	0.926	15.5	0.883		2.09		
	G2	25.0	1.30	25.0	1.35	–	2.10	G	
			1.20	14.3	1.19		1.97		
	D	15.5	0.907		0.876	46.8	3.66	D	
	CO	–		–		–		CO	

Table S3.1 Continued

divamide A		divamide A [a]		divamide A [b]		divamide B [c]		divamide B	
Residue	Position	δ_C	δ_H	δ_C	δ_H	δ_C	δ_H	Position	Residue
Val11	N		8.41		8.26		8.27	N	Ile11
	A	64.4	4.00	64.0	3.94	60.4	4.15	A	
	B	29.9	2.32	29.9	2.28	35.6	2.13	B	
	G	20.5	1.05	20.0	1.01	15.7	0.966	G1	
						25.6	1.54	G2	
							1.24		
						8.8	0.85	D	
	CO	177.1		–		–		CO	
Thr12	N		8.48		8.28		8.42	N	Thr12
	A	62.8	4.36	62.5	4.37	61.3	4.54	A	
	B	50.5	3.35	50.0	3.38	–	3.50	B	
	G	20.7	0.972	20.3	1.00	17.5	1.16	G	
	CO	–		–		–		CO	
Ile13	N		8.05		7.68		7.69	N	Ala13
	A	61.1	4.28	60.5	4.35	50.9	4.54	A	
	B	38.7	1.74	38.4	1.78	18.4	1.38	B	
	G1	19.4	0.912	19.1	0.885				
	G2	25.9	1.51	25.6	1.49				
			1.08		1.06				
	D	10.6	0.838	10.9	0.817				
	CO	174.7		–		174.0		CO	
Val14	N		9.20		8.46		7.91	N	Ile14
	A	61.6	3.85	60.9	3.91	57.4	4.22	A	
	B	30.4	2.04	30.6	2.03	36.7	1.88	B	
	G1	20.3	1.04	20.0	0.980	14.5	0.866	G1	
	G2	12.5	0.915	11.8	0.883	25.3	1.54	G2	
							1.24		
						9.52	0.859	D	
	CO	–		–		–		CO	
Cys15	N		8.98		8.87		9.01	N	Cys15
	A	58.1	3.53	57.8	3.55	57.4	3.62	A	
	B	35.0	1.98	35.3	2.93	–	2.99	B	
					2.84		2.85		
	CO	174.3		–		–		CO	
Hya16	N		7.50		7.37		7.55	N	Hya16
	A	59.8	4.96	58.9	5.03	57.9	5.01	A	
	B	73.1	4.74	–	–	72.3	4.58	B	
	G	–		–		–		G	
	CO	–		–		–		CO	
Gly17	N		7.52		7.45		7.59	N	Gly17
	A	44.1	4.53	43.9	4.42	43.0	4.39	A	
			4.08		4.02		4.03		
	CO	169.5		–		–		CO	
Thr18	N		8.28		8.20		8.22	N	Thr18
	A	57.8	5.10	57.7	5.11	57.5	5.11	A	
	B	69.3	4.15	69.0	4.00	68.3	4.01	B	
	G	18.5	1.20	17.9	1.14	19.4	1.14	G	
	CO	–		–		–		CO	
Thr19	N		7.93		7.80		7.87	N	Thr19
	A	60.8	4.45	60.4	4.41	59.7	4.43	A	
	B	41.9	3.62	42.0	3.56	41.7	3.62	B	
	G	21.3	1.40	20.9	1.35	20.3	1.34	G	
	CO	–		–		–		CO	
Lys20	N		8.82		8.74		8.72	N	Lys20
	A	55.7	4.05	54.9	4.06	53.8	4.12	A	
	B	30.8	2.00	30.6	1.95	29.5	1.97	B	
			1.35		1.38		1.51		
	G	23.3	1.37	23.1	1.32	–	1.38	G	
			1.28				1.37		
	D	26.5	1.70	26.3	1.70	28.8	1.75	D	
							1.34		
	E	48.5	2.86	48.2	2.81	47.9	2.94	E	
			2.72		2.74				
	CO	–		–		–		CO	

Unassigned CO chemical shifts for divamide A[b]: 176.0, 175.7, 175.1, 174.8, 174.7, 174.5, 174.1, 174.0, 173.8, 173.7, 173.5, 173.3, 173.2, 171.9, 171.8, 171.7, 169.8, 169.7, 169.5, 167.1, 167.1 ppm

Table S3.2. NOE and HMBC correlations for divamides A and B. Correlations observed in NOESY and HMBC NMR spectra for divamides A (in D_2O or 90% H_2O/ 10% D_2O) and B (in 70% CD_3OD/ 30% D_2O/ 0.1 % TFA or 70% CD_3OH/ 30% H_2O/ 0.1 % TFA) are summarized.

