Dissertation Advisor: Professor Jon Clardy Alexandra Masu Cantley

Uncovering bacterial metabolites involved in eukaryotic development

Abstract

Microorganisms are ubiquitous. They crowd the soils, the oceans, even deserts and glaciers, and they make up the microbiota that live in and on the bodies of animals. Crammed into their environment with many other competing microorganisms, they form alliances and disputes, often mediated by secreted molecules. For evolutionary biologists, such interactions can reveal how organisms compete for resources, how finely tuned association may aid their fitness in diverse environments, and whether such interactions are conserved in other organisms. Chemists have traditionally focused on the mediation of such symbioses via the production of, or response to, secondary metabolites. Such studies further our understanding of the natural world and may also reveal novel biologically active molecules that can be used to treat human disease.

The focus of this thesis is the relationship between bacteria and eukaryotic organisms. While traditionally bacteria were associated with pathogenesis, they are increasingly viewed as essential contributors to eukaryotic health. Chapter 1 discusses a handful of symbiotic relationships that have provided chemical ecologists with opportunities to discover new chemistry and increase their understanding of the impact of bacteria on eukaryotes. In Chapters 2 and 3 both chemical and evolutionary questions are addressed through the study of chemical communication between bacteria and the choanoflagellate, *Salpingoeca rosetta.* Choanoflagellates are aquatic eukaryotes and are the closest living relatives of animals. This

unique phylogenetic positioning and their ability to switch between a unicellular state and multicellular “rosettes” make choanoflagellates an ideal model organism for studying the emergence of multicellularity in animals. We found that the co-isolated bacteria, *Algoriphagus machipongonensis*, produces lipids that affect multicellular development in choanoflagellates through a complex set of synergistic and inhibitory activities.

Finally, Chapter 4 focuses on a symbiotic relationship between a social amoeba, *Dictyostelium discoideum,* and co-isolated bacteria. Certain *D. discoideum* isolates have been shown to engage in “primitive farming,” in which *D. discoideum* engulf bacteria, which are subsequently released upon spore dispersal. While some of the farmed bacteria are a food source for the amoeba, the rest do not serve a readily apparent function in the symbiosis. Additionally, not all of the *D. discoideum* isolates are capable of farming. We examine how small molecules isolated from the farmed bacteria affect *D. discoideum* “fitness” and we also develop a screen for bacterial genes that have toxic effects on *D. discoideum*.

The overarching theme of the research presented herein is that relationships are complicated. Of the two major symbiotic systems explored, we have found that bacterial-eukaryotic interactions cannot be defined as obviously negative or positive; usually they are both. While developing model systems for studying symbiosis often requires simplicity, the underlying intricacies should not be ignored, as these are what often provide the clearest insight into the nature of the relationship.

Table of Contents

Abstract ..... **iii**

Table of Contents ..... **v**

Acknowledgments ..... **vii**

Chapter 1: Animals in a Bacterial World **1**

1.1 Introduction 2

1.2 Evolution 4

1.3 Defense 6

1.4 Development 8

1.5 Future prospects 11

1.6 References 12

Chapter 2: Isolation and synthesis of a bacterially produced inhibitor of rosette development in choanoflagellates **15**

2.1 Contributions and Acknowledgments 16

2.2 Abstract 17

2.3 Introduction 18

2.4 Results 19

2.5 Experimental 29

2.6 References 41

Chapter 3: Bacteria regulate choanoflagellate development with lipid activators, inhibitors, and synergists **44**

3.1 Contributions and Acknowledgments 45

3.2 Abstract 46

3.3 Introduction 47
3.4 Results 48
3.5 Discussion 61
3.6 Experimental 64
3.7 References 71

Chapter 4: Development of a screen for pathogenic effectors from a bacterial affiliate of *D. discoideum* 75

4.1 Contributions and Acknowledgments 76
4.2 Abstract 77
4.3 Introduction 78
4.4 Results 81
4.5 Discussion 91
4.6 Experimental 93
4.7 References 98

Chapter 5: Future directions and concluding remarks 103

Appendix 1: Supplementary materials for Chapter 2 107

Appendix 2: Supplementary materials for Chapter 3 133

Appendix 3: Supplementary materials for Chapter 4 155

Acknowledgments

I would first like to thank my advisor Jon Clardy for accepting me into his laboratory and providing me with invaluable advice and enlightening anecdotes. I greatly admire his unique ability to cultivate a lab that is intellectually rigorous, supportive and also quite fun; the Clardy lab was an all around lovely place to spend a periodically stressful five years of graduate school. I would also like to thank my previous mentors: Brent Stockwell, who encouraged me to go to graduate school and helped me get there, and Angelica Cibrian, who expanded my appreciation for ecology.

Of course I must also acknowledge every single one of the numerous Clardy lab members with whom I have overlapped. They have all been a delight - both personally and professionally. In particular I want to thank Pierre Stallforth – who was my first mentor in the Clardy lab – and continues (tirelessly) to offer me advice on science and life. Emily Derbyshire – my first “baymate” – who I will always consider to be an incredible scientist and overall role model; David Dietrich, who essentially taught me how to do synthesis; and Elissa Hobert and Tony Ruzzini, who made intra-lab collaborations both fun and educational. I would also like to acknowledge Mo Seyedsayamdost, Christine Beemelmanns, Ethan Van Arnam and Margaret Wong, who all contributed to my scientific growth.

I would also like to thank my dissertation advisory committee: Eric Rubin, Roberto Kolter and Suzanne Walker, who offered me constructive and thought-provoking feedback every step of the way; and Michael Gilmore who generously agreed to be a part of my thesis defense. I would also like to send a hefty thanks to Jason Millberg, who keeps all of us in the ChemBio Program on track.

Many of my projects have been the result of fruitful collaborations that contributed immensely to any success I had in graduate school. In particular Nicole King and Arielle Woznica played fundamental roles in the formation and progression of the choanoflagellate project. Working with them has provided me with a paradigm of how collaborations should operate.

Since coming to Harvard I have met a number of lovely people who have made the last five years truly enjoyable. Carolyn Brotherton, Nitzan Koppel, Kristen Seim, Jack Nicoludis, and Abe Waldman – thanks for inviting me into your clan. Margie Li, Lynn McGregor, Sixun Chen and Jasmina Marjanovic, you ladies have provided me with so many fond memories. Josh Paulk, you have single handedly managed to keep me (relatively) sane. So many thanks for the laughs, coffees and therapy sessions.

Non-scientists - I don't know what I would do without you: Arabella Beatty, Beryl Crofton Atkins, Kate Cetrulo, Julia McDonald and Eric Sturgill (technically a scientist). You have all filled my days with so much laughter, love and support.

I would like express my boundless love and gratitude for my wonderful family: Mariko, I will probably never forgive you for being so much smarter than I am – but I do admire it. I really couldn't have asked for a better older sister. Vicki and Lewis, a.k.a parents, thank you for always motivating me to pursue my interests, rarely criticizing my oddities, and providing me with all the love and support a girl could ask for. Finally, I would like to dedicate all my thesis work to Herb and Marge Sato – my wonderful grandparents – I miss you both every day.

Chapter 1: Animals in a Bacterial World[1]

[1] This chapter was reproduced in part from: Cantley, A. M.; Clardy, J. Animals in a bacterial world; opportunities for chemical ecology. *Nat. Prod. Rep.* 32, 888-892 (2015) Co-written with Jon Clardy.

1.1 Introduction

All animals originated and evolved on a planet already teeming with bacteria, and the two have been competing, co-existing, and cooperating ever since. Most research on the relations between animals and bacteria has focused on pathological interactions - the ways bacteria cause disease. Natural products chemistry has played a decisive role in these studies through defining bacterial virulence factors and discovering naturally occurring antibacterial agents. The pioneering studies leading to penicillin and streptomycin ushered in the antibiotic

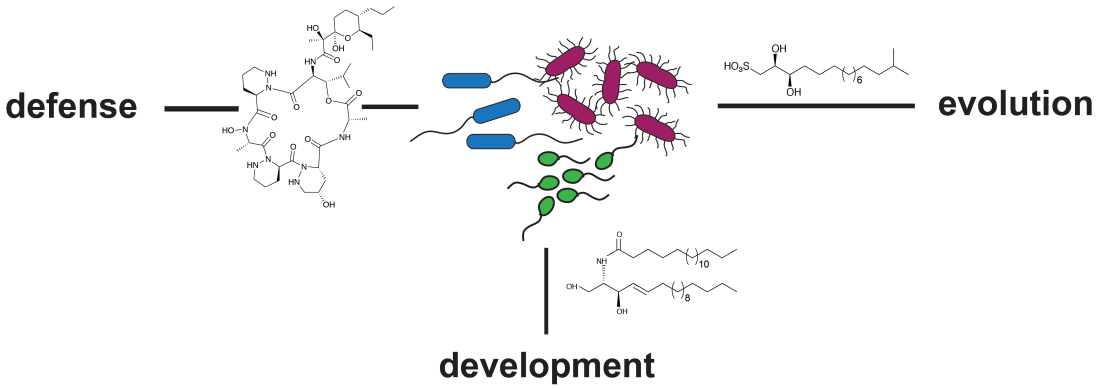


Figure 1.1 Bacterially produced small molecules play essential roles in a large number of eukaryotic processes.

era, and even in the current era (1981-2010) the number of new small molecule antibacterial agents developed from natural sources outnumbered those developed from synthetic molecules by 2:1.[1,2] New technological and bioinformatic approaches to natural product discovery will likely increase their contributions to new drugs.[3-6] The biological motivations for these studies have been almost exclusively medical, not ecological, and the roles of these antibiotics in the lives of their producers is even today very imperfectly understood.[7]

In the last few years studies on the non-pathogenic interactions between animals and bacteria have become increasingly frequent as biologists have begun to pose and answer questions dealing with the ways in which bacteria facilitated the origin, evolution, and

development of animals (Figure 1.1).[8] As bacteria largely sense and respond to the world around them with molecules, a complete answer to these questions requires a full description of the chemical ecology underlying bacteria-animal interactions, and providing this description creates greatly expanded opportunities for natural products chemists to deploy their skills on a fresh set of significant questions.

A chemical ecology approach to natural products has several noteworthy features. It inverts an increasingly common procedure in natural products chemistry by putting biological function ahead of chemical identification. Many current studies begin by identifying a molecule through metabolomic and/or bioinformatic analyses and then searching for a biological function. In contrast, an ecological approach begins with a function, and then identifies the responsible molecule(s) - an approach that reprises the procedure that led to many of our most useful drugs. An ecological approach also studies molecules in the physiological and ecological contexts in which they evolved, and knowing the relevant context enables approaches such as identifying inducers for triggering cryptic metabolite production, unraveling the evolutionary history of biosynthetic pathways, and suggesting medically relevant assays for further exploration and possible exploitation. This chapter will highlight some recent studies that illustrate the role bacterial small molecules played in metazoan evolution as well as how they influence developmental decisions and provide chemical defenses for animals.

1.2 Evolution

Animals are multicellular, and the development of multicellularity was a major evolutionary step in the animal lineage.[9] Multicellularity has evolved at least 25 times on Earth, but only once in animals. Since the 19th century, choanoflagellates have been considered a fitting candidate for understanding the transition to multicellularity, as phylogenetically they represent the last branch of unicellular organisms before multicellular animals emerged.[10] Choanoflagellates, which subsist on bacteria, are found in fresh, brackish and marine

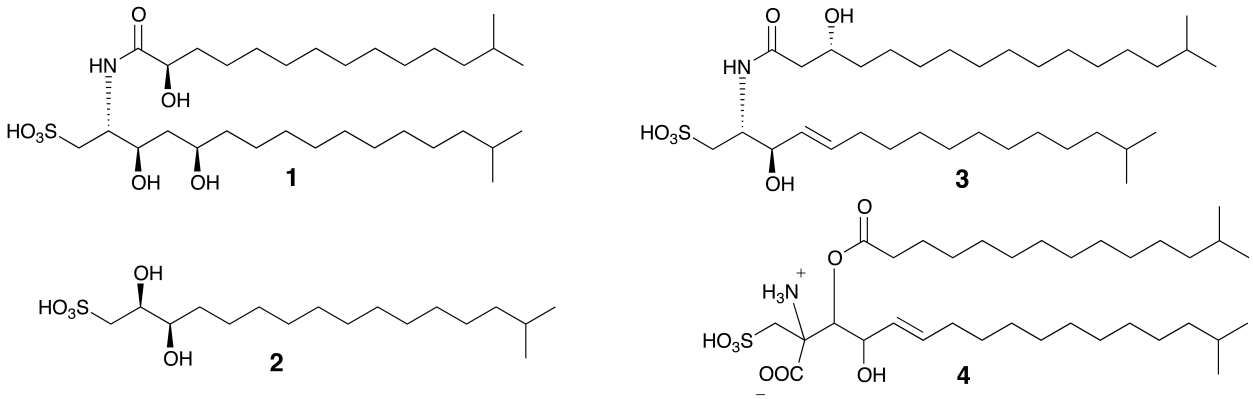


Figure 1.2 Bacterially produced lipids implicated in eukaryotic evolution.
The sulfonolipid RIF-1 (**1**) induces multicellularity in choanoflagellates, while IOR-1 (**2**) inhibits rosette development in choanoflagellates (**3**) and (**4**) are also sulfonolipids produced by marine bacteria, but their functions are unknown.

environments. Some, most notably *Salpingoeca rosetta*, occur in both single-cell and colonial forms; the colonial form, which is called a rosette after its shape, is formed by incomplete cell division from a single founding cell.[11] Surprisingly, the transition from the unicellular to colonial phenotype is induced by marine bacteria called *Algoriphagus machipongonensis*. Using rosette formation as an assay, the inducing molecule, rosette-inducing factor 1 (RIF-1, Figure 1.2) was identified as a sulfonolipid, a rare class of lipids that resemble sphingolipids.[12] The complete stereostructure of RIF-1 had to be defined through total synthesis as its femtomolar potency made isolation of significant quantities problematic. The modular synthesis coupled with further isolation studies produced roughly a dozen RIF analogs, *none* of which had any discernible

biological activity. This remarkably tight structure-activity relationship suggests a very restricted set of interactions between RIF-1 and its receptor.[13]

Further investigation into this system has now revealed that a more complex set of interactions is at work between choanoflagellates and their bacterial prey. Another sulfonolipid dubbed RIF-2, which is a close analogue of RIF-1, was revealed to be a more effective inducer of rosette development than RIF-1, though the mechanism behind the increased efficacy is not understood. RIF-2 also synergizes with lysophospholipids called lysophosphatidylethanolamines (LPEs). This synergistic activity serves to further increase not only the percentage of choanoflagellates in rosettes, but also the number of cells per rosette. This surprising finding suggests that there are two mechanisms relevant to rosette formation: initiation of rosette colony formation and stabilization of the colony.[14] Finally, *A. machipongonensis* also produces a single-digit nanomolar inhibitor of sulfonolipid induced rosette formation – the capnine diol IOR-1 (Figure 1.2). Like RIF-1, only one stereoisomer of IOR-1 is active, furthering the assumption that these lipid-receptor interactions are quite specific.[15] The discovery of the multiple activities of lipids produced by *A. machipongonensis* depicts the molecular and biological complexity, which can be easily overlooked, of this interspecies interaction.

Characterizing the mechanism of action of RIF-1, IOR-1 and related sulfonolipids could reveal homologous signaling pathways in other multicellular organisms, and the mechanism of action may even be general enough that examples could be found throughout the animal lineage. Additionally, while sulfonolipids are not well-studied molecules, they are produced by a number of different marine organisms (Figure 1.2).[16,17] Investigation of these other sulfonolipid producers and their associations with marine eukaryotes, especially sponges, may reveal additional functions.

Not all signals produced by bacteria that play roles in influencing the evolutionary steps along the animal lineage are small molecules. It has been known for several decades that

bacterially produced signals induce larval settling and the initiation of cell differentiation in the marine invertebrate *Hydroides elegans* – a process that has fascinated developmental biologists and is also implicated in biofouling.[1,2,18] Recently, the larval settlement inducer was identified as phage tail-like bacteriocins, which are contractile proteinaceous structures.[3-6,19] While these sorts of molecules had previously been shown to have antibacterial, insecticidal, and anti-feeding activity, they had never been associated with an essential morphological change in an organism's life history.[7,20-22] Further study of how *H. elegans* came to depend on a bacterial signal, while other closely related cnidarians settle in the absence of bacterial biofilms, will greatly enhance our understanding of the first steps in the evolution of the animal lineage.

1.3 Defense

As noted in the introduction, bacteria produce a staggering array of antibiotics, and humans are not the only animals that have benefited from their biosynthetic fecundity.[8,23] Beewolf digger wasps host symbiotic *Streptomyces* bacteria in specialized female glands, and they provide these bacteria to their larvae as they spin their protective cocoons.[9,24,25] Examination of the cocoons revealed that the incorporated *Streptomyces sp*. produce a diverse set of antibiotics that serve to protect the cocoon, and more importantly its inhabitant, from a variety of microbial predators. While each of the antibiotics alone had moderate activity against a range of predators, the antibiotic cocktail produced by multiple bacterial species (Figure 1.3), created a potent broad spectrum antibiotic activity, which argues that insect-bacterial systems evolved not only the use of bacterially produced antibiotics but also combination therapy long before humans.

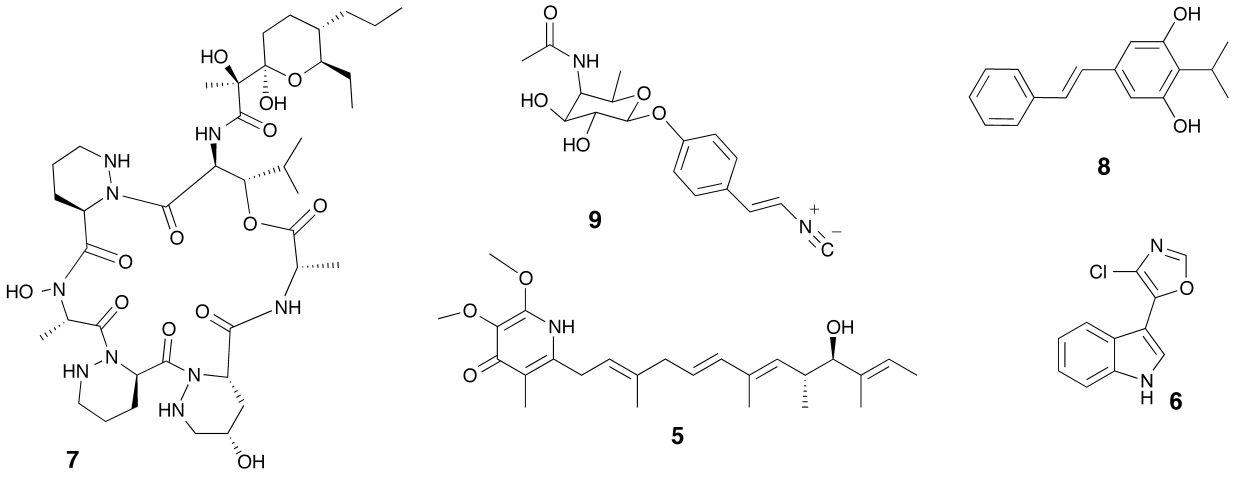


Figure 1.3 Bacterially produced molecules that mediate defense in eukaryotes
Piericidin (**5**) and streptochlorin (**6**) are antibiotics produced by *Streptomyces*. Dentigerumycin is an antifungal involved in resource protection in an ant-bacterial symbiosis (**7**). Stillbene-3 (**8**) is an antibiotic and inhibitor of insect innate immunity, and isocyanide rhabduscin (**9**) is a cryptic metabolite inhibits an essential enzyme in insect innate immune response.

Fungus-farming ants provide a variation on the beewolf system. As the name implies, fungus-farming ants cultivate a fungal food source that provides all of their nutrition. These cultivated fungi, which are grown in underground gardens by the ants, are plagued by a specialized pathogenic fungus that can overwhelm the fungal gardens and destroy the colony.[10,26] The ants host a single strain of Actinobacteria, often housed and fed in highly derived anatomical features called crypts, which provide chemical defenses against the fungal pathogen. In an early study on this system, dentigerumycin (Figure 1.3), was isolated from the bacterial symbionts (*Pseudonocardia* sp.) of the ant *Apterostigma dentigerum*, and dentigerumycin selectively killed the *Escovopsis* sp. pathogen rather than the ants' fungal cultivar.[11,27]

The ants and wasps in the first two examples benefited from symbiotic bacteria, but other animals, like insectivorous *Heterorhabditis* nematode worms, use symbiotic bacteria (*Photorhabdus luminescens*) to prey upon insects.[12,28] The bacteria live peacefully within their nematode host while it searches for insect larvae in the soil, but when the worm enters an insect

larva, the bacteria emerge and begin producing insect toxins, an array of degrading proteases and esterases, antibiotics, and developmental signals to initiate feeding and reproduction in the worms. These nematodes are used as agricultural control agents, and the system attracted the attention of both biologists and chemists.[13,29-31]

One barrier to discovering the antibiotic (and other) molecules being produced was the differential lifestyles of the *P. luminescens* symbionts. While the pathogenic bacteria in the insect produced interesting molecules, the quiescent ones living in the worms were not nearly as prolific. In laboratory culture, the bacteria displayed little of their biosynthetic potential. The productive lifestyle could be triggered by a factor in insect hemolymph (L-proline), which could be used to induce the production of formerly cryptic metabolites in laboratory cultures. L-proline induction led to the identification of several upregulated metabolites - including stillbene-3 (Figure 1.3), an antibiotic and inhibitor of the insect innate immune system, the antibacterial nematophin, as well as a number of cryptic metabolites, such as the isocyanide rhabduscin (Figure 1.3), which disables a key enzyme in the insect's innate immune response.[16,17,32,33] Cryptic metabolites – metabolites that are not produced in standard laboratory settings – are typically cryptic because their production is tightly regulated. In some cases, an environmental trigger like L-proline in the above example is sufficient to upregulate production, but in other cases the regulation have additional layers of repression that need to be lifted.[34] Investigation of these types of interactions can not only give us access to novel natural products, but can further our understanding of how these molecules are regulated in the environment.

1.4 Development

Animal microbiomes, the microbial population living on or in an animal, have been largely studied using massive sequencing efforts. The data from the human microbiome project,

for example, has already generated a million times more data than the initial human genome project, and these data have been useful in generating hypothesis free analyses including the discovery of new natural products. This approach is exemplified by a recent study on the biosynthetic potential of the microbiome, which used a bioinformatics driven approach to reveal that biosynthetic gene clusters encoding thiopeptide antibiotics are widely distributed in the human microbiota. Further, this study also reported a novel thiopeptide, lactocillin, that preferentially targets gram positive vaginal pathogens over commensal vaginal strains. While additional investigation is required to determine the activity of lactocillin *in vivo* – this study suggests that the human microbiome could be a reservoir of novel therapeutics.[35]

Analyses springing from observational hypotheses are much rarer, but the sphingolipids that have been shown to mediate the interactions between *Bacteroides*, an abundant member of the human gut microbiome, and the human immune system form a very interesting exception. Multiple studies pointed to the ability of *Bacteroides* to antagonize invariant natural killer T-cells (iNKT), and later studies pinpointed bacterially produced sphingolipids as the relevant signal. Sphingolipids are important structural and signaling molecules in mammals, including humans – and ubiquitous sphingolipids such as sphingosine-1-phosphate (Figure 1.4) or ceramide (Figure 1.4) have been shown to regulate processes related to cell senescence, apoptosis, cell motility and inflammation.[36] Sphingolipid diversity and function in bacteria, however, is largely unknown.[37] These studies showed that bacterial sphingolipids regulate the iNKT cells through lipid-antigen presentation by the major histocompatibility complex protein, CD1d. [38-40]

This immunomodulatory activity can have important implications in the management, or exacerbation of conditions characterized by hyperactive immunological responses such as autoimmune disorders, or cell-mediated immunity against pathogens. While the general scheme of iNKT regulation was understood, linking particular glycosphingolipids to specific immune responses had not been done. Recently, it was shown that a pervasive human (and mouse) gut

microbiome member, *Bacteroides fragilis,* produces a glyocosphingolipid (α-GalCer) that protects against chemically induced colitis by restricting iNKT population size. This study also revealed that in mice, pre-natal exposure to these bacterially produced glycosphingolipids is necessary for their full anti-proliferative effects, suggesting that exposure to certain bacterial species during early development is an important feature of the mammalian immune response.[41,42]

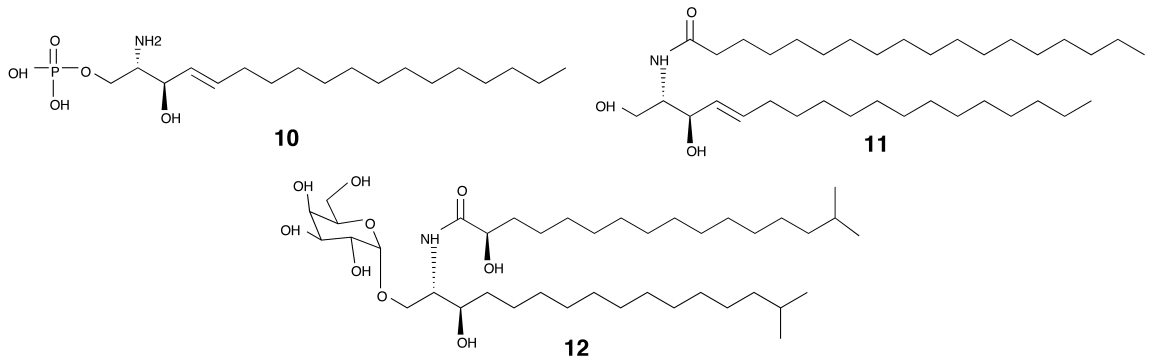


Figure 1.4 Bacterial sphingolipids mediate eukyarotic immune responses.
Sphingosine-1-phosphate (**10**) and ceramide (**11**) regulate essential mammalian cellular processes. α-GalCer (**12**) is involved in iNKT cell expansion.

In a similar study also investigating the sphingolipid repertoire of *B. fragilis*, an α-GalCer was identified that acts as an agonist of iNKTs (Figure 1.4) While the net effect of these *B. fragilis* glycosphingolipids on iNKT population size appears to vary between these studies, it is clear that these molecules are potent regulators of iNKT activation and that minor structural differences between these glycosphingolipids may lead to significant changes to their biological activity.[43] It is also interesting to note that these lipid signals resemble those discussed in the earlier section on evolution. Sphingolipids are ubiquitous molecules in both bacteria and eukaryotes so it is likely not a coincidence that they would serve as excellent interkingdom signaling molecules.

Another very intriguing example of the ability of the human gut microbiome to influence human development comes from a study on the corrective effects of *B. fragilis* in a maternal

immune activation mouse (MIA) – a model that recapitulates several key features of autism spectrum disorders (ASD). This study revealed that *B. fragilis* colonization of the gut could modulate the levels of several key metabolites known to be altered in the ASD mouse model.[44] It would be interesting to see if *B. fragilis* small molecule metabolites could also generate the same metabolomic regulation – and such a study, or a similar study, would likely require a natural products chemist as part of the interdisciplinary team.

1.5 Future prospects

Since every animal – not to mention every plant and fungus – has its own microbiome, the number of possible interactions in these multilaterial systems is effectively unlimited. Study of these interactions will undoubtedly reveal dynamic chemical conversations – such as the production of metabolites in response to inducer molecules from another organism. These inducer molecules alter secondary metabolite expression to reveal previously "cryptic" molecules – expanding opportunities for novel structure discovery and enhancing our understanding of how ecological cues can regulate expression of secondary metabolites. A particularly promising set of interactions exists in the human microbiome and as the first round of DNA sequencing draws to a close, the task of annotating the incredibly complex but important set of chemical interactions that literally make our life on Earth possible now begins.

1.6 References

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75,** 311–335 (2012).

2. Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **1830,** 3670–3695 (2013).

3. Carter, G. T. NP/MS since 1970: from the basement to the bench top. *Nat Prod Rep* **31,** 711–717 (2014).

4. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158,** 412–421 (2014).

5. Charlop-Powers, Z., Owen, J. G., Reddy, B. V. B., Ternei, M. A. & Brady, S. F. Chemical-biogeographic survey of secondary metabolism in soil. *Proc. Natl. Acad. Sci. U.S.A.* (2014). doi:10.1073/pnas.1318021111

6. Bouslimani, A., Sanchez, L. M., Garg, N. & Dorrestein, P. C. Mass spectrometry of natural products: current, emerging and future technologies. *Nat Prod Rep* **31,** 718–729 (2014).

7. Davies, J. Specialized microbial metabolites: functions and origins. *J. Antibiot.* **66,** 361–364 (2013).

8. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 3229–3236 (2013).

9. Rokas, A. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu. Rev. Genet.* **42,** 235–251 (2008).

10. Richter, D. J. & King, N. The Genomic and Cellular Foundations of Animal Origins. *Annu. Rev. Genet.* **47,** 130919210319006 (2012).

11. Fairclough, S. R., Dayel, M. J. & King, N. Multicellular development in a choanoflagellate. *Curr. Biol.* **20,** R875–6 (2010).

12. Alegado, R. A. *et al.* A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *elife* **1,** (2012).

13. Beemelmanns, C. *et al.* Synthesis of the Rosette-Inducing Factor RIF-1 and Analogs. *J. Am. Chem. Soc.* **136,** 10210–10213 (2014).

14. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. USA* **113,** 7894–7899 (2016).

15. Cantley, A. M., Woznica, A., Beemelmanns, C., King, N. & Clardy, J. Isolation and

Synthesis of a Bacterially Produced Inhibitor of Rosette Development in Choanoflagellates. *J. Am. Chem. Soc.* **138,** 4326–4329 (2016).

16. Corcelli, A. *et al.* Novel Sulfonolipid in the Extremely Halophilic Bacterium Salinibacter ruber. *Appl. Environ. Microbiol.* **70,** 6678–6685 (2004).

17. Takikawa, H., Nozawa, D., Kayo, A., Muto, S.-E. & Mori, K. Synthesis of sphingosine relatives. Part 22. Synthesis of sulfobacin A, B and flavocristamide A, new sulfonolipids isolated from Chryseobacterium sp. *J. Chem. Soc., Perkin Trans. 1* 2467–2477 (1999).

18. Hadfield, M. G. Biofilms and Marine Invertebrate Larvae: What Bacteria Produce That Larvae Use to Choose Settlement Sites. *Annu. Rev. Marine. Sci.* **3,** 453–470 (2011).

19. Shikuma, N. J. *et al.* Marine Tubeworm Metamorphosis Induced by Arrays of Bacterial Phage Tail-Like Structures. *Science* (2014). doi:10.1126/science.1246794

20. Jabrane, A. *et al.* Characterization of serracin P, a phage-tail-like bacteriocin, and its activity against Erwinia amylovora, the fire blight pathogen. *Appl. Environ. Microbiol.* **68,** 5704–5710 (2002).

21. Hurst, M. R. H., Glare, T. R. & Jackson, T. A. Cloning Serratia entomophila antifeeding genes--a putative defective prophage active against the grass grub Costelytra zealandica. *J. Bacteriol.* **186,** 5116–5128 (2004).

22. Strauch, E. *et al.* Characterization of enterocoliticin, a phage tail-like bacteriocin, and its effect on pathogenic Yersinia enterocolitica strains. *Appl. Environ. Microbiol.* **67,** 5634–5642 (2001).

23. Bode, H. B. Insects: true pioneers in anti-infective therapy and what we can learn from them. *Angew. Chem. Int. Ed. Engl.* **48,** 6394–6396 (2009).

24. Kroiss, J. *et al.* Symbiotic Streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat Chem Biol* **6,** 261–263 (2010).

25. Seipke, R. F., Kaltenpoth, M. & Hutchings, M. I. Streptomyces as symbionts: an emerging and widespread theme? *FEMS Microbiology Reviews* **36,** 862–876 (2012).

26. Currie, C. R. A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. *Annu. Rev. Microbiol.* **55,** 357–380 (2001).

27. Oh, D.-C., Poulsen, M., Currie, C. R. & Clardy, J. Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nat Chem Biol* **5,** 391–393 (2009).

28. Waterfield, N. R., Ciche, T. & Clarke, D. Photorhabdus and a host of hosts. *Annu. Rev. Microbiol.* **63,** 557–574 (2009).

29. Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in Xenorhabdus and Photorhabdus: two roads to the same destination. *Mol. Microbiol.* **64,** 260–268 (2007).

30. Schild, H.-A., Fuchs, S. W., Bode, H. B. & Grünewald, B. Low-molecular-weight metabolites secreted by Paenibacillus larvae as potential virulence factors of American foulbrood. *Appl. Environ. Microbiol.* **80,** 2484–2492 (2014).

31. Brachmann, A. O. *et al.* A type II polyketide synthase is responsible for anthraquinone biosynthesis in Photorhabdus luminescens. *Chembiochem* **8,** 1721–1728 (2007).

32. Crawford, J. M., Portmann, C., Zhang, X., Roeffaers, M. B. J. & Clardy, J. Small molecule perimeter defense in entomopathogenic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 10821–10826 (2012).

33. Crawford, J. M., Kontnik, R. & Clardy, J. Regulating alternative lifestyles in entomopathogenic bacteria. *Curr. Biol.* **20,** 69–74 (2010).

34. Kontnik, R., Crawford, J. M. & Clardy, J. Exploiting a global regulator for small molecule discovery in Photorhabdus luminescens. *ACS Chem. Biol.* **5,** 659–665 (2010).

35. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158,** 1402–1414 (2014).

36. Hannun, Y. A. & Obeid, L. M. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nat. Rev. Mol. Cell Biol.* **9,** 139–150 (2008).

37. Olsen, I. & Jantzen, E. Sphingolipids in Bacteria and Fungi. *Anaerobe* **7,** 103–112 (2001).

38. Kinjo, Y. *et al.* Natural killer T cells recognize diacylglycerol antigens from pathogenic bacteria. *Nat. Immunol.* **7,** 978–986 (2006).

39. Brennan, C. A. *et al.* A model symbiosis reveals a role for sheathed-flagellum rotation in the release of immunogenic lipopolysaccharide. *elife* **3,** e01579 (2014).

40. Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434,** 520–525 (2005).

41. An, D. *et al.* Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* **156,** 123–133 (2014).

42. Olszak, T. *et al.* Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. *Science* **336,** 489–493 (2012).

43. Wieland Brown, L. C. *et al.* Production of α-Galactosylceramide by a Prominent Member of the Human Gut Microbiota. *PLoS Biol* **11,** e1001610 (2013).

44. Hsiao, E. Y. *et al.* Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. *Cell* **155,** 1451–1463 (2013).

Chapter 2: Isolation and synthesis of a bacterially produced inhibitor of rosette development in choanoflagellates[1]

[1] This chapter was adapted from: *Cantley, A. M., *Woznica, A., Beemelmanns, C., King, N. & Clardy, J. Isolation and Synthesis of a Bacterially Produced Inhibitor of Rosette Development in Choanoflagellates. *J. Am. Chem. Soc.* **138,** 4326–4329 (2016).

2.1 Contributions and Acknowledgments

Contributions

Besides myself, Arielle Woznica, Christine Beemelmanns, Nicole King and Jon Clardy contributed to the research described in this chapter. I performed the synthesis, isolation and characterization of IOR-1. Arielle performed the rosette induction and inhibition bioassays, and Christine Beemelmans performed initial identification of IOR-1. Nicole King and Jon Clardy provided research guidance and funding.

Acknowledgments

C.B. was supported by a postdoctoral fellowship from the German National Academy of Science Leopoldina (LPDS2011-2). N.K. is an investigator in the Howard Hughes Medical Institute and a Senior Scholar in the Integrated Microbial Biodiversity Program of the Canadian Institute for Advanced Research. We would like to thank Furong Sun at the University of Illinois at Urbana-Champaign Mass Spectrometry Laboratory. The authors would also like to thank Drs. D. Dietrich and P. Stallforth for consultations on synthetic strategy and manuscript preparation.

2.2 Abstract

The choanoflagellate, *Salpingoeca rosetta*, is a microbial marine eukaryote that can switch between unicellular and multicellular states. As one of the closest living relatives of animals, this organism has become a model for understanding how multicellularity evolved in the animal lineage. Previously, our labs isolated and synthesized a bacterially-produced sulfonolipid that induces *S. rosetta* to form multicellular "rosettes". In this study we report the identification of a bacterially-produced Inhibitor of Rosettes (IOR-1) as well as total synthesis of this molecule and all of its stereoisomers. Our results confirm the previously noted specificity and potency of rosette-modulating molecules, expand our understanding of the complex chemical ecology between choanoflagellates and rosette-inducing bacteria, and provide a synthetic probe template for conducting further mechanistic studies on the emergence of multicellularity.

2.3 Introduction

The foundational event in animal origins, the transition to multicellularity,[1-3] occurred in oceans filled with diverse bacteria.[4-7] There is a growing appreciation that specific bacteria direct diverse animal developmental processes, including light organ development in the Hawaiian bobtail squid and immune system development and maturation in organisms as diverse as cnidaria and mammals.[8-20] However, the multicellularity of animals and the complex communities of bacteria with which they often interact hinder the complete characterization of many host-microbe dialogues.[2]

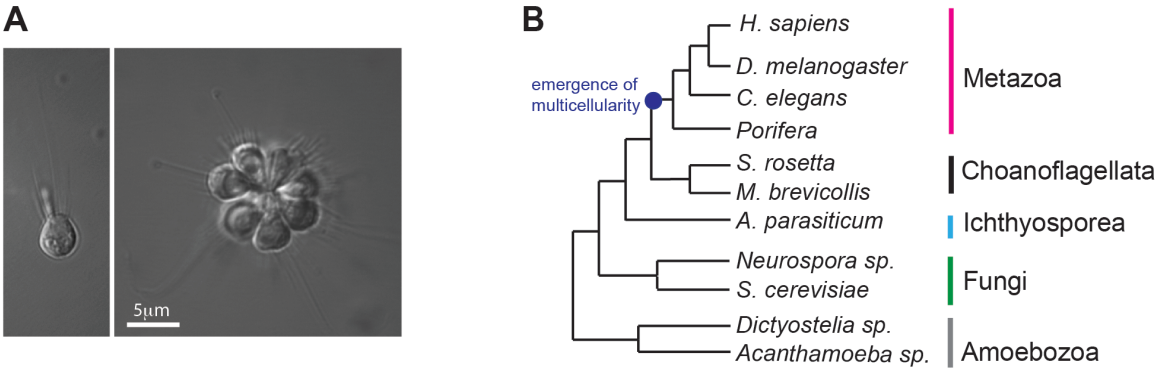


Figure 2.1 Choanoflagellates and the origins of multicellularity.
(A) *S. rosetta* in its unicellular and multicellular forms. **(B)** Phylogenetic tree depicting the divergence of the choanoflagellata just prior to the emergence of multicellular metazoan.

Choanoflagellates are motile microbial eukaryotes that reside in aquatic environments and feed on bacteria. Much like the collar cells of sponges, these microscopic organisms use a single apical flagellum to sweep surrounding bacteria into their actin-rich collar where the bacteria are phagocytosed.[21] Choanoflagellates, which are the closest living relatives of animals, express diverse genes, such as C-type lectins, cadherins and tyrosine kinases, that are known to regulate multicellular processes in animals (Figure 2.1).[22-25] While predominately unicellular,

[2] This initial paragraph was reproduced from: *Woznica, A. & *Cantley, A.M.*et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. USA* **113,** 7894–7899 (2016).

several species of choanoflagellate, including *Salpingoeca rosetta*, alternate between unicellular and multicellular states (Figure 2.1). In an embryogenesis-like process, the multicellular form, known as a "rosette", arises through multiple rounds of cell division in which the sister cells do not completely separate from each other.[26,27] While full mechanistic understanding of rosette development is yet to be achieved, further study of the transition to multicellularity in this ancient organism could provide meaningful insights into how multicellularity evolved in the animal lineage.

2.4 Results

We previously showed that the transition between the unicellular form and the multicellular "rosette" is induced by a sulfonolipid produced by *Algoriphagus machipongonensis* ('Algoriphagus,' for short) – a marine bacterium, originally co-isolated with *S. rosetta*, that serves as prey for the choanoflagellate.[28,29] Subsequent synthesis of the inducing molecule, termed rosette-inducing factor-1 (RIF-1) revealed the absolute configuration of the molecule, as well as the strict stereochemical requirements for activity (Figure 2.2).[30]

However, while RIF-1 could induce a small percentage of cells to form rosettes, the activity of RIF-1 alone did not faithfully recapitulate the activity observed with live bacteria or conditioned medium. Additionally, we noted apparent fluctuations in the activity of isolated (natural) RIF-1, as well as sphingolipid-enriched extracts, leading us to hypothesize that Algoriphagus produces additional choanoflagellate-modulating molecules that could serve as alternative inducers, synergists, or possibly even inhibitors. In this report we describe the isolation and synthesis of a bacterially-produced sulfonate-containing lipid that inhibits sulfonolipid-induced rosette formation in *S. rosetta*.

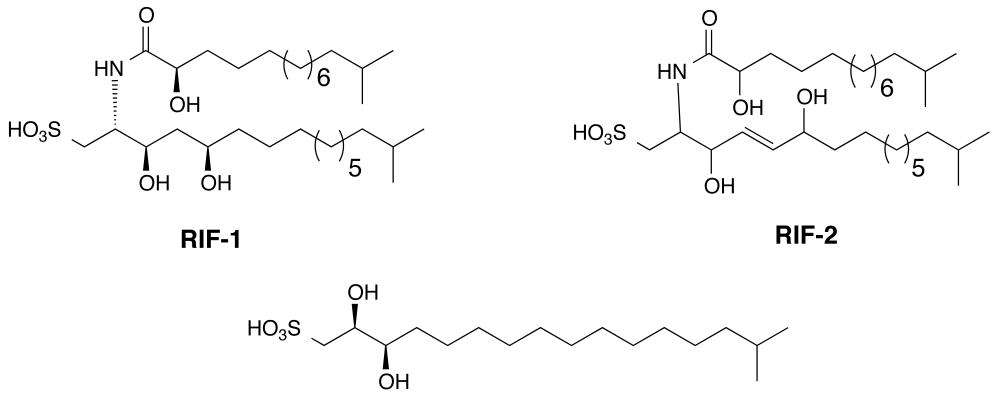


Figure 2.2 Rosette inducing molecules RIF-1 and RIF-2 and an inhibitor of rosettes, IOR-1. RIF-1 was the first isolated inducer of rosette formation in *S. rosetta*.[28] RIF-2, a close analogue of RIF-1, was recently found to be more effective inducer of rosette development.[31] IOR-1 is a capnine that inhibits sulfonolipid induced rosette formation.

We performed a chloroform/methanol extraction on the cell pellet of Algoriphagus and fractionated the extract by reverse phase (C-18) HPLC, using a broad elution range in order to expand our search beyond sulfonolipids.[28] We then tested each fraction in combination with inducers of rosette development to determine if any fractions contained molecules with inhibitory activity. As inducers we used either a sulfonolipid-enriched fraction (RIF-mix) that elicits high levels of rosette formation (with up to 30% of cells in rosettes), or a purified sulfonolipid (RIF-2) - a close structural analogue of RIF-1 whose complete stereostructure remains to be fully elucidated. We identified two adjacent fractions that reduced rosette formation when treated in combination with either the RIF-mix or RIF-2.