Position	divamide A		divamide B	
	NOE	HMBC	NOE	HMBC
1	GLU1HB1-HA, GLU1HB2-HA, GLU1HB2-HB1, GLU1NMeH-HA, GLU1NMeH-CYS2HA, GLU1HA-CYS2H, GLU1HA-CYS2H, GLU1HA-THR19H, GLU1NMeH-CYS6H	GLU1CA-NMeH, GLU1CD-HG1, GLU1CD-HG2, GLU1NMeC-NMeH	GLU1HB1-HA, GLU1HB2-HB1, GLU1NMeH-HA, GLU1HA-CYS2H	GLU1CA-NMeH, GLU1NMeC-NMeH
2	CYS2HB1-HA, CYS2HB2-HA, CYS2HB2-HB1, CYS2HA-H, CYS2HA-ALA3H, CYS2HA-THR19H, CYS2HB1-ALA3H, CYS2HB1-LYS20H		CYS2HB1-HA, CYS2HB2-HA, CYS2HB2-HB1, CYS2HB2-THR19HB, CYS2HA-H, CYS2HA-ALA3H, CYS2HB1-ALA3H, CYS2HB1-LYS20H	
3	ALA3HB-HA, ALA3HB-SER4HA, ALA3H-LYS20H, ALA3HA-H, ALA3HA-SER4H, ALA3HB-SER4H	ALA3CA-HB, ALA3CB-HA, ALA3CO-HA, ALA3CO-HB	ALA3HB-HA, ALA3HA-H, ALA3HA-SER4H, ALA3HA-THR18H, ALA3HB-H, ALA3HB-SER4H	ALA3CA-HB, ALA3CO-HB
4	SER4HA-ALA3H, SER4HB1-ALA3H, SER4HB1-HA, SER4HB1-THR18HA, SER4HB2-ALA3H, SER4HB2-HA, SER4HB2-HB1, SER4HB2-THR18HA, SER4HA-H, SER4HA-THR5H, SER4HB2-H, SER4HB2-THR5H		SER4HA-THR5H, SER4HB1-HA, SER4HB1-HA, SER4HB2-HA, SER4HB2-HB1, SER4HA-H, SER4HA-THR5H, SER4HB2-H, SER4HB2-SER5H	
5	THR5HA-ALA3H, THR5HA-SER4HA, THR5HB-HA, THR5HB-CYS15HA, THR5HG-ALA3H, THR5HG-SER4HB1, THR5HG-HA, THR5HG-HB, THR5HG-VAL14HA, THR5HG-CYS15H, THR5H-GLY17H, THR5HA-H, THR5HA-CYS6H, THR5HB-CYS2H, THR5HB-H, THR5HG-H	THR5CA-HG, THR5CB-HG, THR5CB-CYS15HB, THR5CG-HB	THR5HG-HB, THR5HA-H, THR5HA-CYS6H, THR5HB-CYS6H	THR5CA-HG
6	CYS6HB1-HA, CYS6HB1-THR12HA, CYS6HB1-THR12HB, CYS6HB1-CYS15HA, CYS6HB2-HA, CYS6HB2-HB1, CYS6HB2-THR12HA, CYS6HB2-THR12HB, CYS6HB2-CYS15HA, CYS6HA-H, CYS6HB1-H, CYS6HB2-H	CYS6CB-HA	CYS6HB1-HA, CYS6HB2-THR12HB, CYS6HB1-H, CYS6HB2-H	
7	SER7HB1-HA, SER7HB2-HA, SER7HB2-HB1, SER7HB2-PHE8HD, SER7HA-THR18H, SER7HB1-PHE8H, SER7HB1-PHE8HD, SER7HB2-PHE8H, SER7HB2-PHE8HD		SER7HB-HA, SER8HB-HA, SER7H-SER8H	
8	PHE8HA-HD, PHE8HB1-HA, PHE8HB1-HD, PHE8HB2-HA, PHE8HB2-HB1, PHE8HB2-HD, PHE8HD-HE, PHE8HA-H, PHE8HA-HD, PHE8HA-GLY9H, PHE8HB1-HD, PHE8HB1-GLY9H	PHE8CA-HB1, PHE8CA-HB2, PHE8CB-HD, PHE8CD-HB1, PHE8CD-HB2, PHE8CD-HD, PHE8CE-HD, PHE8CE-HE, PHE8CG-HB1, PHE8CG-HB2, PHE8CG-HE, PHE8CO-HA, PHE8CZ-HD	SER8HA-H, SER8HA-GLY9H, SER8HB-GLY9H	
9	GLY9HA2-HA1, GLY9HA1-H, GLY9HA1-ILE10H, GLY9HA2-H, GLY9HA2-ILE10H	GLY9CO-HA1, GLY9CO-HA2	GLY9HA1-H, GLY9HA2-H	
10	ILE10HB-HA, ILE10HD-HB, ILE10HD-HYA16HA, ILE10HG1-HA, ILE10HG21-HA, ILE10HG21-HB, ILE10HG22-HA, ILE10HG22-HB, ILE10HA-H, ILE10HA-VAL11H, ILE10HB-H, ILE10HG1-H, ILE10HG1-H, ILE10HG21-H, ILE10HG22-H	ILE10CA-HG1, ILE10CB-HD, ILE10CG2-HD	PRO10HB1-HA, PRO10HB2-HA, PRO10HB2-HB1, PRO10HD-HA, PRO10HG1-HD, PRO10HG2-HB1, PRO10HG2-HD, PRO10HG2-HG1, PRO10HA-ILE11H	

Table S3.2 Continued

Position	divamide A		divamide B	
	NOE	HMBC	NOE	HMBC
11	VAL11HB-HA, VAL11HG-HA, VAL11HG-HB, VAL11HG-THR12HA, VAL11HA-H, VAL11HB-H, VAL11HB-THR12H, VAL11HB-ILE13H, VAL11HG-H, VAL11HG-THR12H	VAL11CA-HG, VAL11CB-HA, VAL11CB-HG, VAL11CG-HG, VAL11CO-HA	ILE11HD-HA, ILE11HD-HB, ILE11HG1-HA, ILE11HG1-HB, ILE11HA-H, ILE11HA-THR12H, ILE11HB-H, ILE11HB-THR12H, ILE11HG1-ALA13H	ILE11CA-HG1, ILE11CB-HD, ILE11CB-HG1, ILE11CG2-HD, ILE11CG2-HG1
12	THR12HB-HA, THR12HG-GLU1HB1, THR12HG-ILE10HA, THR12HG-HA, THR12HG-HB, THR12H-ILE13H, THR12HA-H, THR12HG-H	THR12CA-HG, THR12CB-HG	THR12HB-HA, THR12HG-HA, THR12HG-HB, THR12HA-H, THR12HG-H	
13	ILE13HB-HA, ILE13HD-HA, ILE13HD-HB, ILE13HD-HG21, ILE13HG1-HA, ILE13HG1-HB, ILE13HG21-HA, ILE13HG21-HB, ILE13HG22-HA, ILE13HG22-HB, ILE13HG22-HG21, ILE13H-THR12H, ILE13H-VAL14H, ILE13HA-H, ILE13HB-H, ILE13HB-VAL14H, ILE13HG21-H, ILE13HG22-H	ILE13CA-HG1, ILE13CB-HD, ILE13CB-HG1, ILE13CG2-HD, ILE13CG2-HG1, ILE13CO-HA	ALA13HG-HA, ALA13H-THR12H, ALA13HG-H	ALA13CA-HB, ALA13CO-HB
14	VAL14HB-HA, VAL14HG1-ILE13HB, VAL14HG1-HA, VAL14HG1-HB, VAL14HG2-HA, VAL14HG2-HB, VAL14H-ILE13H, VAL14HA-H, VAL14HA-CYS15H, VAL14HB-H, VAL14HG1-H, VAL14HG2-H, VAL14HG2-CYS15H	VAL14CA-HG1, VAL14CA-HG2, VAL14CB-HG1, VAL14CB-HG2, VAL14CG1-HG2	ILE14HB-HA, ILE14HG1-HA, ILE14HG1-HB, ILE14HG1-HG21, ILE14HG1-HG22, ILE14HG21-HA, ILE14HG22-HA, ILE14HG22-HG21, ILE14HA-H, ILE14HA-CYS15H, ILE14HB-H	ILE14CA-HG1, ILE14CB-HG1, ILE14CG2-HG1
15	CYS15HA-CYS6HA, CYS15HB-THR5HB, CYS15HB-HA, CYS15HA-CYS6H, CYS15HA-H, CYS15HA-HYA16H, CYS15HB-H	CYS15CO-HB	CYS15HB1-HA, CYS15HB2-THR5HB, CYS15HB2-HA, CYS15HB2-HB1, CYS15HA-H, CYS15HA-HYA16H, CYS15HB1-H, CYS15HB2-H	
16	HYA16HA-H, HYA16HA-GLY17H	HYA16CB-HA	HYA16H-SER5H, HYA16HA-H, HYA16HA-GLY17H, HYA16HB-H	
17	GLY17HA2-HA1, GLY17H-THR5H, GLY17HA1-HYA16H, GLY17HA1-H, GLY17HA1-THR18H, GLY17HA2-H, GLY17HA2-THR18H	GLY17CO-HA1, GLY17CO-HA2	GLY17HA-HA1, GLY17HA2-HA1, GLY17HA1-H, GLY17HA1-THR18H, GLY17HA2-H, GLY17HA2-THR18H	
18	THR18HA-ALA3H, THR18HA-SER4HA, THR18HB-HA, THR18HG-SER4HB1, THR18HG-VAL14HA, THR18HG-HA, THR18HG-HB, THR18HA-GLY17H, THR18HA-H, THR18HA-THR19H, THR18HB-H, THR18HG-H, THR18HG-THR19H	THR18CA-HG, THR18CB-HG	THR18HA-SER4HA, THR18HA-H, THR18HB-H, THR18HG-HA, THR18HG-HB, THR18HA-H, THR18HA-THR19H, THR18HG-H	THR18CB-HG
19	THR19HB-HA, THR19HG-CYS2HA, THR19HG-THR18HA, THR19HG-HA, THR19HG-HB, THR19H-LYS20H, THR19HA-H, THR19HA-LYS20H, THR19HB-THR5H, THR19HB-H, THR19HB-LYS20H, THR19HG-H	THR19CA-HG, THR19CB-HG	THR19HB-HA, THR19HG-HA, THR19HG-HB, THR19H-LYS20H, THR19HA-H, THR19HB-H, THR19HB-LYS20H, THR19HG-H	THR19CA-HG, THR19CB-HG
20	LYS20HB1-HA, LYS20HB1-HE1, LYS20HB1-HE2, LYS20HD-SER4HA, LYS20HD-HE1, LYS20HD-HE2, LYS20HE2-SER7HB1, LYS20HE2-HE1, LYS20HG1-HA, LYS20HG1-HB1, LYS20HG1-HD, LYS20HG1-HE1, LYS20HG1-HE2, LYS20HG2-SER4HA, LYS20HG2-HA, LYS20HG2-HB1, LYS20HG2-HD, LYS20HG2-HE1, LYS20HG2-HE2, LYS20H-THR19H, LYS20HA-THR19H, LYS20HA-H, LYS20HG1-H, LYS20HG2-H		LYS20HB1-HA, LYS20HB1-HE, LYS20HB2-HB1, LYS20HE-SER7HB, LYS20HG1-HB1, LYS20HG1-HB1, LYS20HG1-HD1, LYS20HG1-HE, LYS20HG2-HA, LYS20HA-H, LYS20HB2-H, LYS20HG2-H	