High-resolution mass spectrometry revealed that both fractions predominately contained a molecule with a mass of [M-H] 351.2216 Da, matching a predicted formula of $C_{17}H_{35}O_5S$. 1D and 2D NMR experiments permitted us to propose the planar structure for this molecule, which we have named Inhibitor of Rosettes (IOR-1) (Figure 2.2). IOR-1 is optically active ($[\alpha]D22$ = +24 c 0.125, MeOH), and its absolute configuration was ultimately determined through synthesis as described below. Dose response curves using purified IOR-1 showed an optimal inhibitory

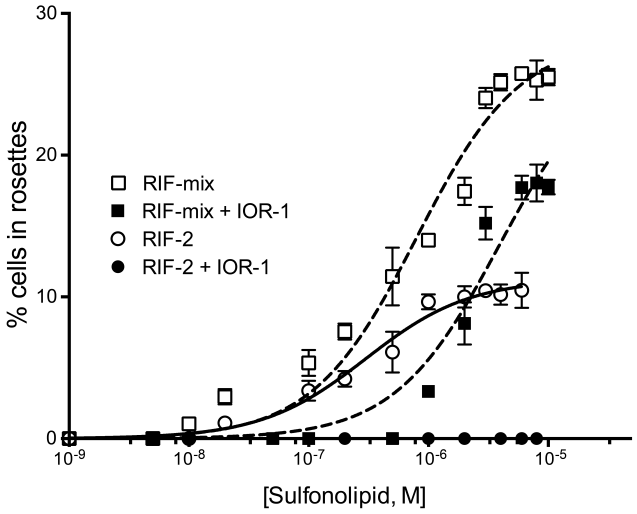


Figure 2.3 Co-treatment of IOR-1 (2.5 nM) with RIF-2 and RIF-mix.
Graphs were generated using GraphPad Prism 6 statistical software. Rosette induction data was analyzed using a one site (specific binding) model.

concentration of 2.5 nM (Figure 2.3), which corresponds with our observation of IOR-1's single digit nanomolar concentration in Algoriphagus conditioned medium.

We were intrigued by the structure of IOR-1 for several reasons. It resembles the capnine base found in bacterially-produced sulfonolipids, especially in that it contains the sulfonic acid head-group present in the previously identified RIF-1 and RIF-2. As capnine bases, like the analogous sphingoid bases, are biosynthetically derived from amino acids, the 2-position typically has a $–NH_2$ substituent, so the –OH group at this position on IOR-1 is a notable modification. In general little is known about capnines, and while they have been postulated to facilitate bacterial gliding, their functions are not well understood and their distribution is quite limited.[32,33] The more common class of sphingosine bases (or lysosphingolipids) act through G-protein coupled receptors to modulate diverse biological processes including triggering apoptosis and mediating inflammation.[34-36] The structural similarity between IOR-1 and these signaling molecules suggests that they may also share functional similarities.

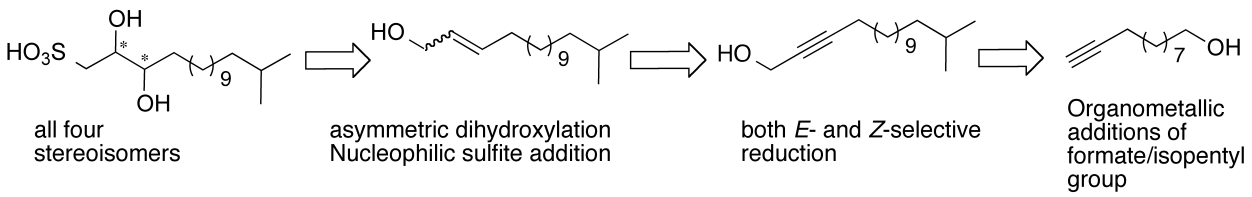
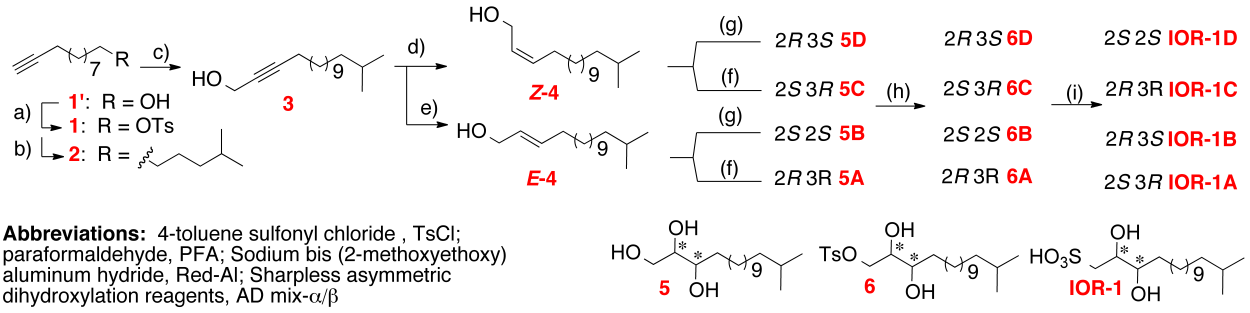


Figure 2.4 Retrosynthetic scheme for IOR-1 and stereoisomers.

We synthesized IOR-1 both to establish its absolute stereostructure and to determine whether it shared the same strict stereochemical requirements seen in RIF-1. Additionally, the relatively simpler synthesis of IOR-1, compared to that of RIF-1, makes IOR-based probes potentially valuable tools for identifying the host targets of rosette-modulating molecules.

As we needed to access all four possible configurations of the hydroxyl groups at the C2 and C3 positions, we reasoned that we could reduce an alkyl chain ending in a propargylic alcohol to either the corresponding *cis* or *trans* alkenes, and perform a Sharpless asymmetric dihydroxylation on both alkenes, using either the α or β mix, to yield all four stereoisomers. In a final step, the sulfonic acid moiety could be added to each purified stereoisomer through nucleophilic substitution (Figure 2.4)

To reach 15-methylhexadec-2-yn-1-ol (3) we started with commercially available 10-undecyne-1-ol (**1'**). We elongated the acyl chain and added the isopropyl tail through a Grignard reaction with isopentyl-MgBr in the presence of Li_2CuCl_4 to yield **2** in a manner similar to previously described methods.[37] The propargylic alcohol was obtained through acetylide formation and subsequent nucleophilic addition to paraformaldehyde.[38] At this stage our synthetic strategy diverged to obtain both the *cis* and *trans* alkenes. We used Lindlar's catalyst to reduce alkyne 3 to the *cis* alkene ***Z*-4** in the presence of H_2, and Red-Al® to obtain the trans analogue ***E*-4**. These reductions were achieved in acceptable yields of 74% and 70% respectively (Figure 2.5)



Abbreviations: 4-toluene sulfonyl chloride , TsCl; paraformaldehyde, PFA; Sodium bis (2-methoxyethoxy) aluminum hydride, Red-Al; Sharpless asymmetric dihydroxylation reagents, AD mix-α/β

Figure 2.5 Synthetic scheme for synthesis of IOR-1 stereoisomers A-B.
Conditions: (a) TsCl, pyridine, CH_2Cl_2, 4 °C, 10 h. (b) *i*-pentylMgBr, THF, Li_2CuCl_4 (cat.), 0 °C - RT, overnight , 42% (over 2 steps). (c) THF, *n*-BuLi, PFA, 0 °C - RT, 2.5 h, 66%. (d) Lindlar catalyst, MeOH, H_2, RT, overnight, 74% or (e) Red-Al, ether, 0 °C - RT, overnight, 70%. (f) AD mix-β, H_2O, *t*-butanol, methanesulfonamide, 0° C, 6-24 h, 77% - 87% or (g) AD mix-α, H_2O, *t*-butanol, methanesulfonamide, 0 °C, 6 - 24 h, 73% - 82%. (h) TsCl, pyridine, DCM, 4 °C, 10 h. Note: at this step tosylated compounds are purified to pure enantiomers by chiral HPLC. (i) Na_2SO_3, H_2O, ethanol, 62 °C, overnight, 14%-19%.

From this branch point we could access each diol configuration pattern through asymmetric bishydroxylation using the Sharpless reagents (AD mix-α and AD mix-β) in the presence of methanesulfonamide, which afforded yields in the 70 - 80% range.[39] This stage proved suitable to purify the diols via chiral chromatography yielding enantiopure **6A**, **6B**, **6C**, and **6D** (Figure 2.5)

Although we previously introduced the sulfonic acid moiety of RIF-1 through a Mitsonobu reaction using thioacetic acid followed by oxidation, with tosylates **6A-D** in hand, a simple nucleophilic displacement strategy at this position would be more efficient. While substitution with thioacetic acid and subsequent oxidation yielded IOR-1, side product formation frustrated the final purification. Addition of sodium sulfite in a heated biphasic solution of water and ethanol yielded fewer side-products, and while giving a low yield of final product (14 - 20%), allowed for a much simpler purification process and higher overall conversion.[40]

Table 2.1 ^{1}H NMR shifts of isolated IOR-1 vs. synthesized IOR-1A and IOR-1B

Atom	Isolated IOR-1		IOR-1A		IOR-1B	
	δ_H, m, H (J in Hz)	δ_C	δ_H, m, H (J in Hz)	δ_C	δ_H, m, H (J in Hz)	δ_C
1	3.08, dd, 1H (13.8, 3.1) 2.99, dd, 1H (13.9, 8.6)	55.23	3.09, dd, 1H (14.0, 3.5) 2.98, dd, 1H (14.0, 8.5)	55.28	3.07, dd, 1H (14.0, 3.2) 3.00, dd, 1H (14.0, 8.8)	54.90
2	4.04, dt, 1H (8.5, 3.2)	71.31	4.04, dt, 1H (8.4, 3.3)	71.36	4.05, dt, 1H (8.9, 3.3)	71.09
3	3.59, dt, 1H (7.6, 3.7)	74.25	3.60, dt, 1H (7.9, 3.6)	74.27	3.57, dt, 1H (7.8, 3.8)	74.09
4	1.58 – 1.50, m, 2H	33.69	1.61 – 1.49, m, 2H	33.73	1.58 – 1.47, m, 2H	33.45
5-13	1.37 – 1.27, m, 18H	28.4-30.9	1.42 – 1.28, m, 18H	28.49-31	1.37 – 1.23, m, 18H	28.22-33.45
14	1.22 – 1.17, m, 2H	40.12	1.24 – 1.17, m, 2H	40.19	1.20-1.15, m, 2H	39.90
15	1.58 – 1.50, m, 1H)	27.02	1.61 – 1.49, m, 1H	27.10	1.58 – 1.47, m, 1H	26.80
16, 16’	0.90, d, 6H (7.1)	23.02	0.91, d, 6H (6.5)	23.03	0.88, d, 6H (6.5)	22.97

Table 2.2 ^{1}H NMR shifts of isolated IOR-1 vs. synthesized IOR-1C and IOR-1D

Atom	Isolated IOR-1		IOR-1C		IOR-1D	
	δ_H, m, H (J in Hz)	δ_C	δ_H, m, H (J in Hz)	δ_C	δ_H, m, H (J in Hz)	δ_C
1	3.08, dd, 1H (13.8, 3.1) 2.99, dd, 1H (13.9, 8.6)	55.23	3.17, dd, 1H (14.0, 2.1) 2.89, dd, 1H (14.1, 9.6)	54.08	3.14, dd, 1H (14.1, 2.0) 2.89, dd, 1H (14.1, 9.7)	54.25
2	4.04, dt, 1H (8.5, 3.2)	71.31	3.96, ddd, 1H (9.5, 5.5, 2.0)	72.26	3.95, ddd, 1H (9.8, 5.4, 2.1)	72.36
3	3.59, dt, 1H (7.6, 3.7)	74.25	3.54, ddd, 1H (8.8, 5.6, 3.2)	74.61	3.54, ddd, 1H (8.8, 5.3, 3.2)	74.71
4	1.58 – 1.50, m, 2H	33.69	1.64 – 1.50, m, 2H	33.54	1.61 – 1.49, m, 2H	33.62
5-13	1.37 – 1.27, m, 18H	28.4-30.9	1.40 – 1.24, m, 18H	28.33-30.83	1.36 – 1.23, m, 18H	28.40-30.91
14	1.22 – 1.17, m, 2H	40.12	1.23 – 1.16, m, 2H	40.02	1.18 – 1.12, m, 2H	40.10
15	1.58 – 1.50, m, 1H)	27.02	1.64 – 1.50, m, 1H	26.69	1.61 – 1.49, m, 1H	26.76
16, 16’	0.90, d, 6H (7.1)	23.02	0.89, d, 6H (6.6)	23.00	0.86, d, 6H (6.6)	23.02

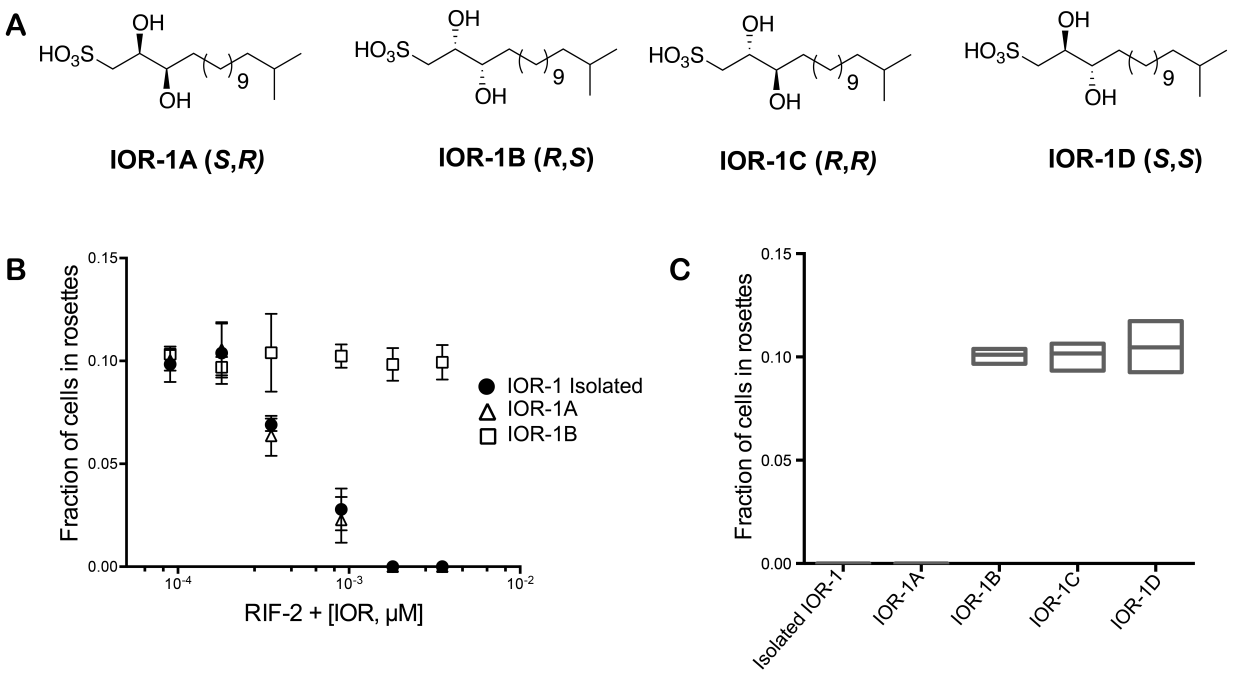


Figure 2.6 Rosette inhibiting activity of synthetic versus purified IOR-1.
(A) Structures of synthesized IOR-1 stereoisomers. **(B)** Dose curves of isolated IOR-1 versus synthetic IOR-1A and IOR-1B. RIF-2 treated at 2 μM. **(C)** Maximum inhibitory activity of IOR-1 isolated versus synthetic stereoisomers. Cells were induced to form rosettes with 2 μM RIF-2 and treated with 2.5 nM of indicated IOR-1 analogue.

^{1}H-NMR spectra of compounds IOR-1A and IOR-1B were identical to the isolated IOR-1 (Table 2.1), whereas compounds IOR-1C and IOR-1D exhibited different chemical shifts of protons at positions C1, C2 and C3 (Table 2.2). Determination of the optical rotations for these molecules revealed matching signs and value for IOR-1A and IOR-1, suggesting IOR-1A is likely a match to the isolated molecule.

To verify the activity and specificity of IOR-1, we tested each of the synthetic stereoisomers in our rosette inhibition assay. A full dose response curve revealed almost identical activity for IOR-1A as compared to the isolated inhibitor, whereas IOR-1B displayed no activity (Figure 2.6); unsurprisingly IOR-1C and 1D were also inactive. Given both the spectroscopic and biological data we were able to determine the absolute configuration of IOR-1 as 2S, 3R. Significantly, only one stereoisomer of inhibitor is active – reprising the theme that

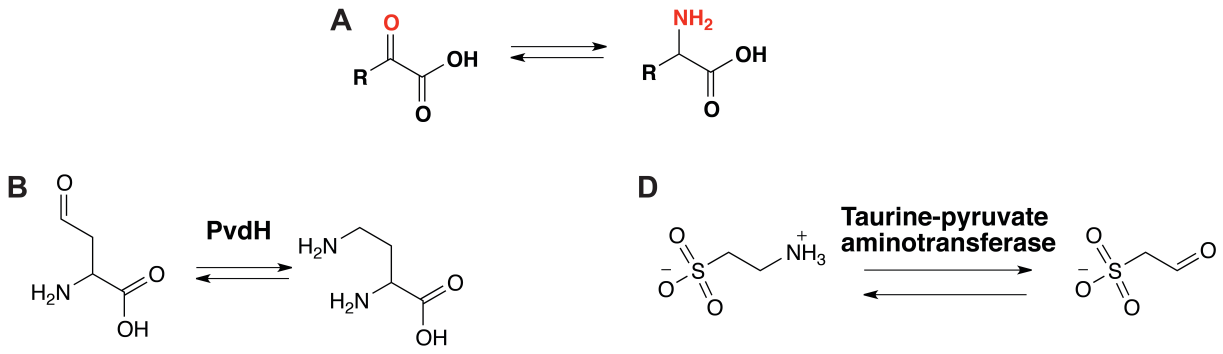
these molecules are interacting in a highly specific manner with their target. We further validated this specificity by testing a handful of commercially available IOR-1 analogs, and none were active at concentrations ranging from 0.1 ng/ml – 1 μg/ml (Table 2.3).

Table 2.3 Commercial compounds tested for inhibition of rosette formation

Compound name	Rosette inhibition
Safingol	No activity
D-erythro-sphinganine	No activity
Sphinganine-1-phosphate	No activity

Furthermore this specificity suggests IOR-1 (IOR-1A) is an appropriate starting point for the development of a bioaffinity probe that could be used to investigate the choanoflagellate target and mechanism of rosette-modulating molecules. Its straightforward synthesis and scalability allows for quick access to modified versions of IOR-1, and its potency (2.5 nM) would minimize the likelihood of non-specific interactions even if the probes were of somewhat lower potency. Our synthetic route also supplies us with inactive stereoisomers of IOR-1 - useful negative controls for target identification.

The assignment of the hydroxyls of IOR-1 in the *syn* configuration was unexpected; we had predicted that the hydroxyls would have the same relative configuration as the C2 and C3-position amine and hydroxyl groups observed in RIF-1, which is by far the most common stereochemistry for sulfonolipids and sphingolipids.[34] While not unheard of, it is quite rare for capnine bases to exhibit the *syn* configuration and biosynthesis of the *syn* diol has not been reported.[41,42] Exploration of the fully annotated genome of Algoriphagus confirmed the presence of a number of transaminases, which could invert the configuration of the hydroxyl group at the C2-position during conversion from an amino group (Figure 2.7).[43] As the biosynthesis of IOR-1 clearly has components that are distinct from the known sulfonolipids (*cf*. RIF-1), this molecule is unlikely to be either a degradation product or a precursor to the more standard sphingolipids and sulfonolipids. Further investigation into the biosynthesis and regulation of IOR-1 are



PvdH [*Pseudomonas aeruginosa*]
Gbk ID: BAT65564.1
Length: 469 aa
Locus: BAT65564

Taurine-pyruvate aminotransferase
[*Bilophila wadsworthia*]
Gbk ID: AAG50296.1
Length: 456 aa
Locus: AAG50296

C

RAST annotation	AA length	Identity (%)
Acetylornithine aminotransferase	397	31
Acetylornithine aminotransferase	393	30
Acetylornithine aminotransferase	376	28
Ornithine aminotransferase	411	26
Aminotransferase, class III	757	28
Glutamate-1-semialdehyde aminotransferase	430	28

E

RAST annotation	AA length	Identity (%)
Acetylornithine aminotransferase	397	34
Acetylornithine aminotransferase	393	29
Ornithine aminotransferase	411	28
Aminotransferase, class III	757	29
Acetylornithine aminotransferase	376	29

Figure 2.7 Putative aminotransferases from *A. machipongonensis* genome.
(A) Conversion performed by canonical aminotransferase. **(B)** Previously characterized diaminobutyrate – 2 – oxoglutarate aminotransferase (PvdH) isolated from *P. aeruginosa*. **(C)** PvdH homology search against *A. machipongonensis* genome **(D)** Taurine-pyruvate aminotransferase is a previously characterized enzyme from *B. wadsworthia*. **(E)** Taurine-pyruvate aminotransferase homology search against *A. machipongonensis* genome. The *A machipongonensis* genome (NCBI refseq: NZ_CM001023.1) was re-annotated using Rapid Annotation using Sybsystem Technology (RAST). BLAST+ was used for sequence comparison to previously characterized PvdH and taurine-pyruvate aminotransferases.

ongoing and will be of great interest in understanding the ecological context in which these molecules are produced.

From an ecological perspective the isolation and characterization of IOR-1 raises a number of interesting questions about the choanoflagellate-bacterial predator-prey relationship.

The isolation of both an inducer and an inhibitor from the same bacterium highlights the complexity of the relationship between *Algoriphagus* and *S. rosetta*. Our current hypothesis is that rosette formation improves bacterial prey capture by choanoflagellates, and if true, production of factors that attenuate rosette-colony formation would confer an apparent benefit to the producing bacteria.[44] Understanding how IOR-1 and RIFs are produced and regulated should begin to reveal how *Algoriphagus* could use both sets of molecules to manipulate its predators. More generally, examining the complex phenotypic effects triggered by these bacterially-produced small molecules will increase our understanding of the role of bacteria in the evolution of multicellular organisms.

In summary, we have isolated, characterized and synthesized an atypical sulfonolipid that potently inhibits the conversion from a unicellular to a multicellular morphology in choanoflagellates. Through synthesis we were able to confirm that this lipid has the rare *syn* diol configuration, and that 2S, 3R stereochemistry is necessary for activity. The discovery of this molecule reveals that the chemical interaction between choanoflagellates and rosette-inducing bacteria is more complex than previously imagined and argues that further investigation is warranted. Finally IOR-1 provides a starting point for pathway identification in this important model system.

2.5 Experimental

Instrumentation

All HPLC was performed on Agilent 1100 or 1200 series instruments. Specific columns used are specified in methods. LCMS was performed on an Agilent 1200 series HPLC with 6130 ESI mass spectrometer. High resolution mass spectrometry was performed on an Agilent 6530 QTOF LCMS (ESI) or a Waters Micromass (EI) 70-VSE (EI experiments conducted at the University of Illinois at Urbana-Champaign Mass Spectrometry Laboratory). Optical rotations were performed on a Jasco P-2000 polarimeter with a sodium lamp. NMRs were performed on the following instruments: Varian Inova 400 MHz, Varian Unity Inova 600 MHz, or a 500 MHz Oxford magnet with Varian Inova consul, equipped with a Varian HCN coldprobe.

Choanoflagellate husbandry

Salpingoeca rosetta strain SrEpac (Levin et al. 2013) was propagated in 5% Sea Water Complete media. 5% Sea Water Complete (SWC) media (250 mg/L peptone, 150 mg/L yeast extract, 150mL/L glycerol in artificial sea water) was made by diluting SWC to 5% (vol/vol) in artificial sea water. SrEpac was passaged 1:10 into fresh medium once a day to stimulate rapid growth. For all bioassays, lipids were added to SrEpac shortly after passaging, at a density of approximately 10^4-10^5 cells/mL. Rosettes were quantified 22-25 hours post induction.

Activity profile of lipids

Lipid inducing/inhibitory activity was determined using a quantitative bioassay for rosette development. Lipid samples were resuspended in DMSO to a concentration of 2mg/mL. Lipids were first pre-mixed in 5% SWC to avoid precipitation of the sample, and then added to 100mL SrEpac, aliquoted into 96-well plates (Corning Costar), to yield the desired concentration.

To quantify rosette development, SrEpac was pipetted vigorously and fixed in 1% formaldehyde immediately before counting (Bright-Line hemacytometer, Hausser Scientific). To determine the fraction of cells in rosettes, single cells and cells within rosettes were scored until 1000 total cells had been counted. A group of four or more cells qualified as a rosette if the cells maintained an organized polarity (each cell oriented with the apical flagellum pointing outward) after vigorous physical perturbation. At least three biological replicates were performed for each assay. Graphs were generated using GraphPad Prism 6 statistical software. Curves were fit to data using non-linear regression (curve fit, one site total).

Isolation of IOR-1 from *A. machipongonensis*

A. machipongonensis was grown in multi-liter scale in marine broth, shaking at 200 rpm at 30^{o}C for 3 days. Cells were pelleted by centrifugation and then extracted with 2:1 chloroform/methanol 2x and 1:1 chloroform/methanol 1x as previously described[30]. All phases were recombined and cell debris was removed by filtration. After drying, crude extract was fractionated by preparatory scale reverse phase HPLC using a Phenomenex Gemini NX C18 column (10μ, 110 A, 250 x 21.2 mm). Compounds were eluted at 10 ml/min in a gradient of solvents A (water + 0.1% NH_4OH) and B (MeOH + 0.1% NH_4OH) using the following method: 30% - 100% solvent B for 30 minutes, isocratic at 100% solvent B for 8 minutes, and ramp back down to 30% B over 2 minutes. Fractions containing IOR-1 eluted at around 80% B.

Purification of IOR-1

As this molecule is undetectable by UV, presence of IOR-1 was determined by LCMS on an Agilent 1200 series HPLC with 6130 series ESI mass spectrometer, injecting onto a Phenomenex Gemini NX-C18 column (110 A, 5 μm, 100 x 4.6 mm). Method for IOR-1 detection: compounds eluted at 0.5 ml/min using a gradient of solvents A (water + 0.1%

NH_4OH) and B (methanol + 0.1% NH_4OH), starting from 65% solvent B and increasing to 100% solvent B over 20 minutes.

To remove fatty acid impurities, fractions containing IOR-1 were resuspended in methanol and treated with TMS-diazomethane (Sigma Aldrich 362832), which was added until mixture turned yellow. Reaction was stirred vigorously at room temperature for ~15 minutes. Acetic acid was added dropwise to quench reaction. After drying *in vacuo*, entire mixture was purified by semi-prep HPLC using a Phenomenex Gemini NX-C18 (100A, 5μm, 250 x 10 mm) and the following method: With a flow rate of 2.4 ml/min, elute compounds with a gradient of 65 - 90% solvent B (MeOH + 0.1% NH4OH) over 20 minutes, 90-100% solvent B over 2 minutes, and isocratic at 100% B for 8 minutes. Compound was detected using an evaporative light scattering device (Agilent 1200 series ELSD).

To determine the concentration of IOR-1 in conditioned medium, 1 L *A. machipongonensis* was grown in marine broth for 3 days at 30 ^{o}C, shaking at 200 rpm. Culture was spun down and sterile filtered to remove cells. 500 mls of conditioned medium was lyophilized and resuspended in 20 ml methanol. 50 μl of suspension was injected onto the LC/MS for comparison with a IOR-calibration curve. Calculated equation for calibration: y = 17415x-29972, revealed that based on integration of the MS spectrum, 19.9 ng of 351 were injected, which back-calculated to 536 ng/500 ml. We were thus able to determine the concentration of IOR-1 in *A. machipongonensis* conditioned medium was calculated to be 2.84 nM.

Structure elucidation of IOR-1

High-resolution mass spectrometry (negative mode) of IOR-1 revealed a mass of 351.2216 Da, corresponding to a molecular formula of $C_{17}H_{35}O_5S^-$. Both 1D and 2D NMR spectroscopy were performed on IOR-1 to elucidate the structure of the molecule. 1H and 13C

NMR allowed us to determine the core structure of the molecule. ^{1}H NMR (600 MHz, $CD_3OD/CDCl_3$) δ 4.04 (dt, *J* = 8.5, 3.2 Hz, 1H), 3.59 (dt, *J* = 7.6, 3.7 Hz, 1H), 3.08 (dd, *J* = 13.8, 3.1 Hz, 1H), 2.99 (dd, *J* = 13.9, 8.6 Hz, 1H), 1.58 – 1.50 (m, 2H), 1.37 – 1.27 (m, 18H), 1.22 – 1.17 (m, 2H), 0.90 (d, *J* = 7.1 Hz, 6H); ^{13}C NMR (126 MHz, $CD_3OD/CDCl_3$) δ 74.25, 71.31, 55.23, 40.12, 33.69, 30.73, 30.67, 30.65, 28.42, 27.02, 23.02.

A suite of 2D NMR experiments (gCOSY, TOCSY, HSQCAD and HMBC) allowed us to confirm the placement of the diol and the sulfonic acid headgroups (Figure 2.8).

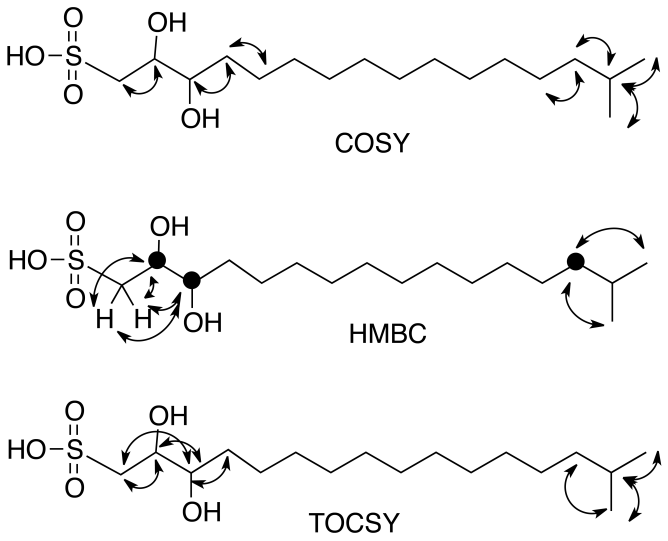


Figure 2.8 Key H-H and C-H correlations determined by 2D NMR

Synthesis of IOR-1A-1D

Undec-10-yn-1-yl toluene-p-sulfonate (**1**)

To a solution of undec-10-yn-1-ol (Alfa Aesar L11807, **1'**) (29.5 mmol, 5 g) in pyridine (11.8 mL) and dichloromethane (17.7 mL) at 0 ^{o}C, 1.5 eq. toluene-p-sulfonyl chloride was added. Mixture was stirred overnight at 4^{o} C. Mixture was then diluted with water and extracted with hexanes.

Extract was washed with water, aq. HCl and brine, dried over sodium sulfate and concentrated under vacuum. Moved to next step without further purification[37]. MS (EI) observed: 345.1498, expected: 345.1500 ($C_{18}H_{26}O_3SNa$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.79 (d, J = 8.3 Hz, 2H), 7.34 (d, J = 8.0 Hz, 2H), 4.02 (t, J = 6.5 Hz, 2H), 2.45 (s, 3H), 2.17 (td, J = 7.0, 2.7 Hz, 2H), 1.93 (t, J = 2.6 Hz, 1H), 1.67 – 1.59 (m, 2H), 1.56 – 1.46 (m, 3H), 1.30 – 1.16 (m, 9H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 144.90, 133.54, 130.03, 128.09, 84.95, 70.95, 68.38, 29.48, 29.18, 29.11, 29.08, 28.91, 28.69, 25.57, 21.95, 21.86, 18.64.

14-methylpentadec-1-yne (**2**)

To a solution of **1** (15.5 mmol, 5 g) in THF (17.85 ml), 2.1 eq. i-pentylMgBr (31 mmol, 23.2 ml) was added dropwise at 0 ^{o}C followed by immediate addition Li_2CuCl_4 (3.2 mL, 0.31 mmol). Mixture was allowed to return to room temperature overnight. Mixture was quenched with NH4Cl, extracted with hexanes and washed with water, sodium bicarbonate and brine.[37] Extract was dried over sodium sulfate and concentrated under vacuum yielding **2** (1.4 g, 42%) as a yellow oil. Product was purified by silica column using 100% hexanes yielding. Unable to obtain HRMS for this compound. Expected mass: 222.2348 ($C_{16}H_{30}$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.18 (td, J = 7.1, 2.7 Hz, 2H), 1.93 (d, J = 2.6 Hz, 1H), 1.56 – 1.48 (m, 3H), 1.41 – 1.36 (m, 2H), 1.27 (dd, J = 6.4, 3.2 Hz, 14H), 1.18 – 1.12 (m, 2H), 0.86 (d, J = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 85.17, 68.36, 39.42, 30.30, 30.06, 30.01, 29.97, 29.87, 29.48, 29.13, 28.87, 28.33, 27.78, 23.05, 23.01, 18.77.

15-methylhexadec-2-yn-1-ol (**3**)

To a solution of **2** (2 g, 9 mmol) in THF (14 mL) at 4 $^{\circ}$C, *n*-BuLi (20 mmol) was added. After 30 minutes, paraformaldehyde (10.8 mmol, 324 mg) was added in portions. Mixture was warmed to RT over 2 hours. Reaction quenched with 1:1 water/sat. NH_4Cl and extracted with hexanes.[38]

Crude material was purified on silica gel using 100% hexanes followed by 3:2 hexanes/ethyl acetate to obtain pure **3** as a light yellow solid (1.4g, 66%). Expected mass 252.2453 ($C_{17}H_{32}O$). Unable to obtain HRMS for this compound. ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.25 (t, *J* = 2.2 Hz, 2H), 2.20 (tt, *J* = 7.1, 2.2 Hz, 1H), 1.56 – 1.45 (m, 3H), 1.39 – 1.33 (m, 2H), 1.33 – 1.18 (m, 14H), 1.17 – 1.08 (m, 2H), 0.86 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 86.67, 78.23, 51.42, 39.04, 29.92, 29.69, 29.65, 29.62, 29.50, 29.13, 28.87, 28.59, 27.95, 27.40, 22.64, 18.72.

(*Z*)-15-methylhexadec-2-en-1-ol (***Z*-4**)

To a solution of **3** (1 g, 4 mmol) in methanol (10 mL), Lindlar reagent was added (500 mg) and stirred for 30 minutes. Flask was charged with H_2 and stirred overnight at room temperature. Reaction mixture was filtered over celite and concentrated under vacuum yielding ***Z*-4** as a white solid (0.73 g, 74%). Crude extract was purified by HPLC (C18) using a gradient of 85-100% acetonitrile. MS (EI) observed: 254.2615, expected: 254.2610 ($C_{17}H_{34}O$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.64 – 5.50 (m, 2H), 4.19 (d, *J* = 6.1 Hz, 2H), 2.07 (q, *J* = 7.0 Hz, 2H), 1.58 – 1.45 (m, 1H), 1.34 – 1.20 (m, 18H), 1.17 – 1.11 (m, 2H), 0.86 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 133.29, 128.26, 58.61, 39.04, 29.92, 29.69, 29.65, 29.59, 29.58, 29.47, 29.21, 27.95, 27.42, 27.39, 22.64.

(*E*)-15-methylhexadec-2-en-1-ol (***E*-4**)

To a solution of **3** (1 g, 4 mmol) in ether (20 mL), Red-Al was added at 0 °C and stirred for 2 hours, then stirred at RT overnight. Reaction quenched with sat. potassium sodium tartrate at 0 °C and extracted with hexanes affording ***E*-4** as a white solid (0.68 g, 70%). Crude extract was purified by HPLC (C18) using a gradient of 85-100% acetonitrile. MS (EI) observed: 254.2610, expected: 254.2610 ($C_{17}H_{34}O$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.74 – 5.58 (m, 2H), 4.08 (d, *J* = 5.5 Hz, 2H), 2.03 (q, *J* = 6.9 Hz, 2H), 1.51 (dp, *J* = 13.3, 6.6 Hz, 1H), 1.38 – 1.22 (m, 18H), 1.14

(q, *J* = 6.9 Hz, 2H), 0.86 (d, *J* = 6.7 Hz, 6H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 133.55, 128.78, 63.80, 39.04, 32.19, 29.93, 29.69, 29.65, 29.59, 29.48, 29.17, 29.12, 27.95, 27.40, 22.63.

(2*R*,3*R*)-15-methylhexadecane-1,2,3-triol (**5A**)

A round bottom flask was charged with 10 ml water and 10 ml tert-butanol. 2.8 grams AD-mix (β) was added and stirred at room temperature until two phases appeared. Methanesulfonamide (180 mg, 1.9 mmol) was added and mixture cooled to 0 °C until salts start to precipitate out (~15 minutes)[39]. **E-4** (500 mg, 1.9 mmol) was added and stirred vigorously at 0 °C for 6-24 hours until production of 5A (0.44 g, 77%), a white solid. Reaction progress monitored by TLC (1:1 hexanes/ethyl acetate). MS (EI) observed: 270.2559 ($C_{17}H_{34}O_2$ – loss of water), expected: 288.2664 ($C_{17}H_{36}O_3$); ^{1}H NMR (400 MHz, CD_3OD) δ 3.63 (dd, *J* = 11.3, 4.6 Hz, 1H), 3.58 – 3.51 (m, 2H), 3.46 (dt, *J* = 6.2, 4.3 Hz, 1H), 1.51 – 1.43 (m, 3H), 1.26 (d, *J* = 11.3 Hz, 18H), 1.17 – 1.10 (m, 2H), 0.84 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 74.73, 72.32, 64.34, 49.00, 39.66, 33.84, 30.49, 30.28, 30.27, 30.23, 28.54, 27.98, 26.37, 22.89.

(2*S*,3*S*)-15-methylhexadecane-1,2,3-triol (**5B**)

Same protocol as above using α-mix. 5B was produced as a white solid (0.45 g, 82%). MS (EI) observed: 270.2559 ($C_{17}H_{34}O_2$), expected: 288.2664 ($C_{17}H_{36}O_3$); ^{1}H NMR (400 MHz, CD_3OD) δ 3.63 (dd, *J* = 11.1, 4.8 Hz, 1H), 3.60 – 3.51 (m, 2H), 3.47 (dt, *J* = 6.5, 4.4 Hz, 1H), 1.58 – 1.45 (m, 3H), 1.36 – 1.22 (m, 18H), 1.21 – 1.12 (m, 2H), 0.87 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 75.34, 72.58, 64.52, 40.10, 34.16, 30.90, 30.69, 30.67, 30.64, 28.99, 28.40, 26.86, 23.01.

(2*S*,3*R*)-15-methylhexadecane-1,2,3-triol (**5C**)

A round bottom flask was charged with 10 ml water and 10 ml tert-butanol. 2.8 grams AD-mix (β) was added and stirred at room temperature until two phases appeared. Methanesulfonamide (180 mg, 1.9 mmol) was added and mixture cooled to 0 °C until salts start to precipitate out (~15 minutes). **Z-4** (500 mg, 1.9 mmol) was added and stirred vigorously at 0 °C for 6 - 24 hours until production of 5C (0.49 g, 87%), a white solid. Reaction progress monitored by TLC (1:1 hexanes/ethyl acetate). MS (EI) observed: 270.2564 ($C_{17}H_{34}O_2$), expected: 288.2664 ($C_{17}H_{36}O_3$); ^{1}H NMR (400 MHz, CD_3OD) δ 3.72 (dd, *J* = 11.3, 3.7 Hz, 1H), 3.56 (dd, *J* = 11.3, 6.6 Hz, 1H), 3.49 (td, *J* = 6.3, 3.2 Hz, 1H), 3.42 (td, *J* = 6.6, 3.7 Hz, 1H), 1.71 – 1.63 (m, 1H), 1.58 – 1.47 (m, 2H), 1.40 – 1.24 (m, 18H), 1.21 – 1.15 (m, 2H), 0.88 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 76.35, 73.70, 64.77, 40.25, 34.14, 31.05, 30.86, 30.81, 30.79, 29.16, 28.54, 26.79, 23.08, 22.99.

(2*R*,3*S*)-15-methylhexadecane-1,2,3-triol (**5D**)

Same protocol as above using AD-mix (α). Reaction yielded **5D** (0.41, 73%), a white solid. MS (EI) observed: 270.2562 ($C_{17}H_{34}O_2$), expected: 288.2664 ($C_{17}H_{36}O_3$); ^{1}H NMR (400 MHz, CD_3OD) δ 3.71 (dd, *J* = 11.3, 3.7 Hz, 1H), 3.55 (dd, *J* = 11.3, 6.5 Hz, 1H), 3.47 (td, *J* = 8.5, 7.7, 2.8 Hz, 1H), 3.42 (td, *J* = 6.5, 3.6 Hz, 1H), 1.69 – 1.60 (m, 1H), 1.56 – 1.46 (m, 2H), 1.29 (s, 18H), 1.20 – 1.13 (m, 2H), 0.87 (d, *J* = 6.5 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 76.28, 73.69, 64.77, 40.25, 34.13, 31.05, 30.86, 30.79, 29.16, 28.54, 26.79, 23.09, 22.99.

(2*R*,3*R*)-2,3-dihydroxy-15-methylhexadecyl 4-methylbenzenesulfonate (**6A**)

To a solution of **5A** (100 mg, 0.34 mmol) in pyridine (1.4 ml) and dichlororomethane (2 ml) stirring at O °C, toluene-*p*-sulfonyl chloride (1.5 eq, 0.52 mmol, 100 mg) was added. Reaction was stirred overnight at 4 °C. Mixture was then diluted with water and extracted with ethyl

acetate. Extract was washed with water, 1M HCl and brine, dried over sodium sulfate and concentrated under vacuum. Crude extract was purified by cellulose column using 20% ethanol + 0.1% diethylamine and 80% hexanes + 0.1% diethylamine. Overall reaction and purification yielded enantiopure 6A (63 mg, 41%, ee 72%). HRMS – ESI (M+H) observed: 443.2821 ($C_{24}H_{43}O_5S^+$), (M+H) expected: 443.2831 ($C_{24}H_{43}O_5S^+$); ^{1}H NMR (400 MHz, CD_3OD) δ 7.80 (d, *J* = 8.4 Hz, 2H), 7.44 (d, *J* = 8.0 Hz, 2H), 4.10 (dd, *J* = 10.1, 4.4 Hz, 1H), 3.99 (dd, *J* = 10.1, 6.9 Hz, 1H), 3.61 (dt, *J* = 7.4, 4.0 Hz, 1H), 3.47 (dt, *J* = 8.2, 3.9 Hz, 1H), 2.46 (s, 3H), 1.59 – 1.47 (m, 1H), 1.44 – 1.37 (m, 3H), 1.34 – 1.23 (m, 17H), 1.21 – 1.14 (m, 2H), 0.88 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 146.46, 134.39, 131.13, 129.00, 72.90, 72.54, 72.12, 40.24, 33.83, 31.04, 30.81, 30.77, 30.72, 29.15, 28.54, 26.87, 23.09, 22.97, 21.65.

(2*S*,3*S*)-2,3-dihydroxy-15-methylhexadecyl 4-methylbenzenesulfonate (**6B**)

Followed same protocol as above using a solution of **5B** instead of 5A. Reaction + purification yielded enantiopure 6B (51 mg, 36%, ee 75%) HRMS – ESI (M+H) observed: 443.2827 ($C_{24}H_{43}O_5S^+$), (M+H) expected: 443.2831 ($C_{24}H_{43}O_5S^+$); ^{1}H NMR (400 MHz, CD_3OD) δ 7.80 (d, *J* = 8.3 Hz, 2H), 7.44 (d, *J* = 8.1 Hz, 2H), 4.10 (dd, *J* = 10.1, 4.3 Hz, 1H), 3.99 (dd, *J* = 10.1, 6.9 Hz, 1H), 3.61 (dt, *J* = 7.0, 4.0 Hz, 1H), 3.47 (dt, *J* = 8.1, 3.9 Hz, 1H), 2.46 (s, 3H), 1.57 – 1.46 (m, 1H), 1.45 – 1.35 (m, 3H), 1.29 (m, 17H), 1.22 – 1.13 (m, 2H), 0.88 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 146.48, 134.38, 131.12, 129.02, 72.90, 72.56, 72.13, 40.25, 33.83, 31.04, 30.80, 30.77, 30.71, 29.15, 28.53, 26.86, 23.07, 23.00, 21.64

(2*S*,3*R*)-2,3-dihydroxy-15-methylhexadecyl 4-methylbenzenesulfonate (**6C**)

Followed same protocol as above using a solution of **5C**. Reaction + purification yielded enantiopure 6C (36 mg, 26%, ee 68%) – a white solid. HRMS – ESI (M+H) observed: 443.2838 ($C_{24}H_{43}O_5S^+$), (M+H) expected: 443.2831 ($C_{24}H_{43}O_5S^+$); ^{1}H NMR (400 MHz, CD_3OD)

δ 7.81 (d, *J* = 8.4 Hz, 2H), 7.44 (d, *J* = 8.0 Hz, 2H), 4.20 (dd, *J* = 10.1, 2.9 Hz, 1H), 3.99 (dd, *J* = 10.1, 6.9 Hz, 1H), 3.52 (td, *J* = 7.0, 2.9 Hz, 1H), 3.39 (dq, *J* = 8.5, 3.1 Hz, 1H), 2.46 (s, 3H), 1.66 – 1.57 (m, 1H), 1.56 – 1.46 (m, 2H), 1.35 – 1.22 (m, 18H), 1.21 – 1.14 (m, 2H), 0.88 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 146.43, 134.37, 131.02, 129.11, 73.74, 73.38, 72.66, 40.25, 34.30, 31.04, 30.81, 30.78, 30.74, 29.15, 28.53, 26.53, 23.03, 21.57.

(2*R*,3*S*)-2,3-dihydroxy-15-methylhexadecyl 4-methylbenzenesulfonate (**6D**)

Followed same protocol as above using a solution of **5D**. Reaction + purification yielded enantiopure 6C (52 mg, 38%, ee 78%) – a white solid. HRMS – ESI (M+H) observed: 443.2832 ($C_{24}H_{43}O_5S^+$), (M+H) expected: 443.2831 ($C_{24}H_{43}O_5S^+$); ^{1}H NMR (400 MHz, CD_3OD) δ 7.81 (d, *J* = 8.3 Hz, 2H), 7.44 (d, *J* = 8.2 Hz, 2H), 4.21 (dd, *J* = 10.2, 2.9 Hz, 1H), 3.99 (dd, *J* = 10.1, 6.8 Hz, 1H), 3.52 (td, *J* = 7.0, 2.9 Hz, 1H), 3.39 (dq, *J* = 8.3, 3.0 Hz, 1H), 2.46 (s, 3H), 1.65 – 1.57 (m, 1H), 1.55 – 1.44 (m, 2H), 1.36 – 1.23 (m, 18H), 1.22 – 1.13 (m, 2H), 0.88 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (100 MHz, CD_3OD) δ 146.43, 134.37, 131.02, 129.11, 73.74, 73.38, 72.66, 40.25, 34.30, 31.04, 30.81, 30.78, 30.74, 29.15, 28.53, 26.53, 23.04, 21.57.

(2*S*,3*R*)-2,3-dihydroxy-15-methylhexadecane-1-sulfonic acid (**IOR-1A**)

Sodium sulfite was dissolved in 1.5 ml water and heated to 62 °C. To this solution, **6A** (30 mg, 0.067 mmol) dissolved in ethanol (0.6 ml) was added. Reaction was stirred at 62 °C for 10 hours. Reaction was then extracted with ethyl acetate and washed with water and brine affording IOR-A – a white solid (3.3 mg, 14%). **IOR-1A** was purified by RP-HPLC (C18) using a gradient of Methanol/water + 01% ammonium hydroxide. Optical rotation: $[\alpha]_D^{21.8}$ +20 (*c* 0.25, MeOH); HRMS – ESI (M-H) observed: 351.2211 ($C_{17}H_{35}O_5S^-$), (M-H) expected: 351.2211 ($C_{17}H_{35}O_5S^-$); ^{1}H NMR (500 MHz, $CD_3OD/CDCl_3$) δ 4.04 (dt, *J* = 8.4, 3.3 Hz, 1H), 3.60 (dt, *J* = 7.9, 3.6 Hz, 1H), 3.09 (dd, *J* = 14.0, 3.5 Hz, 1H), 2.98 (dd, *J* = 14.0, 8.5 Hz, 1H), 1.61 – 1.49 (m,

3H), 1.42 – 1.28 (m, 18H), 1.24 – 1.17 (m, 2H), 0.91 (d, *J* = 6.5 Hz, 6H); ^{13}C NMR (126 MHz, $CD_3OD/CDCl_3$) δ 74.27, 71.36, 55.28, 40.19, 33.73, 31.00, 30.74, 29.10, 28.49, 27.10, 23.03.

(2*R*,3*S*)-2,3-dihydroxy-15-methylhexadecane-1-sulfonic acid (**IOR-1B**)

Same protocol as above using **6B** afforded IOR-1B (4.1 mg, 17%). Optical rotation: $[\alpha]_D^{22}$ -12 (*c* 0.5, MeOH) HRMS – ESI (M-H) observed: 351.2214 ($C_{17}H_{35}O_5S^-$), (M-H) expected: 351.2211 ($C_{17}H_{35}O_5S^-$); ^{1}H NMR (500 MHz, $CD_3OD/CDCl_3$) δ 4.05 (dt, *J* = 8.9, 3.3 Hz, 1H), 3.57 (dt, *J* = 7.8, 3.8 Hz, 1H), 3.07 (dd, *J* = 14.0, 3.2 Hz, 1H), 3.00 (dd, *J* = 14.0, 8.8 Hz, 1H), 1.58 – 1.47 (m, 3H), 1.37 – 1.23 (m, 18H), 1.20-1.15 (m, 2H), 0.88 (d, *J* = 6.5 Hz, 6H); ^{13}C NMR (126 MHz, $CD_3OD/CDCl_3$) δ 74.09, 71.09, 54.90, 39.90, 33.45, 30.72, 30.47, 28.79, 28.22, 26.80, 22.97.

(2*R,* 3*R*)-2,3-dihydroxy-15-methylhexadecane-1-sulfonic acid (**IOR-1C**)

Same protocol as above using **6C** yielded a white solid (4.5 mg, 19%). Optical rotation: $[\alpha]_D^{22}$ +10 (*c* 0.5, MeOH); HRMS – ESI (M-H) observed: 351.2215 ($C_{17}H_{35}O_5S^-$), (M-H) expected: 351.2211 ($C_{17}H_{35}O_5S^-$); ^{1}H NMR (500 MHz, $CD_3OD/CDCl_3$) δ 3.96 (ddd, *J* = 9.5, 5.5, 2.0 Hz, 1H), 3.54 (ddd, *J* = 8.8, 5.6, 3.2 Hz, 1H), 3.17 (dd, *J* = 14.0, 2.1 Hz, 1H), 2.89 (dd, *J* = 14.1, 9.6 Hz, 1H), 1.64 – 1.50 (m, 3H), 1.40 – 1.24 (m, 18H), 1.23 – 1.16 (m, 2H), 0.89 (d, *J* = 6.6 Hz, 6H) ; ^{13}C NMR (126 MHz, $CD_3OD/CDCl_3$) δ 74.61, 72.26, 54.08, 40.02, 33.54, 30.83, 30.57, 28.91, 28.33, 26.69, 23.00.

(2*S*, 3*S*)-2,3-dihydroxy-15-methylhexadecane-1-sulfonic acid (**IOR-1D**)

Same protocol as above using **6D** yielded a white solid (4.3 mg, 18%). Optical rotation: $[\alpha]_D^{22}$ -4 (*c* 0.5, MeOH); HRMS – ESI (M-H) observed: 351.2210 ($C_{17}H_{35}O_5S^-$), (M-H) expected: 351.2211 ($C_{17}H_{35}O_5S^-$); ^{1}H NMR (500 MHz, $CD_3OD/CDCl_3$) δ 3.95 (ddd, *J* = 9.8, 5.4, 2.1 Hz, 1H), 3.54 (ddd, *J* = 8.8, 5.3, 3.2 Hz, 1H), 3.14 (dd, *J* = 14.1, 2.0 Hz, 1H), 2.89 (dd, *J* = 14.1, 9.7

Hz, 1H), 1.61 – 1.49 (m, 3H), 1.36 – 1.23 (m, 18H), 1.18 – 1.12 (m, 2H), 0.86 (d, *J* = 6.6 Hz, 6H); ^{13}C NMR (126 MHz, $CD_3OD/CDCl_3$) δ 74.71, 72.36, 54.25, 40.10, 33.62, 30.91, 30.70, 29.00, 28.40, 26.76, 23.02.

2.6 References

1. Carroll, S. B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature* **409,** 1102–1109 (2001).

2. Tyler, S. Epithelium—The Primary Building Block for Metazoan Complexity. *Integr. Comp. Biol.* **43,** 55–63 (2003).

3. King, N. The unicellular ancestry of animal development. *Dev. Cell* **7,** 313–325 (2004).

4. Stanley, S. M. An Ecological Theory for the Sudden Origin of Multicellular Life in the Late Precambrian. *Proc. Natl. Acad. Sci. U.S.A.* **70,** 1486–1489 (1973).

5. McFall-Ngai, M. J. The Development of Cooperative Associations Between Animals and Bacteria: Establishing Détente Among Domains. *American Zoologist* **38,** 593–608 (1998).

6. Narbonne, G. M. The Ediacara Biota: Neoproterozoic Origin of Animals and Their Ecosystems. *Annu. Rev. Earth Planet Sci.* **33,** 421–422 (2005).

7. Knoll, A. H. The Multiple Origins of Complex Multicellularity. *Annu. Rev. Earth Planet Sci.* 217–239 (2011).

8. Hooper, L. V. & Gordon, J. I. Commensal host-bacterial relationships in the gut. *Science* **292,** 1115–1118 (2001).

9. Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 15451–15455 (2002).

10. Koropatnick, T. A. Microbial Factor-Mediated Development in a Host-Bacterial Mutualism. *Science* **306,** 1186–1188 (2004).

11. Webster, N. S. *et al.* Metamorphosis of a scleractinian coral in response to microbial biofilms. *Appl. Environ. Microbiol.* **70,** 1213–1221 (2004).

12. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).

13. Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* **456,** 507–510 (2008).

14. Cheesman, S. E., Neal, J. T., Mittge, E., Seredick, B. M. & Guillemin, K. Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88. *Proc. Natl. Acad. Sci. USA* **108 Suppl 1,** 4570–4577 (2011).

15. Weiss, B. L., Maltz, M. & Aksoy, S. Obligate symbionts activate immune system

development in the tsetse fly. *J. Immunol.* **188,** 3395–3403 (2012).

16. Bosch, T. C. G. Cnidarian-microbe interactions and the origin of innate immunity in metazoans. *Annu. Rev. Microbiol.* **67,** 499–518 (2013).

17. An, D. *et al.* Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* **156,** 123–133 (2014).

18. Shikuma, N. J. *et al.* Marine Tubeworm Metamorphosis Induced by Arrays of Bacterial Phage Tail-Like Structures. *Science* (2014). doi:10.1126/science.1246794

19. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 3229–3236 (2013).

20. McFall-Ngai, M. J. & Ruby, E. G. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science* **254,** 1491–1494 (1991).

21. Hibberd, D. J. Observations on the ultrastructure of the choanoflagellate Codosiga botrytis (Ehr.) Saville-Kent with special reference to the flagellar apparatus. *Journal of Cell Science* **17,** 191–219 (1975).

22. Ruiz-Trillo, I., Roger, A. J., Burger, G., Gray, M. W. & Lang, B. F. A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol.* **25,** 664–672 (2008).

23. Geijtenbeek, T. B. H. & Gringhuis, S. I. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* **9,** 465–479 (2009).

24. Manning, G., Young, S. L., Miller, W. T. & Zhai, Y. The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl. Acad. Sci. USA* **105,** 9674–9679 (2008).

25. Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. *Genome Biol.* **14,** R15 (2013).

26. Dayel, M. J. *et al.* Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. *Dev. Biol.* **357,** 73–82 (2011).

27. Fairclough, S. R., Dayel, M. J. & King, N. Multicellular development in a choanoflagellate. *Curr. Biol.* **20,** R875–6 (2010).

28. Alegado, R. A. *et al.* A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *elife* **1,** (2012).

29. Alegado, R. A. *et al.* Algoriphagus machipongonensis sp. nov., co-isolated with a colonial choanoflagellate. *Int. J. Syst. Evol. Microbiol.* **63,** 163–168 (2013).

30. Beemelmanns, C. *et al.* Synthesis of the Rosette-Inducing Factor RIF-1 and Analogs. *J. Am. Chem. Soc.* **136,** 10210–10213 (2014).

31. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. USA* **113,** 7894–7899 (2016).

32. Godchaux, W. & Leadbetter, E. R. Sulfonolipids of gliding bacteria. Structure of the N-acylaminosulfonates. *J. Biol. Chem.* **259,** 2982–2990 (1984).

33. Abbanat, D. R., Godchaux, W., III & Leadbetter, E. R. Surface-induced synthesis of new sulfonolipids in the gliding bacterium Cytophaga johnsonae. *Arch. Microbiol.* **149,** 358–364 (1988).

34. Merrill, A. H. De novo sphingolipid biosynthesis: a necessary, but dangerous, pathway. *J. Biol. Chem.* **277,** 25843–25846 (2002).

35. Salcedo, M. *et al.* The marine sphingolipid-derived compound ES 285 triggers an atypical cell death pathway. *Apoptosis* **12,** 395–409 (2007).

36. Hannun, Y. A. & Bell, R. M. Lysosphingolipids inhibit protein kinase C: implications for the sphingolipidoses. *Science* **235,** 670–674 (1987).

37. Takikawa, H., Nozawa, D., Kayo, A., Muto, S.-E. & Mori, K. Synthesis of sphingosine relatives. Part 22. Synthesis of sulfobacin A, B and flavocristamide A, new sulfonolipids isolated from Chryseobacterium sp. *J. Chem. Soc., Perkin Trans. 1* 2467–2477 (1999).

38. Romuald, C., Cazals, G., Enjalbal, C. & Coutrot, F. Straightforward synthesis of a double-lasso macrocycle from a nonsymmetrical [c2]daisy chain. *Org. Lett.* **15,** 184–187 (2013).

39. Sharpless, K. B. *et al.* The osmium-catalyzed asymmetric dihydroxylation: a new ligand class and a process improvement. *J. Org. Chem.* **57,** 2768–2771 (1992).

40. Sacoman, J. L. & Hollingsworth, R. I. Synthesis and evaluation of an N-acetylglucosamine biosynthesis inhibitor. *Carbohydr. Res.* **346,** 2294–2299 (2011).

41. Molinski, T. F. *et al.* Halisphingosines A and B, Modified Sphingoid Bases from Haliclona tubifera. Assignment of Configuration by Circular Dichroism and van't Hoff's Principle of Optical Superposition. *J. Nat. Prod.* **76,** 374–381 (2012).

42. Pruett, S. T. *et al.* Biodiversity of sphingoid bases ('sphingosines') and related amino alcohols. *J. Lipid Res.* **49,** 1621–1639 (2008).

43. White, R. H. Biosynthesis of the sulfonolipid 2-amino-3-hydroxy-15-methylhexadecane-1-sulfonic acid in the gliding bacterium Cytophaga johnsonae. *J. Bacteriol.* **159,** 42–46 (1984).

44. Roper, M., Dayel, M. J., Pepper, R. E. & Koehl, M. A. R. Cooperatively Generated Stresslet Flows Supply Fresh Fluid to Multicellular Choanoflagellate Colonies. *Phys. Rev. Lett.* **110,** 228104 (2013).

Chapter 3: Bacteria regulate choanoflagellate development with lipid activators, inhibitors, and synergists[1]

[1] This chapter was adapted from: *Woznica, A., *Cantley, A.M., Beemelmanns, C., Freinkman, E., Clardy, J. & King, N. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. USA* **113,** 7894–7899 (2016).

3.1 Contributions and Acknowledgments

Contributions

Contributors to the research described in this chapter: Myself, Arielle Woznica, Christina Beemelmanns, Elizaveta Freinkman, Jon Clardy and Nicole King. I performed the bacterial growth/extraction and purification of RIFs, LPEs and IOR-1, as well as the characterization of the lysophosphatidylethanolamines. Arielle Woznica performed the rosette induction assays and choanoflagellate staining and microscopy. Christina Beemelmanns characterized the structure of RIF-2. Elizaveta Freinkman performed high-resolution mass spectrometry on the LPEs. Jon Clardy and Nicole King provided research guidance and funding.

Acknowledgments

We thank Monika Abedin, David Booth, and Brad Davidson for critical reading of the manuscript, Rosie Alegado for many early discussions, Laura Wetzel for help constructing the 'violin' plots, and Jennifer Wang at the Harvard Small Molecule Mass Spectrometry Facility for her help with high resolution mass spectrometry. Nicole King is a Senior Fellow in the Integrated Microbial Biodiversity Program of the Canadian Institute for Advanced Research. This work was funded by the National Institute of Health (GM099533).

3.2 Abstract

In choanoflagellates, the closest living relatives of animals, multicellular “rosette” development is regulated by environmental bacteria. The simplicity of this evolutionarily-relevant interaction provides an opportunity to identify the molecules and regulatory logic underpinning bacterial regulation of development. We find that the rosette-inducing bacterium *Algoriphagus machipongonensis* produces three structurally divergent classes of bioactive lipids that, together, activate, enhance, and inhibit rosette development in the choanoflagellate *S. rosetta.* One class of molecules, the lysophosphatidylethanolamines (LPEs), elicits no response on its own, but synergizes with activating sulfonolipid rosette inducing factors (RIFs) to recapitulate the full bioactivity of live *Algoriphagus.* LPEs, while ubiquitous in bacteria and eukaryotes, have not previously been implicated in the regulation of a host-microbe interaction. This study reveals that multiple bacterially produced lipids converge to regulate a developmental switch using a logic resembling that of animal developmental gene regulatory networks.

3.3 Introduction

Choanoflagellates, a group of microbial eukaryotes that are the closest living relatives of animals,[1-4] promise to help illuminate the mechanisms by which bacteria influence animal development. As did cells in the first animals, choanoflagellates use a distinctive collar of actin-filled microvilli surrounding a flow-generating apical flagellum to capture bacteria as prey.[5-7] Indeed, choanoflagellate-like cells likely formed the basis for the evolution of animal epithelial cells that today provide a selective barrier for mediating interactions with bacteria.[5,8,9]

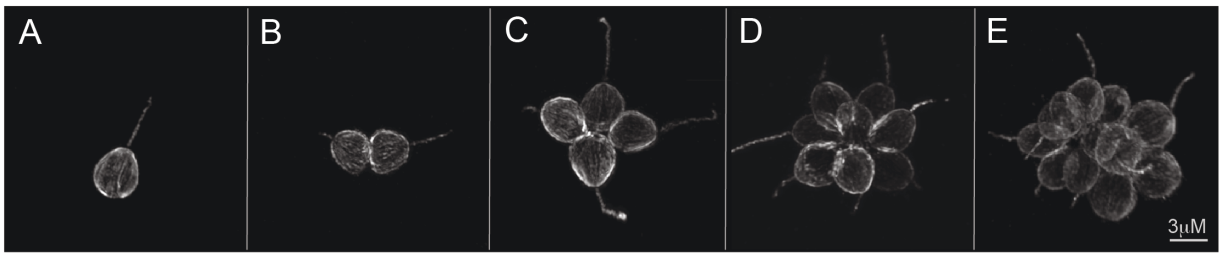


Figure 3.1 Stages of rosette development in *S. rosetta*.
During rosette development, a single founding cell undergoes serial rounds of cell division resulting in a structurally integrated rosette. Importantly, rosette development does not involve cell aggregation. Shown are a single cell **(A),** a pair of cells **(B),** a 4-cell rosette **(C)**, and 8-cell rosette **(D)** and a 16-cell rosette **(E)**

In many choanoflagellates, including *Salpingoeca rosetta*, a developmental program can be initiated such that single cells develop into multicellular "rosettes." Importantly, rosette development does not occur through cell aggregation. Instead, as in the development of an animal from a zygote, rosettes develop from a single founding cell that undergoes serial rounds of oriented cell division, with the sister cells remaining stably adherent (Figure 3.1). The orientation of the nascently divided cells around a central focus, the production of extracellular matrix, and the activity of a C-type lectin called Rosetteless ultimately result in the formation of spherical, multicellular rosettes.[10-12] Rosettes resemble morula stage embryos and the transition to multicellularity in *S. rosetta* evokes ancestral events that spawned the first animals.[5,13]

The initiation of rosette development was recently found to be induced by a co-isolated

environmental bacterium, *Algoriphagus machipongonensis* (phylum Bacteroidetes).[14,15] The ecological relevance of the *Algoriphagus* - *S. rosetta* interaction is evidenced by the co-existence of these organisms in nature,[35] and the predator-prey relationship between choanoflagellates and bacteria.[6,16] Indeed, rosettes likely have a fitness advantage over single cells in some environments, as multicellular choanoflagellates are predicted to produce increased flux of water past each cell,[17] and prey capture studies reveal that rosettes collect more bacterial prey/cell/unit time than do single cells.[18] However, in other environments, rosette development would likely reduce fitness as rosettes have reduced motility relative to single cells. Therefore, we hypothesize that choanoflagellates utilize bacterially produced molecules to identify environments in which rosette development might provide a fitness advantage.

The simplicity of the interaction between *S. rosetta* and *A. machipongonensis* (hereafter, '*Algoriphagus*'), in which both members can be cultured together or independently, offers a biochemically tractable model for investigating the chemical bases of bacterial-eukaryotic interactions. Using rosette development as a bioassay, the first rosette-inducing molecule, Rosette Inducing Factor-1 (RIF-1), was isolated from *Algoriphagus*. The observation that RIF-1 fails to fully recapitulate the bioactivity of the live bacterium, raises the possibility that additional molecules might be required.[35] To gain a more complete understanding of the molecules and regulatory logic by which bacteria regulate rosette development, we set out to identify the minimal suite of *Algoriphagus* molecules that are necessary and sufficient to regulate *S. rosetta* rosette development.

3.4 Results

A newly discovered sulfonolipid activates the rosette development pathway.

To identify the minimal set of *Algoriphagus* molecules required for full rosette induction,

we used a bioassay based on a co-culture of *S. rosetta* with the non-rosette inducing prey bacterium *Echinocola pacifica.* This culture, called 'SrEpac' (for *S. rosetta* + *E. pacifica*)[19] reproducibly yields high percentages of cells in rosettes (>80%) in response to live *Algoriphagus*, *Algoriphagus* outer membrane vesicles (OMVs) isolated from conditioned medium, and *Algoriphagus* bulk lipid extracts (Figure 3.2). In addition, incubation of SrEpac with the only previously known Rosette Inducing Factor, the sulfonolipid RIF-1, results in low but reproducible levels of rosette development (~1.5% of cells in rosettes), consistent with previous results using a different *S. rosetta* culture. [35]

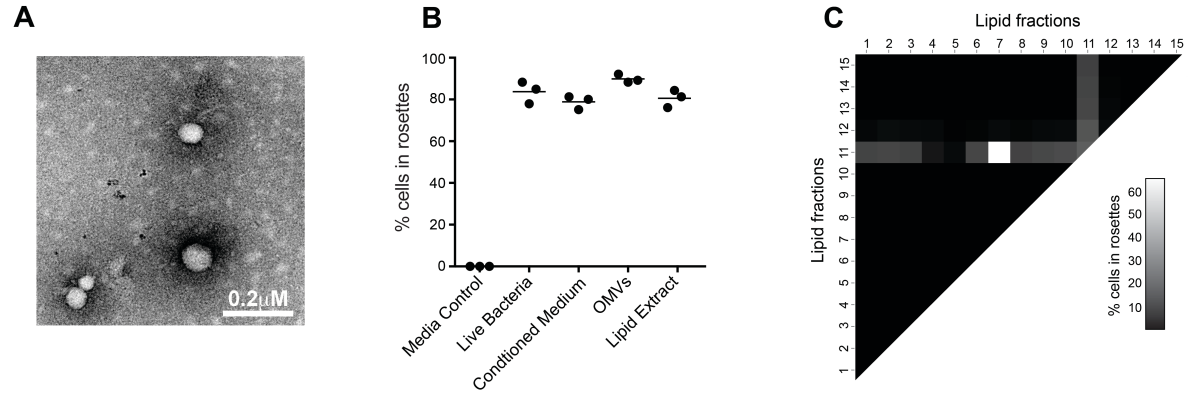


Figure 3.2 Maximal rosette development requires lipid co-factor interactions.
(A) TEM of OMVs from *Algoriphagus* conditioned media. **(B)** *S. rosetta* does not form rosettes when treated with media that lacks additional bacterial signals (Media Control). Maximal (~90% of cells in rosettes) or near-maximal levels of rosette development are induced by live *Algoriphagus*, *Algoriphagus* conditioned media, *Algoriphagus* OMVs, and *Algoriphagus* bulk lipid extract. **(C)** A heat map depicts the rosette-inducing activity of *Algoriphagus* lipid fractions used to treat SrEpac, either in isolation or in combination, at a final lipid concentration of 2 μg/mL. Sulfonolipid-enriched fraction 11 was the only fraction sufficient to induce rosette development when tested alone (30% of cells in rosettes). Tests of each the lipid fractions in combination with fraction 11 (and all other fractions) revealed previously unidentified inhibitory and enhancing co-factor activity. Fractions 4 and 5 decreased rosette development (to 12% and 8%, respectively), whereas fraction 7 increased rosette development to 65%.

Because *Algoriphagus* bulk lipid extracts elicit the same rosette development response as live bacteria (Figure 3.2), we began by fractionating a bulk extraction of *Algoriphagus* lipids by reversed-phase high performance liquid chromatography (HPLC) and testing the resulting 15 lipid fractions in SrEpac. Only fraction 11 was sufficient to induce rosette development, whereas

all other lipid fractions lacked rosette-inducing activity at all concentrations tested (Figure 3.2). To further separate and isolate the active molecules in fraction 11, we performed a subsequent round of reversed-phase HPLC and tested the resulting sub-fractions for activity in SrEpac. The rosette-inducing activity tracked with one sub-fraction (hereafter, "RIF mix") that induced rosette development in 23.5% of cells (Figure 3.3; Figure 3.4). Structural analysis by NMR, high resolution mass spectrometry (HRMS), and tandem mass spectrometry (MSMS) revealed that

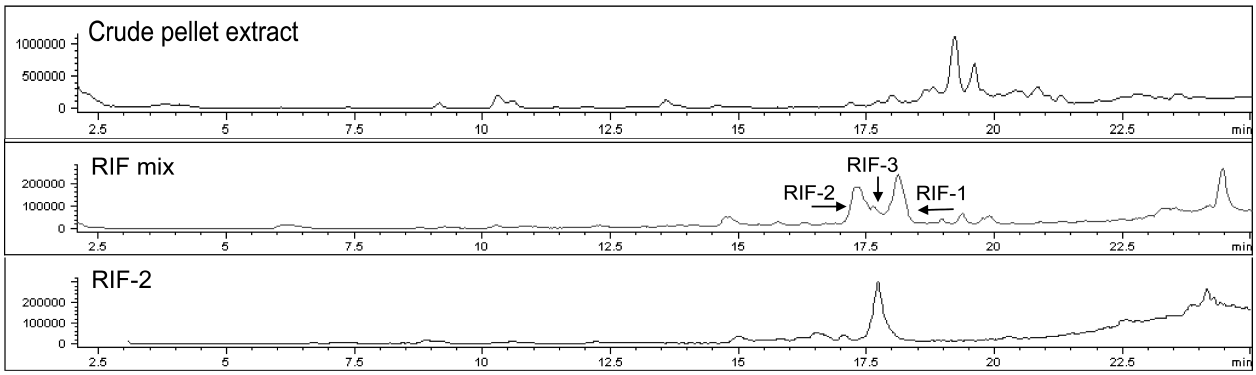


Figure 3.3 LCMS spectra depicting RIF mix composition.
Extracts were resuspended in methanol and injected onto a C18 column. Compounds were eluted using a gradient of 65%-100% methanol + 0.1% NH_4OH. X-axis shows retention time (minutes) and Y-axis depicts total ion count (TIC).

the RIF mix contained RIF-1 and two structurally related but previously uncharacterized sulfonolipids with approximate molecular weights of 605 Da and 593 Da. Sulfonolipids are a largely uncharacterized class of molecules that are structurally similar to sphingolipids, a diverse group of molecules based on sphingoid bases that play structural roles in cell membranes and important non-structural roles in signal transduction.[20] Although sulfonolipids have been reported to contribute to the gliding motility of Bacteroidetes bacteria, almost nothing is known regarding their potential roles as signaling molecules.[21,22]

Additional activity-guided fractionation by HPLC allowed us to isolate pure samples of RIF-1 and of the 605 Da sulfonolipid.[15,23] Purified RIF-1 induced maximal (~1.5%) rosette development at femtomolar to nanomolar concentrations (Figure 3.4). In contrast, the purified

605 Da sulfonolipid (hereafter “RIF-2”) elicited 10-fold higher levels of rosette development (10.5% of cells in rosettes; Figure 3.4) than RIF-1, although at micromolar concentrations. The planar structure of RIF-2 (Figure 3.5) was determined by one and two-dimensional NMR and found to closely resemble RIF-1, with the exception of slight structural variations of the capnoid base, which contains a double bond at the position C-4 and a hydroxyl group at the position C-6.

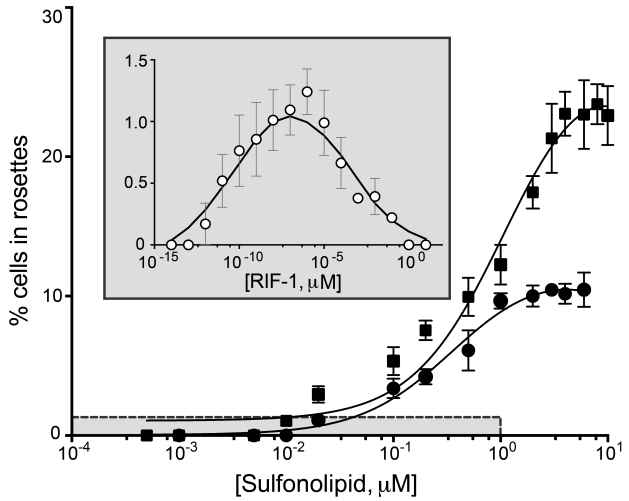


Figure 3.4 Rosette inducing activity of RIF-2

The RIF mix (solid square) and purified RIF-2 (solid circle) induce rosette development at micromolar concentrations. Grey inset: RIF-1 (open circle) is active at femtomolar to nanomolar concentrations, but induces 10-fold lower levels of rosette development than RIF-2. Grey box in the main graph corresponds to the range of concentrations at which RIF-1 is active, and the range of rosette-inducing activity of RIF-1 compared with the data from the RIF-mix and RIF-2. Rosette development was quantified 24 hours after induction. Minor ticks on X-axis are log spaced

Previously, we determined the stringent stereochemical requirements for activity of RIF-1;[23] therefore, it is likely that the hydroxyl groups at the C-2’ and C-3 position of RIF-2 share the same configuration as RIF-1 (Figure 3.5). However, although our previous structure-activity relationship study explored several double-bond containing RIF-1 analogs (e.g.,Figure 3.5), we did not previously investigate the effect of a hydroxyl group at the 6-position of the capnoid base. Thus, either the double bond or C-6 position –OH, or both, could be responsible for the enhanced activity of RIF-2. Ongoing experiments to assign the absolute stereochemistry of RIF-2 and explore the activity of RIF-2 analogs will greatly inform our understanding of the

structure-activity relationship of rosette-inducing sulfonolipids.

The remaining 593 Da sulfonolipid in the RIF mix is produced by *Algoriphagus* at low levels (approximately 1/5th the amount of RIF-2) and elutes closely to RIF-2 during fractionation (Figure 3.3). Although HRMS data suggests that this molecule is a sulfonolipid similar to RIF-1 low levels of production and co-elution with RIF-2 have prevented us from fully isolating and

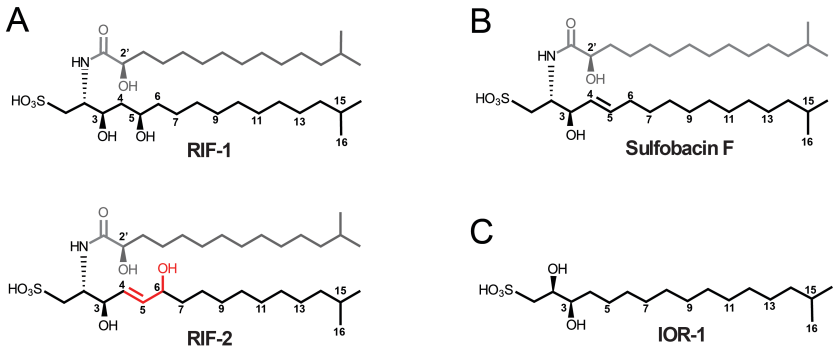


Figure 3.5 Structural similarities and differences among RIFs, an inactive sulfonolipid, and the inhibitory capnine IOR-1.'

(A) The three dimensional structure of RIF-1 (previously determined by total synthesis)[23], compared to the proposed molecular structure of RIF-2, and **(B)** the structure of an inactive *Algoriphagus* sulfonolipid, Sulfobacin F.[23] Shared features of *Algoriphagus* sulfonolipids include a fatty acid chain (shown in grey), and a capnoid base (shown in black). Distinguishing features between RIF-1 and RIF-2 (highlighted in red) include a double bond at position 4, and a hydroxyl group at position 6. The tight structure-activity relationships of RIF-1 and RIF-2 suggest a restricted set of interactions between these molecules and a binding target. Interestingly, there are no features shared by RIF-1 and RIF-2 to the exclusion of Sulfobacin F. **(C)** IOR-1 is a capnine that antagonizes the rosette inducing activity of the RIFs. Like the capnoid base of the RIF sulfonolipids, IOR-1 is composed of a sulfonic acid head group and a branched chain containing two hydroxyl groups. Furthermore, the carbon chain length and branching pattern of IOR-1 is the same as that of the capnoid base in each of the sulfonolipid RIFs. The similarities between IOR-1 and the RIFs raise the possibility that IOR-1 competitively inhibits RIF binding to a target receptor.

characterizing the activity of the 593 Da sulfonolipid. However, because the combination of RIF-2 and the 593 Da sulfonolipid induced rosettes at levels indistinguishable from those of RIF-2 alone, we infer that the rosette-inducing activity of the RIF mix is largely the product of RIF-2. Nonetheless, we note that the maximal level of rosette development induced by the RIF mix (Figure 3.4, Figure 3.6) is greater than the sum of purified RIF-1 + RIF-2 for reasons that we do not yet understand.

The discovery of RIF-2 revealed that RIF-1 is not the sole *Algoriphagus* determinant of *S. rosetta* rosette development. However, even the RIF mix, which contains both RIF-1 and RIF-2, failed to recapitulate the full level of rosette induction elicited by either intact *Algoriphagus* or *Algoriphagus* bulk lipid extract. Therefore, we hypothesized that additional molecular cues are required to fully potentiate the rosette-inducing activities of RIF-1 and -2.

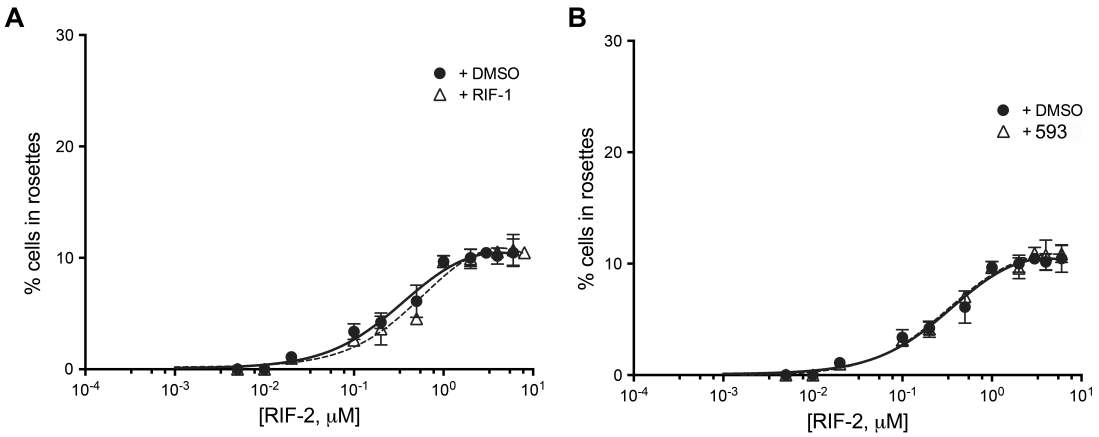


Figure 3.6 Combined activity of RIF-1 + RIF-2 does not recapitulate that of the RIF mix.
(A) Addition of 10^{-7} μM RIF-1 (the concentration of RIF-1 at which maximal activity was observed) to RIF-2 (triangle) does not increase levels of rosette induction compared to RIF-2 alone (circle). Rosette development was quantified 24 hours after induction. **(B)** RIF-2 + sulfonolipid 593 (triangle) induces similar levels of rosette development compared to RIF-2 alone (circle). Rosette development was quantified 24 hours after induction.

Lipid cofactors inhibit and enhance RIF activity

To identify potential cofactors of the RIFs, we mixed each of the 15 *Algoriphagus* lipid fractions in pairwise combinations and tested the mixtures at several concentrations in SrEpac (Figure 3.2). We observed two types of cofactor activity: enhancing activity in fraction 7 and, unexpectedly, inhibitory activity in fractions 4 and 5. Importantly, the activities of these cofactor-containing fractions were only evident when tested in combination with fraction 11, which contained both RIF-1 and RIF-2.

The inhibitory activity observed in fractions 4 and 5 is the first example of a compound(s) – either isolated from *Algoriphagus* or commercially available – that specifically reduces levels

of rosette development at concentrations that do not otherwise inhibit growth. Therefore, we used bioactivity-guided fractionation in the presence of RIF-2 to determine the molecular basis for inhibition. High-resolution mass spectrometry, together with 1D and 2D NMR experiments, allowed us to propose the planar structure of a molecule with the approximate mass of 351 Da (hereafter referred to as Inhibitor of Rosettes-1, "IOR-1"; Figure 3.5). Comprehensive methods detailing the IOR-1 isolation and structure determination, along with dose-response curves of IOR-1 in the presence of the RIF mix and RIF-2, are described in chapter 2.[24]

Nanomolar concentrations of IOR-1 completely inhibited the ability of RIF-2 to induce rosette development, and reduced rosette development in the presence of the RIF mix (Figure

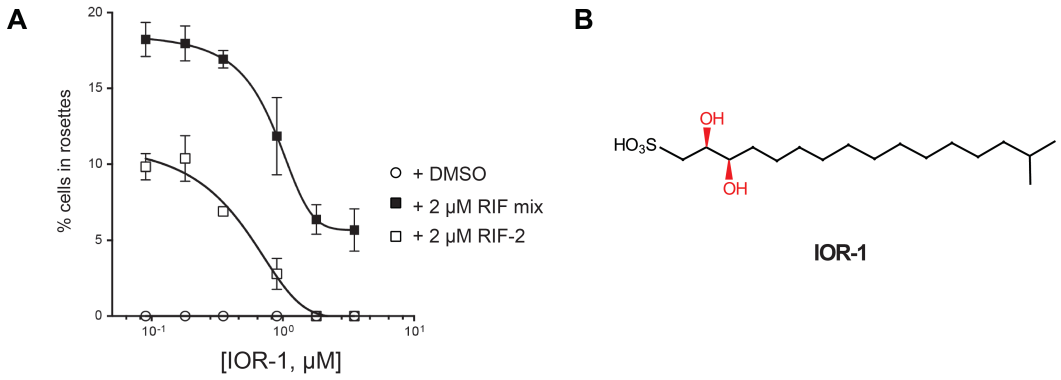


Figure 3.7 IOR-1 inhibits RIF-2 induced rosette formation.
(A) To determine the range of concentrations at which IOR-1 displayed inhibitory activity, IOR-1 was titrated and applied to SrEpac alone (circle), with 2 μM RIF mix (black square) or SrEpac + 2 μM RIF-2 (white square). Rosette development was quantified 25 hours after induction. **(B)** Structure of IOR-1. Stereochemical assignment of 2 hydroxyl groups (in red) determined by total synthesis as previously described.[24]

3.7). IOR-1 is a capnine lipid that resembles the capnoid backbone of *Algoriphagus* RIFs (Figure 3.5). Thus, we hypothesize that IOR-1 antagonizes rosette development by competitively binding a RIF-2 target receptor. Because the RIF-mix induces low levels of rosette formation in the presence of IOR-1 (Figure 3.7), we infer that the combined effects of the RIFs are sufficient to partially overcome the presence IOR-1.