Table S3.3. Amino acid standards for chiral GC/MS. Standards were used to determine divamide A stereocenter configurations by comparative analysis via chiral GC/MS. Lal could not be detected using the chiral stationary phase, but diastereomers could be resolved using an achiral column (*). Fragmentation patterns of noncanonical amino acids, including Hya (a),[36] MeLan (b),[37] and Lal (c),[38] were previously known.

Standard	RT (min)	Fragment ions (*m/z*)
L-Ala	3.748	281.1, 190.2, 119.1
L-Val	4.967	218.2, 203.2, 119.1
D-Thr	5.47	381.1, 366.1, 235.2, 230.2, 202.2, 119.1
Gly	5.495	176.2, 119.1
L-Thr	5.935	381.1, 366.1, 235.2, 230.2, 202.2, 119.1
D-*allo*-Ile	6.005	235.2, 232.2, 203.2, 119.1
L-*allo*-Ile	6.646	235.2, 232.2, 203.2, 119.1
L-Ile	7.011	235.2, 232.2, 203.2, 119.1
D-*allo*-Thr	9.033	381.1, 366.1, 235.2, 230.2, 202.2, 119.1
D-Ser	9.058	352.1, 247.2, 215.2, 188.2, 160.2, 119.1, 100.2
L-Ser	9.596	352.1, 247.2, 215.2, 188.2, 160.2, 119.1, 100.2
D-*threo*-Hya [a]	9.613	246.0, 214.0, 154.0, 119.1
L-*allo*-Thr	9.841	381.1, 366.1, 235.2, 230.2, 202.2, 119.1
L-*threo*-Hya [a]	10.514	246.0, 214.0, 154.0, 119.1
D-Asp	11.097	248.2, 228.2, 216.2, 206.2, 119.1
L-Asp	11.382	248.2, 228.2, 216.2, 206.2, 119.1
L-*erythro*-Hya [a]	12.3	246.0, 214.0, 154.0, 119.1
D-Glu	15.559	262.1, 230.2, 202.2, 119.1
L-Glu	16.265	262.1, 230.2, 202.2, 119.1
D-Phe	17.063	266.1, 162.2, 131.1, 119.2, 103.2, 91.2, 77.2
L-Phe	17.71	266.1, 162.2, 131.1, 119.2, 103.2, 91.2, 77.2
L-Lys	27.317	420.1, 230.2, 176.2, 119.1
DL-MeLan [b]	29.138	379.1, 262.1, 248.0, 234.1, 202.1, 188.1, 119.1
LL-MeLan [b]	29.399	379.1, 262.1, 248.0, 234.1, 202.1, 188.1, 119.1
L-*allo*-MeLan [b]	29.779	379.1, 262.1, 248.0, 234.1, 202.1, 188.1, 119.1
D-*allo*-MeLan [b]	29.959	379.1, 262.1, 248.0, 234.1, 202.1, 188.1, 119.1
LL-Lal [*,c]	28.261	640.2, 465.2, 405.1, 230.1, 190.1, 119.0, 67.0
LD-Lal [*,c]	28.334	640.2, 465.2, 405.1, 230.1, 190.1, 119.0, 67.0

Table S3.4. Gene annotations for the *div$_A$* biosynthetic gene cluster. CDSs of the *div$_A$* biosynthetic gene cluster (GenBank: KY115608) from the E11-036 metagenome are listed alongside their most similar BLAST search results. Conserved protein domains and biochemical characterizations are included.

Gene	Length (bp)	Closest BLAST hit	% ID	Annotated function (Organism)	Conserved protein domains (accession)	Functional characterization
-2	369	WP_007353413.1	54	hypothetical protein (*Kamptonema*)	none	not characterized
-1	321	XP_010382366.1	32	uncharacterized protein LOC104678724 (*Rhinopithecus roxellana*)	none	not characterized
divA	276	WP_017718744.1	63	hypothetical protein (*Oscillatoria sp.* PCC 10802)	none	precursor gene
divM	3255	WP_017718743.1	70	hypothetical protein (*Oscillatoria sp.* PCC 10802)	LanM-like (cd04792)	class II lanthionine synthetase
divX	918	WP_017718741.1	58	hypothetical protein (*Oscillatoria sp.* PCC 10802)	Nif11 (pfam07862)	putative Asp ß-hydroxylase (Okesli, 2011)
divY	780	WP_038094010.1	45	hypothetical protein (*Tolypothrix bouteillei*)	Het-C (pfam07217)	not characterized
divMT	798	WP_015192250.1	50	SAM-dependent methyltransferase (*Stanieria cyanosphaera*)	Methyltransf_18 (pfam12847)	methyltransferase
divT	1941	WP_008178015.1	64	ABC transporter permease and ATPase (*Moorea producens*)	ABC_membrane_2 (pfam06472), ABC_tran (pfam00005)	not characterized
divU	858	WP_069349017.1	63	hypothetical protein (*Scytonema millei*)	FGE-sulfatase (pfam03781)	not characterized
divLA	351	WP_008178005.1	53	hypothetical protein (*Oscillatoria sp.* PCC 10802)	none	putative role in Lal formation (Okesli, 2011)
+1	1500	WP_015145144.1	60	phosphopeptide-binding protein (*Pleurocapsa minor*)	CHAT (pfam12770), FHA (pfam00498)	not characterized
+2	1905	WP_015185269.1	54	serine/threonine protein phosphatase (*Microcoleus sp.* PCC 7113)	PP2Cc (cd00143)	not characterized

Table S3.5. Comparison of the partial *div$_B$* biosynthetic gene cluster to *div$_A$*. CDSs of the partial *div$_B$* biosynthetic gene cluster (GenBank: KY115609) from the E11-037 metagenome are compared with their respective *div$_A$* homologues.

Gene ID	Length (bp)	Coverage with corresponding *div$_A$* gene	% ID to corresponding *div$_A$* gene
-1	286 (partial)	88	100
divA	276	100	93
divM	111 (partial)	3	100
divX	604 (partial)	66	100
divY	272	37	86
divMT	603 (partial)	76	99

Table S3.6. Summary of quantitative NMR experiments. All samples were suspended in 120 µl of D_2O. A known concentration of L-Tyr was used to assess the quality of the L-Trp standard curves. A difference of 7.6% was calculated for the NMR-determined concentration relative to the UV-determined concentration. All divamide samples were quantified with (top row) and without (bottom row) inclusion of the methyl region integral because this integral accounts for a large number of protons, which may reduce the accuracy of integration. Desmethyl divamide A, divamide B, and divamide C were quantified using integrals for only two proton signals such that numbers obtained with exclusion of methyl integrals represent a single integral. Yields were calculated from the average of the two concentrations.

Sample	MW (g/mol)	UV mg/ml	NMR mg/ml	SD	mM	SD	yield (µg)
L-Tyr	181.19	0.326	0.351	0.00950	1.94	0.0524	42.1
E11-036 divamide A	2020.92	–	0.692	0.167	0.342	0.0824	77.8
			0.605	0.120	0.299	0.0596	
E. coli divamide A	2020.92	–	1.34	0.251	0.664	0.124	157
			1.27	0.198	0.630	0.0980	
desmethyl divamide A	1977.92	–	3.85	2.48	1.95	1.25	326
			1.59	0.0881	0.802	0.0445	
divamide B	1930.83	–	0.983	0.328	0.509	0.170	100
			0.691	0.0779	0.358	0.0403	
*divamide C	2077.95	–	0.272	0.0736	0.131	0.0354	31.7
			0.256	0.0797	0.123	0.0383	

Table S3.7. Summary of quantitative LC/MS experiments. Concentrations were determined from the standard curve of desmethyl divamide A shown in Figure S3.9. For mixtures of compounds, the concentrations in mg/ml were calculated based on individual molecular weights before adding together. The concentration of *E. coli* divamide A determined by LC/MS was in agreement with the NMR concentration (Δ -5.92%).

Sample	MW (g/mol)	mM	SD	mg/ml	SD	yield (μg)
E. coli divamide A	2020.92	0.624	0.0363	1.26	0.0733	**151.4**
extraproteolytic divamide A	1977.86	0.193	0.114	0.381	0.226	45.7
	2048.9	0.0212	0.00269	0.0433	0.00550	5.2
	2119.94	0.0232	0.00798	0.0491	0.0169	5.9
	2233.02	0.369	0.132	0.825	0.295	99.0
	mix	0.606	0.257	1.30	0.544	**155.7**
deslysinoalanine divamide A	1977.86	0.440	0.0848	0.871	0.168	104.5
	2048.9	0.173	0.0373	0.355	0.0764	42.6
	2119.94	0.0701	0.0199	0.149	0.0422	17.8
	mix	0.684	0.148	1.37	0.286	**164.9**
E11-036 divamide A degradation product	2020.92	0.567	0.0343	1.15	0.0694	137.4
	2034.9	0.409	0.0666	0.832	0.136	99.8
	mix	0.976	0.0324	1.98	0.205	**237.3**
cinnamycin	2040.85	0.448	0.0666	0.914	0.136	109.7
	2056.85	0.450	0.0624	0.925	0.128	111.1
	2072.84	0.148	0.0253	0.306	0.0526	36.7
	mix	1.05	0.129	2.15	0.317	**257.5**
biotinylated divamide A	2430.11	0.103	0.0354	0.250	0.0861	30.0
	2684.26	0.115	0.0384	0.309	0.103	37.1
	mix	0.218	0.0737	0.559	0.189	**67.1**
biotinylated cinnamcyin	2491.08	0.435	0.0491	1.08	0.122	129.9
	2507.08	0.203	0.0290	0.508	0.0726	61.0
	mix	0.637	0.0781	1.59	0.195	**190.9**

Table S3.8. Divamide-like masses observed from *Didemnum molle* extracts. Ions observed by LC/MS resembled those of characterized divamides in *m/z* and isotope distribution. While some masses appear to be derivatives of the characterized divamides A-C, others likely reflect alternative amino acid sequences at various stages of post-translational modification.

Observed *m/z* ($z = 2$)	Source	ID
1010.9	E11-036	divamide A
880.5	E11-037	?
893.5	E11-037	?
900.5	E11-037	?
935.9	E11-037	?
942.9	E11-037	?
944.9	E11-037	divamide B - 3 x Me
951.9	E11-037	divamide B - 2 x Me
953.9	E11-037	?
958.9	E11-037	divamide B - 1 x Me
965.9	E11-037	divamide B
973.9	E11-037	divamide B + 1 x O
974.9	E11-037	?
982.9	E11-037	?
1010.5	E11-037	?
1016.5	E11-037	?
1018.5	E11-037	?
1025.5	E11-037	?
1026.5	E11-037	?
1027.5	E11-037	?
1039.5	E11-037	divamide C
1047.5	E11-037	divamide C + 1 x O
1053.9	E11-037	?
1089.5	E11-037	?
1091.9	E11-037	?
1098.9	E11-037	?
1099.9	E11-037	?
1127.6	E11-037	?
1146.0	E11-037	?
1185.5	E11-037	?
1203.6	E11-037	?