In contrast with the inhibitory activity associated with IOR-1, the *Algoriphagus* lipid fraction 7 greatly enhanced rosette development when used in combination with the RIF-containing fraction 11 (Figure 3.2). Notably, fraction 7 did not contain any sulfonolipids, the only class of molecules previously known to regulate rosette development. After separating the components of fraction 7 by HPLC, we treated SrEpac with each subfraction in combination with

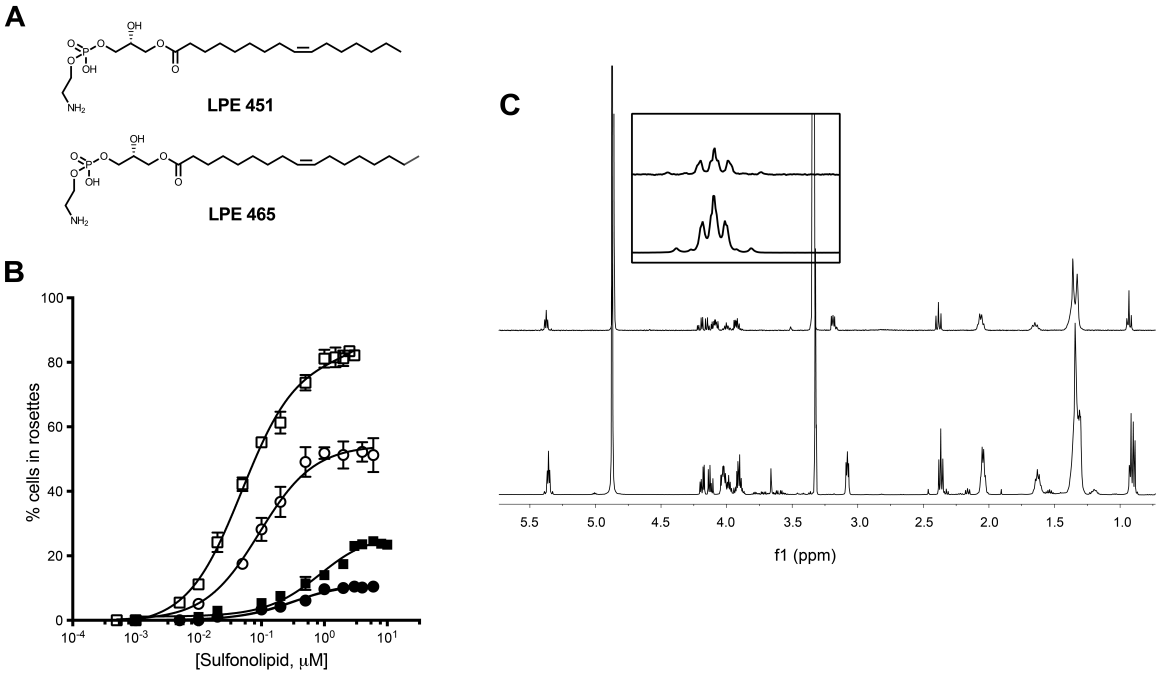


Figure 3.8 LPEs synergize with RIFs to enhance rosette development
(A) The structures of LPE 451 and LPE 465 as determined by NMR and tandem mass spectrometry. LPE 451 and LPE 465 differ from each other by only one methyl group along the fatty acid chain. **(B)** The addition of 2 mM LPE mix increases the maximal percentage of cells in rosettes in RIF-2-induced SrEpac from 10.5% (solid circle) to 52% (open circle) and the maximal inducing activity of the RIF mix from 23.5% (solid square) to 82% (open square) of cells in rosettes. Minor ticks on X-axis are log spaced. **(C)** 1H NMR spectra of commercial LPE (18:1(9Z)) standard (top) overlaid with LPEs isolated from *Algoriphagu*s (bottom). Inset: *cis* vicinal protons of commercial LPE: J = 5.0 Hz (top). Vinylic protons of isolated LPE – 450/464: J = 4.9 Hz (bottom).

the RIF mix and quantified the level of rosette development. The subfractions that enhanced rosette development in the presence of the RIF mix contained one or both of two lysophosphatidylethanolamines (LPEs) with molecular weights of 451 Da and 465 Da (hereafter referred to as LPE 451 and LPE 465, respectively; Figure 3.8).

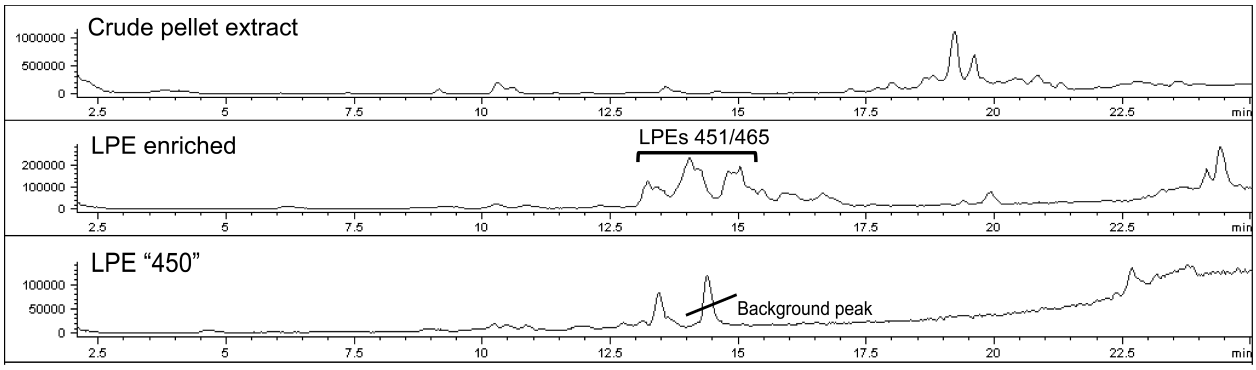


Figure 3.9 LCMS spectra depicting contents of LPE-enriched fractions isolated from *Algoriphagus*
Extracts were resuspended in methanol and injected onto a C18 column. Compounds were eluted using a gradient of 65%-100% methanol + 0.1% NH_4OH. X-axis shows retention time (minutes) and Y-axis depicts total ion count (TIC).

As this class of molecules is well known, literature precedence allowed us to confirm the core LPE structure by NMR and tandem mass spectrometry (Figure 3.8). We performed an olefin metathesis on the most active LPE fractions[25] to determine that the major species present (in both LPE 451 and LPE 465) contains a double bond between carbons 9 and 10, which is common for fatty acid chains of this length. Due to the difficulties associated with purifying these types of molecules, we were unable to completely exclude other LPE isoforms (which can differ in double bond location or position on the glycerol backbone); however, multiple iterations of bioassay-guided fractionation consistently yielded a fraction from the purification process (hereafter, the "LPE mix") in which 98% of the fraction was composed of LPEs 465 and 451, with the remaining 2% of the sub-fraction containing trace amounts of other structurally related LPE analogs and a stearic acid contaminant (Figure 3.9).

We ruled out the possibility that the stearic acid was bioactive by applying it to SrEpac both alone and in combination with the RIF mix. No rosette inducing or RIF-enhancing activity was observed in response to stearic acid. Moreover, no commercially available LPEs tested in combination with the RIF mix either activated or enhanced rosette development (Table 3.1). Therefore, we infer that LPE 451, LPE 465, or both are responsible for the synergistic RIF-

enhancing activity of the LPE mix. Furthermore, as with the RIFs [23] and IOR-1[24], it appears that the enhancing activity of the LPEs results from a highly specific structure-activity relationship.

LPEs belong to a large and diverse class of deacylated phospholipids, called lysophospholipids, that include structural components of cellular membranes as well as biologically active lipid mediators.[26,27] While LPEs are found in most bacterial and eukaryotic cell

Table 3.1 Commercial compounds tested for rosette inducing/enhancing activity and inhibitory activity. No activity was detected for any of these compounds.

Name	Description
Oleoylethanolamine	Ethanolamide
Oleic acid	Fatty acid
Monosialoganglioside	Glycosphingolipid
1-myristoyl-2-hydroxy-glycero-3-phosphoethanolamine	Lysophosphatidylethanolamine
1-octadecanoyl-*sn*-glycero-3-phosphoethanolamine	Lysophosphatidylethanolamine
1-oleoyl-2-hydroxy-*sn*-glyero-3-phosphoethanolamine	Lysophosphatidylethanolamine
1-palmitoyl-2-hydroxy-*sn*-glyero-3-phosphoethanolamine	Lysophosphatidylethanolamine
1-tetradecanoyl -*sn*-glyero-3-phosphoethanolamine	Lysophosphatidylethanolamine
L-α-lysophophatidylethanolamine	Lysophosphatidylethanolamine
Sphingosylphosphorylcholine	Lysosphingolipid
Sphingomyelin	Sphingolipid
Ceramide-1-phosphate	Sphingolipid
Sphingosine-1-phosphate	Sphingolipid
Ceramide phosphorylethanolamine	Sphingolipid
Dihydrosphingosine	Sphingosine precursor

membranes and present in somewhat elevated concentrations in many marine and estuarine bacteria,[28] little is known about how and in what contexts LPEs might act as signaling molecules.[27,29]

To characterize how LPEs regulate rosette development, we started by investigating the concentrations at which the LPE mix displayed maximal enhancing activity (Figure 3.10). In contrast with the 10.5% of cells in rosettes induced by 2 μM RIF-2 alone, treatment of SrEpac with 2μM RIF-2 and micromolar concentrations of the LPE mix increased rosette development

five-fold to 53% (Figure 3.8, Figure 3.10). Furthermore, maximal levels of rosette development elicited by the RIF mix + the LPE mix matched those induced by the *Algoriphagus* lipid extract (Figure 3.2;Figure 3.8).

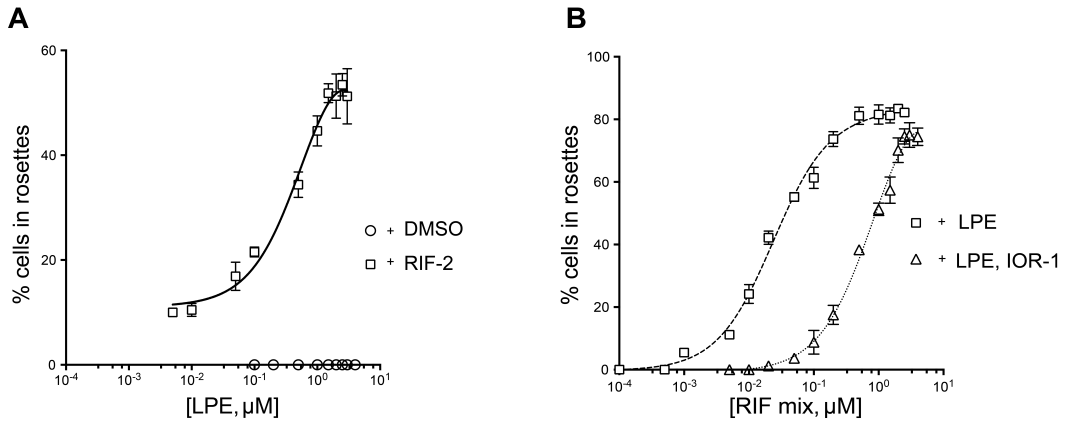


Figure 3.10 Lipid interactions enhance and inhibit rosette inducing activity.
(A) To determine the concentrations at which the LPE mix maximally enhanced rosette development, the LPE mix was titrated and applied to SrEpac alone (circle), or SrEpac incubated with 1 μM RIF-2 (square). Rosette development was quantified 25 hours after induction. **(B)** Synergistic activity between the RIFs and the LPEs induce rosette development in the presence of IOR-1. To assess the combined activity of the RIFs, LPEs, and IOR-1, concentrations of the LPE mix (2 μM) and IOR-1(2 nM) were kept constant and added to SrEpac alongside titrating concentrations of the RIF mix.

Finally, we observed that LPEs also influence RIF potency. In bioassays in which the concentration of the LPE mix was held stable at 2 μM and the RIF mix or RIF-2 was titrated, the sensitivity of *S. rosetta* to the RIFs increased such that 25-fold less RIF mix and 3-fold less RIF-2 was required to achieve half-maximal induction (Figure 3.8). Furthermore, when the inhibitor of rosette formation IOR-1 is added to the mix, the potency of RIF-2 decreases but the net affect of this lipid mix (RIFs + LPEs + IOR-1) is still rosette induction in ~80% of cells (Figure 3.10). Not only does this recapitulate our observations in live *Algoriphagus*, it also hints at the idea that while IOR-1 can inhibit RIF-2-induced rosette formation, it does not interfere with LPE-mediated synergy. Overall these data reveal that the rosette inducing activity of *Algoriphagus* can be largely recapitulated with specific representatives from just two different classes of lipids: sulfonolipids and LPEs.

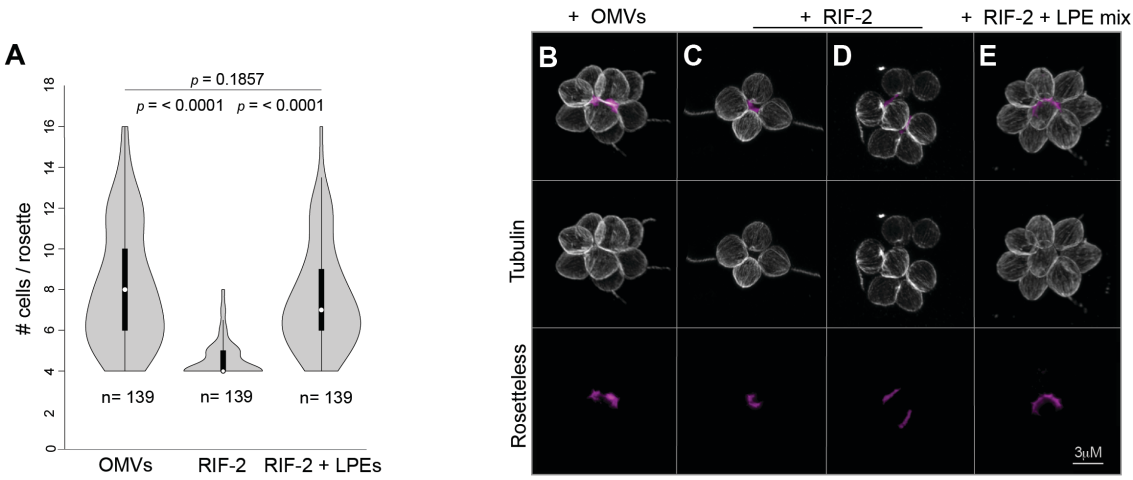


Figure 3.11 LPEs promote proper rosette development and maturation.

(**A**) Frequency distribution of rosette size in SrEpac incubated with OMVs, RIF-2, and RIF-2 + LPE mix after exposure to shear stress by pipetting. Rosettes induced with RIF-2 alone contained fewer cells on average and reached a smaller maximal size than rosettes induced with *Algoriphagus* OMVs. The addition of the LPE mix to RIF-2 increased the median rosette size and frequency distribution to levels that recapitulated induction by OMVs. Rosette size was assessed 22 hours after induction (n=139 for each condition). Data are presented as 'violin' box plots, showing the median cell number (white circle), 75% quartile (thick line), and range excluding outliers (thin line). Surrounding the box plot is a kernel density trace, plotted symmetrically to show the rosette size frequency distribution. P values (unpaired t-tests) were calculated using GraphPad Prism **(B-E)** Rosette morphology, cell packing, and localization of Rosetteless protein in rosettes induced by **(B)** OMVs, **(C and D)** RIF-2 alone, and **(E)** RIF-2 + LPEs. (**B**) Cells in OMV-induced rosettes express Rosetteless and are tightly packed. Anti-tubulin (white) highlights the cell body and anti-Rosetteless antibodies (magenta) stain Rosetteless, a specific marker of rosette induction that localizes to ECM in the center of rosettes (Levin et al., 2014). **(C)** 4-celled rosettes induced by RIF-2 are tightly packed, whereas larger rosettes induced by RIF-2 alone **(D)** appear "loose", with cells spaced farther apart. **(E)** Rosettes induced by RIF-2 + LPE mix are large and closely packed, and phenocopy rosettes induced by OMVs. All rosettes were fixed 22 hours after treatment with 1 μM RIF-2, 1μM RIF-2 + 2uM LPEs, or OMVs.

LPEs promote a previously unidentified maturation step in rosette development

Rosettes induced by live *Algoriphagus* bacteria or *Algoriphagus* OMVs, lipid-rich vesicles that fully recapitulate the inducing activity of live bacteria, are remarkably resistant to shear and can range in size from 4 cells, the minimum number of cells required to confirm the organized polarity of a rosette to as many as 50 cells. Because the rosette-inducing activity of OMVs is stable, highly reproducible, and equivalent to that of live bacteria, we used it as a positive control for the study of rosette cell number. Within just 22 hours after treatment, OMV-induced rosettes were resistant to shear introduced by pipetting, and the median cell number per rosette

was 8 cells, although some grew to as large as 16 cells/rosette (Figure 3.11). In contrast, treatment with purified RIF-2 resulted in rosettes that were sensitive to mechanical disruption; after pipetting, the median cell number per rosette was significantly smaller (4 cells/rosette) than that induced by *Algoriphagus* OMVs (8 cells/rosette, Figure 3.11). Furthermore, the size frequency distribution for RIF-2-induced rosettes was restricted to small rosettes, ranging from the minimum size of 4 cells up to 8 cells, compared to *Algoriphagus*- and OMV-induced cultures in which larger rosettes of 10-16 cells were frequently observed. Because the combinatorial activity of RIF-2 + LPE mix resulted in elevated percentages of cells in rosettes, we hypothesized that LPEs might promote rosette stability and therefore protect larger rosettes when exposed to shear. Indeed, the median cell number (7 cells/rosette) and size frequency distribution of SrEpac induced by RIF-2 + LPE mix was statistically indistinguishable from OMV-induced cultures (Figure 3.11).

The hypothesis that RIF-2 induced rosettes exhibit less structural integrity than rosettes induced by either OMVs or RIF-2 + LPEs was supported by observations made using high-resolution microscopy (Figure 3.11). Cells in OMV-induced rosettes were tightly packed and properly localized a specific marker of rosette development, the C-type lectin protein Rosetteless[12] to the extracellular matrix-rich center of the rosette (Figure 3.11). While 4-celled rosettes induced by RIF-2 alone showed close cell packing, cells in all larger rosettes induced by RIF-2 (e.g. those with 5-7 cells/rosette) were spaced farther apart than those in OMV-induced rosettes of equivalent size (Figure 3.11). Despite a 'loose' morphology, RIF-2-induced rosettes secreted Rosetteless protein, demonstrating that they had properly initiated rosette development. Importantly, induction with RIF-2 + the LPE mix restored a robust rosette morphology, with the cells tightly packed together, phenocopying OMV-induced rosettes. Thus, although RIFs alone are sufficient to initiate rosette development, LPEs promote structural stability during rosette development, and thereby facilitate rosette maturation (Figure 3.12).

3.5 Discussion

Animals rely on bacteria for everything from proper metabolism to the stimulation of immune system development to the regulation of gut morphogenesis.[30,31] Bacterial cues even direct major life history transitions in animals, with many marine invertebrates producing motile larvae that will not settle and undergo morphogenesis until they encounter the "proper" environmental bacteria.[32] In one of the most dramatic examples of cross-talk between bacteria and an animal, *Vibrio fischeri* bacteria induce the development of a "light organ" in the Hawaiian bobtail squid, a structure which the bacteria later colonize.[33] The widespread phylogenetic distribution of bacterially-regulated developmental processes in animals suggests that such interactions may have been pivotal during the origin and early evolution of animals.[31,34]

As the number of animal developmental processes influenced by bacteria grows, detailed molecular characterization of the relevant bacterially-produced cues promises to reveal the regulatory logic underlying host-microbe interactions. Through the study of rosette development in a close relative of animals, *S. rosetta*, we have found three classes of structurally distinct lipids produced by *Algoriphagus* that are interpreted by *S. rosetta* as activators, synergistic enhancers, and inhibitors of development (Figure 3.12). When tested alone, activating RIFs elicit relatively low levels of rosette development and the synergistic LPEs have no detectable activity. However, when used in combination, the activating RIFs + synergizing LPEs induce levels of rosette development in *S. rosetta* that recapitulate those induced by live *Algoriphagus.* Moreover, while the *Algoriphagus* capnine IOR-1 is a potent antagonist of the RIFs,[24] the synergistic activities of the RIFs and LPEs overcome the inhibitory activities of IOR-1, potentially explaining why endogenous IOR-1 does not prevent robust rosette induction.

We hypothesize that the reliance of *S. rosetta* on multiple inputs from *Algoriphagus* prevents the developmental switch to rosette development under suboptimal conditions. The

commitment to rosette development requires a trade-off; rosette development is a lengthy process and while rosettes are potentially more efficient than single cells in the capture of planktonic bacteria, they are poor swimmers[35] and therefore likely to be less effective at dispersal and escape from certain predators (e.g. amoebae). Moreover, the aquatic world in which choanoflagellates live is patchy,[36] with the diversity and density of bacteria dramatically

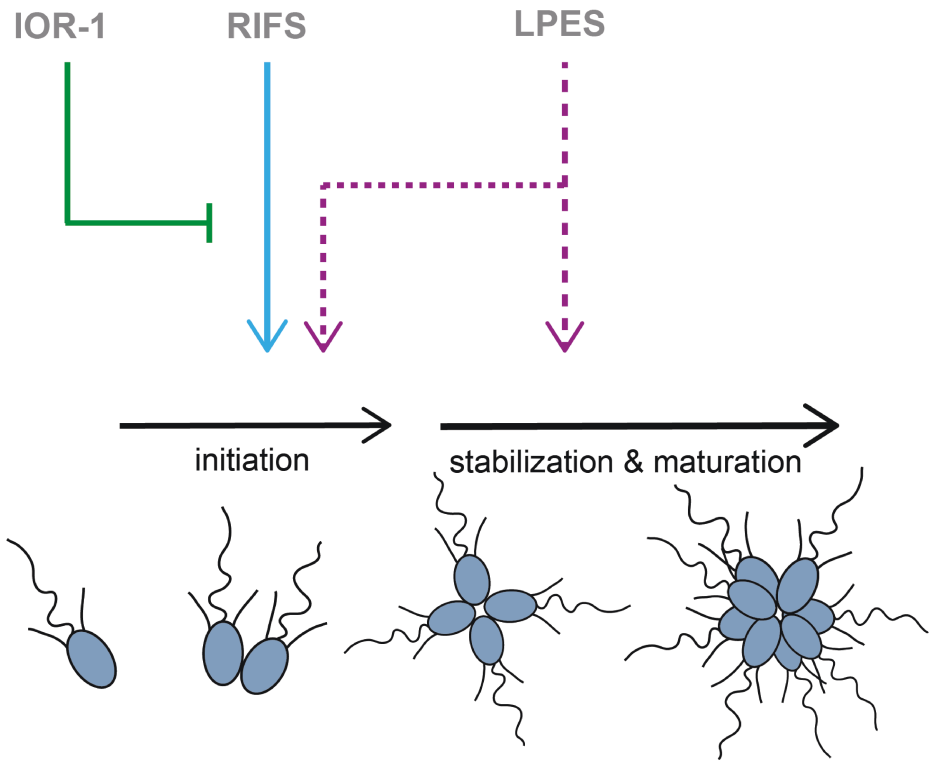


Figure 3.12 Multiple bacterial inputs regulate rosette development in *S. rosetta*.
Algoriphagus produces three chemically distinct classes of lipids – sulfonolipids, LPEs, and a capnine – that interact to alternately induce, enhance, or inhibit rosette development in *S. rosetta.* The sulfonolipids RIF-1 and RIF-2 are sufficient to initiate rosette development in *S. rosetta*, although rosettes induced by RIFs alone are restricted in size, potentially because of their sensitivity to shear. Complete rosette maturation requires the synergistic activities of RIFs and LPEs. Although LPEs have no detectable activity on their own, they enhance RIF activity and facilitate the growth of larger and more structurally stable rosettes, perhaps by regulating downstream pathways important for rosette maturation. While the molecular mechanisms by which LPEs regulate rosette development are unknown (indicated by dotted lines), multiple lines of evidence (see main text) suggest that LPEs act both to promote the initiation of rosette development and, separately, to promote the subsequent maturation of rosettes. *Algoriphagus* also produces the inhibitory molecule IOR-1, which inhibits rosette-inducing activity of RIFs.[24] Importantly, when *S. rosetta* is exposed simultaneously to RIFs and the synergistic LPEs, mature rosettes develop even in the presence of IOR-1.

varying between local microenvironments. In animals, the integration of multiple signals is fundamental to the robustness of many developmental decisions, including the establishment of the body axis during early embryogenesis,[37-40] and the progressive specification of cell fates.[41-43] Likewise, the multi-input regulatory module that controls *S. rosetta* development may act to ensure that rosette development is not initiated under the wrong environmental conditions or in response to the wrong bacterial cues.

The integration of multiple bacterial inputs is also essential for proper animal development in two well-studied host-microbe models. In the Hawaiian bobtail squid, two molecules (LPS and TCT) produced by *Vibrio fischeri* act synergistically to induce organogenesis,[44] and in mice, several bacterial cell wall molecules (LPS, PGN, and polysaccharide A) together shape the development of the immune system of the gut.[45-47] The finding that rosette development in *S. rosetta* requires the integration of a network of bacterial lipids extends this phenomenon to the closest living relatives of animals. Ultimately, as the molecular-underpinnings of more host-microbe interactions are fully elucidated, the mechanisms by which bacteria influence their animal hosts may be found to be as intricate and complex as those regulating animal development, with microbial communities providing cocktails of activating, enhancing, and inhibitory cues.

3.6 Experimental

A new bioassay for Rosette Inducing Factors

S. rosetta was originally isolated from nature as a rosette and, along with co-isolated environmental bacteria, was expanded and cryogenically preserved as strain ATCC50818[11]. Treatment of ATCC50818 with an antibiotic cocktail killed a subset of the co-isolated bacteria, including *Algoriphagus*, and yielded a strain in which rosettes failed to form. This strain, named RCA (for 'Rosette Colonies Absent'), was subsequently used as a bioassay in which bacterial lysates and fractions were tested for their ability to induce rosette development.[15] Although the RCA-based bioassay allowed the identification of *Algoriphagus* as a rosette-inducing bacterium and RIF-1 as a rosette-inducing factor, the competency of RCA strains to form rosettes upon supplementation with *Algoriphagus* was variable and dependent upon the underlying rate of *S. rosetta* cell proliferation. Moreover, maximal rosette formation in *Algoriphagus*-treated RCA (~50% of cells in rosettes) did not approach that achieved when *S. rosetta* was grown solely in the presence of *Algoriphagus* (~75% of cells in rosettes).[15]

Therefore, to improve the sensitivity and reproducibility of the bioassay, the studies presented here were performed in a recently-established strain of *S. rosetta*, 'SrEpac',[19] in which *S. rosetta* is grown in the presence of the non-rosette inducing Bacteroidetes bacterium *Echinicola pacifica.*[48] In contrast with RCA, SrEpac supplemented with *Algoriphagus* forms rosettes consistently and at uniformly high levels (approaching 90%), thus allowing for a robust bioassay with increased sensitivity. Furthermore, low concentrations of *Algoriphagus* conditioned media (0.5% vol/vol), *Algoriphagus* OMVs ($1{:}10^6$ dilution), and *Algroiphagus* bulk lipid extract 0.1μg/mL elicit levels of rosette induction that recapitulate live bacteria.

Choanoflagellate husbandry

SrEpac was propagated in 5% Sea Water Complete media at 22°C. Sea Water Complete (SWC) media (250 mg/L peptone, 150 mg/L yeast extract, 150μL/L glycerol in artificial sea water) was diluted to 5% (vol/vol) in artificial sea water to make 5% Sea Water Complete media. Artificial sea water was made by adding 32.9 g Tropic Marin sea salts (Wartenberg, Germany) to 1L distilled water to a salinity of 32-27 parts per thousand.

SrEpac was passaged 1:10 into 9mL fresh 5% SWC once a day to stimulate rapid growth (cells were grown in 25cm^2 Corning cell culture flask). For all rosette development bioassays, cultures of single cells were induced shortly after passaging at a density of approximately 10^4-10^5 cells/mL.

Quantifying rosette development

For all assays, rosette development was quantified approximately 22-24 hours post-induction. In untreated cultures, *S. rosetta* can sometimes form small clumps of cells that break apart upon pipetting. Therefore, to quantify rosette development, 100μL of treated SrEpac was first pipetted vigorously to disrupt clusters of cells that were not in rosettes, then fixed in 1% formaldehyde immediately before counting (Bright-Line hemacytometer, Hausser Scientific). To determine the fraction of cells in rosettes, single cells and cells within rosettes were scored until 500 total cells had been counted (per technical replicate). To quantify rosette size, the number of cells in each rosette were counted and recorded. A group of four or more cells qualified as a rosette if the cells maintained an organized polarity relative to a central focus (with each cell oriented with the apical flagellum pointing outward) after vigorous physical perturbation, in this case from pipetting up and down or vortexing. At least three biological replicates were performed for each assay.

Isolation of Algoriphagus OMVs

Algoriphagus was grown by shaking in 200 mL SWC media for 48H at 30 °C, and pelleted. Cell-free supernatant was filtered twice through a 0.2 μM filter, and then spun at 36,000 x *g* for three hours at 4 °C (Type 45 Ti rotor, Beckman Coulter). OMVs were resuspended in 2 mL 50 mM HEPEs, pH 7.4. For rosette induction bioassays, OMVs were first diluted 1:1000 in ASW, and then added to SrEpac at a 1:1000 dilution (a final dilution of $1:1e^6$).

General information for isolation/characterization of lipids

All NMR experiments were carried out on a Varian INOVA 400 MHz, 600 MHz NMR spectrometer, a Bruker Advance (sgu) 900 MHz or a Varian Unity Inova 600 MHz equipped with a cryoprobe. Chemical shifts are reported in ppm from tetramethylsilane. Data in table form are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, br = broad, m = multiplet), coupling constants, and integration. Integrals are in accordance with assignments, coupling constants are given in Hz. All ^{13}C-NMR spectra are proton-decoupled. For detailed peak assignments 2D spectra were measured (COSY, HMQC, HMBC, NOESY and NOE if necessary). HPLC purifications were carried out on an Agilent 1100 or 1200 Series HPLC system (Agilent Technologies) equipped with a photo diode array detector. LC-MS analysis was performed on an Agilent 1200 Series HPLC system equipped with a diode array detector and a 6130 Series ESI mass spectrometer. High resolution mass spectrometry (HR-MS) analysis was performed on a Waters Micromass Q-ToF Ultima ESI-TOF mass spectrometer at the University of Illinois Urbana-Champaign, School of Chemical Sciences Mass Spectrometry Laboratory, or was carried out by Ted Voss at the WM Keck Foundation Biotechnology Resource Laboratory at Yale University on a Bruker 9.4T FT-ICR MS. All

solvents and reagents were purchased from Sigma, Aldrich, Alfa Aesar or VWR and used without further purification.

Cultivation conditions

In short, *A. machipongonensis* PR1 was grown in seawater complete media at 30 °C (16 x 1L) for 2 days as previously described.[15,23] The cells were harvested by centrifugation and extracted with $CHCl_3$:MeOH (2:1, 4L). The organic extract was filtered and concentrated to give approximately 4 g crude lipid extract. The crude extract was dissolved in a minimum amount of methanol, and purified by preparative reversed phase HPLC (RP-HPLC) using a preparative C18 column (Phenomenex Luna C18(2), 5 μm, 100 Å, 250 x 21.2 mm). Compounds were eluted at 10 ml/min in a gradient of solvents A (water + 0.1% NH_4OH) and B (MeOH + 0.1% NH_4OH) using the following method: 30 - 100% solvent B for 30 min, isocratic at 100% solvent B for 8 minutes, and ramp back down to 30% B over 2 min. Fractions were dried, weighed. Fractions were either resuspended to 5 mg/mL in DMSO to test for rosette-inducing activity, or further analyzed and purified as described below.

Purification of RIFs

Fractions containing sulfonolipids were purified as previously described.[15,49] Briefly, preparative HPLC purifications were carried out using a preparative C18 column (Phenomenex Luna C18(2), 5 μm, 100 Å, 250 x 21.2 mm) and the following a gradient of solvents [A (0.1% NH_4OH in water) and B (0.1% NH_4OH in methanol)] at a flow rate of 10 ml/min: 65% B increasing to 100% B over 30 min, isocratic at 100% B for 10 min before returning to 65% B and re-equilibrating over 10 min. If necessary, HPLC was repeated until analytical pure compounds were obtained. Fractions were analyzed by LC-MS using a gradient of solvents at a flow rate of 0.5 ml/min [A (0.1% NH_4OH in water) and B (0.1% NH_4OH in methanol)]: 65% B increasing to 100% B over 30 min,

isocratic at 100% B for 1 min before returning to 65% B and re-equilibrating over 3 min. The obtained pure isolates were repeatedly dissolved in MeOH and dried under vacuum to remove traces of NH_4OH to give pure sulfonolipids as white solids. Structure elucidation was performed by NMR (d_6-DMSO, d_4-MeOD) and HRMS measurements.

Purification of LPEs

HPLC fractions containing LPEs were eluted at around 85% solvent B. LPEs were then further purified by semi-preparatory HPLC using a semipreprative C-18 column (Phenomenex Gemini NX-C18, 100 Å, 5 μm, 250 x 10 mm) and the following method: With a flow rate of 2.4 ml/min, compounds were eluted using a gradient of 65 - 100% solvent B (MeOH + 0.1% NH_4OH) over 20 minutes, and isocratic at 100% B for 4 minutes. LPEs were detected by ELSD and known derivatives verified by LC/MS and HRMS. Final purification of LPEs was achieved using an analytical C18 column (Phenomenex Gemini NX-C18, 110 Å, 5 μm, 250 x 4.6 mm) with a flow rate of 0.5 ml/min and a gradient of 70%-95% solvent B over 30 minutes. To elucidate the structure of LPEs, Grubb's metathesis and MS/MS analysis was performed on the most active fractions.

Grubb's metathesis

100 μg LPE was resuspended in 250 μl CH_2Cl_2 (anhydrous). After addition of 10 μl methyl acrylate (Sigma Aldrich M27301) and 30 μg of second generation Hoveyda-Grubbs catalyst (Sigma Aldrich 569755), the reaction was stirred at room temperature for ~3 h. For analysis, 10 μl of reaction was quenched with 10 μL methanol, and the crude mixture analyzed by LC/MS.[25]

MSMS analysis

Dried samples were dissolved in 100 μL of 65/30/5 (v/v/v) acetonitrile/isopropanol/water and 1

μl was analyzed by LC-MS/MS with data-dependent fragmentation as described previously.[49] Data were analyzed manually using XCalibur QualBrowser v2.2 and were compared to literature MS/MS spectra available at LipidMaps (lipidmaps.org).

Testing Algoriphagus lipid extracts for activity in isolation and in combination

SrEpac was cultured as described above, and aliquoted into 24-well plates (Corning Costar). *Algoriphagus* lipid fractions were first pre-mixed in 5% SWC to avoid precipitation of the sample, and then added to SrEpac to yield the desired concentration. Each sample was tested in isolation for cell viability and rosette-inducing activity at concentrations of 0.05, 0.2, 2, 5, and 10 μg/mL. Samples were tested in pairwise combinations by first mixing two lipid fractions in equal parts, and then testing the mixture at final concentrations of 0.1, 0.5, 2, and 5 μg/mL. Rosette development was quantified as described above. Heat map was generated using Plotly.

Activity profile of purified lipids

The potency of purified lipids was determined using a quantitative bioassay for rosette development. SrEpac was cultured as described above, and aliquoted into 96 well plates (Corning Costar). Lipid samples were resuspended in DMSO to a concentration of 2 mg/mL. Lipids were first pre-mixed in 5% SWC to avoid precipitation of the sample, and then added to SrEpac to yield the desired concentration. Rosette development was quantified as described above. Graphs were generated using GraphPad Prism 6 statistical software. Rosette induction data was analyzed using a one site (specific binding), non-linear regression model.

Testing commercial lipids for activity

The inducing and enhancing activity of commercially available lipids was determined

using a quantitative bioassay for rosette development. SrEpac was cultured as described above, and aliquoted into 96 well plates (Corning Costar). Lipid samples were resuspended according to product specification. To test for inducing activity, lipids were first pre-mixed in 5% SWC to avoid precipitation of the sample, and then added to SrEpac to concentrations of 5, 10, 20, and 50μg/mL. To test for enhancing activity, samples were tested in combination with 2μM RIF mix at concentrations of 5 and 20μg/mL. Rosette development was quantified as described above.

Immunofluorescence microscopy

Live cells were allowed to settle for 30 min onto poly-L-lysine coated coverslips (BD Bioscience) and fixed in two steps: 5 min in 6% acetone followed by 10 minutes in 4% formaldehyde. Cells were stained with an anti-Rosetteless antibody at 1.25 ng/μL (1:1000), E7 anti tubulin antibody (1:1000; Developmental Studies Hybridoma Bank), Alexa fluor 488 anti-rabit and Alexa fluor 647 anti-mouse (1:1000 each; Molecular Probes), and .01mg/mL Hoechst 3342 (Thermo Fischer) before mounting in Prolong Gold antifade reagent (Molecular Probes). Cells were imaged at 63x using a Zeiss LSM 880 AxioExaminer with Airyscan.

3.7 References

1. Lang, B. F., O'kelly, C., Nerad, T., Gray, M. W. & Burger, G. The Closest Unicellular Relatives of Animals. *Curr. Biol.* **12,** 1773–1778 (2002).

2. Carr, M., Leadbeater, B. S. C., Hassan, R., Nelson, M. & Baldauf, S. L. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc. Natl. Acad. Sci. USA* **105,** 16641–16646 (2008).

3. King, N. *et al.* The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451,** 783–788 (2008).

4. Ruiz-Trillo, I., Roger, A. J., Burger, G., Gray, M. W. & Lang, B. F. A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol.* **25,** 664–672 (2008).

5. Richter, D. J. & King, N. The Genomic and Cellular Foundations of Animal Origins. *Annu. Rev. Genet.* **47,** 527–555 (2012).

6. Leadbeater, B. & Morton, C. A microscopical study of a marine species of Codosiga James‐Clark (Choanoflagellata) with special reference to the ingestion of bacteria. *Biol. J. Linn. Soc.* **6,** 337–347 (1974).

7. Nielsen, C. Six major steps in animal evolution: are we derived sponge larvae? *Evol. Dev.* **10,** 241–257 (2008).

8. Artis, D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat Rev Immunol* **8,** 411–420 (2008).

9. Arendt, D., Benito-Gutierrez, E., Brunet, T. & Marlow, H. Gastric pouches and the mucociliary sole: setting the stage for nervous system evolution. *Phil. Trans. R. Soc. B* **370,** 20150286 (2015).

10. Fairclough, S. R., Dayel, M. J. & King, N. Multicellular development in a choanoflagellate. *Curr. Biol.* **20,** R875–6 (2010).

11. Dayel, M. J. *et al.* Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. *Dev. Biol.* **357,** 73–82 (2011).

12. Levin, T. C., Greaney, A. J., Wetzel, L. & King, N. The rosetteless gene controls development in the choanoflagellate S. rosetta. *elife* **3,** (2014).

13. Haeckl, E. The gastrea-theory, the phylogenetic classification of the animal kingdom and the homology of the germ-lamelle. *Q J Microscop Sci.* 142–165 (1874).

14. Alegado, R. A. *et al.* Algoriphagus machipongonensis sp. nov., co-isolated with a colonial choanoflagellate. *Int. J. Syst. Evol. Microbiol.* **63,** 163–168 (2013).

15. Alegado, R. A. *et al.* A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *elife* **1,** (2012)

16. Dayel, M. J. & King, N. Prey capture and phagocytosis in the choanoflagellate Salpingoeca rosetta. *PLoS ONE* **9,** e95577 (2014).

17. Roper, M., Dayel, M. J., Pepper, R. E. & Koehl, M. A. R. Cooperatively Generated Stresslet Flows Supply Fresh Fluid to Multicellular Choanoflagellate Colonies. *Phys. Rev. Lett.* **110,** 228104 (2013).

18. Kreft, J. M. Effects of forming multicellular colonies or attaching to surfaces on feeding rates of the choanoflagellate *Salpingoeca rosetta*. Thesis (M.A. Integrative Biology) University of California, Berkeley, 2010.

19. Levin, T. C. & King, N. Evidence for Sex and Recombination in the Choanoflagellate Salpingoeca rosetta. *CURBIO* **23,** 2176–2180 (2013).

20. Bartke, N. & Hannun, Y. A. Bioactive sphingolipids: metabolism and function. *J. Lipid Res.* **50,** S91–S96 (2008).

21. Abbanat, D. R., Godchaux, W., Polychroniou, G. & Leadbetter, E. R. Biosynthesis of a sulfonolipid in gliding bacteria. *Biochem. Biophys. Res. Commun.* **130,** 873–878 (1985).

22. Godchaux, W. & Leadbetter, E. R. Sulfonolipids are localized in the outer membrane of the gliding bacterium Cytophaga johnsonae. *Arch. Microbiol.* **150,** 42–47 (1988).

23. Beemelmanns, C. *et al.* Synthesis of the Rosette-Inducing Factor RIF-1 and Analogs. *J. Am. Chem. Soc.* **136,** 10210–10213 (2014).

24. Cantley, A. M., Woznica, A., Beemelmanns, C., King, N. & Clardy, J. Isolation and Synthesis of a Bacterially Produced Inhibitor of Rosette Development in Choanoflagellates. *J. Am. Chem. Soc.* **138,** 4326–4329 (2016).

25. Kwon, Y., Lee, S., Oh, D.-C. & Kim, S. Simple determination of double-bond positions in long-chain olefins by cross-metathesis. *Angew. Chem. Int. Ed. Engl.* **50,** 8275–8278 (2011).

26. Isao Ishii, Nobuyuki Fukushima, Xiaoqin Ye, A. & Chun, J. Lysophospholipid Receptors: Signaling and Biology. *Annu. Rev. Biochem.* 321–354 (2004).

27. Makide, K. *et al.* Novel lysophosphoplipid receptors: their structure and function. *J. Lipid Res.* **55,** 1986–1995 (2014).

28. Moat, A. G., Foster, J. W. & Spector, M. P. *Lipids and Sterols*. *Microbial Physiology* 450–474 (John Wiley & Sons, Inc., 2002).

29. Grzelczyk, A. & Gendaszewska-Darmach, E. Novel bioactive glycerol-based lysophospholipids: New data – New insight into their function. *Biochimie* **95,** 667–679 (2013).

30. Sommer, F. & Bäckhed, F. The gut microbiota - masters of host development and

physiology. *Nat. Rev. Microbiol.* **11,** 227–238 (2013).

31. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 3229–3236 (2013).

32. Hadfield, M. G. Biofilms and Marine Invertebrate Larvae: What Bacteria Produce That Larvae Use to Choose Settlement Sites. *Annu. Rev. Marine. Sci.* **3,** 453–470 (2011).

33. McFall-Ngai, M. J. & Ruby, E. G. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science* **254,** 1491–1494 (1991).

34. Alegado, R. A. & King, N. Bacterial Influences on Animal Origins. *Cold Spring Harb Perspect Biol* **6,** a016162 (2014).

35. Kirkegaard, J. B., Marron, A. O. & Goldstein, R. E. Motility of Colonial Choanoflagellates and the Statistics of Aggregate Random Walkers. *Phys. Rev. Lett.* **116,** 038102 (2016).

36. Stocker, R. Marine microbes see a sea of gradients. *Science* **338,** 628–633 (2012).

37. Kimelman, D. & Kirschner, M. Synergistic induction of mesoderm by FGF and TGF-β and the identification of an mRNA coding for FGF in the early Xenopus embryo. *Cell* **51,** 869–877 (1987).

38. Altaba, A. R. I. & Melton, D. A. Interaction between peptide growth factors and homoeobox genes in the establishment of antero-posterior polarity in frog embryos. *Nature* **341,** 33–38 (1989).

39. Bachiller, D. *et al.* The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature* **403,** 658–661 (2000).

40. Niehrs, C. Regionally specific induction by the Spemann|[ndash]|Mangold organizer. *Nat Rev Genet* **5,** 425–434 (2004).

41. Carmena, A., Gisselbrecht, S., Harrison, J., Jiménez, F. & Michelson, A. M. Combinatorial signaling codes for the progressive determination of cell fates in the Drosophila embryonic mesoderm. *Genes Dev.* **12,** 3910–3922 (1998).

42. Davidson, E. H., Cameron, R. A. & Ransick, A. Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125,** 3269–3290 (1998).