Table S3.9. DNA sequences used to construct pDiv. gBlocks purchased from Integrated DNA Technologies.

DNA	Sequence (5'-3')	Length (bp)
1a	ctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacaATGCCGACCACCCTGGAAAAACCGAGCGTGGCCTATCTGGAGA AACTGTTTCACCAGACCGCCATCGACAGCGAATTTCGTAGCGAACTGCAGAGCCATCCGGAAGCCTTTGGTATTAGCGCCGATCTGGAACTGCC GCAGAGCGTGGAGAAACAGGACGAAAGCTTCGTGGAACTGCTGAACAACGCCCTGGGCGAAATCGATATTGCCGCCGAATGCGCCAGCACCTGT AGCTTTGGCATCGTGACCATCGTGTGCGATGGTACCACCAAATAAGTTGTTTTACTAGCATATTATATTATTAATATGTCTGGGCTAGCACCAC ATTGTGAAGCTAGCCTCATTAAATTGATTGTTGGGCATGGTTAGAAAGGAACTAAGTACAAGACAAAACTTGGATGAGAAGCGTAGTAAAATGT CCGAAACTATTGGCTCATAAAGTTTCGAGCAATATTTTCCATCTCTCTTGCTCCATATGAGTTTCAGTTGTTTCAGCTG	549
1b	AATCCGGAGACCAAGGAACTGGTGACCCTGGCCGAGAAACCGAAAATCGTGCGCTTCTTCAAAATCGAAAGCGAGCCGAGCGAAAATTCTGTGGC ACTGCATTGTTAACACCACCAAGCTGAATCGCTGGAGCTTCGGTTTCAACGTGAGCAGCAAGTGGATGGACAAACTGCTGACAGTGTAAGGAGA ATAAAATGAAAATTTTCCTGACATGCCTGCTGGCCCTGGCCCTGCTGTTAGGTATGCCGAGCAGTGCCTTTGCCTTCAAAGTGCCGATCCACGA AGAGATCACCCGCGAAGTGTTTGAGGATTTCCAGGTGGTGGTGGAAGGCGAGACCTTTAAGTTCACCGACTATGCCATCGACCAGATCGTGAAG GCCAACAAGGATACCGATGACCTGCCGAACCAGTTCAATACCGAGATGCACTTCGACGGCGAAGACTTTAGCGGTGGCAGCAATCGTGTGATGT TCCTGAAGGAGCGCACCATTACCAAAGTGACCGACCCTCAGGATCCGCAGGGCACAAGCGCACGTAATGATCTGGGTACCGCCCTGCATACCGT GCAGGACTTCTATGCCCATAGCAATTGGGTGGAACTGGGTCACAGCAGCAGCGATATCAACACCAAGATCGGCCGCGAGGTGTTCAGCGGCGCC GATAAAAACACCGCCACCTGCCCGAACGATCCGGGCATTCTGGGTGGTGCCGGCCTGACCGAACTGACCAGTGGCTACTTCACCTTCATCGGCG TTGTGCCTAGCTGCGATGTTCCTGAGGGTAAATGCCGCCATGGCGTTCCGATTGTGTGCCCGGATGGCCTGAACAAGGACGATAATAGTCGCCC GGGTTTTCCGACCGCCCGTGCATTAGCCGTTAAAGCCACCGAGGACTTCCTGAATCAGATCTTCAGCGACAGCCGTATGGACGGCAACGTGGAT GCCATCAAACTGCTGATGCGCATCCGTAATTAATCCCTTTAACAAGTAAATTGTACCTCGCCTAAACCATACCCGACAGGATTGGTTTAGGCTT ATCTGCTGCTTTATAAAAGACAGATAGAGATTAGAATTATCGCAATGGAAAATAATAATTATCCGTTCGAACTGAAAGCCTATGACTTTAGCTT TAAAATCTTCAAGGAGATCATTTTCGCCCTGAACGCCTTCTTCATTGGCTTTGGGTGTGCTGAAACGCGAACATTACCACCTGGTGGAT AGCATTTACTACAACCAGACCGAAATGTACCGCGACGAGAACTACAACAAACGCGGCCTGTGGGACTGGGAAGAGAAAGTTCTGGCCCAGTACT TCCAGCAGTGCCATAATCTGCTGGTGGTTGCAGCAGGCGGTGGTCGTGAGGTTCTGGCCTTATGCAAACGCGGCTATGAAGTGGACGGCTTTGA GTGCAATGCCAACCTGCTGAAATTCGCCAACAACCTGATCAAGCAAGAGGAGTTTGCCAGCCATATCAAACTGGCACCGCGTGATCAGTGCCCG GATAGCCAGAAGGAGTATGATGGCCTGATCGTTGGCTGGGGTGCCTATATGCTGATCCAGGGTAAGGAACGCCGCATTGAATTCCTGCGTCAGC TGCGTACCCAGGCCAAGAAGAACAGTCCGGTGCTGCTGAGCTTCTTCTGCTACAGCGAGACCACCGGCGGCCGTGATTTCAAAGCCATCGCCAT GATCGGCAATGCCTTTCGTCGCCTGCTGGGTCGTGAATGCCTGGAAGTTGGCGATAATCTGGCCCCGAACTACGTGCACTACTTCACCAAGGAC GAGATCGCCAAGCGAACTGCAGGCAGGCGGCTTTGAACTGAAGATCTACTGCACCAATCAGTATGGCCACGCCGTGGGTATCGCAGTGTAAcacc tttaagcctatagtctaggaaaaagaggttgaattagagccaacttggtcatctgaaaccctcatgctgtcgtttcccctgaacagttacgaa actacgaagatagagaagctttATGTATAAAGGCCTGGATCGCGACGAAATTCGCCAGATTCAGGTGCTGATGCTGCTGTGCCTGTGCCTGAGC CCGCAGAGCAAACTGCGCCAGCTGCTGGAAATTGCACTGGCCGCCAGCGAAACCCAGATCATGACCCGCATGACACCGTGCGATGATGTGAATG TGGACGGCCTGTTTACCTGGGTGCAGAGCCTGTTTGCCCAGGGTGGTCTGACCGAAGAAGAGAAACGCCTGCTGAAGTGGCAGAACGAAAGCCG TAACATGCTGCCGGCCATTGACGAACTGAAAACCATTGAGAAGAAGCTGGGCTTCAAGATCCGCATTCAGAAGCTGCAGAGCCACAATTAAtta ttggggtgcggtcaaaaacagttgaattttaagtcgcggcttacccg	2397
2	ATTTTCCATCTCTCTTGCTCCATATGAGTTTCAGTTGTTTCAGCTGCACCGAGCGCAAGGAGAATAACGAACCTTGTTTATTTTTACTGTTAAA GCAAAACACCAATTTCACACTGCGCACCAGCACCATGCTGGACGACTTCAGCCTGTTAAAGTTAGCCAGTCGCGCAAGCAACCTGTGTGAGCAA ACCCTGATTGTGAAAGAGCTGGCAAAGAGCAAAGCACCGATCGCCAGCACAACACAACTGAGTCCTGTTGACAGCTGGAAAATTAAGAAACTGA CAGGCAAACTGGCCGTGCAGCCGTTCAAAGAGAGCTACGAGCAGGGCACAATCAGTCAGACGTGATTGAGGACCTGCGTAAACTGCTGATCGA TTACAAACTGTACGAACTGAACCTGGCCAAACTTAAGCGAGAGCGACCGCCTGGAATTTATTAAGCCTCACAGCCAGTGGTTAAAGGCATACCAG GCCGCCATGGCAACCCTGGATCTGCCGCGTGAAAAGTTCAGCGGTAGTTGTTGGGGCGAACCGGATATTTACTACGGTAAATTCGCCAAAGTGT GTGAACCGTTTCTGCGCCTGCTGCACCAAACCCTGCGCGGTACCGGCGACGCAATTAACGCCACCGCCGACAATTACCGCATTAATCCGCAGGT GGCCATCGACATCGAGCTGCATCTGCTGAACCGCTTCGAATTAGCCCTGGCCTGGGCCTTAGAGGCCAACATCAATGTTTATTGTAGTCAGAAA GCAATCGCCAAGAGCGAGGATGATGAAGCCTACATCGCATACCTGGAAGGAGACCTTCGACCGCAAGCAAGCATATCATGATTTTTATTGTC GTTTCCCGGTGCTGGCCCGTTGGTTAGCCCAAGTGACCTATTTCCTGTGCAATTTTGGCGAGGAAACCCTGCAACGCCTGACCAGCGATCGTGA GCAGATTGGCGCCACCTTCTTCGGCAGCAAGCCGATCAGTCAAATCAAAAGTTTTAAACTGGGTAAAAGCGACTACCATGCCGGTGCCAAGAGC GTGGTGATCGTGGAGCTGGAGCTGGCCAACAGCGAACCTGCCACCCTGGTGTATAAGCCTCGCAGCATTCAGAGCGAAGCGGCATGCAGGGCC TGCTGGCACAGCTGAACCAAGATAAGGTGGTTCGCTTTGCCCACTATCAGGTGCTGTGTCGCGATGGTTATGGCTATGCAGAGTTCATCCCGAG CGGCAAAAACCAGGTTCAGAATAAAGAAGATCTGAAGAAATTCTACCAACAGCTGGGCGGCTTTTTAAGCATCTTCCATATCCTGGGTGGCGGC GATCTGCACCATGAAAACATCCTGGTTGCAGATGGTAACGCCTTCATCTGTGACTGCGAGACCGTGCTGGAGGTGCTGCCGCAAGGCATGGATA AACTGCCTGGTACAGTTTTAGATAGTGTGTTTAAAACAGCCATGCTGGACTGGCCTCGTGATAGCGCAAGCCCGGAGAACAGCGAGATGATGAG CATTAGTGGTTACAGCGGTGGCGAAAGCTATGAGGTTGCATTCACCGTGCCGTGGTATCACCGTGAGATTGTGGATCCGCAGGACTATAAGGATAGCATTG TGGACGGTTTTAACCAGGTTTATACATGGTTTCAGCAGCACCCGACCAAGGCAATTACCCGTATTAAGGAGCTGTTTAGCAGTAGTTTAGTGCG CTTTATTAATTGGGGCACCCAAGCCTACGCCAAGAGCATTGTGGCCGTTCGTCACCCTAAGTGCCTGACCCTCTCTGGAGGTGGACCTGATC TTTAATAGCCTGAAAGAGCATAAACGCCAGTGGGACAAAAAGGGCGAACTGGCAGAGTTAGAGCTGGGCAGCCTGTGGCAACTGGATATCCCGA TTTTCACCGCCTTAGCCGCCGAAAGCAAAGACTTAATCTTCAATTATCAGTATAGCGTTAGTGATACCTTAGCCATCAGCCCGTTAGACAATGC CAAACGCCGTCTCGAGCAGCTGAGCACCGAGAACCG	2104
3	TAGACAATGCCAAACGCCGTCTCGAGCAGCTGAGCACCGAGAACCGTATCCGTCAGAATCAATACATCTATACCAGCCTGAGCACCGACAGAGAT CAACAGTCCGTACTTTATCGCCGCAGCCGTTAACTATGCCCAGCAGATCGGCTGCAGCTGTGTGAACAGCTGAGTAGCGATAGTAGTAGCAAAGCA CCTTGGCAGACATGGGACTACACCGCAACCGGCAAGCGCTTAGTGGATATCAGCGGCGACCTGTATGATGGCAGCGCCGGCATTTGTCTGTTCC TGGCCTACCTGGATGCCATCAAACCGCAAGTGGAATTCCGTCAGGCCGCCGAACGCGCCTTAGAATACAGTATCGAGAAACGTAACACAACCCT GATCGGTGCATTCCAGGGCGAAACAGGTCTGATTTATCTGCTGACCCATCTGGCACAGTTATGGGACAAACCGGCATTACTGGACCTGGCAGTT GACCTGAGCGACGAGCTGCTGCCGCGCATCAAACAGGACATCTACTTCGATATCCTGCATGGCGTTGCCGGTATCATTCCGGTTATGCTGGGCC TGGCCGAAGCAACCGGTGGTAAAGGCATTGATTGCGCATTACAGTGTGCCGAGCACCTGCTGGAGCAGGGTATTTACCAGGATAACACCCTGAG CTGGCCTCCGGGTCGTCCTGACCTGGTGCGCGGCAATTTTACAGGCTTCAGCCACGGTGCAAGCGGCATCGGTTGGGCCTTAATTATGCTGGGC TGCCACAGCAATAAGAGCGAGTATATTGAGGCAGGCCGTCAGGGCTTTGCCTACGAAGCCACACAGTTTGATGAGGAACAGCGCGATTGGTACG ACTTACGCAAGAGCGTTACCACCGCAGATAGCAACGAACCTCACTTTGCCAACGCATGGTGCAATGGTGCCGCCGGTATTGGCCTGAGTCGCAT CATTAGTTGGGCCGCCCTGGGCAAAACAGACGACGACATTCTGCGCGATGCCTACACCGCACTGAAGTGCAACCTTACGCAATTTCAACAAGCTG GGCAACGATAGCCTGTGTCATGGCAAGAGTGGTAACGCCGAGTTATTCCTGCGTTTTGCACAGCTGCGCGATACCCCGTATCTGCAAATGGAGG CAAACGTGCAGGCCCAGGCACAATGGCGCAACTTTGAAAAGGCACGCCGTTGGATGTGCGGTAGCACAAGGCAACGATGTTTTCCCGGATTTAAT GCTGGGCCTGGCCGGCATTGGCATGCACTTCCTGCGCCTGGCCTACCGTGAACGCGTTCCGAGCCCGTTATTATTAGATCCGCCTCCGCCGCCC ATCGACTAAGTTTTGAGTTTTTTAATTCAATATTCAACACCCAAGAGTAATATAGTTTAGTTTGTTTATTAGGATCGAGATGAGCAAAGAAACC GTGATTGAATTTTATGAAGCAATTTTTGAAAGTCCGGAATTCATTCAGGAAATTAAAGCCATTACCCGCCAAGAGGAACTGATCAAACTGGGTG CCCGCAACGGCTACCATTTTACCATGGAAGACCTGGCCCAGGCCGACGCCAGCTATATCCCGAAGAATAACCAGCCGCTGATCAGCATTGATAG CGACGACCGTGCCCGCGAAGAACTGCCGCGCCCGTATCACTATGAATTCGAGTTCAGTGAGATCCCGGGCTTCGAAGAGATCGATCGTGAACTG AAAAAACTGCAGATCAAACCGAACACCGTGGATCTGGACCTGTACGAGAAGAGTTTCCGCGAAGAAGATTTTAAATTTAACTATATTAGCCCGA CCGTGCCGGGCTTCCGCCAGTATTACTACAAGAGCCTGAAAAGCTATCTGGATCTGCCTAGCCCTCAGCCGGAATATGCCTGGCGTCCGTTTCA CCTGATCAATCTGGATTGCCACGTGGAGGATCCGCTGTACGAGGATTATTTCCAGACCAAGGTGCGTCTGCTGAAGCTGCTGGAGAATTGCCTG GAGACAGAGCTGCGCTTTAGTGGCAGCCTGTGGTATCCGCCGAACGCCTATCGCCTGTGGCACACCAATGAAACCCAGCCGGGCTGGCGCATGT ACCTGGTGGATTTCGACAATTTTGACGACAACCAGGAAGGCGAGGTGTTCTTCCGCTACATGAATCCGGAGACCAAGGAACTGGTGACCCTGGC CGAGAAACCGAAAAT	2177
divA-divamideB	GTGGAACTGCTGAACAACGCCCTGGGCGAAATCGATATTGCCGCCGAATGCGCGTCGACCTGTAGCAGCGGCCCGATCACCGCCGATCTGCGATG GTACCACCAAATAAGTTGTTTTACTAGCATATTATATTATTAATATGTCTGGGCTAcCCACCACATTGTGAAGCTAGCCTCATTAAATTGATTGT TGGGCATGGTTAG	