43. Halfon, M. S. *et al.* Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors. *Cell* **103,** 63–74 (2000).

44. Koropatnick, T. A. Microbial Factor-Mediated Development in a Host-Bacterial Mutualism. *Science* **306,** 1186–1188 (2004).

45. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An

immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).

46. Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* **456,** 507–510 (2008).

47. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis. *Cell* **118,** 229–241 (2004).

48. Nedashkovskaya, O. I. Echinicola pacifica gen. nov., sp. nov., a novel flexibacterium isolated from the sea urchin Strongylocentrotus intermedius. *Int. J. Syst. Evol. Microbiol.* **56,** 953–958 (2006).

49. Tafesse, F. G. *et al.* Disruption of Sphingolipid Biosynthesis Blocks Phagocytosis of Candida albicans. *PLoS Pathog.* **11,** e1005188 (2015).

Chapter 4: Development of a screen for pathogenic effectors from a bacterial affiliate of *D. discoideum*

4.1 Contributions and Acknowledgments

Contributions

Several researchers contributed to the work presented in this chapter. I performed extraction, purification and characterization of small molecules from Bx11, synthesis of the DKPs, isolation of Bx11 gDNA and construction of the cosmid library. The high throughput plaque assay was developed and executed with Antonio Ruzzini, who also cloned the pyrrolnitrin gene cluster and assembled the Bx11 genome. Debra Brock performed the sporulation assays on the Bx11 extracted material. Jon Clardy, Joan Strassmann and David Queller provided research guidance and financial support.

Acknowledgments

We would like to thank Dr. Pierre Stallforth for experimental advice and critical reading of this manuscript, as well as Raizza Carminato for assistance in the lab. We would also like to acknowledge the Duke University Center for Genomic and Computation Biology where the PacBio sequencing of the Bx11 genome was performed. This work was funded by NIH GM086258.

4.2 Abstract

The act of primitive farming by *Dictylostelium discoideum* is ridden with complex inter- and intra-species interactions. The farmer clones of *D. discoideum* carry bacteria with them that can either serve as a food source, or secrete toxic molecules preventing non-farmer *D. discoideum* from usurping the farmer's resources. Here we examine how a bacterial co-isolate of a farmer amoeba causes decreased sporulation in the non-farmer *D. discoideum*. In our approach we first examined the small molecule repertoire of this bacterial isolate through bioactivity-guided fractionation. Concurrently we aimed to identify genes contributing to the toxic effects of these bacteria through heterologous expression of the fragmented bacterial genome, and screening for toxicity in the non-farmer amoeba. To accomplish this we developed a novel high throughput screen for toxicity in *D. discoideum* and identified a set of hits highly enriched in genes predicted to be involved in pathogenesis. This screen serves as not only a new approach to further understand the relationship between bacterial co-isolates of farmer amoeba and the non-farmers to which they are toxic, it is also a generalizable high throughput assay for bacterial toxicity or pathogenicity against amoebae.

4.3 Introduction

Dictyostelium discoideum, also known as a social amoeba, is a soil-dwelling eukaryotic organism that has become a model for studying a number of biological and chemical processes including cell differentiation, signaling and intraspecies interactions.[1-7] Much of the suitability of

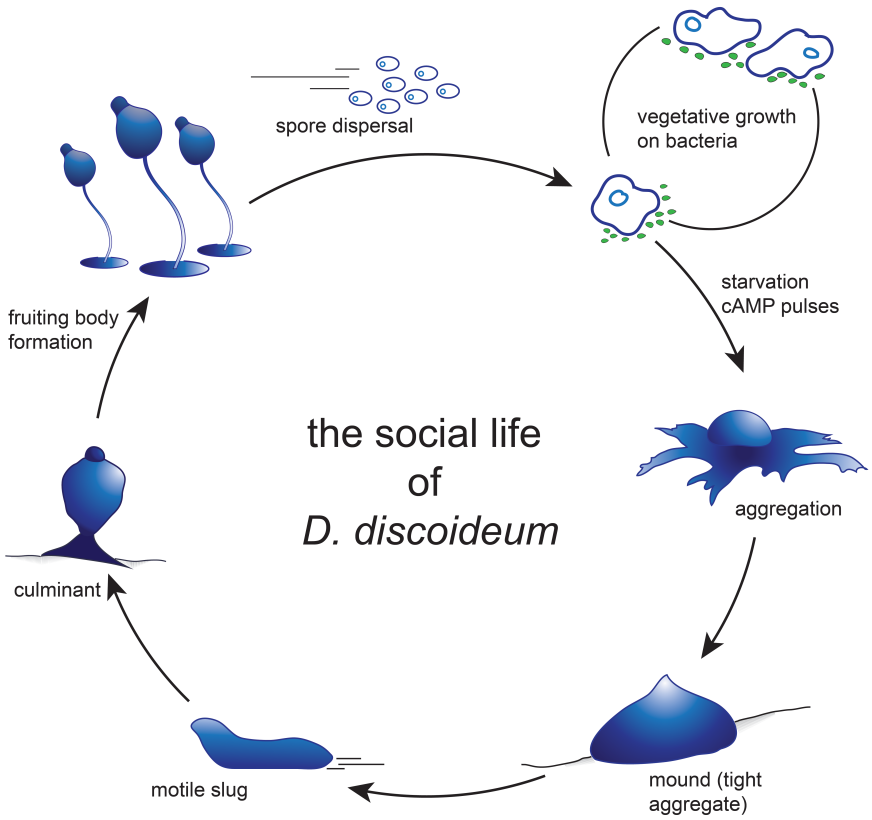


Figure 4.1 The life cycle of *D. discoideum*
During vegetative growth *D. discoideum* feed on bacteria, but upon starvation they release cAMP pulses that signal for aggregation. The amoeba then progress through several aggregation stages including the "mound" and the "motile slug". The release of small molecule signals (such as cAMP or DIF-1), and each cell's response to these signals, determine where in the aggregate each amoeba will end up. The final stage of development is the formation of the fruiting body, which is composed of a stalk of cells topped with a sorus, which is essentially a ball of *D. discoideum* spores. When the sorus bursts, the spores are dispersed and the cycle starts anew.

D. discoideum as a model organism is due to its complex and fascinating life cycle (Figure 4.1). The unicellular amoebae feed on bacteria during the vegetative growth stage. Upon starvation the amoebal cells start to aggregate, forming a series of structures and culminating in the fruiting body – a sorus full of spores perched a top a stalk. When these sori break open and the

spores disperse, the dictyostelia are free to colonize a new, and hopefully bacteria-rich environment.[8]

To trigger the transition from one stage of their life cycle to the next, social amoebae produce and respond to small molecule signals. Cyclo adenosine monophosphate (cAMP), for example, is produced by the vegetative cells when they run out of their bacterial food source. This signal induces a complex set of signaling cascades leading to chemotaxis and cell polarization, and these processes are what ultimately determine where in the aggregate structures each amoebal cell will play out the rest of its life.[9-11] Other interesting small molecules, such as differentiation inducing factor 1 (DIF-1) and the structurally related DIF-2 and DIF-3, have also been shown to be essential regulators of dictyostelium development; and many other polyketide synthase derived small molecules have been isolated from dictoystelia, though the function of most of these molecules remains unknown.[12-16]

To an evolutionary biologist the process of cell differentiation and aggregation in the social amoeba is of utmost significance. The amoeba that end up in the sori of the fruiting body get to pass on their genes to the next generation; those that end up in the stalk, however, do not. This inherent inequality among the positions within the fruiting body shows the importance of altruism during the multicellular stages of the amoebal life cycle, but also presents the opportunity for intraspecies competition and cheating.[17,18] This concept of competition within the fruiting body is even further emphasized by the occurrence in nature of “clonal chimeras” – or fruiting bodies that are made up of multiple clones; this finding suggests that amoeba may have evolved methods for competition or kin selection during the multicellular development process.[19]

Recently it was reported that environmental isolates of *D. discoideum* can engage in a symbiotic relationship with co-isolated bacteria in which certain clones of *D. discoideum* (called “farmers”) carry bacteria from aggregation through to sorus formation, so when the spores are disseminated, they bring an inoculum of their bacterial food source with them; this behavior was

designated as “primitive farming”. However, not all clonal isolates of *D. discoideum* engage in primitive farming (termed “non-farmers”) – though it is not yet understood what inherently differentiates the farmer amoeba from the non-farmers.[20]

While it is obvious why the farmer would farm their food source, much of the bacterial inoculum is made up of bacteria upon which they will not feed. Follow-up experiments on this system showed that when the farmer and non-farmer are mixed and allowed to develop as chimeric fruiting bodies, the presence of certain strains of non-food bacteria favors the predominance of the farmer over the non-farmer amoeba.[21] These findings suggest that these non-food bacteria somehow provide a selective advantage to the farmer during multicellular development (Figure 4.2).

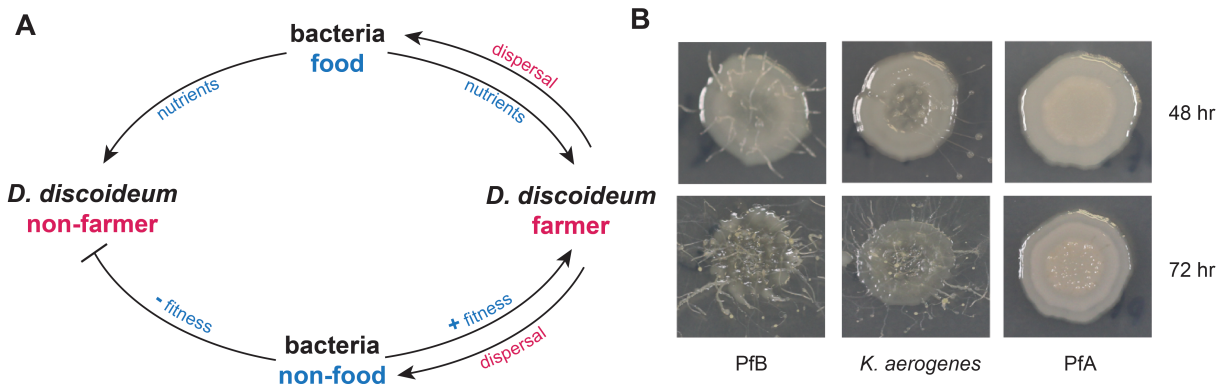


Figure 4.2 Primitive farming in *Dictyostelium discoideum.*
(A) Model depicting the hypothesized relationship between the farmer and non-farmer *D. discoideum*, and co-isolated food and non-food bacteria. **(B)** Images depicting growth of farmer *D. discoideum* on the bacterial food source PfB, and lack of growth on the non-food source PfA. *K. aerogenes* is the standard food source for *D. discoideum* when grown under laboratory conditions.

Our lab was very interested in whether bacterially produced small molecules mediate the disadvantage incurred to non-farmer amoebae by the non-food bacteria. In collaboration with the Strassmann/Queller group (who initially observed the farming behavior), we determined that a non-food *Pseudomonas fluorescens* (PfA) that was co-isolated with a farmer clone, produces small molecules (chromene and pyrrolnitrin) that increase sporulation (or “fitness”) in the farmer

amoeba and decrease sporulation in the non-farmer. As pyrrolnitrin is a commonly isolated antimicrobial molecule with known anti-amoebal activity, it was not surprising that it would have a deleterious effect on the non-farmer amoeba, however its lack of toxicity to the farmer suggests that the farmer amoeba may have evolved resistance to its effects.

We were also interested in understanding how the non-food source bacteria, PfA, differed from the food source bacteria, PfB. These two strains, which were co-isolated with the same farmer amoeba, seemed at first glance, to be genetically identical. Through analysis of the genomic sequence we found that a point mutation in PfB rendered a major regulator of secondary metabolism non-functional; this single change, which had a significant effect on secondary metabolite production, turned out to be the definitive difference between the food source bacteria (PfB) and the non-food source bacteria (PfA).[22] These initial results regarding the nature of the *D. discoideum*-bacterial symbiosis revealed the importance of small molecules in mediating this novel interaction, and alluded to the potential co-evolution of the non-food bacteria and the farmer amoeba.

4.4 Results

In previous studies, it was shown that β-Proteobacteria in the genus *Burkholderia* were one of the most commonly co-isolated groups with the farmer *D. discoideum*.[20] Further exploration of the role of these bacteria in this symbiotic association revealed that, similar to PfA, many of the co-isolated non-food Burkholderia strains promoted sporulation (or "fitness") in the farmer while simultaneously decreasing the "fitness" of the non-farmer amoeba.[21] Our work with *Pseudomonas* symbionts of *D. discoideum* demonstrated the importance of the small molecules pyrrolnitrin and chromene in mediating both the positive interaction with farmer amoeba and the

negative interaction with non-farmer isolates[22] thus naturally, we opted to focus our examination of the associated *Burkholderia* strain (referred to as Bx11) on secondary metabolite production.

To determine if the effect of Bx11 supernatant on the farmer (QS11) and non-farmer (QS9) amoeba is mediated by small molecules, we extracted and fractionated the supernatant of bacterial monocultures. We first grew Bx11 on SM/5 agar plates for 5-7 days, in a manner analogous to how these bacteria are cultured with the farmer *D. discoideum* (QS11). We removed the bacteria from the plate by immersing the agar surface in KK2 buffer and scraping the cells from the agar. We then starved the bacteria in buffer for 1-2 hours prior to spinning down to separate the cell pellet from the supernatant. The supernatant was extracted with ethyl acetate and dried down before testing.

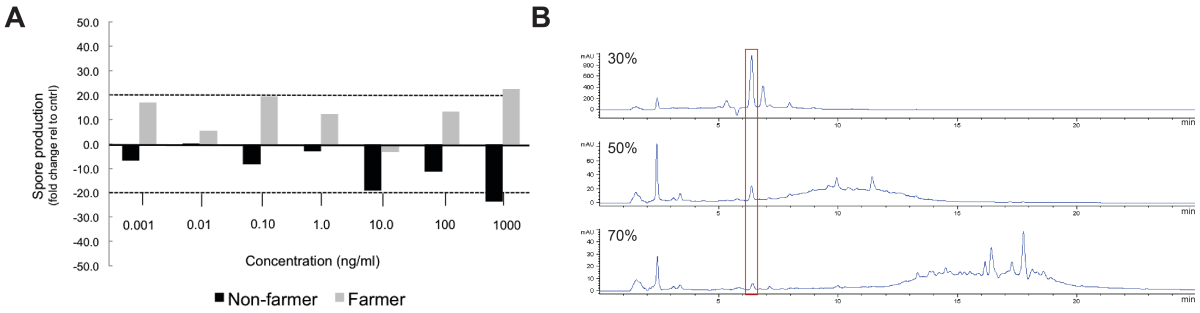


Figure 4.3 Ethyl acetate extracts from Bx11 show differential effects on sporulation in the farmer and non-farmer amoeba.

(A) Fold change in spore production in farmer and non-farmer amoeba treated with a range of concentrations of crude ethyl acetate extract from Bx11. Spores were counted 36 hours after plating and normalized to a DMSO control. **(B)** LCMS trace indicating single peak retained in the three most active ethyl acetate extract fractions. Fractions were run over a C18 column using a gradient of acetonitrile/water supplemented with 0.1% formic acid.

We tested the ethyl acetate supernatant extract in the farmer/non-farmer sporulation assay to determine if it mirrored the effects observed from supernatant and live cells. We found that the ethyl acetate extract could both increase sporulation of the farmer (QS11) and decrease that of the non-farmer (QS9) to an extent that was comparable to Bx11 supernatant (Figure 4.3). To hone in on the small molecule(s) responsible for this effect, we performed a crude

fractionation (Sep-Pak, C18 column) and found that fractions eluted with 30%, 50% or 70% acetonitrile showed both activities observed in the crude ethyl acetate fraction.

Liquid chromatography-mass spectrometry (LC-MS) analysis showed that these fractions shared a major peak, and further purification by reverse phase high performance liquid chromatography (RP-HPLC) allowed us to hone in on a single fraction containing two major components (Figure 4.3). High resolution mass spectrometry (HRMS) along with 1D and 2D NMR spectroscopy led us to determine that this fraction was primarily composed of the riboflavin derivative lumichrome, and the diketopiperazine 3-benzyl-6-isopropylpiperazine-2,5-dione (Figure 4.4). Both of these molecules have been isolated from numerous bacterial sources, and have been ascribed a variety of biological activities, though their ecologically relevant roles are not well understood.

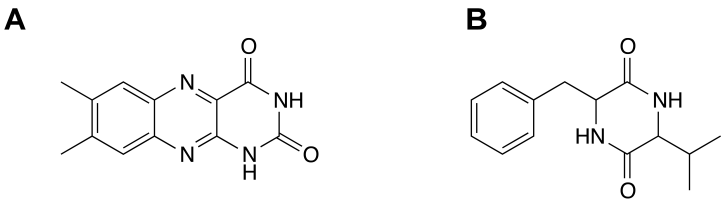


Figure 4.4 Molecules isolated from Bx11 ethyl acetate active fraction
(A) lumichrome and **(B)** Phe-Val-diketopiperazine were the two major components of the most active fraction from Bx11 ethyl acetate extract.

Lumichrome has been isolated from bacteria, plants and algae, and is implicated as an important agent of interkingdom signaling. Of note, its activity is often inextricably linked to that of its precurosor, as riboflavin can spontaneously convert to lumichrome in certain environments, causing some confusion about which molecule is responsible for the observed activity. Both lumichrome and riboflavin have been shown to act as a quorum sensing signal in *Pseudomonas aeruginosa* through direct binding to the LasR acyl homoserine lactone binding pocket. Since these molecules can be produced by plants and algae as well as bacteria, it is believed that

certain plants/algae may use lumichrome as a quorum sensing mimic that helps modulate associated bacterial populations.[23,24] Lumichrome and riboflavin, which are often produced by members of the family *Rhizobiacea* have also been shown to play important roles in modulating plant growth and development, though the exact activity seems to vary from plant to plant, and the mechanism of these diverse activities is not understood.[25,26]

The diketopiperazines (DKPs), or cyclic dipeptides, are also a ubiquitous class of molecules that have been isolated from every kingdom and ascribed an overwhelming number

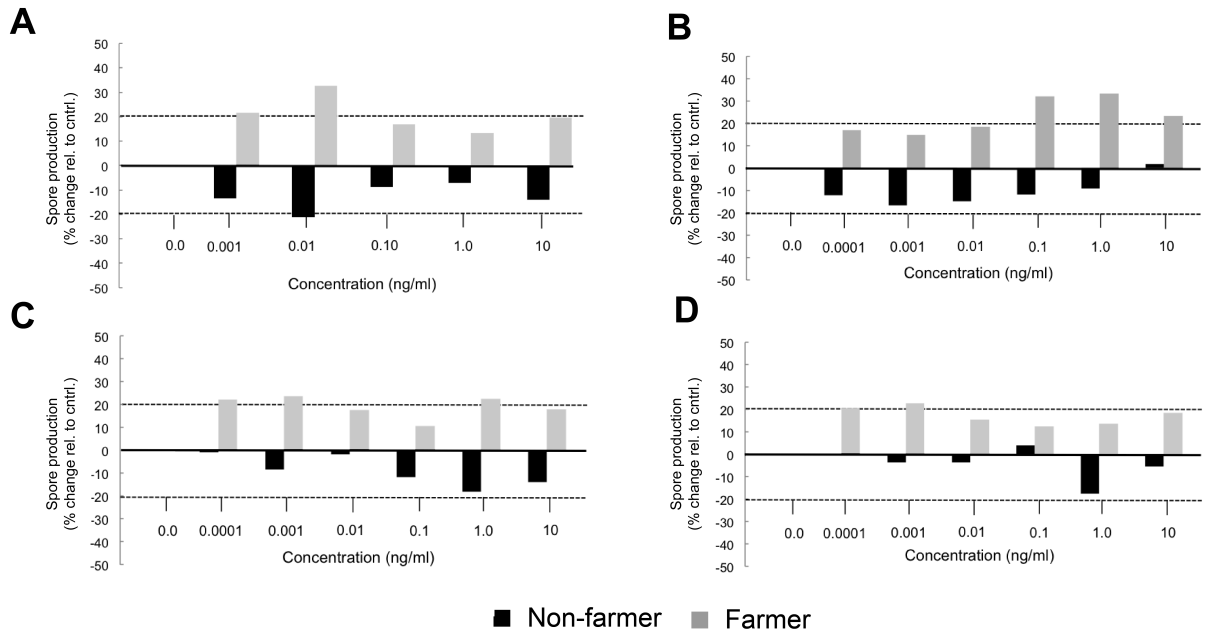


Figure 4.5 Sporulation assays of non-farmer and farmer *D. discoideum*
(A) Bx11 semi-purified fraction containing both lumichrome and val-phe DKP**. (B)** Synthetic L-val-L-Phe DKP. **(C)** Lumichrome (Sigma Aldrich 103217). **(D)** L-val-L-Phe DKP + lumichrome. Dotted line marks 20% change relative to control.

of biological activities, though often with relatively little clarity on mechanism or ecological relevance.[27-32] The molecule we isolated, 3-benzyl-6-isopropylpiperazine-2,5-dione (from this point forward referred to as Val-Phe-DKP), is of the simplest class of these molecules – formed from a condensation of two amino acids – in this case, phenylalanine and valine. While this

molecule has been isolated from a few sources, there have not been many studies examining the biological activity of this particular DKP.[33-35] Val-Phe-DKP is, however, closely related to the cyclic dipeptide Aureusimine B (phevalin), which was originally isolated from *Streptomyces* as a protease inhibitor and was later isolated from *Staphylococcus aureus.*[36] Its function as a potential regulator of virulence factors in this human pathogen has been much disputed.[37]

We synthesized L-Val-L-Phe DKP and D-Val-L-Phe DKP for further testing of these molecules alone and in combination with lumichrome. Using NMR spectroscopy we were able to determine that D-Val-L-Phe is not a component of the Bx11 active fraction, so we did not pursue this compound further. Using our previously described sporulation assay, we tested each molecule on both the non-farmer and farmer dictyostelia, and could not find a condition that could perfectly recapitulate the effects seen from either the Bx11 supernatant, the ethyl acetate crude extract, or even the lumichrome + L-val-L-phe DKP containing fraction. However, we did observe that most of these conditions showed a general growth enhancing effect on the farmer, yet none of them had a major detrimental effect on the non-farmer (Figure 4.5)

Concurrently we took a complementary approach by examining the potential for secondary metabolite production through genome analysis. Previous 16S rRNA gene analysis of Bx11 (using the 16S ribosomal database) showed 98% sequence identity to *Burkholderia xenovorans*, an organism isolated from a landfill in New York, that is noted for its capacity to degrade polychlorinated biphenyls (PCBs).[38] Because *B. xenovorans* and many other *Burkholderia* species have large genomes and are well-known for their prolific secondary metabolism, we expected the Bx11 genome to be large and ridden with (often cryptic) biosynthetic gene clusters.[39-42] We isolated genomic DNA from Bx11 and performed whole genome sequencing using PacBio SMRT sequencing technology. We found that that the genome of Bx11, composed of 2 chromosomes of 3.31 MB and 852 kb respectively, is significantly smaller than that of *B. xenovorans*, and more comparable to parasitic or

endosymbiotic *Burkholderia* species, which often have reduced genomes due to selective pressures to remain associated with the host.[43,44]

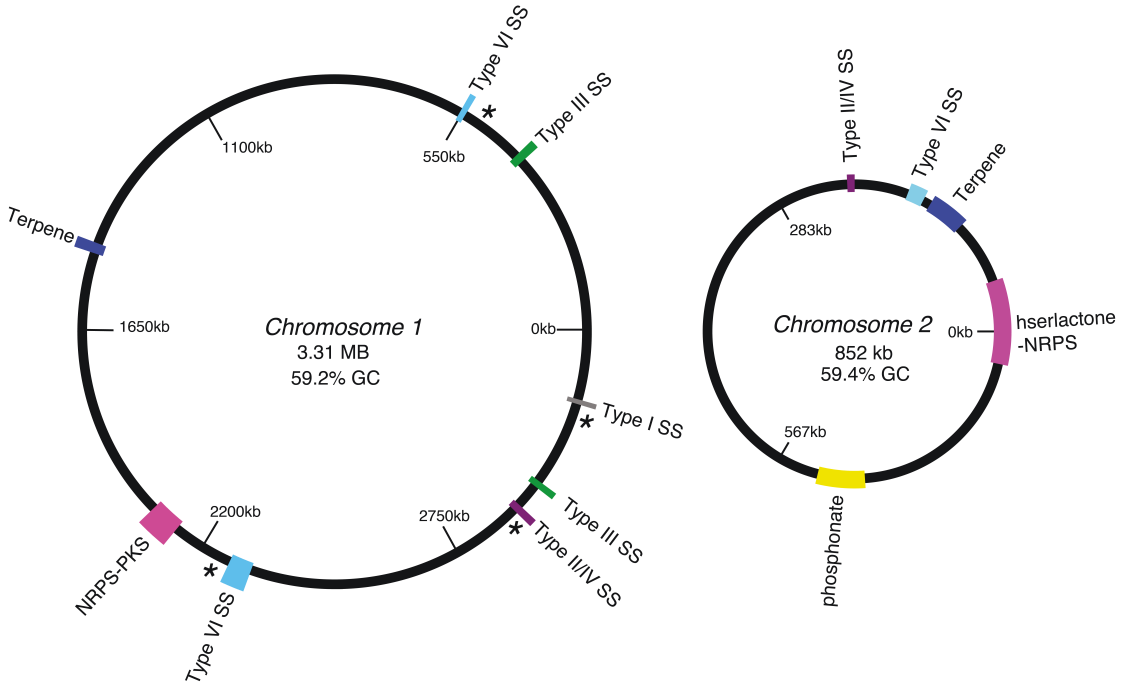


Figure 4.6 Bx11 genome map
Colored markers show locations of putative biosynthetic gene clusters and secretion systems contained within the Bx11 genome. Annotations and predictions were determined using RAST and antiSMASH software. Loci marked with an asterisk were later determined to be "hits" in the cosmid library screen.

To annotate the genome and determine the biosynthetic gene cluster content of Bx11, we used the antibiotics & Secondary Metabolite Analysis SHell (antiSMASH) and Rapid Annotation using Subsystem Technology (RAST). This analysis revealed that the Bx11 genome contains only two PKS-NRPS clusters, both predicted to encode siderophores containing subclusters for catecholamine and hydroxamine functional groups; however, the gene homology for these related molecules is too low to make any reliable structure predictions (Figure 4.6)

While the Bx11 genome contained little potential for new secondary metabolites, the RAST analysis reported a number of genes encoding proteins typically involved in virulence, namely components of Type VI (Rhs protein family/VgrG), Type I, Type III (Ysc protein family)

and Type II/IV (Flp, Rcp and Tad proteins) secretion systems.[45-49] The abundance of these potential virulence factors led us to wonder if Bx11 is potentially acting as an amoebal pathogen, to which the non-farmer is susceptible and farmer is resistant – a model that is alluded to in a recent study from the Strassmann/Queller lab.[50]

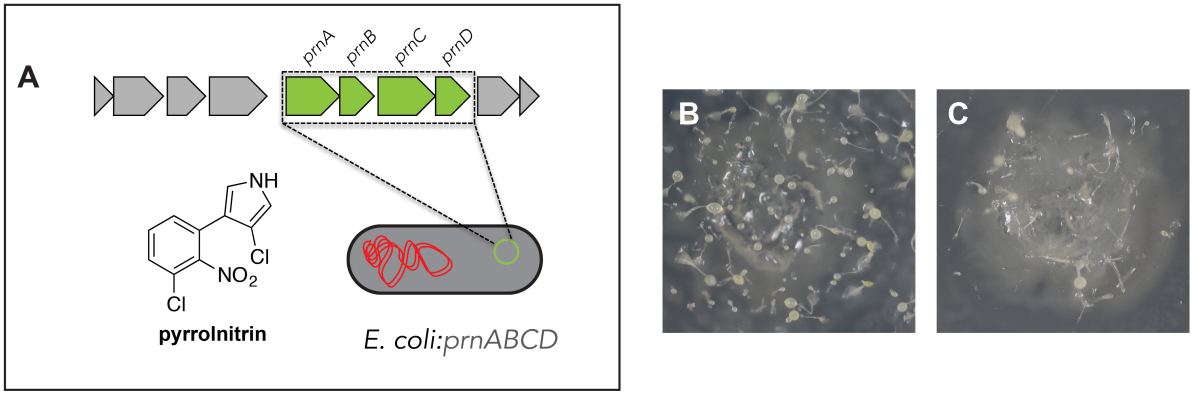


Figure 4.7 Pyrrolnitrin expression in host strain causes decrease in non-farmer feeding
(A) Scheme depicting expression of the pyrrolnitrin gene cluster in the heterologous host (DH10β). **(B)** Non-farmer amoeba (QS9) grown on *K. aerogenes*, the standard laboratory food source. **(C)** Non-farmer grown on DH10β expressing the pyrrolnitrin gene cluster. Images were taken after 36 hours.

Our lack of definitive success in mining for secondary metabolites, and the information provided by the genome sequence, motivated us to change our strategy towards understanding the Bx11 – amoebal interaction. Rather than attempt to screen for molecules that increase farmer sporulation and decrease non-farmer sporulation, we decided to focus on the potential pathogenic effects of Bx11 on the non-farmer amoeba (QS9). We reasoned that *D. discoideum's* distaste for Bx11 as a food source is likely related to the potential toxicity of the bacteria, so we opted to develop a plaque assay in which we can observe whether the amoeba will eat each clone. We were curious as to whether this assay would be sensitive enough to detect our previous finding that the small molecule pyrrolnitrin, produced by PfA, could cause toxicity in the non-farmer amoeba; so we cloned the pyrrolnitrin gene cluster into our host *E. coli* strain (DH10β), and co-plated it with non-farmer *D. discoideum.* We found that compared to our negative control, growth of the non-farmer *D. discoideum* was delayed, but not completely

abolished (Figure 4.7). This is consistent with our previous findings that while pyrrolnitrin displayed some toxicity towards non-farmer amoeba, it was not sufficient to prevent *D. discoideum* from feeding.[22]

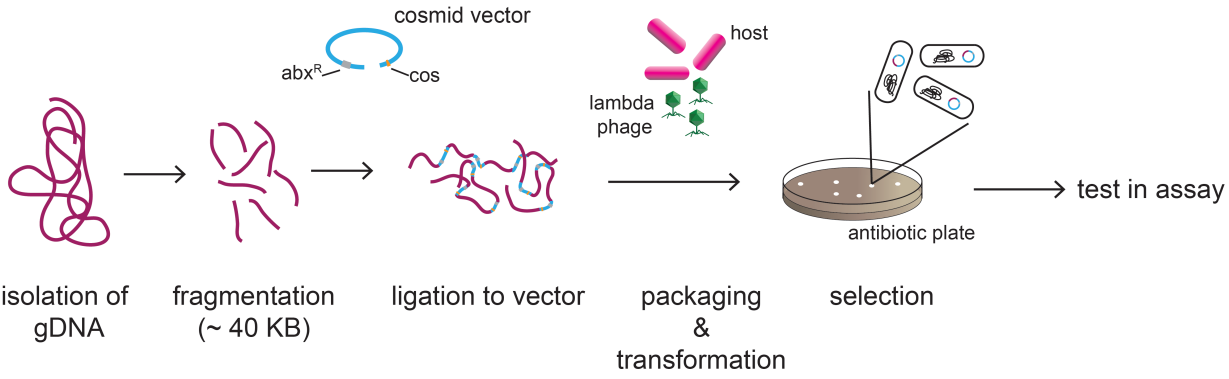


Figure 4.8 Cosmid library construction
For construction of our library we expressed 40 kb fragments of the Bx11 genome in a DH10β host strain. After picking clones from the antibiotic plate, we deposited them in 96-well plates to be efficiently screened for toxicity in the QS9 (non-farmer) plaque assay.

Heartened by these results, we decided to construct a cosmid library from the Bx11 genome. This approach has been extremely effective for discovering novel antibiotics as well as pathogenic effectors, and is particularly powerful when coupled with a robust high throughput screen.[51-54] Additionally, we believed that testing a cosmid library would be an unbiased approach to find the molecules responsible for making Bx11's toxicity to the non-farmer amoeba; since the cosmids have 40 kb inserts, they would be sufficiently large to include any of the putative secretion systems, as well as almost all of the predicted biosynthetic gene clusters from the Bx11 genome (Figure 4.8).

We isolated 1200 *E. coli* DH10β clones carrying Bx11 cosmids and grew them up in 96-well plates overnight for testing in the non-farmer plaque assay. We initially verified that the heterologous host (DH10β) is a suitable food source for the non-farmer, as compared to Bx11, which completely prevents non-farmer growth and sporulation (Figure 4.9). We mixed each clone 1:1 with *D. discoideum* spores (brought to a density of 10,000 spores/μL KK2 buffer) and

plated 1 μL of the amoeba-bacteria mixture onto an SM/5 agar plate (150 x 15 mm). An optimal distance between spots was obtained by fitting 48 tests/plate. Setting up this two organism co-culture assay to test 1200 clones took approximately 2 hours, which is comparable to the much simpler cytotoxicity or disc diffusion assays that are normally performed on cosmid libraries, and significantly shorter than the many weeks or even months it would have taken to generate a library of 1200 different bacterial extracts.

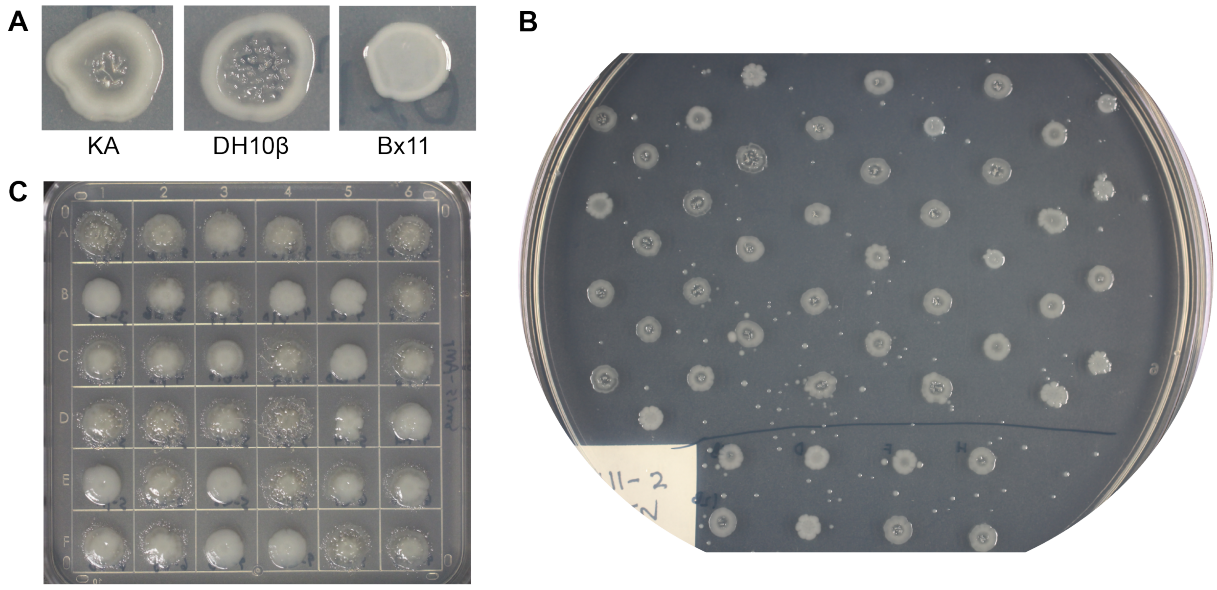


Figure 4.9 Images of non-farmer plaque assay
(A) Negative controls depicting non-farmer grown on *K. aerogenes* (standard lab food strain), the heterologous host strain DH10β, and Bx11. Clearance in the center of the spot shows the non-farmer is eating and starting to aggregate. **(B)** Assay plate of non-farmer co-spotted with cosmid library clones. Image taken after 48 hours. 48 clones can be tested per plate. **(C)** Secondary assay is done on smaller plates with slightly larger spots volumes. Image taken after 48 hours.

The co-cultures are incubated at room temperature, and after 48–72 hours the amoeba start to develop into the fruiting body stage. For retesting we selected any spot that did not develop at all, or showed a significant (>24 hours) delay in development. Using these criteria we assigned 137 of the 1200 clones as “hits” and re-tested them in the same assay using smaller plates and 2 μL of the amoeba-bacteria mixture. Upon completion of the primary assay we had evidence of a certain amount of false positives since some of the bacterial clones did not grow

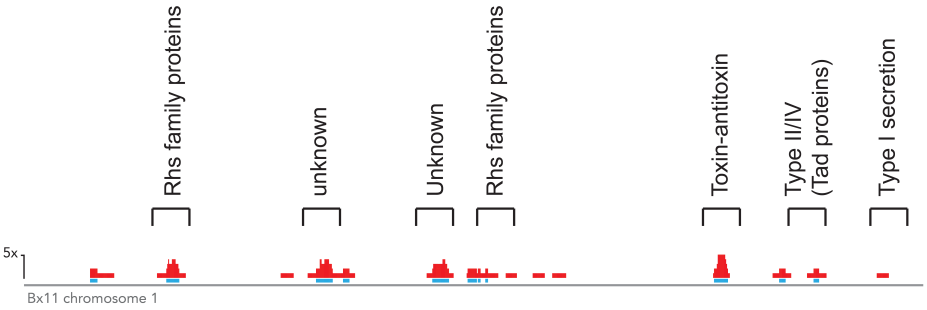


Figure 4.10 Plaque assay hits mapped onto Bx11 chromosome
Sequencing of the hit cosmids revealed a significant amount of clustering around certain genes.

very well after inoculation. The results of the secondary assay, which afforded about a 1/3 re-hit rate, verified this assumption.

We sequenced 38 of the hits that repeated in the secondary assay and mapped them onto the Bx11 genome. These sequencing results revealed a significant amount of clustering around specific genes, most of which have known roles in toxicity or pathogenesis (Figure 4.10). Of note, a number of hits overlapped with significant portions of two different Type VI secretion systems – both including the Rhs proteins and VgrG. Additionally we noted a couple hits on the Type I and Type II/IV secretion systems – though there was not significant clustering of hits around these genes. We also observed two major areas of clustering around a number of genes that do not obviously bear resemblance to toxicity or pathogenesis. Finally, we observed that eight separate hits including a type II toxin-antitoxin pair (TA pair) with a nuclease (PIN) domain. These genes were of particular interest as they are ubiquitous in pathogenic bacteria and are known to be involved in processes related to persistence and virulence – though their effects are usually observed on bacterial cells.[55-57] However, it is likely that TA pairs are also relevant within eukaryotic host organisms, since toxins containing PIN domains have been shown to exhibit the same nuclease-mediated toxicity towards eukaryotic organisms as they do towards other prokaryotes.[58]

Follow-up studies will focus on validation of the hits from the non-farmer plaque assay screen. While these hits provide us with clear segments of the genome to examine, and likely candidates for pathogenic effectors, generating knockouts and performing heterologous expression of these individual genes will be essential to demonstrate their relevance to the non-farmer—Bx11 interaction. Because the toxin-antitoxin pair was the most frequent hit, and also the smallest candidate set of genes, we will likely start our validation process here. We will also get the chance to explore the hits that did not contain genes that are clearly related to pathogenesis – presenting an opportunity to discover unexpected mediators of Bx11 toxicity towards non-farmer amoebae. Finally once we have compiled a validated set of hits, we will test their toxicity towards the farmer clone to determine if the amoebae hosts of Bx11 have indeed evolved resistance to these toxic factors.

4.5 Discussion

We began our investigation into Bx11 and its relationship with *D. discoideum* farmer and non-farmer clones by searching specifically for small molecules that mediate this tripartite relationship. However, when this approach failed to yield conclusive results, the genome of this environmental isolate offered new hypotheses about the nature of the *D.discoideum*-bacterial relationship.

In taking an unbiased approach we were able to develop a novel high throughput assay that can be used to determine food aversion or toxicity in *D. discoideum*, and were thus able to identify a set of hit genes that may be relevant for the toxic effects of Bx11 on the non-farmer amoeba. Follow up on all of the hits from the Bx11 cosmid screen will be essential for determining the molecules responsible for the observed toxicity to non-farmers and may ultimately lead to the characterization of previously unidentified toxic factors. Of note, the *D.*

discoideum plaque assay, which is cheap and requires minimal specialized equipment, can be applied to cosmid libraries from any type of bacterial DNA (including DNA from non-culturables) and is applicable to any strain of *D. discoideum* that can develop on an agar plate – making for a new generalizable assay for amoebacidal agents.

The discovery of bacterial farming behavior in soil isolates of *D. disoideum* marked the first observed instance of a naturally occurring symbiotic interaction between amoeba and *D. discoideum*. As we become increasingly aware of the major role bacterial symbionts play in eukaryotic health and behavior, it's surprising to think that so little is known about the biological interaction between two ubiquitous soil-dwelling organisms. Further investigation of this fascinating system will doubtlessly reveal additional intricacies in the interactions between the farmer amoeba, the non-farmer, and their food/non-food bacterial affiliates.

4.6 Experimental

Culturing and extraction of Bx11

Bx11 were grown in a lawn on SM/5 agar plates for 5-7 days. They were then washed off the plate using KK2 buffer and brought to an OD_{600} of 1.5. They were shaken at room temperature for 2 hours and spun down. The supernatant was extracted with 1 equivalent of ethyl acetate, dried over sodium sulfate and concentrated *in vacuo*. Extracts were resuspended in DMSO for testing in sporulation assays.

Isolation and characterization of diketopiperazines

All HPLC was performed on Agilent 1100 or 1200 series instruments. Specific columns used are specified in methods. LCMS was performed on an Agilent 1200 series HPLC with 6130 ESI mass spectrometer. High resolution mass spectrometry was performed on an Agilent 6530 QTOF LCMS (ESI) or a Waters Micromass (EI) 70-VSE (EI experiments conducted at the University of Illinois at Urbana-Champaign Mass Spectrometry Laboratory). NMRs were performed on the following instruments: Varian Inova 400 MHz or Varian Unity Inova 600 MHz.

Ethyl acetate extracts of Bx11 were resuspended in 30% methanol in water and run over a C18 sep-pak column (Waters) eluting with 15%, 30%, 50%, 70%, 90% and 100% acetonitrile in water. Active fractions were then further fractionated by semi-preparative HPLC using a Phenomenex Kinetex C18 column (110 A, 5 μm, 100 x 4.6 mm). Lumichrome and Val-Phe-diketopiperazine were purified using the following gradient 25%-32% acetonitrile in water with 0.1% formic acid over 20 minutes, isocratic at 32% acetonitrile for 5 minutes and a final wash at 100% acetonitrile + 0.1% formic acid for 5 minutes.