Table S3.10. PCR and sequencing primers. Primers used in the construction and sequencing of pDiv, pDiv2, and pRSFDuet-divMT.

#	Primer name	Sequence (5'-3')
1	div1a-fwd	CTCGTATGTTGTGTGGAATTGTG
2	div2-rev	CGGTTCTCGGTGCTCAG
3	div3-fwd	TAGACAATGCCAAACGCC
4	div1b-rev	CGGGTAAGCCGCGAC
5	divM-203-fwd	CGATCGCCAGCACAAC
6	divM-1424-rev	CCAGTCCAGCATGGCTG
7	divM-2818-fwd	GCAGGCCGTCAGGG
8	divX-650-rev	AGCCCGGCTGGGTTTC
9	divX-580-fwd	GAGACAGAGCTGCGCTTTAG
10	divMT-79-rev	CAATGAAGAAGGCGTTCAG
11	divMT-20-fwd	CGTTCGAACTGAAAGCCTATG
12	divM-203-fwd	CGATCGCCAGCACAAC
13	divM-1977-f	CTTAGCCGCCGAAAGC
14	divX-461-f	TGCCTAGCCCTCAGCC
15	divHypo-780-f	CGCATCCGTAATTAATCCC
16	intergenic-divM-native-fwd	GAGCAATATTTTCCATCTCTCTTGCTCCATATG
17	intergenic-divM-native-rvs	GGGTGTTGAATATTGAATTAAAAAACTCAAAACTCAATC
18	divMp-fwd	GTGGAACTGCTGAACAAC
19	divM-non-optimized-pET16b-Gibson-fwd	cgaaggtcgtcatATGCTTGATGATTTTTCTCTCC
20	divM-native-fwd-1	CTTTAGCCTGGGCTCTTG
21	divM-native-fwd-2	GGACAGTGCTTGATTCAG
22	divM-native-rvs-4	CTTCATCAAATTGGGTCG
23	divMT-MCS1-fwd	TTCGAGCTCGATGGAAAATAATAATTATCCGTTCG
24	divMT-MCS1-rvs	TATGCGGCCGCTTACACTGCGATACCCACG

3.7 References

1. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013**, *30*, 108–160.

2. Knerr, P. J.; van der Donk, W. A. *Annu. Rev. Biochem.* **2012**, *81*, 479–505.

3. Cameron, D. M.; Gregory, S. T.; Thompson, J.; Suh, M. J.; Limbach, P. A.; Dahlberg, A. E. *J. Bacteriol.* **2004**, *186*, 5819–5825.

4. Dai, X.; Otake, K.; You, C.; Cai, Q.; Wang, Z.; Masumoto, H.; Wang, Y. *J. Proteome Res.* **2013**, *12*, 4167–4175.

5. Dognin, M. J. *FEBS Lett.* **1977**, *84*, 342–346.

6. Withers, N.; Vidaver, W.; Lewin, R. A. *Phycologia* **1978**, *17*, 167–171.

7. Donia, M. S.; Fricke, W. F.; Ravel, J.; Schmidt, E. W. *PLoS ONE* **2011**, *6*, e17897.

8. Ökesli, A.; Cooper, L. E.; Fogle, E. J.; van der Donk, W. A. *J. Am. Chem. Soc.* **2011**, *133*, 13753–13760.

9. Tianero, M. D.; Pierce, E.; Raghuraman, S.; Sardar, D.; McIntosh, J. A.; Heemstra, J. R.; Schonrock, Z.; Covington, B. C.; Maschek, J. A.; Cox, J. E.; Bachmann, B. O.; Olivera, B. M.; Ruffner, D. E.; Schmidt, E. W. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 1772–1777.

10. Widdick, D. A.; Dodd, H. M.; Barraille, P.; White, J.; Stein, T. H.; Chater, K. F.; Gasson, M. J.; Bibb, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4316–4321.