Synthesis of diketopiperazines (Figure 4.11)

To a solution of Boc-L/D-phenylalanine (D: Sigma Aldrich 15484, L: Sigma Aldrich 15480) (0.75 mmol, 0.2 g) in dimethylformamide (7.5 ml), 1.1 eq. HBTU (Sigma Aldrich 12804), 3 eq. DIPEA (Sigma Aldrich 387649) and 1 eq. D/L-valine-methyl ester hydrochloride (D: Sigma Aldrich 94665, L: Sigma Aldrich 860271) (0.75 mmol, 0.28 g). Mixture was stirred under argon for 12 h at room temperature. 70 ml ethyl acetate was added to reaction and mixture was washed with sodium bicarbonate, 5% citric acid and brine, dried over sodium sulfate and concentrated *in vacuo* to obtain **3** (either L-val-L-phe dipeptide or D-val-L-phe dipeptide). Structural information for L-val-L-phe stereoisomer: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.23 (m, 2H), 7.16 (m, 3H), 6.61 (d, *J* = 8.9, 1H), 5.25 (d, *J* = 8.1, 1H), 4,42 (dd, *J* = 8.6, 5.2, 1H), 4.36 (m, 1H), 3.64 (s, 3H), 3.04 (m, 2H). Structural information for D-val-L-phe stereoisomer: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.26 (m, 2H), 7.20 (m, 3H), 6.5 (d, *J* = 8.9, 1H), 5.08 (d, *J* = 7.9, 1H), 4.45 (dd, *J* = 8.5, 4.5), 1H), 4.39 (m, 1H), 3.68 (s, 3H), 3.05 (dd, *J* = 7.4, 2.8 Hz), 2.03 (td, 1H), 1.38 (s, 9H), 0.78 (d, *J* =7.0, 3H), 0.75 (d, *J* = 7.0, 3H), 2.06 (m, 1H), 1.35 (s, 9H), 0.84 (d, *J* = 6.9, 3H), 0.81 (d, *J* = 6.9, 3H)

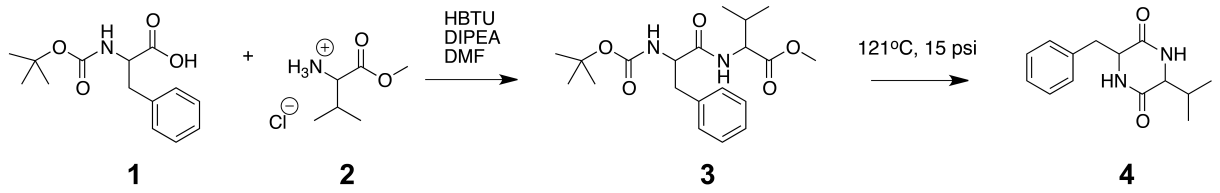


Figure 4.11 Synthetic scheme of val-phe diketopiperazines
(1) Boc-L/D-phenylalanine **(2)** D/L-valine-methyl ester hydrochloride **(3)** methyl 2-(2-((*tert*-butoxycarbonyl)amino)-3-phenylpropanamido)-3-methylbutanoate **(4)** D/L-val-D/L-phe-diketopiperazine. Abbreviations: HBTU, 3-[Bis(dimethylamino)methyliumyl]-3*H*-benzotriazol-1-oxide hexafluorophosphate; DIPEA, *N,N*-diisopropylethylamine; DMF, *N,N*-dimethylformamide

Without further purification, the dipeptide was diluted with water (2 ml) and autoclaved for 3 hours (121°C, 15 psi). Reaction was dried under vacuum and purified by HPLC using the same procedure described for purification of lumichrome and diketopiperazines. Structural information for L-val-L-phe DKP: ^{1}H NMR (400 MHz, CD_3OD) δ 7.33-7.28 (m, 2H), 7.27-7.23, (m, 3H), 4.34 (td, *J* = 5.2, 1.6, 1H), 3.67 (dd, *J* = 4.6, 1.6, 1H), 3.26 (dd, *J* = 13.8, 5.4, 1H), 3.07 (dd, *J* = 13.8, 4.7, 1H), 1.68 (m, 1H), 0.83 (*J* =7.1, d, 3H), 0.47 (d, *J* =6.8, 3H). Structural information for D-val-L-phe DKP: ^{1}H NMR (400 MHz, CD_3OD) δ 7.34-7.29 (m, 3H), 7.28-7.24 (m, 2H), 4.31 (t, *J* = 4.3, 1H), 3.29 (dd, *J* =13.8, 4.1, 1H), 3.05 (dd, *J* =13.8, 4.6, 1H), 2.97 (dd, *J* = 2.9, 0.9, 1H), 2.17 (m, 1H), 0.93 (d, *J* =7.2, 3H), 0.87, (d, *J* = 6.8, 3H).

Sporulation assays

Non-farmer (QS9) and farmer (QS11) sporulation assays with Bx11 extracts were performed as previously described.[20-22]

Genome Isolation and annotation

Bx11 was grown in liquid LB shaking at 30°C for 48 hours. The cells were pelleted by centrifugation and gDNA was isolated, sequenced and assembled using previously described methods.[59] The Bx11 genome was annotated using Rapid Annotation using Subsystem Technology (RAST) and the biosynthetic gene cluster analysis was performed using the antiSMASH2 software.

Cosmid Library Construction

The cosmid library was constructed using the reagents and protocols provided in the pWEB™ cosmid cloning kit (Epicentre PC8805). Clones were picked into sterile 96-well plates

containing 200 μL LB and grown overnight, shaking at 30 °C. Glycerol was added to each well (to make a final stock of 20% glycerol in LB) and plates were stored at –80°C.

High-throughput plaque assay in non-farmer *D. discoideum*

Cosmid library glycerol stocks were thawed and shaken lightly for ~15 minutes. In the meantime, non farmer (QS9) stocks were prepared from glycerol stocks and brought to a concentration of 10,000 spores/μL (in KK2 buffer). 5 μL of QS9 dilution were added to each well of a sterile v-bottom PCR plate. Note: it is important to keep mixing the QS9 spore dilution because these can form clumps and sink to the bottom of the tube.

Using a multichannel pipette, 5 μL from each well of the thawed 96-well cosmid library plate were added to the PCR plate containing QS9 and mixed well. Using an 8-channel with every other tip removed, 1 μL of the dicty-bacteria mixture were transferred to a 150 x 15 mm SM/5 agar plate supplemented with ampicillin (100 μg/ml). This is repeated for each column of the plate, ensuring that there is enough space between spots for growth and development of the bacteria and amoeba. This should allow plating of 48 spots/plate or two agar plates for each 96-well cosmid library plate. Note: the type of antibiotic added to the plate will be dependent on the selection required for the cosmid library. Prior to screening the amoeba should be tested against a dose curve of the antibiotic to ensure that it is not toxic at concentration required for selection.

Spots are allowed to dry completely (~10 minutes) and the plates were left at room temperature for two to three days. Note was taken of bacteria that did not grow or was delayed – these slow-growers can result in false positives. Note: *D. discoideum* can be quite sensitive to temperature, so spikes or dips in temperature may affect the time frame of the assay.

Cloning of pyrrolnitrin gene cluster (PRN)

Isolated gDNA from *Pseudomonas fluorescens* (Pf-5) (Taxonomy ID: 220664) and performed PCR using the primer pair:

F1: ATTAGAGGCCTGTGTCCATGAAAAGGG

R1: TACGCCGTGATCTCCGTGACG

Products were confirmed by enzymatic digest using HindIII or KpnI. 3' A was added to purified PCR product and ligated into pGEMT vector. Ligation product was transformed into DH10β and production of pyrrolnitrin was validated by LCMS.

4.7 References

1. Williams, R. S. B. *et al.* Towards a molecular understanding of human diseases using Dictyostelium discoideum. *Trends Mol Med* **12,** 415–424 (2006).

2. Li, S. I. & Purugganan, M. D. The cooperative amoeba: Dictyostelium as a model for social evolution. *Trends Genet.* **27,** 48–54 (2011).

3. Kikuchi, H. *et al.* Furanodictine A and B: Amino Sugar Analogues Produced by Cellular Slime Mold DictyosteliumdiscoideumShowing Neuronal Differentiation Activity. *J. Org. Chem.* **66,** 6982–6987 (2001).

4. Todoriki, M. & Urabe, I. Induced symbiosis: distinctive Escherichia coli-Dictyostelium discoideum transferable co-cultures on agar. *Symbiosis (Rehovot)* **42,** 135 (2006).

5. Kuzdzal-Fick, J. J., Fox, S. A., Strassmann, J. E. & Queller, D. C. High relatedness is necessary and sufficient to maintain multicellularity in Dictyostelium. *Science* **334,** 1548–1551 (2011).

6. Reid, C. R., Latty, T., Dussutour, A. & Beekman, M. Slime mold uses an externalized spatial 'memory' to navigate in complex environments. *Proc. Natl. Acad. Sci. USA* **109,** 17490–17494 (2012).

7. Hsu, D.-W., Chubb, J. R., Muramoto, T., Pears, C. J. & Mahadevan, L. C. Dynamic acetylation of lysine-4-trimethylated histone H3 and H3 variant biology in a simple multicellular eukaryote. *Nucleic Acids Res.* **40,** 7247–7256 (2012).

8. Gaudet, P., Fey, P. & Chisholm, R. Dictyostelium discoideum: The Social Ameba. *Cold Spring Harbor Protocols* **3,** (2008).

9. Chisholm, R. L. & Firtel, R. A. Insights into morphogenesis from a simple developmental system. *Nat. Rev. Mol. Cell Biol.* **5,** 531–541 (2004).

10. Berzat, A. & Hall, A. Cellular responses to extracellular guidance cues. *The EMBO Journal* **29,** 2734–2745 (2010).

11. Cotter, D. A. *et al.* Patterning of development in Dictyostelium discoideum: factors regulating growth, differentiation, spore dormancy, and germination. *Biochem. Cell Biol.* **70,** 892–919 (1992).

12. Neumann, C. S., Walsh, C. T. & Kay, R. R. A flavin-dependent halogenase catalyzes the chlorination step in the biosynthesis of Dictyostelium differentiation-inducing factor 1. *Proc. Natl. Acad. Sci. USA* **107,** 5798–5803 (2010).

13. Kikuchi, H. *et al.* Anti-leukemic activities of Dictyostelium secondary metabolites: A novel aromatic metabolite, 4-methyl-5-n-pentylbenzene-1,3-diol, isolated from Dictyostelium mucoroides suppresses cell growth in human leukemia K562 and HL-60 cells. *Life Sciences* **80,** 160–165 (2006).

14. Kikuchi, H. *et al.* Biological activities of novel derivatives of DIF-1 isolated from Dictyostelium. *Biochem. Biophys. Res. Commun.* **377,** 1012–1017 (2008).

15. Kikuchi, H., Ishiko, S., Nakamura, K., Kubohara, Y. & Oshima, Y. Novel prenylated and geranylated aromatic compounds isolated from Polysphondylium cellular slime molds. *Tetrahedron* **66,** 6000–6007 (2010).

16. Kikuchi, H. *et al.* Dihydrodictyopyrones A and C: new members of dictyopyrone family isolated from Dictyostelium cellular slime molds. *Tet. Lett.* **48,** 5905–5909 (2007).

17. Kuzdzal-Fick, J. J., Fox, S. A., Strassmann, J. E. & Queller, D. C. High relatedness is necessary and sufficient to maintain multicellularity in Dictyostelium. *Science* **334,** 1548–1551 (2011).

18. Hirose, S., Benabentos, R., Ho, H. I., Kuspa, A. & Shaulsky, G. Self-Recognition in Social Amoebae Is Mediated by Allelic Pairs of Tiger Genes. *Science* **333,** 467–470 (2011).

19. Fortunato, A., Strassmann, J. E., Santorelli, L. & Queller, D. C. Co-occurrence in nature of different clones of the social amoeba, Dictyostelium discoideum. *Mol. Ecol.* **12,** 1031–1038 (2003).

20. Brock, D. A., Douglas, T. E., Queller, D. C. & Strassmann, J. E. Primitive agriculture in a social amoeba. *Nature* **469,** 393–396 (2011).

21. Brock, D. A., Read, S., Bozhchenko, A., Queller, D. C. & Strassmann, J. E. Social amoeba farmers carry defensive symbionts to protect and privatize their crops. *Nat. Comm.* **4,** 1–7 (1AD).

22. Stallforth, P. *et al.* A bacterial symbiont is converted from an inedible producer of beneficial molecules into food by a single mutation in the gacA gene. *Proc. Natl. Acad. Sci. USA* **110,** 14528–14533 (2013).

23. Rajamani, S. *et al.* The vitamin riboflavin and its derivative lumichrome activate the LasR bacterial quorum-sensing receptor. *Mol. Plant Microbe Interact.* **21,** 1184–1192 (2008).

24. Venturi, V. & Subramoni, S. Future research trends in the major chemical language of bacteria. *HFSP J* **3,** 105–116 (2009).

25. Phillips, D. A. *et al.* Identification of lumichrome as a sinorhizobium enhancer of alfalfa root respiration and shoot growth. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 12275–12280 (1999).

26. Dakora, F. D., Matiru, V. N. & Kanu, A. S. Rhizosphere ecology of lumichrome and riboflavin, two bacterial signal molecules eliciting developmental changes in plants. *Front. Plant Sci.* **6,** 1505–11 (2015).

27. Borthwick, A. D. 2,5-Diketopiperazines: Synthesis, Reactions, Medicinal Chemistry, and Bioactive Natural Products. *Chem. Rev.* (2012). doi:10.1021/cr200398y

28. Kumar, S. N., Nambisan, B., Sundaresan, A., Mohandas, C. & Anto, R. J. Isolation and identification of antimicrobial secondary metabolites from Bacillus cereus associated with a rhabditid entomopathogenic nematode. *Ann Microbiol* **64,** 209–218 (2014).

29. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for 'clickable' amino acids. *Angew. Chem. Int. Ed. Engl.* **53,** 10105–10108 (2014).

30. Ström, K., Sjögren, J., Broberg, A. & Schnürer, J. Lactobacillus plantarum MiLAB 393 produces the antifungal cyclic dipeptides cyclo(L-Phe-L-Pro) and cyclo(L-Phe-trans-4-OH-L-Pro) and 3-phenyllactic acid. *Appl. Environ. Microbiol.* **68,** 4322–4327 (2002).

31. Lazos, O. *et al.* Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. *Chemistry & Biology* **17,** 160–173 (2010).

32. Song, F. *et al.* Brevianamides with antitubercular potential from a marine-derived isolate of Aspergillus versicolor. *Org. Lett.* **14,** 4770–4773 (2012).

33. Stark, T. & Hofmann, T. Structures, sensory activity, and dose/response functions of 2,5-diketopiperazines in roasted cocoa nibs (Theobroma cacao). *J. Agric. Food Chem.* **53,** 7222–7231 (2005).

34. Pérez-Picaso, L., Olivo, H. F., Argotte-Ramos, R., Rodríguez-Gutiérrez, M. & Rios, M. Y. Linear and cyclic dipeptides with antimalarial activity. *Bioorg. Med. Chem. Lett.* **22,** 7048–7051 (2012).

35. Li, H. *et al.* Bioactive Cyclic Dipeptides from a Marine Sponge-Associated Bacterium, Psychrobacter sp. *Biomolecules and Therapeutics* **16,** 356–363 (2008).

36. Alvarez, M. E. *et al.* Phevalin, a new calpain inhibitor, from a Streptomyces sp. *J. Antibiot.* **48,** 1165–1167 (1995).

37. Secor, P. R. *et al.* Phevalin (aureusimine B)Production by Staphylococcus aureus Biofilm and Impacts on Human Keratinocyte Gene Expression. *PLoS ONE* **7,** e40973 (2012).

38. Parnell, J. J. *et al.* Coping with polychlorinated biphenyl (PCB) toxicity: Physiological and genome-wide responses of Burkholderia xenovorans LB400 to PCB-mediated stress. *Appl. Environ. Microbiol.* **72,** 6607–6614 (2006).

39. Chain, P. S. G. *et al.* Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 15280–15287 (2006).

40. Franke, J., Ishida, K. & Hertweck, C. Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic Burkholderia species. *Angew. Chem. Int. Ed. Engl.* **51,** 11611–11615 (2012).

41. Biggins, J. B., Gleber, C. D. & Brady, S. F. Acyldepsipeptide HDAC inhibitor production induced in Burkholderia thailandensis. *Org. Lett.* **13,** 1536–1539 (2011).

42. Ishida, K., Lincke, T., Behnken, S. & Hertweck, C. Induced biosynthesis of cryptic polyketide metabolites in a Burkholderia thailandensis quorum sensing mutant. *J. Am. Chem. Soc.* **132,** 13966–13968 (2010).

43. Losada, L. *et al.* Continuing evolution of Burkholderia mallei through genome reduction and large-scale rearrangements. *Genome Biol Evol* **2,** 102–116 (2010).

44. Moran, N. A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108,** 583–586 (2002).

45. D'Cruze, T. *et al.* Role for the Burkholderia pseudomallei type three secretion system cluster 1 bpscN gene in virulence. *Infect. Immun.* **79,** 3659–3664 (2011).

46. Tomich, M., Planet, P. J. & Figurski, D. H. The tad locus: postcards from the widespread colonization island. *Nat. Rev. Microbiol.* **5,** 363–375 (2007).

47. Costa, T. R. D. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13,** 343–359 (2015).

48. Burtnick, M. N. *et al.* The Cluster 1 Type VI Secretion System Is a Major Virulence Determinant in Burkholderia pseudomallei. *Infect. Immun.* **79,** 1512–1525 (2011).

49. Sun, G. W. & Gan, Y.-H. Unraveling type III secretion systems in the highly versatile Burkholderia pseudomallei. *Trends Microbiol.* **18,** 561–568 (2010).

50. DiSalvo, S. *et al.* Burkholderia bacteria infectiously induce the proto-farming symbiosis of Dictyostelium amoebae and food bacteria. *Proc. Natl. Acad. Sci. USA* **112,** E5029–37 (2015).

51. Niepold, F., Anderson, D. & Mills, D. Cloning determinants of pathogenesis from Pseudomonas syringae pathovar syringae. *Proc. Natl. Acad. Sci. U.S.A.* **82,** 406–410 (1985).

52. Brady, S. F. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat Protoc* **2,** 1297–1305 (2007).

53. Brady, S. F., Simmons, L., Kim, J. H. & Schmidt, E. W. Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat Prod Rep* **26,** 1488–1503 (2009).

54. Brady, S. F., Chao, C. J. & Clardy, J. New Natural Product Families from an Environmental DNA (eDNA) Gene Cluster. *J. Am. Chem. Soc.* **124,** 9968–9969 (2002).

55. Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat Chem Biol* **12,** 208–214 (2016).

56. Wang, X. & Wood, T. K. Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Appl. Environ. Microbiol.* **77,** 5577–5583 (2011).

57. Lobato-Márquez, D., Díaz-Orejas, R. & García-Del Portillo, F. Toxin-antitoxins and bacterial virulence. *FEMS Microbiol. Rev.* fuw022 (2016). doi:10.1093/femsre/fuw022

58. Mattison, K., Wilbur, J. S., So, M. & Brennan, R. G. Structure of FitAB from Neisseria gonorrhoeae Bound to DNA Reveals a Tetramer of Toxin-Antitoxin Heterodimers Containing Pin Domains and Ribbon-Helix-Helix Motifs. *J. Biol. Chem.* **281,** 37942–37951 (2006).

59. Sit, C. S. *et al.* Variable genetic architectures produce virtually identical molecules in bacterial symbionts of fungus-growing ants. *Proc. Natl. Acad. Sci. USA* **112,** 13150–13154 (2015).

Chapter 5: Future directions and concluding remarks

The research described in the previous chapters focuses on specific symbiotic interactions between organisms of which most people have never heard. Yet this does not negate the broader lessons that can be gleaned from teasing apart the dynamics of these complex relationships. Evolution often comes to the similar solutions to similar problems – so each symbiotic relationship we truly understand can help generate the tools to recognize patterns in widely diverse organismal interactions; and the characterization of metabolites essential to these interactions allows us to form hypotheses about the potential roles of these molecules in other ecological systems.

Choanoflagellates

In Chapters 2 and 3 we investigate the role of bacterial lipid signals in the development of a multicellular phenotype in choanoflagellates. Not only does this research aid in establishing choanoflagellates as an important model organism for understanding the emergence of multicellularity in the animal lineage, it also implicates bacteria as an essential component of this major evolutionary transition.

As mentioned in Chapter 1, the lipids regulating rosette formation in choanoflagellates share structural similarities to mammalian lipids, like sphingosine-1-phosphate, or ceramide, that bind to human g-protein couple receptors, leading to signaling cascades that have numerous biological affects. Additionally, because choanoflagellates are known to possess numerous genes that are homologous to those regulating processes related to multicellularity in humans, it is highly possible that further elucidation of the mechanism of rosette induction in choanoflagellates may uncover analogous signaling components in animal cells. These similarities present the possibility that rosette-inducing lipids may have binding partners that are expressed in animal cells. We plan to further explore this potential bridge between multicellularity in choanoflagellates and homologous pathways in metazoa through the

investigation into members of the human microbiota that can induce rosette formation, and have already begun exploring the lipid repertoire of *Akkermansia muciniphila*, a member of the human gut that can induce rosette formation in choanoflagellates (unpublished). Ultimately we hope that the identification of rosette inducing molecules produced in the human body will unveil relics from pathways involved in choanoflagellate multicellularity, and lead to the discovery of microbiota-produced molecules involved in human-bacterial signaling.

Similarly, we plan to expand our studies with the inhibitor of rosette formation, IOR-1. The results presented in Chapters 2 and 3 point to the likelihood that IOR-1 is a competitive inhibitor of the target of RIF-2. Because this molecule is structurally simpler than the RIFs and is now accessible by a straightforward and efficient synthesis (Chapter 2), it presents us with a possible tool for pursuing the mechanism of rosette induction. Future studies with IOR-1 will involve the synthesis of derivatives that can be tagged for imaging experiments on IOR-1 localization and choanoflagellate lysate pull-downs for target identification.

Dictyostelium discoideum

Much like the interaction between choanoflagellates and rosette inducing bacteria, the relationship between *D. discoideum* and their bacterial co-isolates has only recently been discovered. This presents the opportunity to expand our perception of how bacteria may use secondary metabolism and/or toxic factors to interact with an amoebal host. Chapter 4 describes the multiple approaches we took to understand how the non-food bacterium, Bx11, contributes to farming behavior in *D. discoideum.* While our search for small molecules that mediate “fitness” in the farmer and non-farmer amoeba led to some intriguing results, our most significant findings were the result of screening a Bx11 cosmid library for toxicity to *D. discoideum*. As detailed in Chapter 4, some of the hits from this screen, like the toxin-antitoxin pair, and the Type VI secretion system, were clearly associated with pathogenic activity, and we

will start by validating these hits through gene deletion and heterologous expression. We also hope to explore the hits that did not have such clear functions, as these could potentially reveal new amoebicidal agents.

Finally, while we developed the QS9 plaque assay screen in the context of a specific ecological interaction, we hope to make this assay a generalizable method to screen bacterial cosmid libraries for amoebicidal activity. While cosmid libraries have the potential to provide a significant amount of information about genes related to specific bioactivities, there are a limited number of ways to efficiently screen them; live bacteria can be a confounding factor in many assays (especially in eukaryotic systems), and growth and extraction of each library clone requires huge amount of time and work. Our assay, which takes advantage of the fact that *D. discoideum* feeds on bacteria, provides an efficient method for testing live bacteria for toxicity in a eukaryotic organism.

Appendix 1: Supplementary materials for Chapter 2

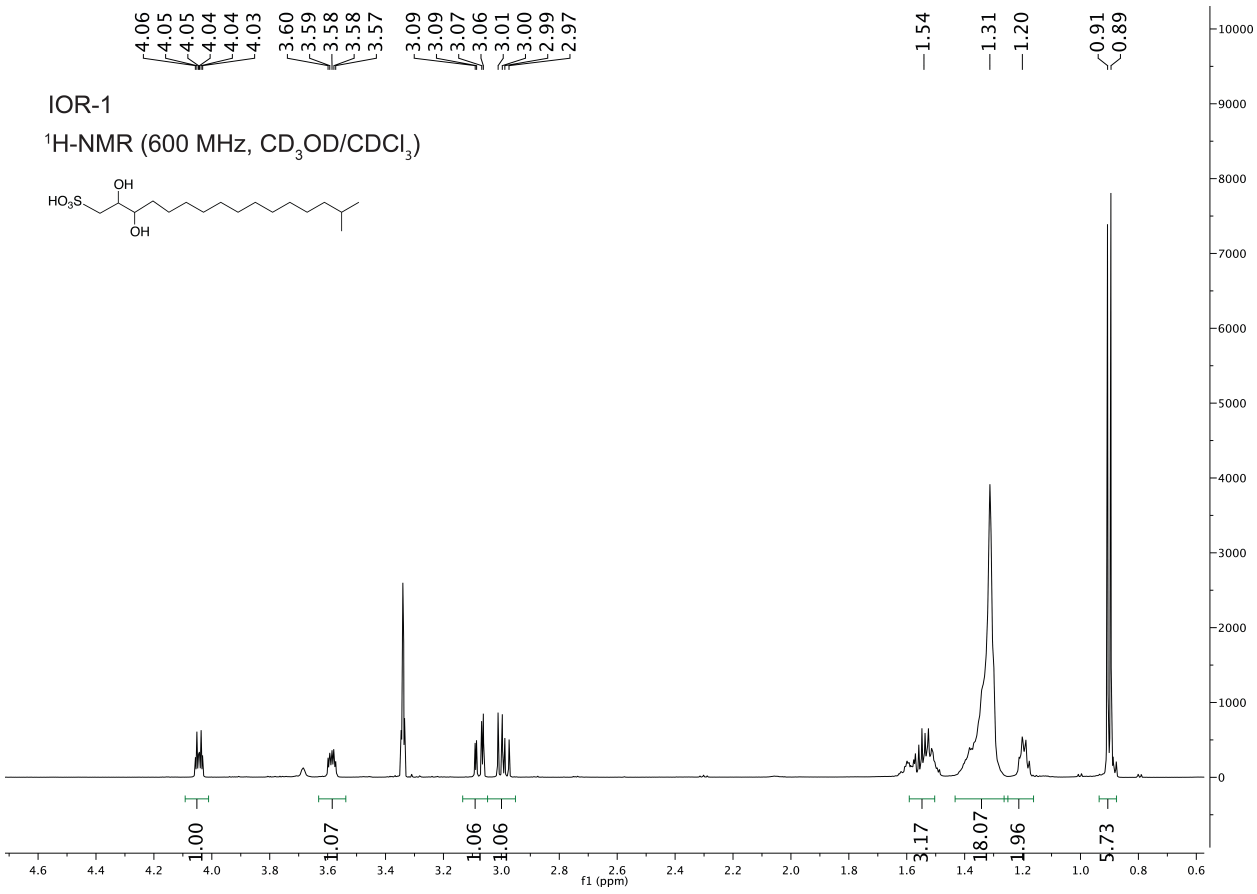


Figure S1.1 IOR-1 (isolated) ^{1}H-NMR spectrum (600 MHz, $CD_3OD/CDCl_3$)

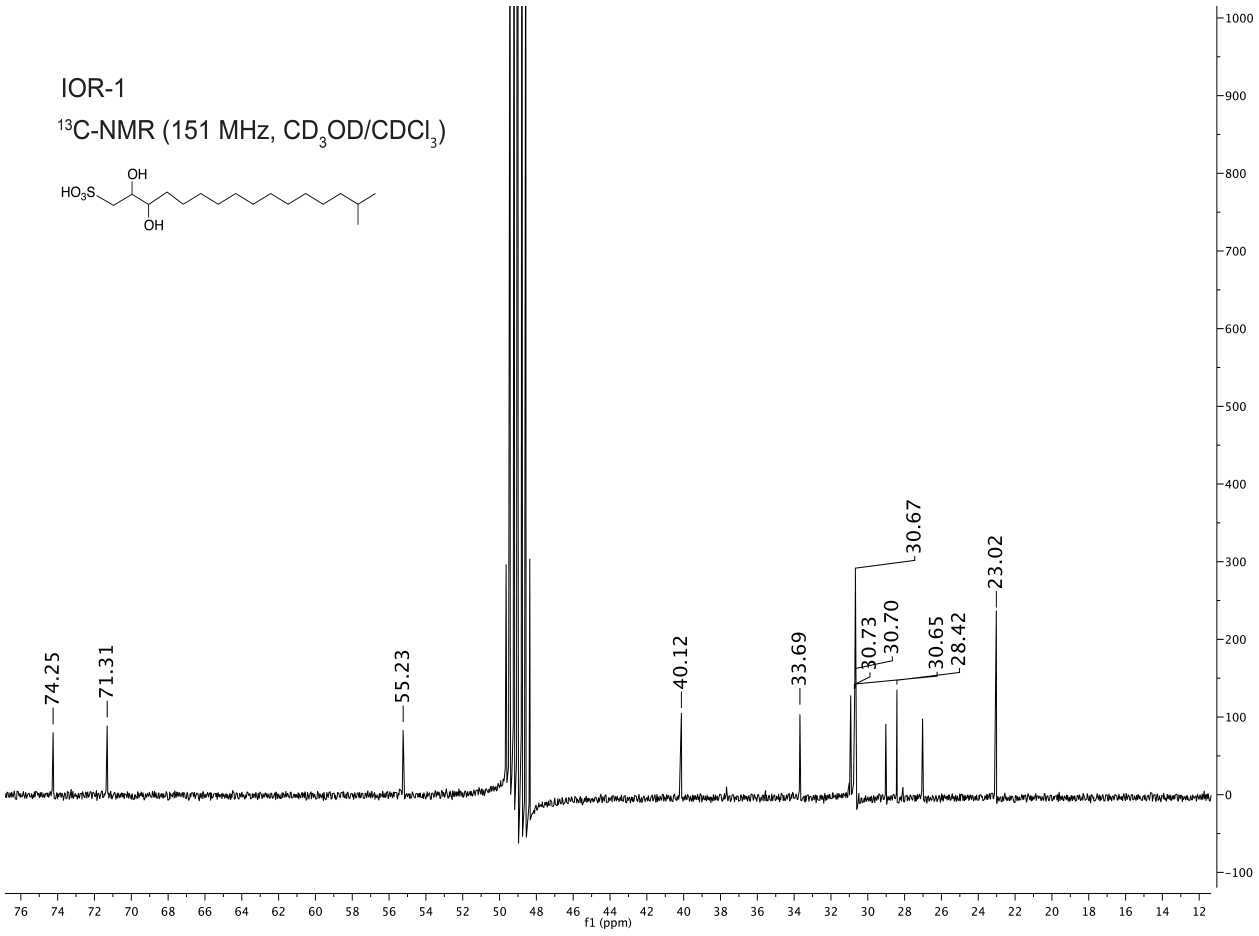


Figure S1.2 IOR-1 (isolated) ^{13}C-NMR spectrum (151 MHz, $CD_3OD/CDCl_3$)

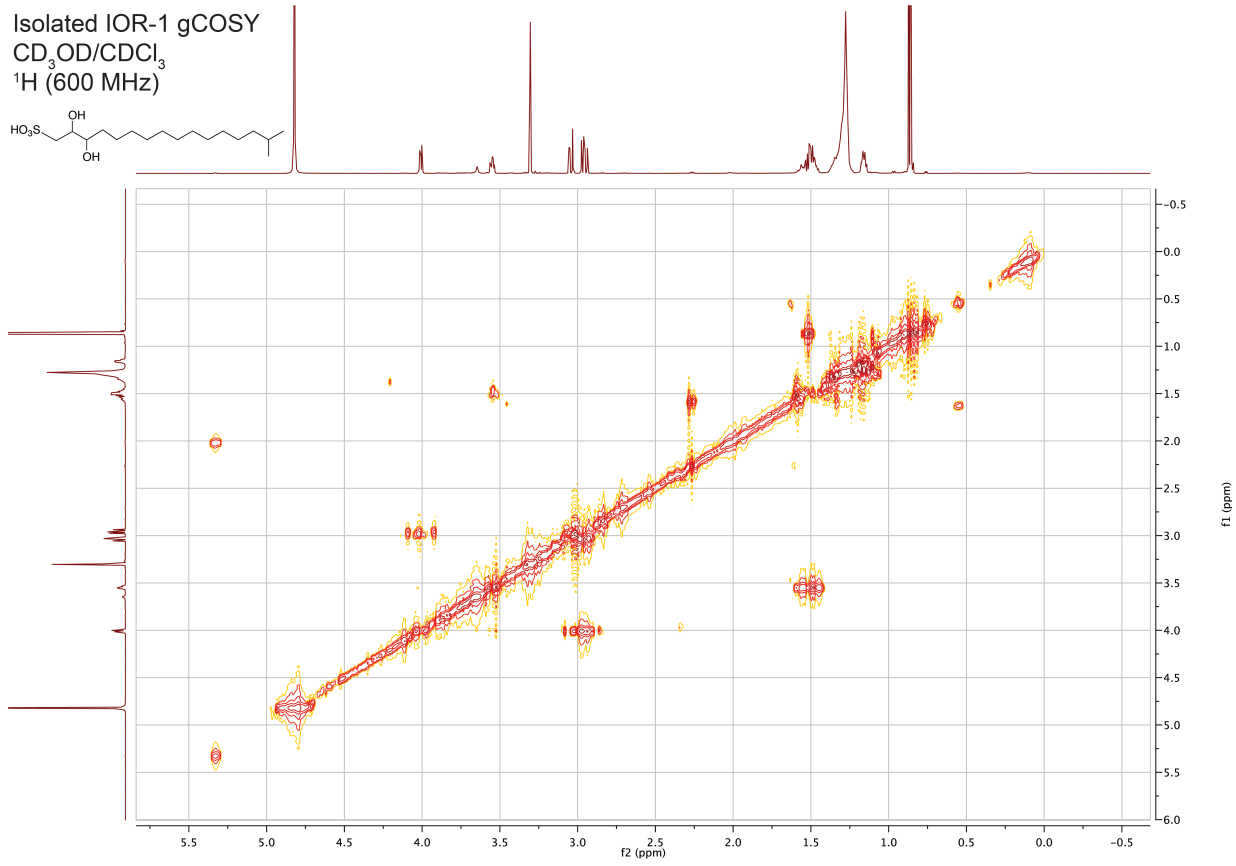


Figure S1.3 IOR-1 (isolated) gCOSY spectrum (600 MHz, $CD_3OD/CDCl_3$)

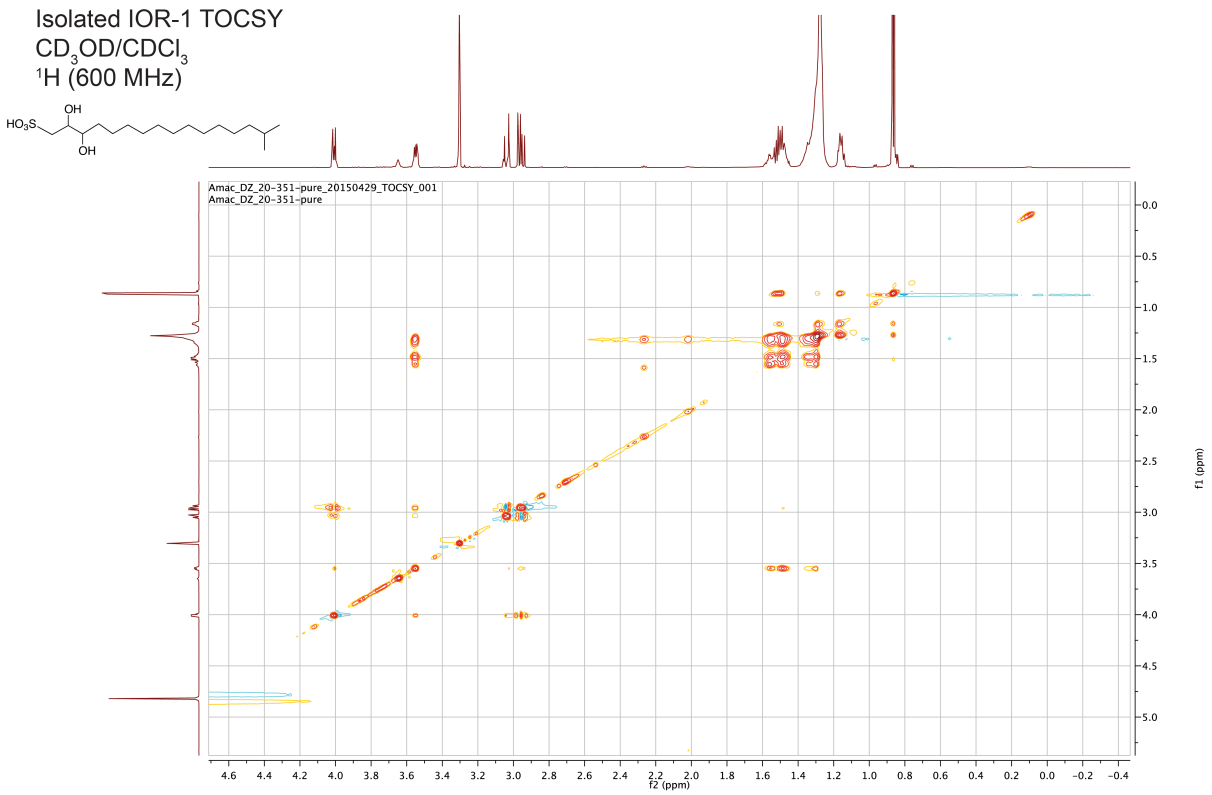


Figure S1.4 IOR-1 (isolated) TOCSY spectrum (600 MHz, $CD_3OD/CDCl_3$)

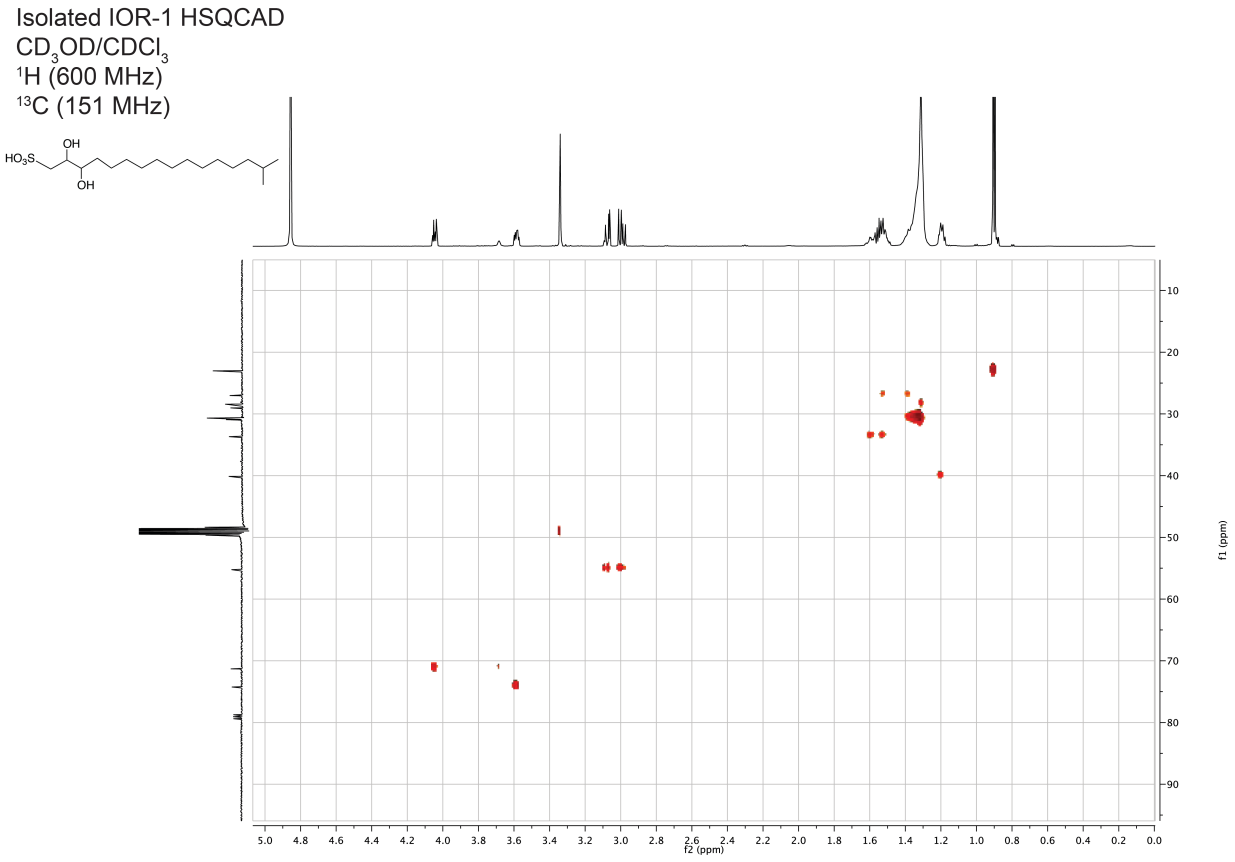


Figure S1.5 IOR-1 (isolated) HSQCAD ((600/151 MHz, $CD_3OD/CDCl_3$)

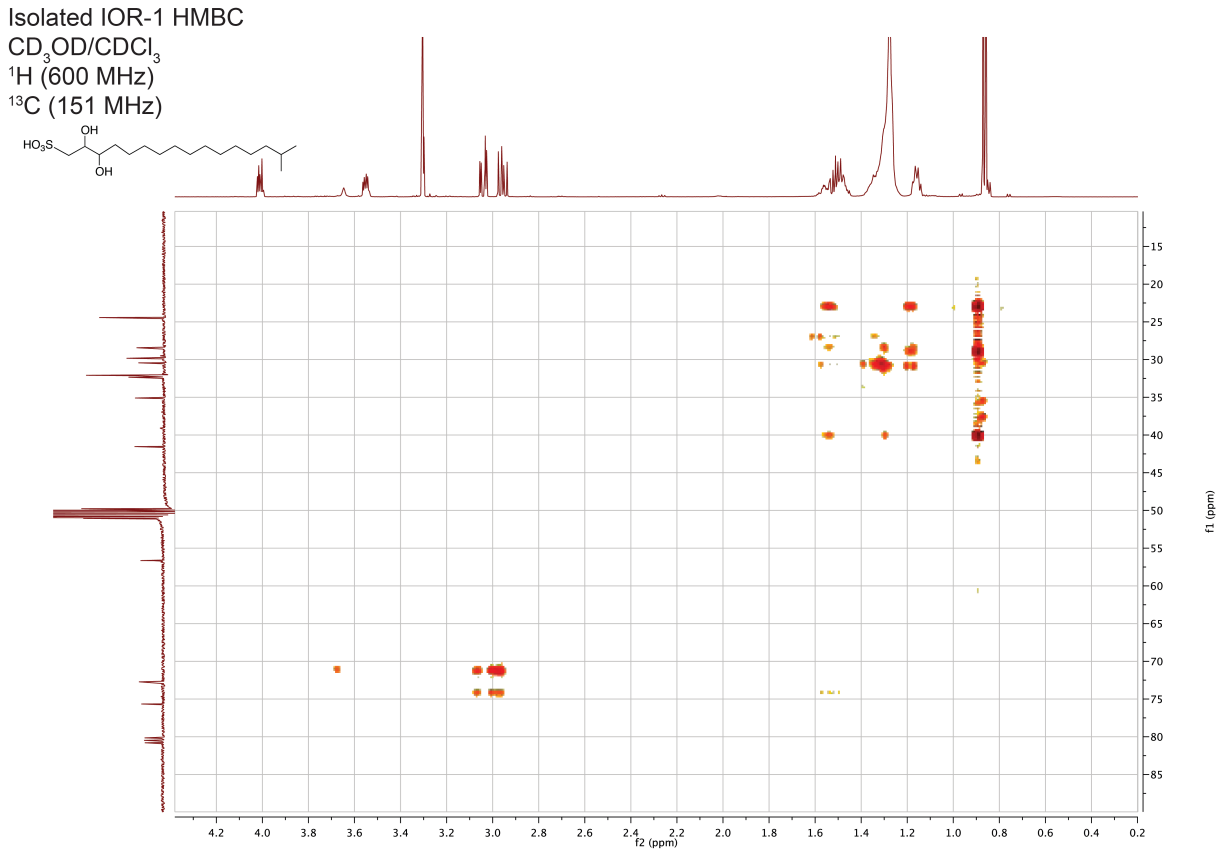


Figure S1.6 IOR-1 (isolated) HMBC spectrum (600 MHz, $CD_3OD/CDCl_3$)

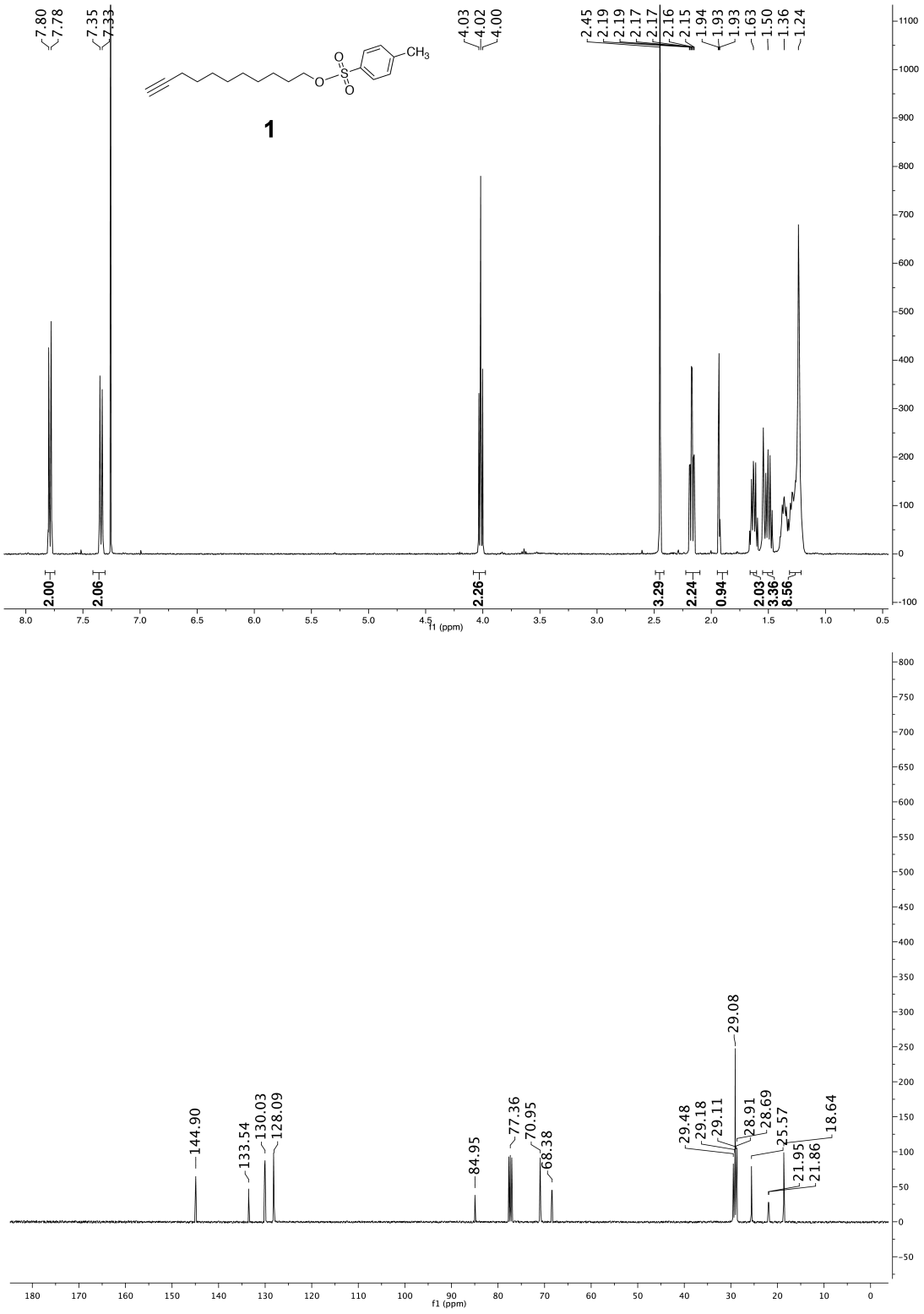


Figure S1.7 NMR spectra for synthetic compound 1
(Top) ^{1}H-NMR (400 MHZ, $CDCl_3$) **(Bottom)** ^{13}C-NMR (100 MHz, $CDCl_3$)

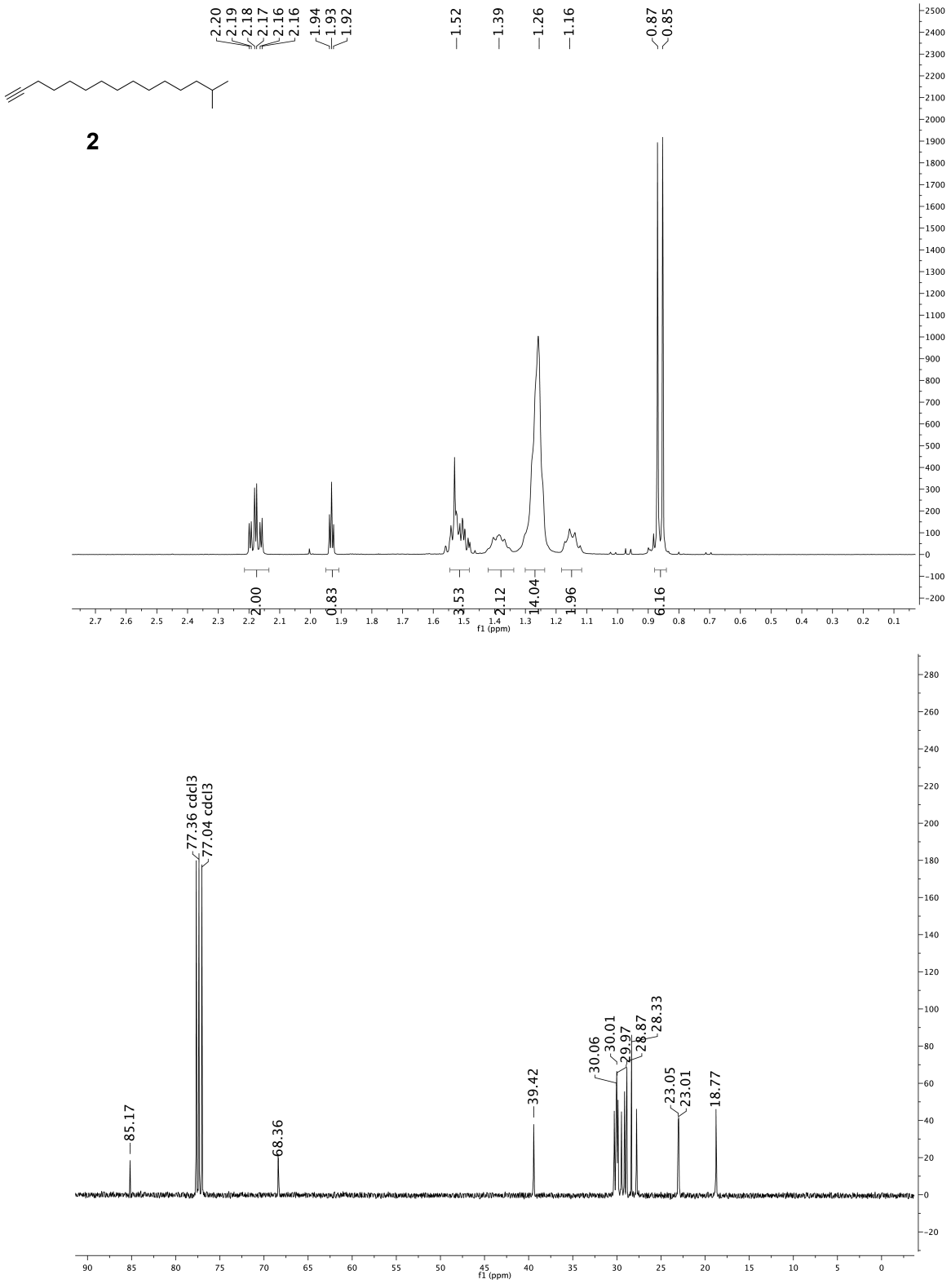


Figure S1.8 NMR spectra for synthetic compound 2
(Top) ^{1}H-NMR (400 MHZ, $CDCl_3$) **(Bottom)** ^{13}C-NMR (100 MHz, $CDCl_3$)

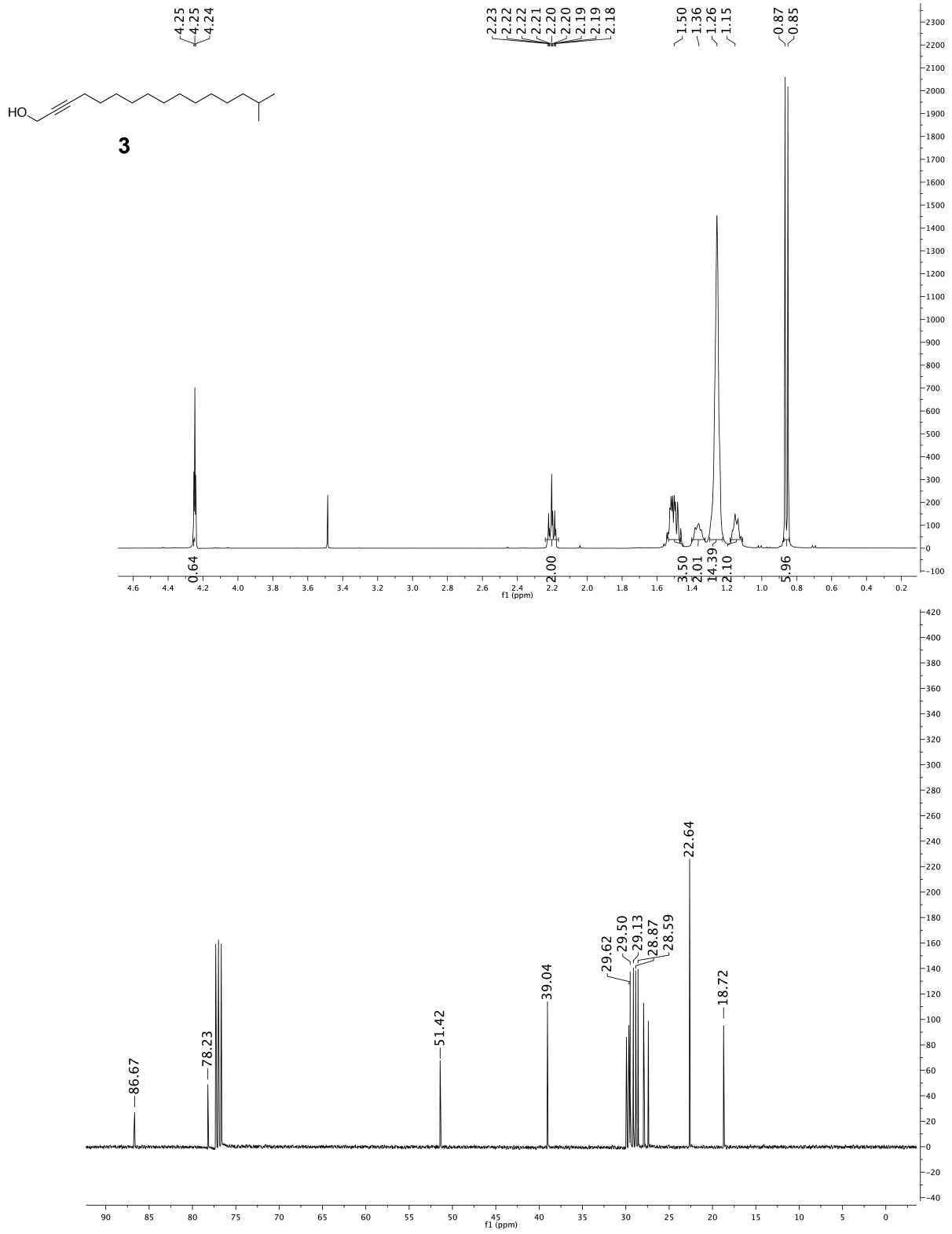


Figure S1.9 NMR spectra for synthetic compound 3
(Top) ^{1}H-NMR (400 MHZ, $CDCl_3$) **(Bottom)** ^{13}C-NMR (100 MHz, $CDCl_3$)

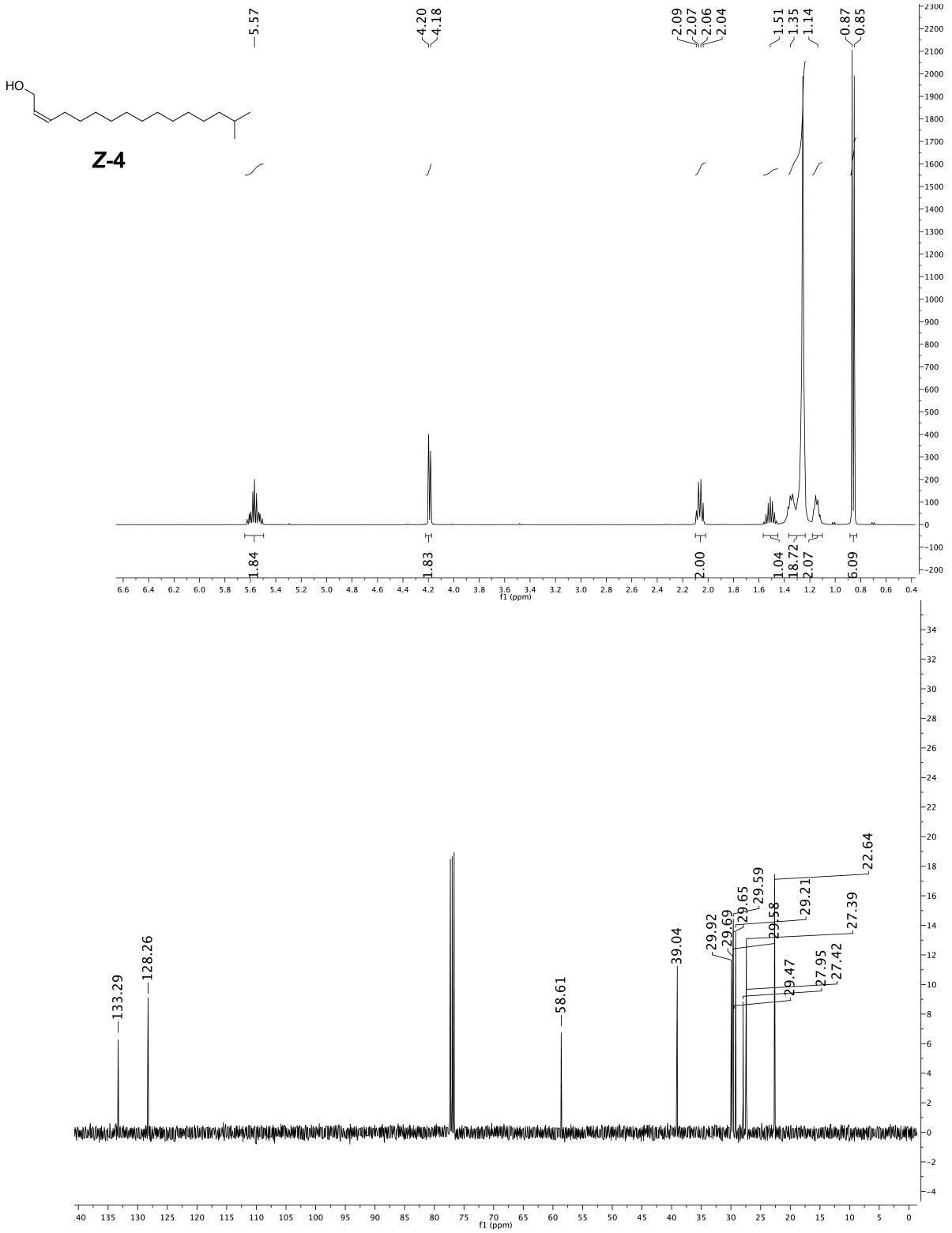


Figure S1.10 NMR spectra for synthetic compound *Z*-4
(Top) ^{1}H-NMR (400 MHZ, $CDCl_3$) **(Bottom)** ^{13}C-NMR (100 MHz, $CDCl_3$)

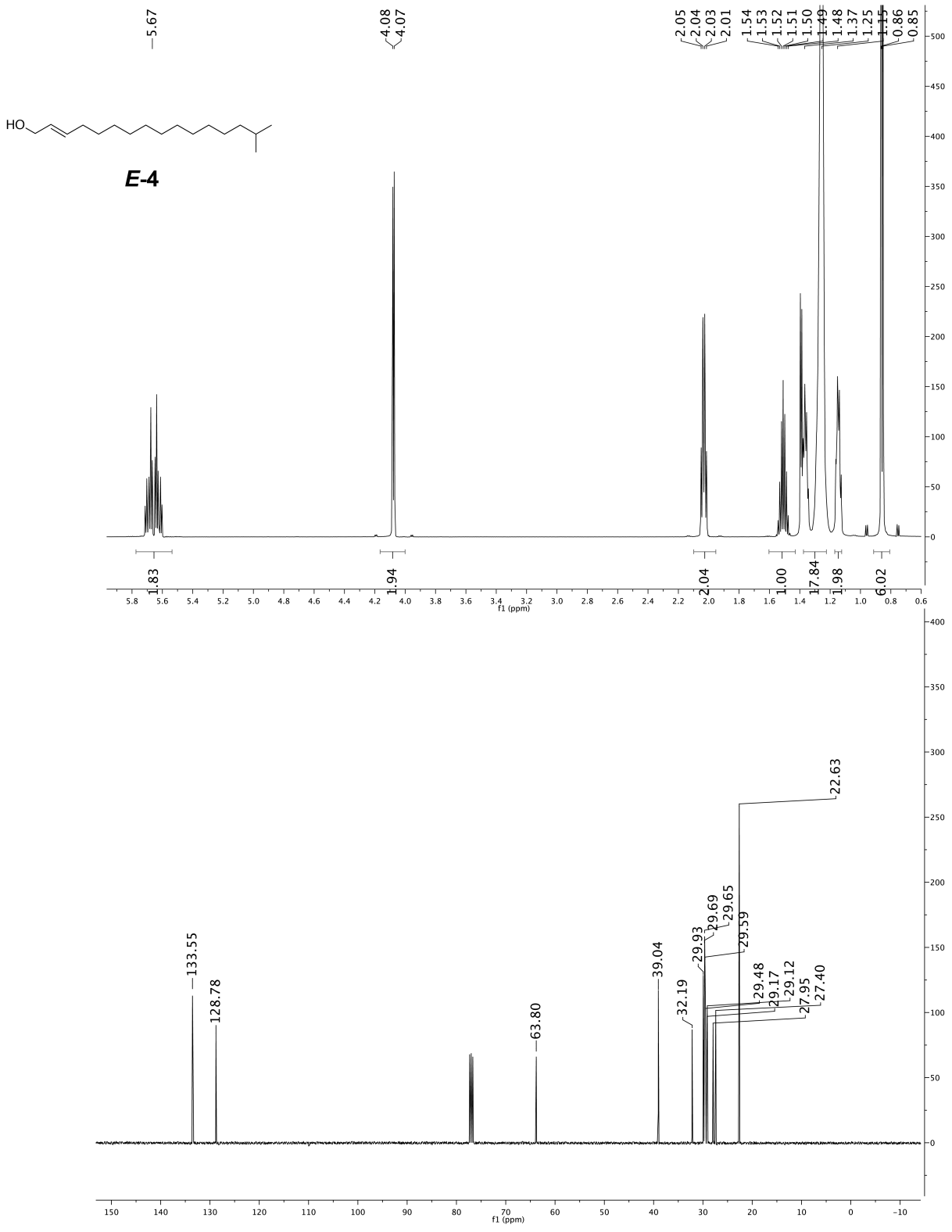


Figure S1.11 NMR spectra for synthetic compound *E*-4
(Top) ^{1}H-NMR (400 MHZ, $CDCl_3$) **(Bottom)** ^{13}C-NMR (100 MHz, $CDCl_3$)

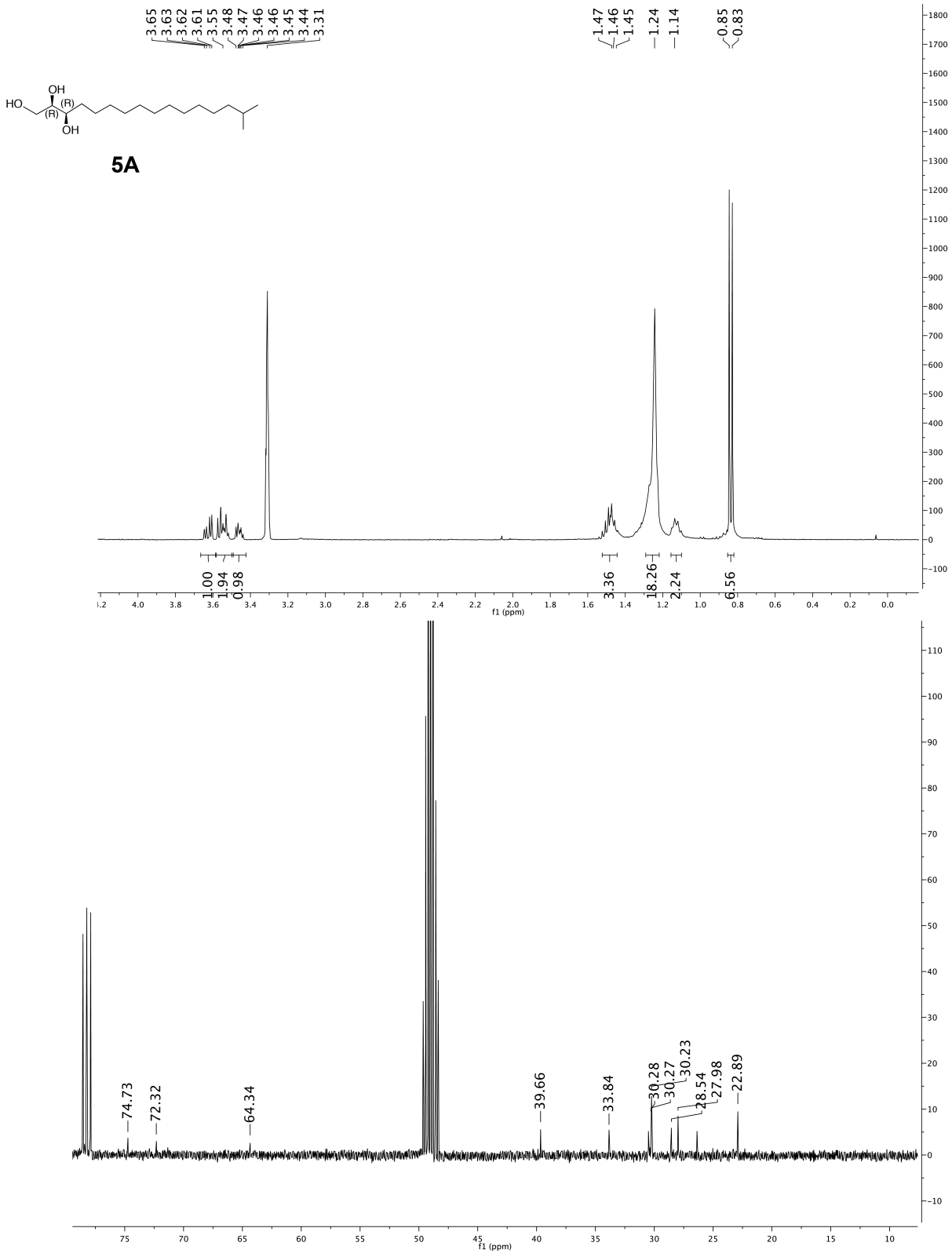


Figure S1.12 NMR spectra for synthetic compound 5A
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

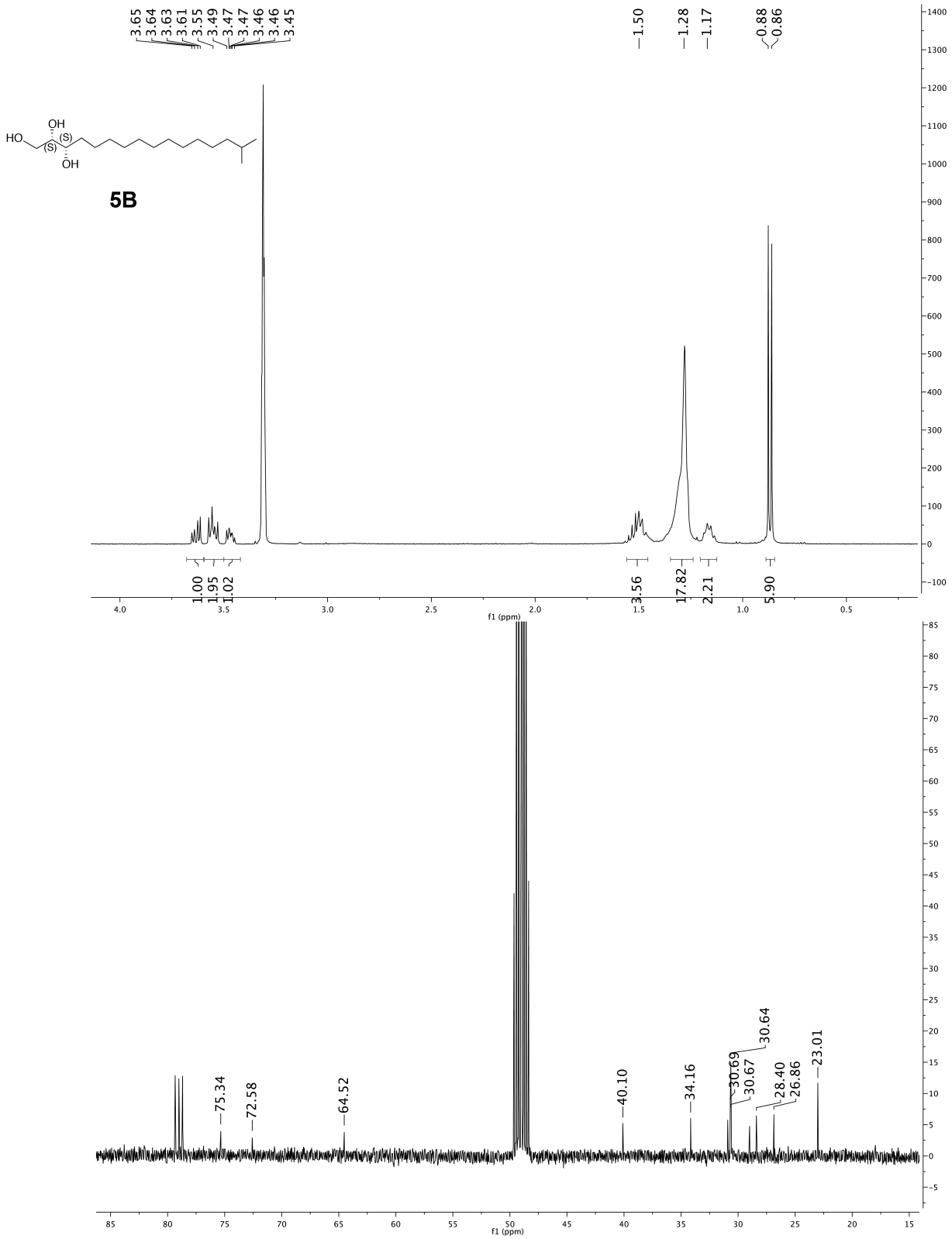


Figure S1.13 NMR spectra for synthetic compound 5B
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

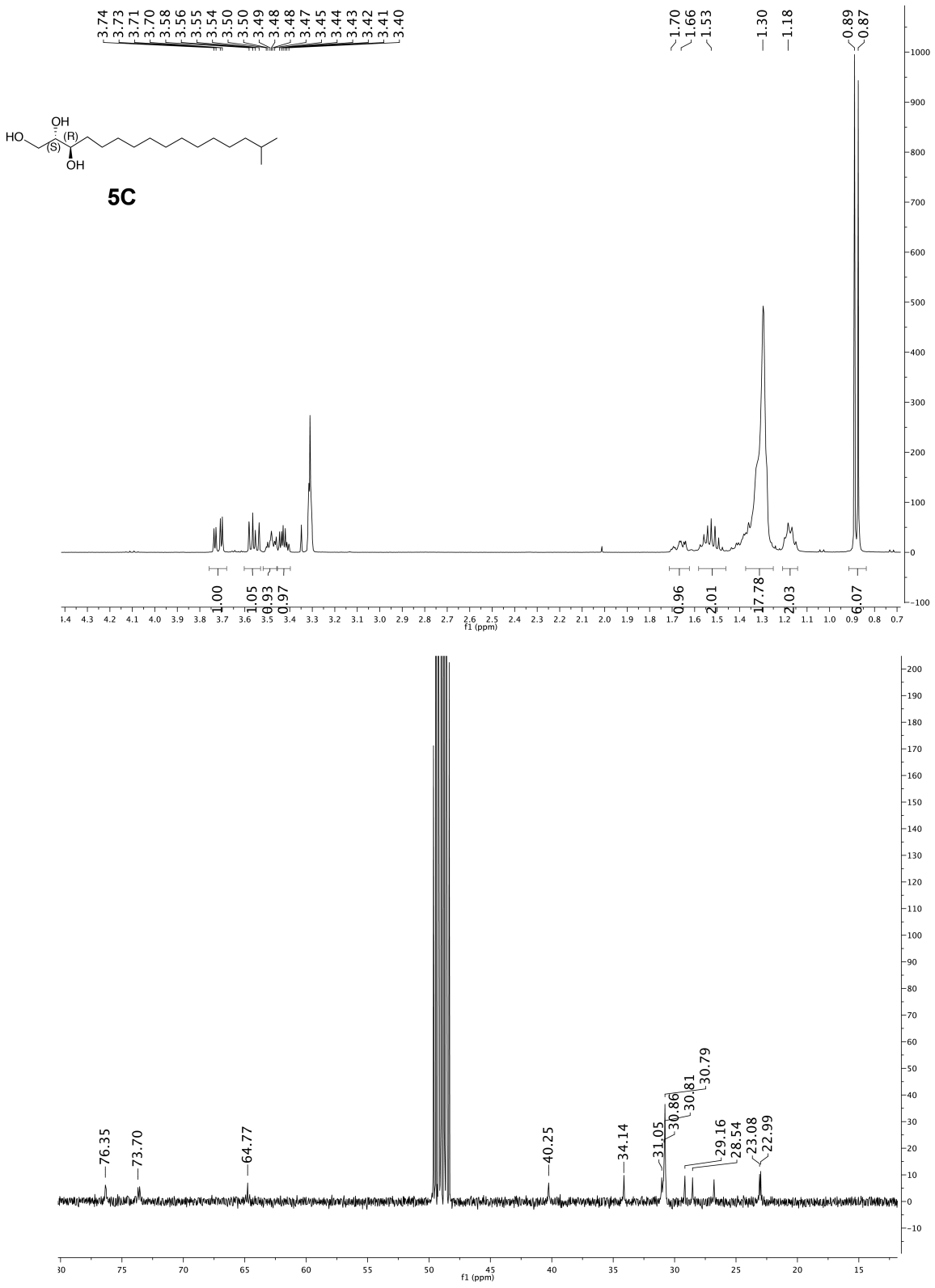


Figure S1.14 NMR spectra for synthetic compound 5C
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

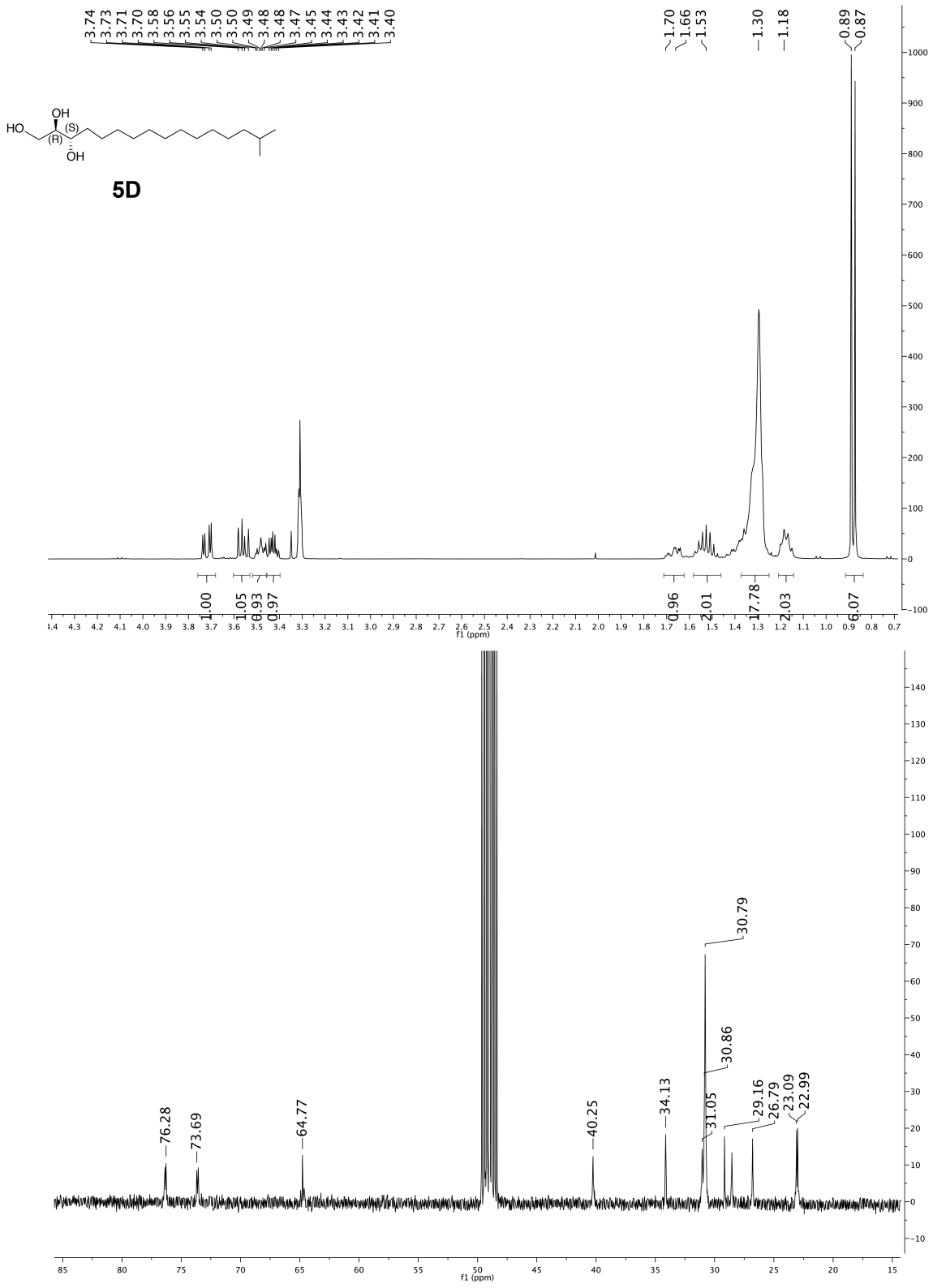


Figure S1.15 NMR spectra for synthetic compound 5D
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

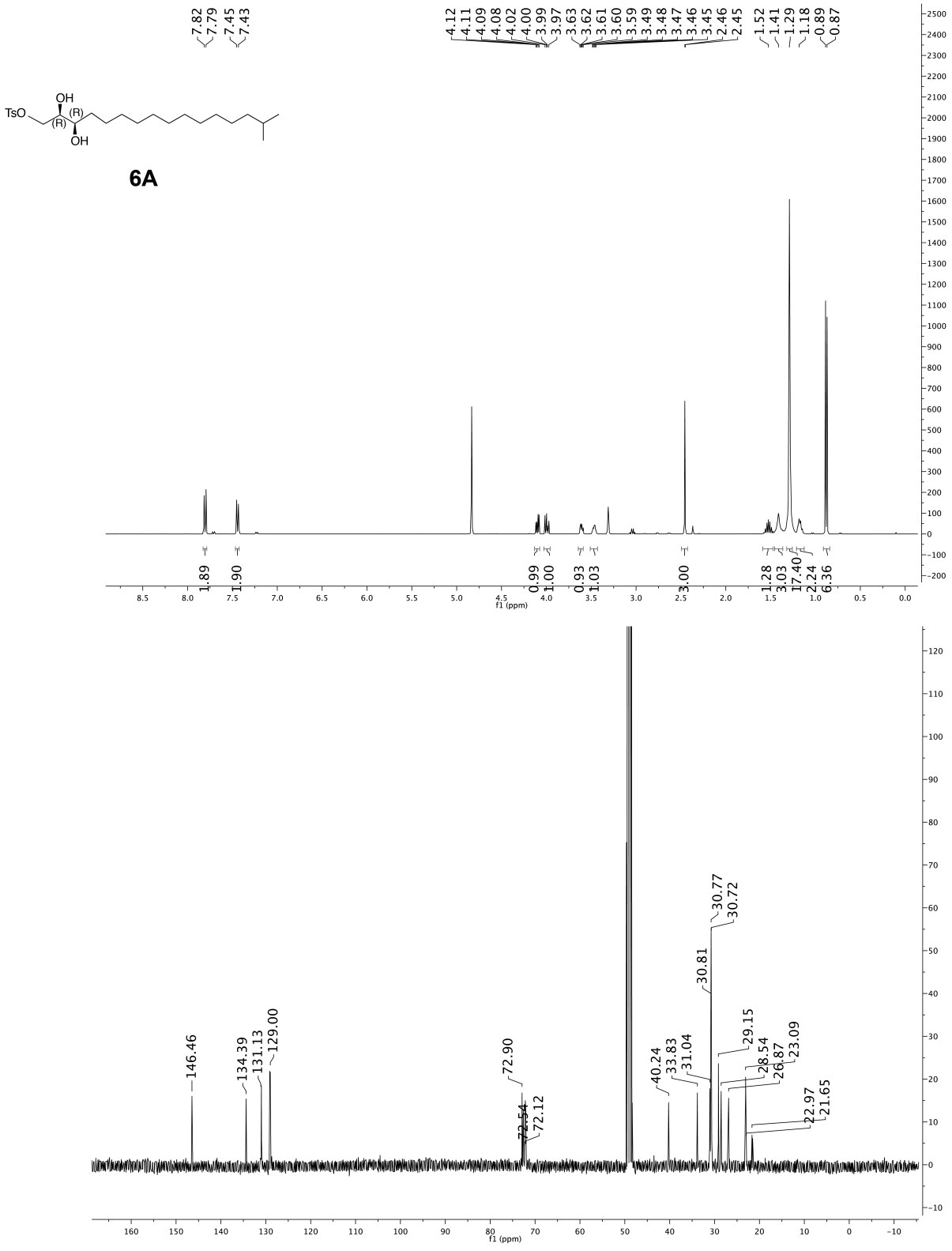


Figure S1.16 NMR spectra for synthetic compound 6A
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

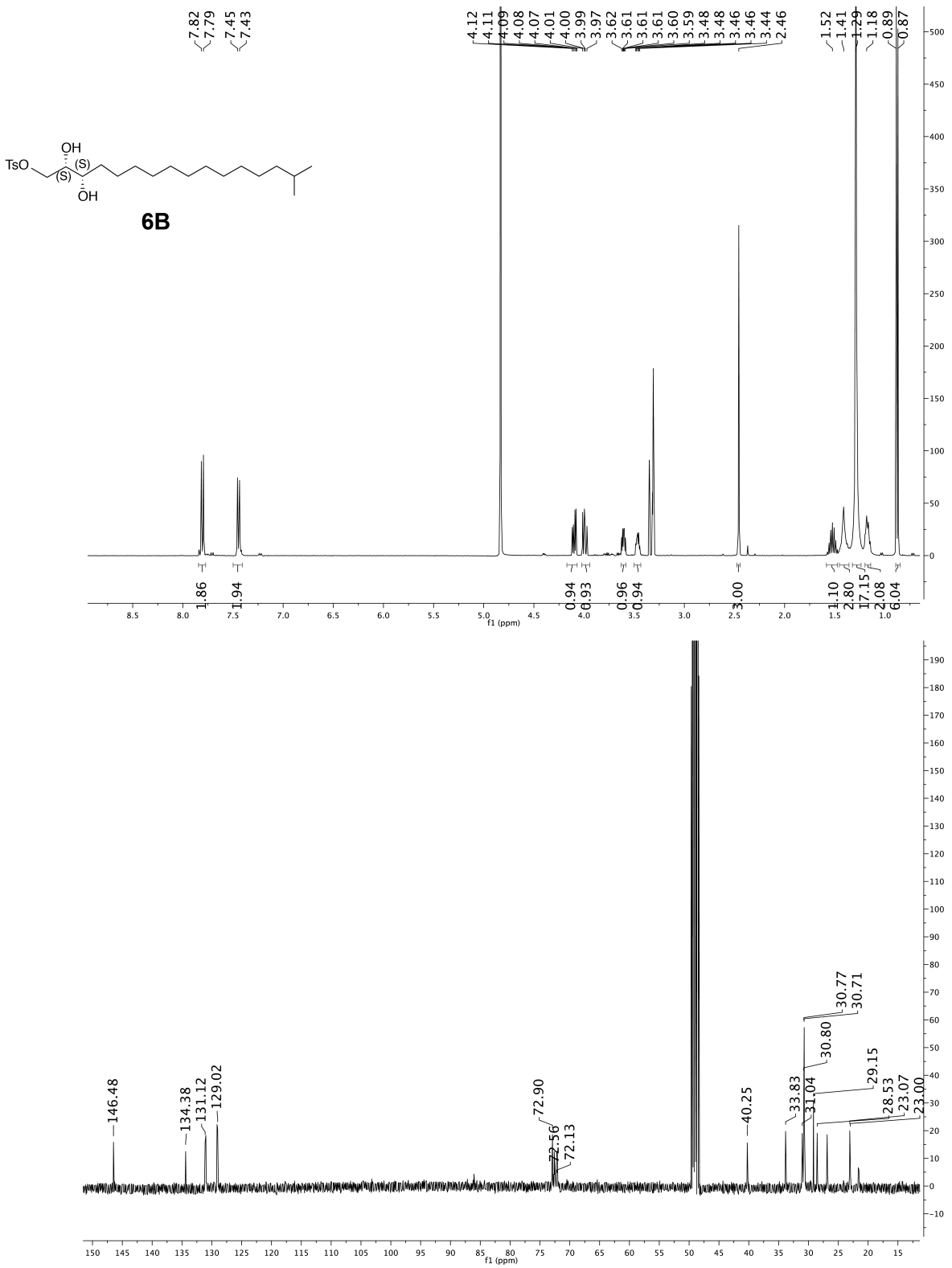


Figure S1.17 NMR spectra for synthetic compound 6B
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