11. Pannecouque, C.; Daelemans, D.; De Clercq, E. *Nat. Protoc.* **2008**, *3*, 427–434.

12. Choung, S.-Y.; Kobayashi, T.; Inoue, J.-i.; Takemoto, K.; Ishitsuka, H.; Inoue, K. *Biochim. Biophys. Acta* **1988**, *940*, 171–179.

13. Yates, K. R.; Welsh, J.; Udegbunam, N. O.; Greenman, J.; Maraveyas, A.; Madden, L. A. *Blood Coagul. Fibrin.* **2012**, *23*, 396–401.

14. Zhang, K.; Yau, P. M.; Chandrasekhar, B.; New, R.; Kondrat, R.; Imai, B. S.; Bradbury, M. E. *Proteomics* **2004**, *4*, 1–10.

15. Lin, Z.; Torres, J. P.; Tianero, M. D.; Kwan, J. C.; Schmidt, E. W. *Appl. Environ. Microbiol.* 2016, *82*, 3450–3460.

16. Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10430–10435.

17. Sardar, D.; Pierce, E.; McIntosh, J. A.; Schmidt, E. W. *ACS Synth. Biol.* **2015**, *4*, 167–176.

18. Meyerson, N. R.; Sawyer, S. L. *Trends. Microbiol.* **2011**, *19*, 286–294.

19. König, S. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 2103–2104.

20. Burton, I. W.; Quilliam, M. A.; Walter, J. A. *Anal. Chem.* **2005**, *77*, 3123–3131.

21. Sokolov, E. P. *J. Moll. Stud.* **2000**, *66*, 573–575.

22. Tianero, M. D.; Kwan, J. C.; Wyche, T. P.; Presson, A. P.; Koch, M.; Barrows, L. R.; Bugni, T. S.; Schmidt, E. W. *ISME J.* **2015**, *9*, 615–628.

23. Joshi, N. A.; Fass, J. N. **2011**, [Software] *(Version 1.33)*, https://github.com/najoshi/sickle.

24. Zerbino, D. R.; Birney, E. *Genome Res.* **2008**, *18*, 821–829.

25. Buhr, F.; Jha, S.; Thommen, M.; Mittelstaet, J.; Kutz, F.; Schwalbe, H.; Rodnina, M. V.; Komar, A. A. *Mol. Cell* **2016**, *61*, 341–351.

26. He, Z.; Kisla, D.; Zhang, L.; Yuan, C.; Green-Church, K. B.; Yousef, A. E. *Appl. Environ. Microbiol.* **2007**, *73*, 168–178.

27. Kawulka, K. E.; Sprules, T.; Diaper, C. M.; Whittal, R. M.; McKay, R. T.; Mercier, P.; Zuber, P.; Vederas, J. C. *Biochemistry* **2004**, *43*, 3385–3395.

28. Kodani, S.; Lodato, M. A.; Durrant, M. C.; Picart, F.; Willey, J. M. *Mol. Microbiol.* **2005**, *58*, 1368–1380.

29. Martin, N. I.; Sprules, T.; Carpenter, M. R.; Cotter, P. D.; Hill, C.; Ross, R. P.; Vederas, J. C. *Biochemistry* **2004**, *43*, 3049–3056.

30. Sebei, S.; Zendo, T.; Boudabous, A.; Nakayama, J.; Sonomoto, K. *J. Appl. Microbiol.* **2007**, *103*, 1621–1631.

31. Chen, H.; Boyle, T. J.; Malim, M. H.; Cullen, B. R.; Lyerly, H. K. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 7678–7682.

32. Hirose, M.; Nozawa, Y.; Hirose, E. *Zoolog. Sci.* **2010**, *27*, 959–964.

33. Yang, J. J., M.S. Thesis, University of Helsinki, **2016**.

34. Kodani, S.; Komaki, H.; Ishimura, S.; Hemmi, H.; Ohnishi-Kameyama, M. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 1159–1165.

35. Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102*, 7315–7320.

36. Fredenhagen, A.; Fendrich, G.; Märki, F.; Märki, W.; Gruner, J.; Raschdorf, F.; Peter, H. H. *J. Antibiot.* **1990**, *43*, 1403–1412.

37. Küsters, E.; Allgaier, H.; Jung, G.; Bayer, E. *Chromatographia* **1984**, *18*, 287–293.

38. De Weck-Gaudard, D.; Liardon, R.; Finot, P. A. *J. Agr. Food Chem.* **1988**, *36*, 717–721.

CHAPTER 4

CONCLUSIONS

4.1. An integrated approach for natural product discovery

Natural products provide a significant source of inspiration for drug design and are, in some instances, drugs themselves.[1] The drug design process, however, relies on the discovery of novel chemical entities and the identification of drug leads targeting unique biological processes. The discovery of novel chemical scaffolds has dwindled since the field's infancy. Particular locations or environments have been sampled redundantly such that more ubiquitous and copious chemistry, the "low-hanging fruit", is commonly re-isolated. However, a vast ocean of uncharted chemistry, or chemical "dark matter", exists in nature, waiting to be discovered but impossible to access using traditional approaches. Modern strategies employ genome mining to identify putative biosynthetic clusters based on comparison to known biosynthetic machinery. This has the advantage of providing the blueprints for biosynthesis, but in the absence of structural or biological characterization it cannot predict the therapeutic value of a pathway.

Here, I have described two instances in which structural and genomic approaches were integrated for the discovery and supply of therapeutically relevant natural products. In both cases, structural features were used to predict specific aspects of the biosynthesis that, at least for the divamides, has led to pathway discovery. The adociasulfate example

shows how an integrated approach can be used to begin to recognize pathways that may not fall within any typical natural product pathway organization or, as for the divamides, can be used to gain access to novel chemistry from small samples. Such a strategy could be applied to even smaller samples by utilizing small-scale, high-throughput LC-MS-SPE-NMR screening to collect structural data for very small amounts of material.[2-3] Additionally, hundreds of natural products with known biological activities exist for which total synthesis is exceedingly difficult or impossible. A biosynthetic analysis of these compounds followed by genomic or metagenomic interrogation may reveal novel routes for natural product construction. Unique pathways interwoven with primary metabolism may exist, or even cooperative pathways in which multiple organisms take part in the production of a compound. Such instances would not be recognized without taking into account both structure and genetics.

To date, natural product discovery has relied heavily on obtaining large quantities of material for structural and biological characterization, and has recently become overwhelmed with expansive genomic and metabolomic data. The union of the two approaches can result in a targeted discovery process, providing a compass with which unexplored chemical space can be probed.

4.2. References

1. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016**, *79*, 629–661.

2. Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T. S. *Anal. Chem.* **2012**, *84*, 4277–4283.

3. Sturm, S.; Seger, C.; Godejohann, M.; Spraul, M.; Stuppner, H. *J. Chromatogr. A* **2007**, *1163*, 138–144.