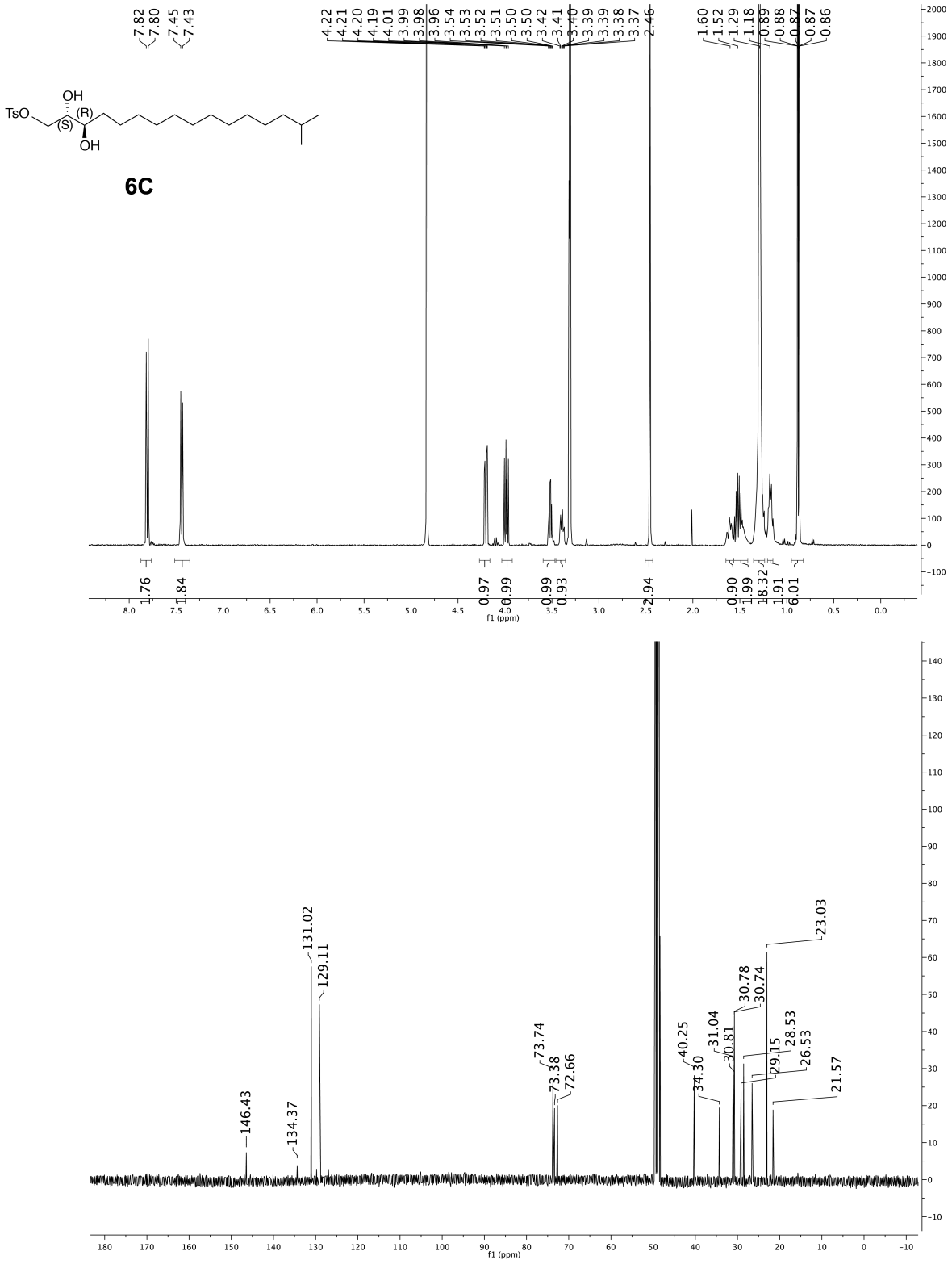


Figure S1.18 NMR spectra for synthetic compound 6C
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

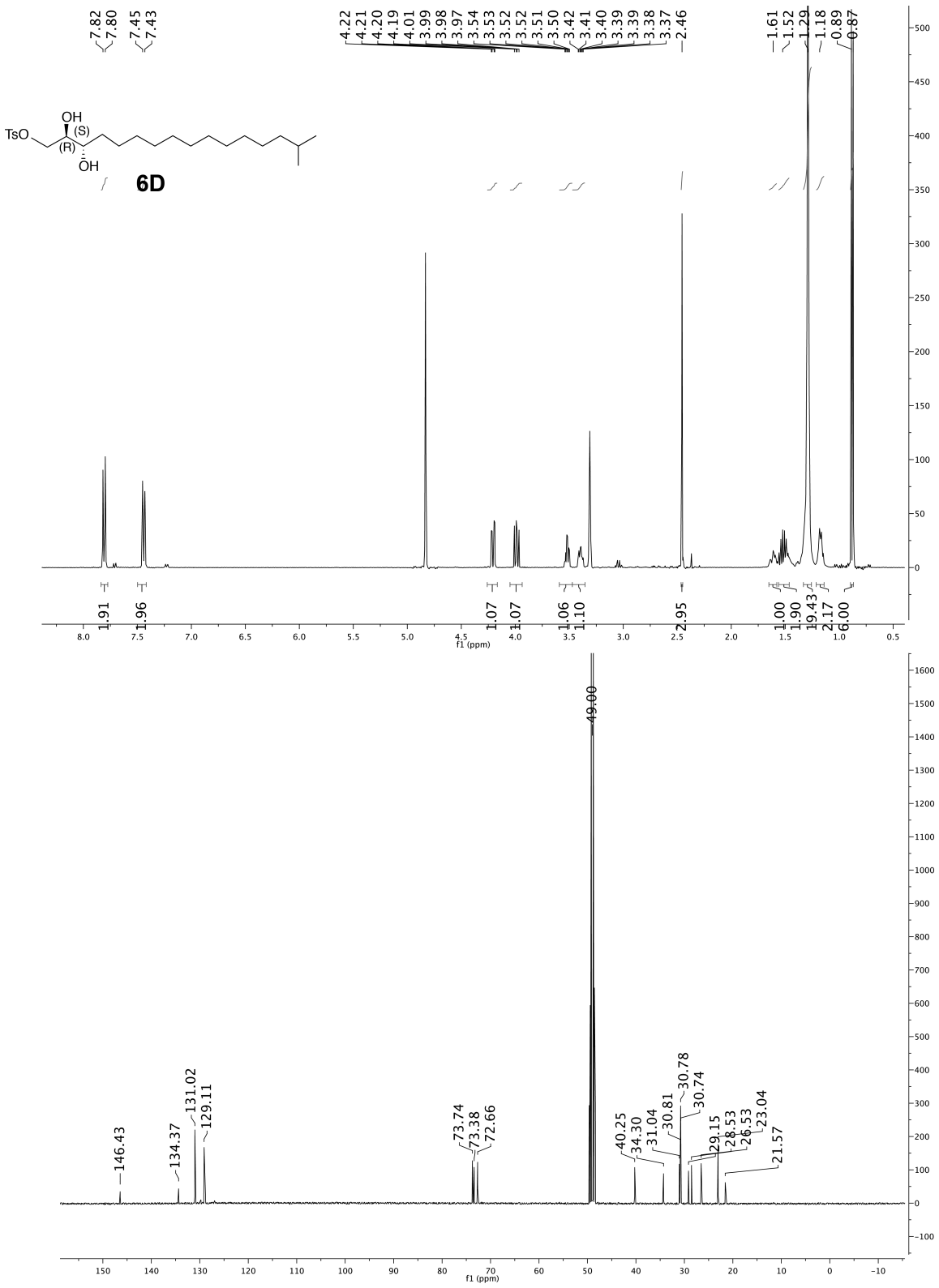


Figure S1.19 NMR spectra for synthetic compound 6D
(Top) ^{1}H-NMR (400 MHZ, CD_3OD) **(Bottom)** ^{13}C-NMR (100 MHz, CD_3OD)

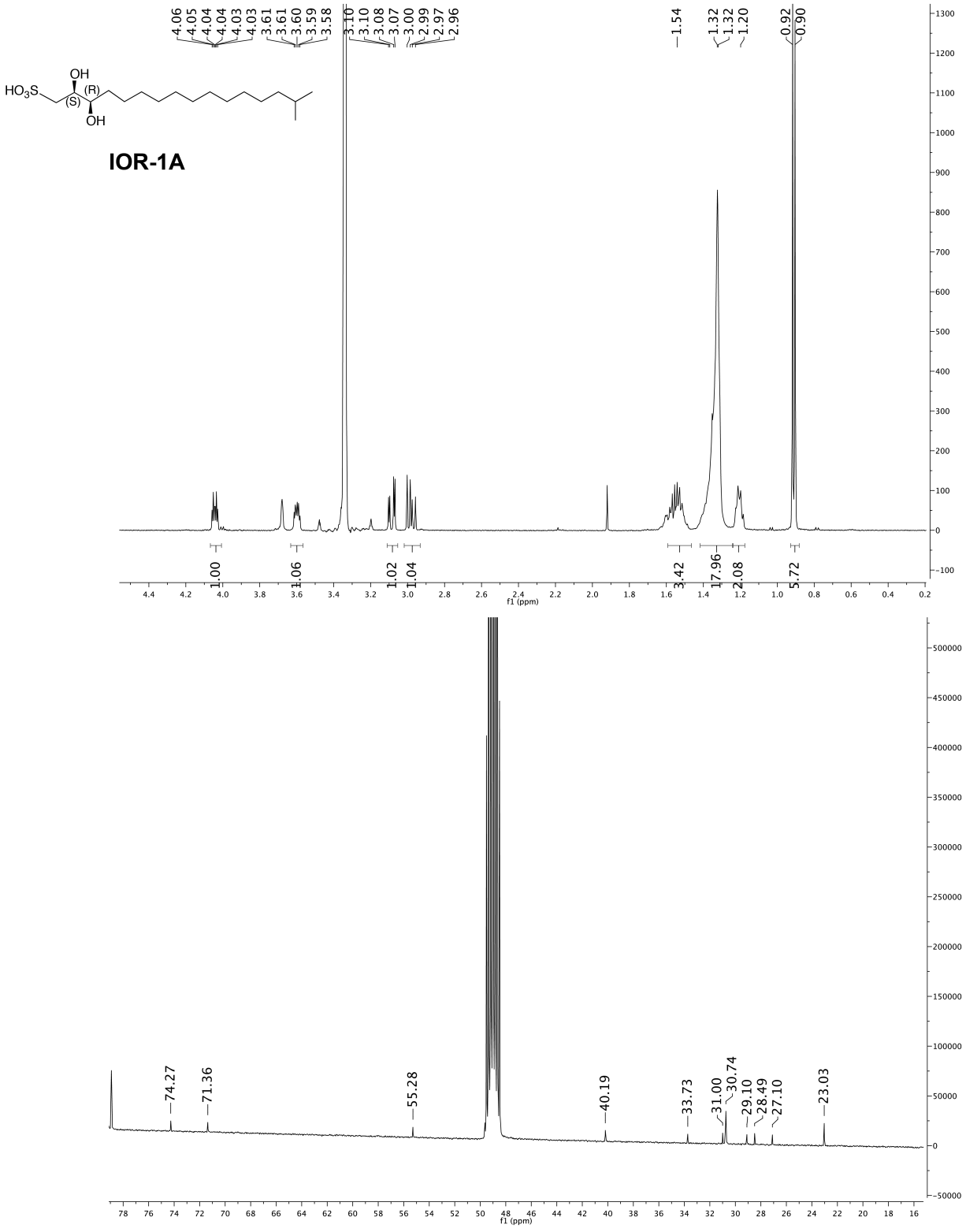


Figure S1.20 NMR spectra for synthetic compound IOR-1A
(Top) ^{1}H-NMR (500 MHZ, $CD_3OD/CDCl_3$) **(Bottom)** ^{13}C-NMR (151 MHz, $CD_3OD/CDCl_3$)

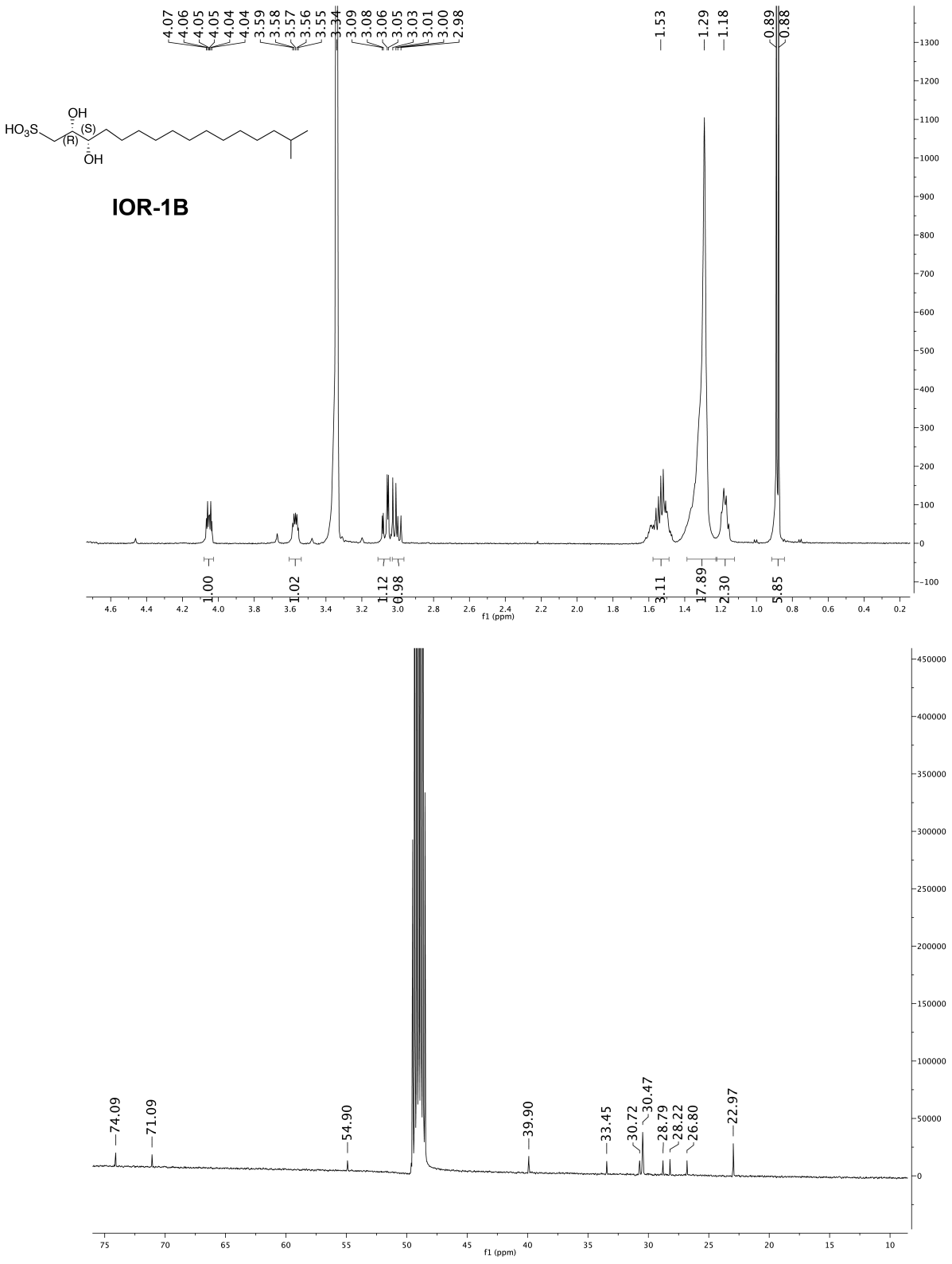


Figure S1.21 NMR spectra for synthetic compound IOR-1B
(Top) ^{1}H-NMR (500 MHZ, $CD_3OD/CDCl_3$) **(Bottom)** ^{13}C-NMR (151 MHz, $CD_3OD/CDCl_3$)

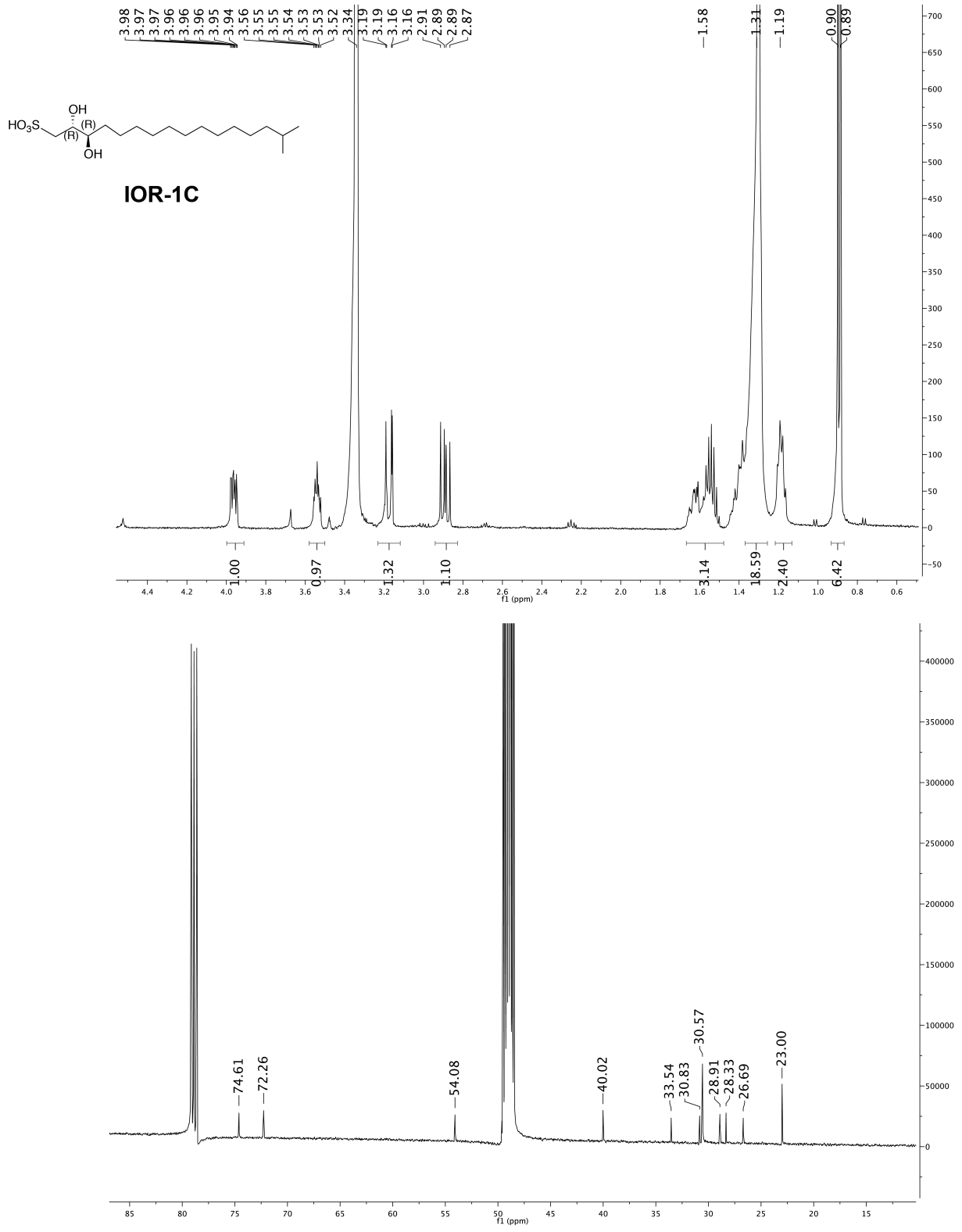


Figure S1.22 NMR spectra for synthetic compound IOR-1C
(Top) ^{1}H-NMR (500 MHZ, $CD_3OD/CDCl_3$) **(Bottom)** ^{13}C-NMR (151 MHz, $CD_3OD/CDCl_3$)

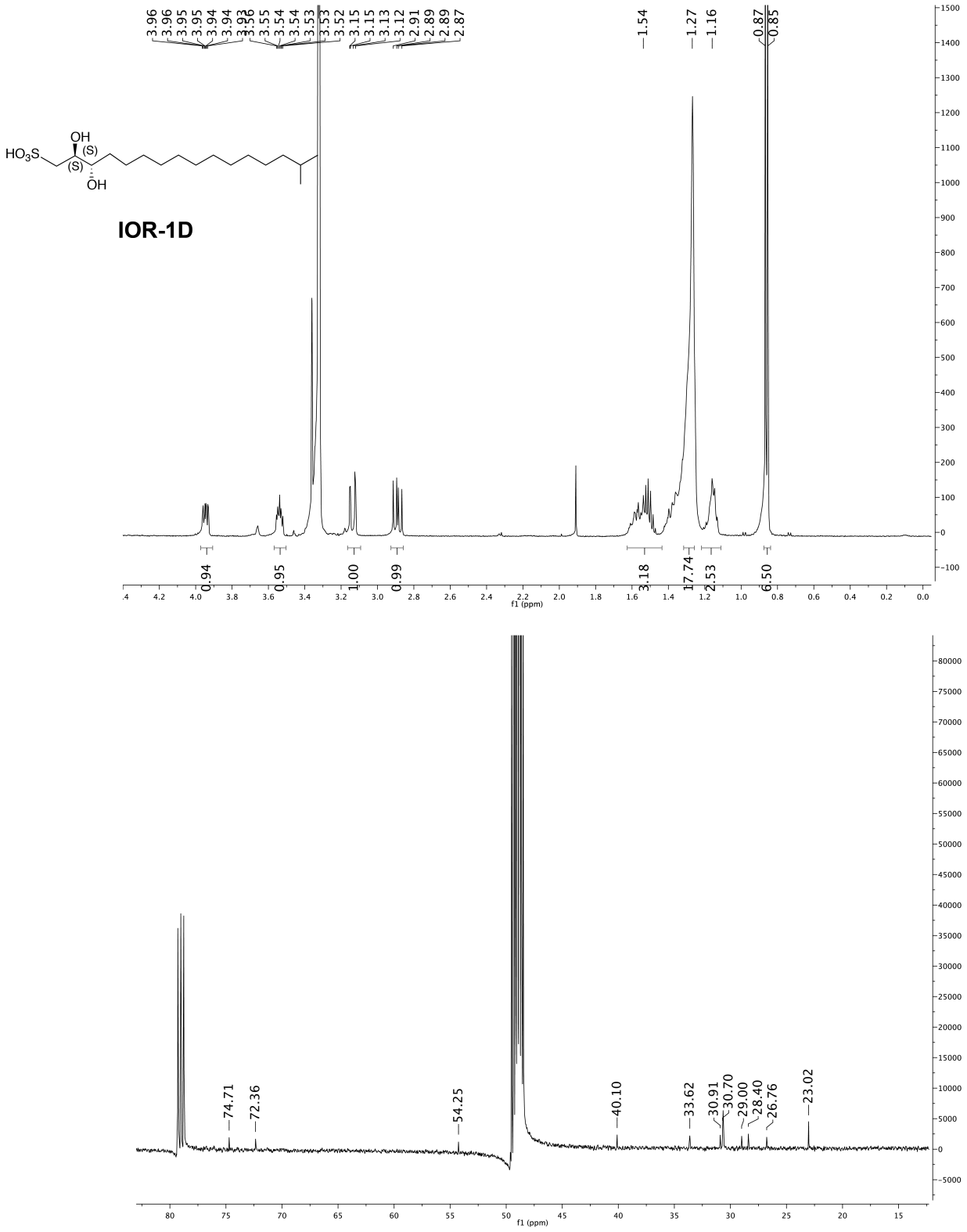


Figure S1.23 NMR spectra for synthetic compound IOR-1D
(Top) ^{1}H-NMR (500 MHZ, $CD_3OD/CDCl_3$) **(Bottom)** ^{13}C-NMR (151 MHz, $CD_3OD/CDCl_3$)

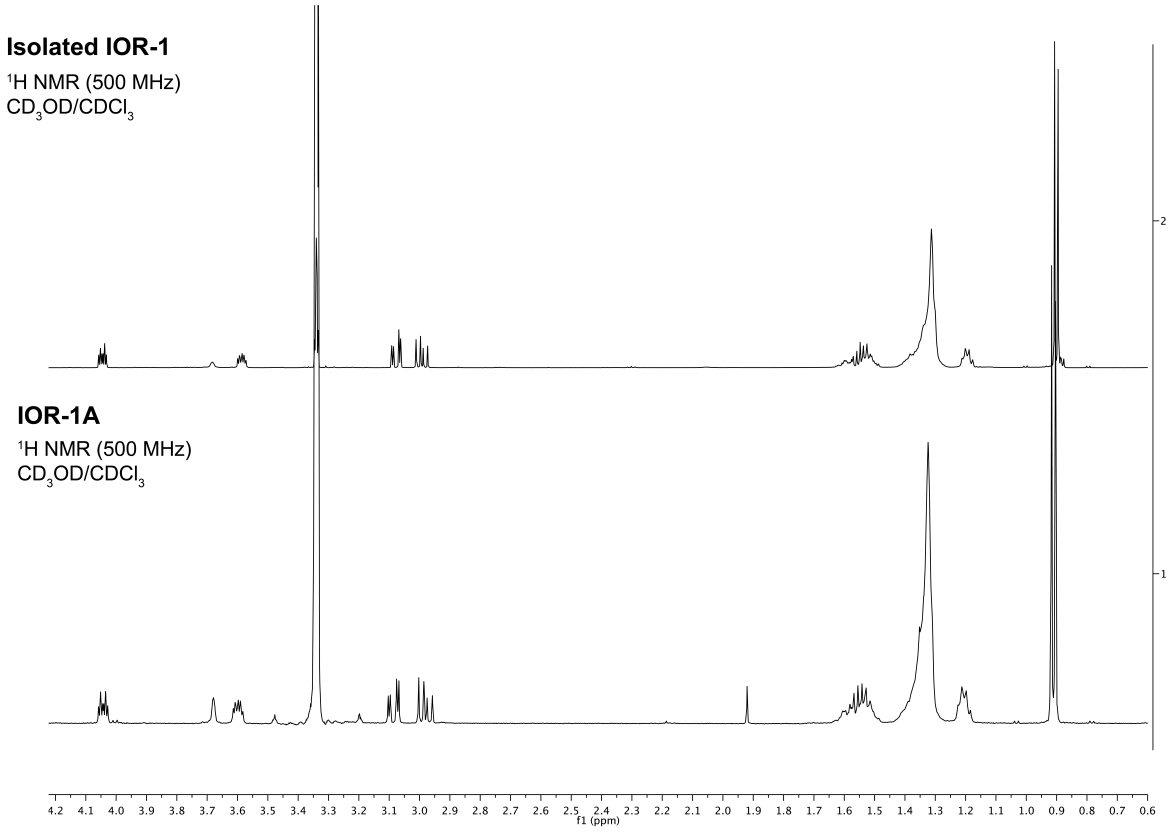


Figure S1.24 ^{1}H-NMR spectra of isolated IOR-1 compared to synthetic IOR-1A
(Top) Isolated IOR-1 (500 MHZ, $CD_3OD/CDCl_3$) **(Bottom)** Synthetic IOR-1A (500 MHz, $CD_3OD/CDCl_3$)

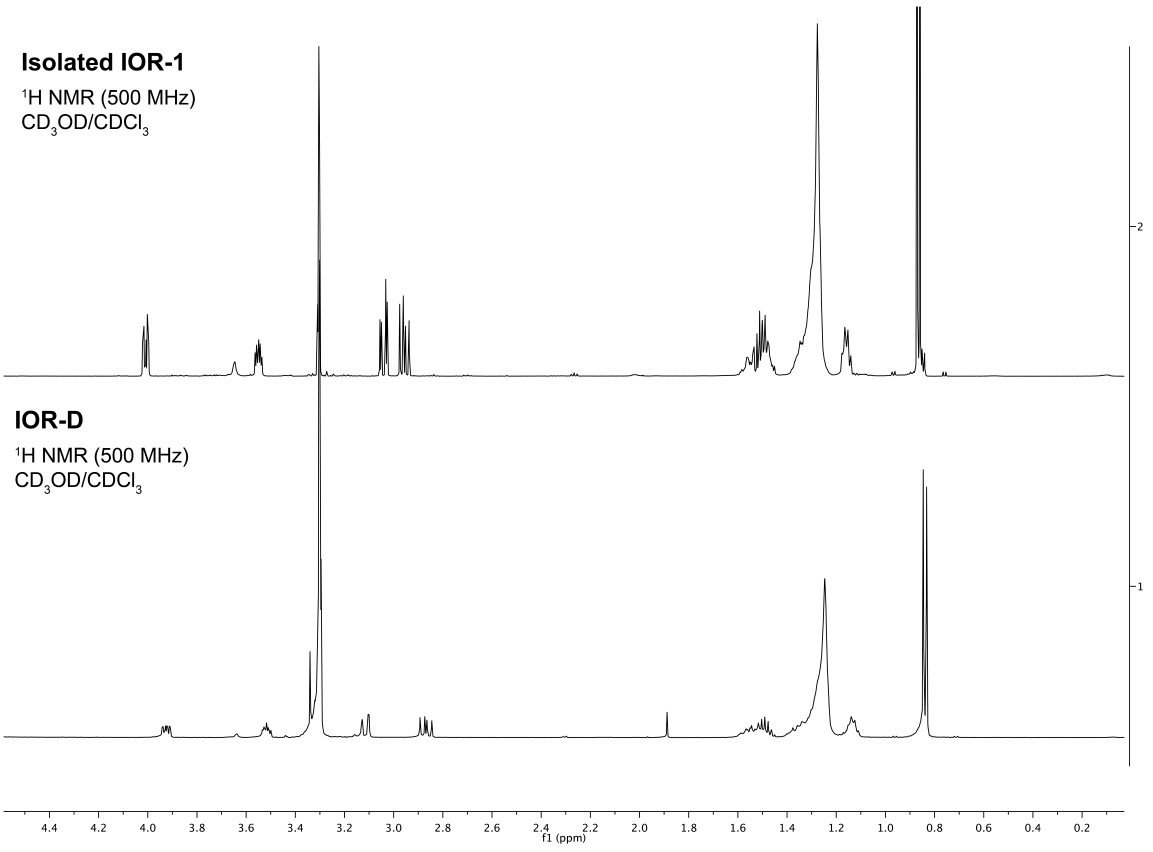


Figure S1.25 ^{1}H-NMR spectra of isolated IOR-1 compared to synthetic IOR-1D
(Top) Isolated IOR-1 (500 MHZ, $CD_3OD/CDCl_3$) **(Bottom)** Synthetic IOR-1D (500 MHz, $CD_3OD/CDCl_3$)

Appendix 2: Supplementary materials for Chapter 3

Table S2.1 ^{1}H and ^{13}C NMR shifts for RIF-2

Position	RIF-2 (600 MHz, d6-DMSO) ^{1}H, m, H (*J* in Hz)	(150 MHz) ^{13}C, m	RIF-2 (600 MHz, d4-MeOD) ^{1}H, m, H (*J* in Hz)	(150 MHz) ^{13}C, m
1	2.57, dd, 1H (14.1, 4.6) 2.86, dd, 1H, (14.1, 6.1)	50.7, t	2.99, dd, 1H (14.4, 3.8) 3.07, dd, 1H (14.4, 8.2)	51.3, t
1'		173.8, s		n.d.
2	3.94, m, 1H	50.4, d	4.23, m, 1H	52.6, d
2'	3.74, m, 1H	70.7, d	4.04, dd, 1H (12.6, 6.2)	73.8, d
2'-OH	5.48, d, 1H (5.2)			
3	4.23, m, 1H	71.0, d	4.38, t, 1H (5.0)	73.8, d
3-OH	5.24, dd, 1H (6.2)		n.d.	
3'	1.40, m, 1H 1.57, m, 1H	n.d.	n.d.	
4	5.52, ddd, 1H (15.6, 5.4, 1.0)	128.6, d	5.68, ddd, 1H (15.4, 5.6, 1.0)	130.6, d
5	5.58, ddd, 1H (15.6, 6.0, 1.0)	134.0, d	5.80, ddd, 1H (15.6, 6.1, 1.2)	137.5, d
6	3.85, m, 1H	70.2, d	3.98, dd, 1H (8.4, 3.6)	73.8, d
6-OH	4.50, d, 1H (4.8)		n.d.	n.d.
7	1.30-1.34, m, 1H	37.0, t	1.40-1.44	38.9, t
8-13 **4'-11'**	1.17 - 1.27, br s, 30H	22.5-30.0, t	1.16-1.31, br s, 30H	28-34, t
14, 12'	1.11 - 1.16, m, 2H	38.0, t	1.08-1.13, m, 2H	40.8, t
15, 13'	1.49, m, 2H	27.0, d	1.46, m, 2H	29.6, d
16, 14'	0.84, d, 12H (6.6)	22.1, q	0.81, d, 12H (6.6)	23.9, q
SO_3H	n.d	n.d.	n.d.	n.d.
NH	8.02, d, 2H (8.1)	n.d.	n.d.	n.d.

IR (ATR): υ = 3300, 2955, 2860, 1640, 1555, 1465 cm^{-1}; MS (ESI-TOF): calcd. for $C_{32}H_{63}NO_7SNa$: 628.4223; found 628.4229

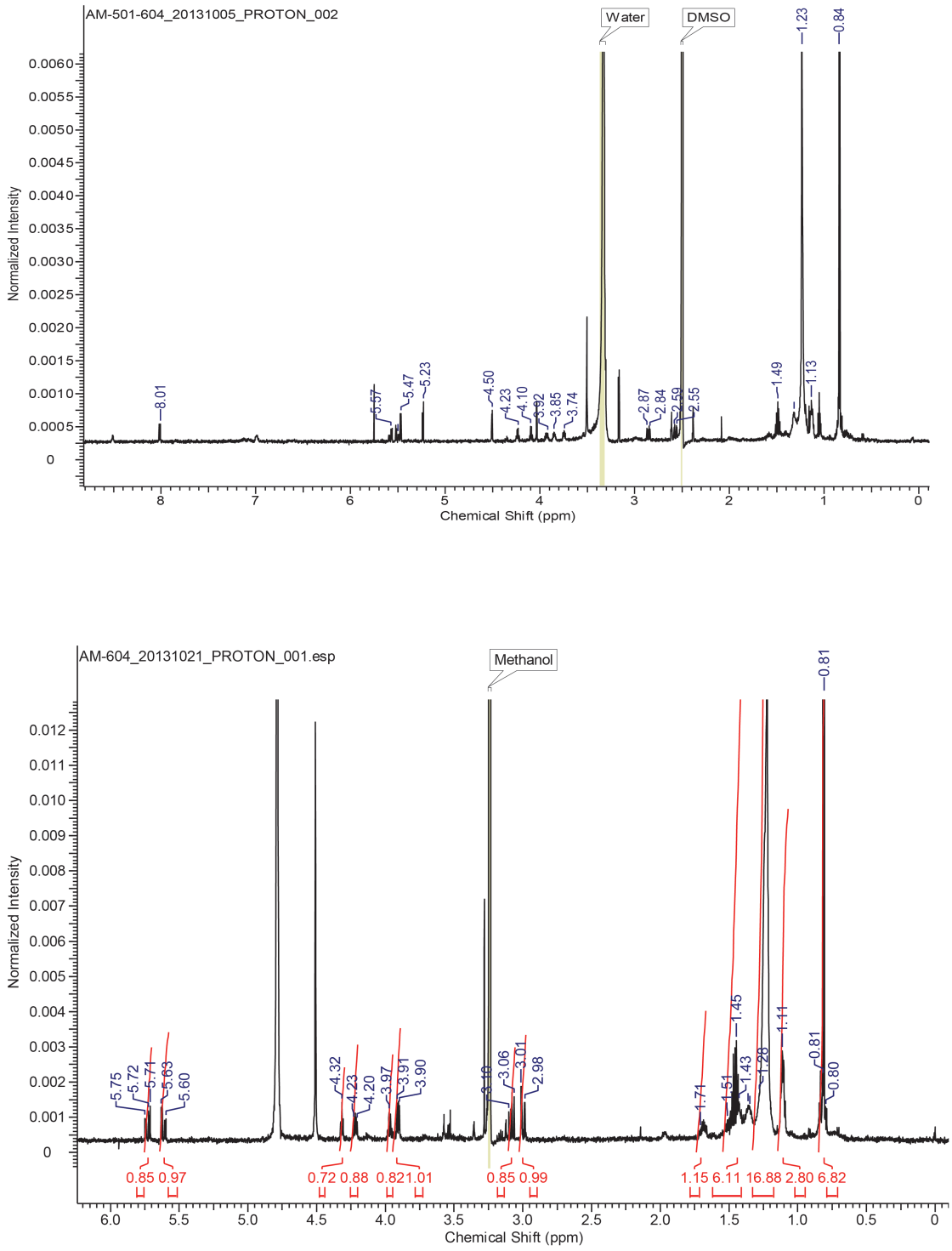


Figure S2.1 ^{1}H-NMR spectrum of RIF-2
(Top) 500 MHz, d6-DMSO (Bottom) 500 MHz, d4-MeOD

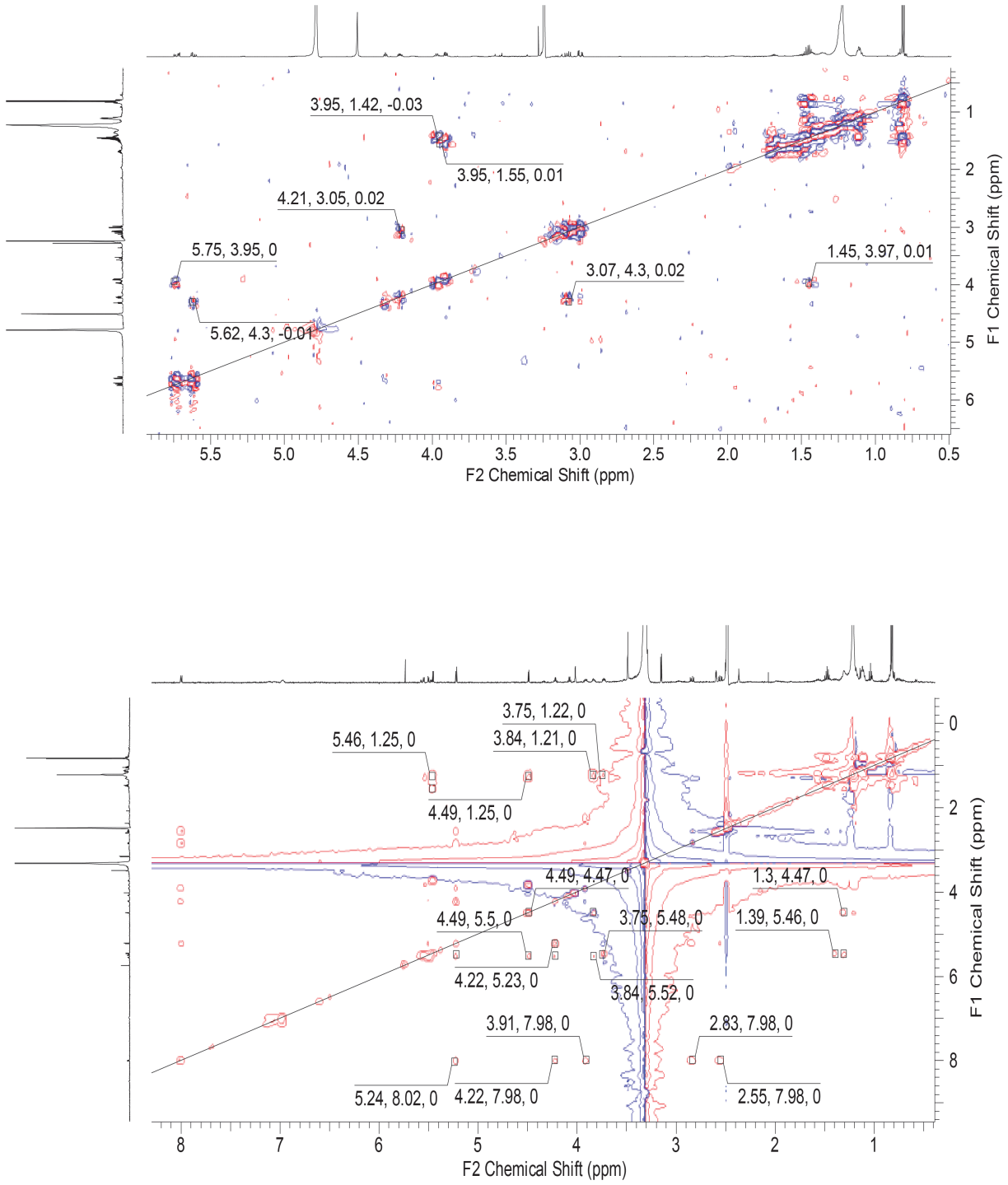
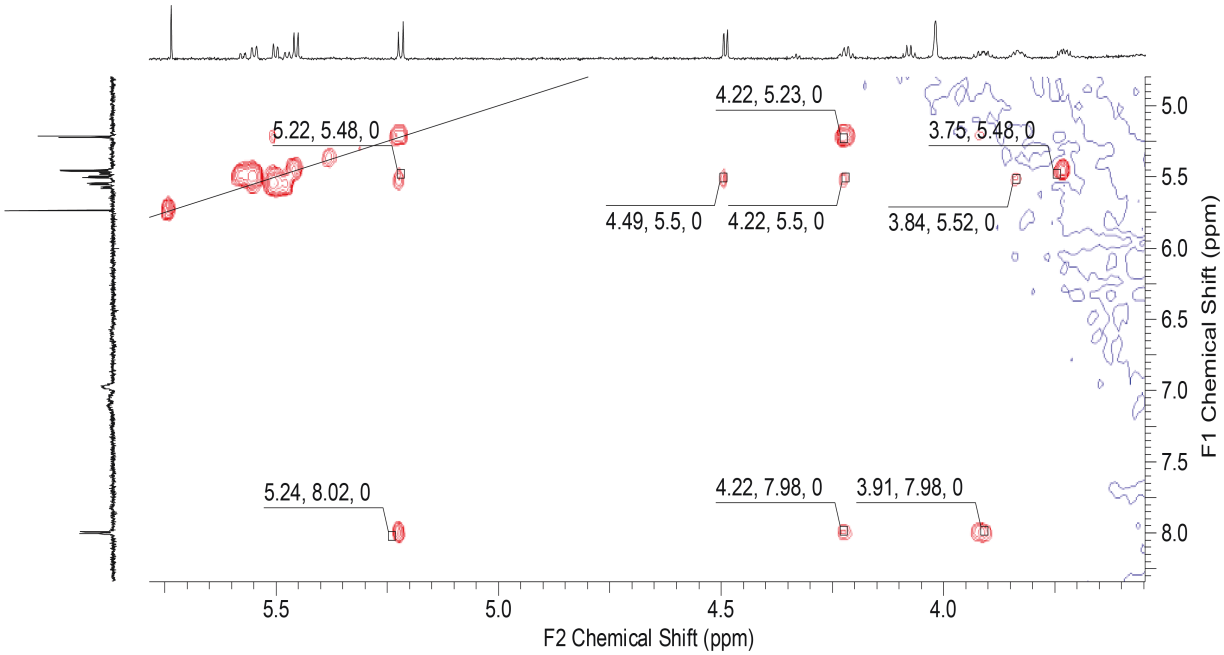


Figure S2.2 2D NMR spectra of RIF-2

(Top) COSY spectrum, 500 MHz, d4-MeOD. (Bottom) TOCSY spectrum, 500 MHz, d6-DMSO



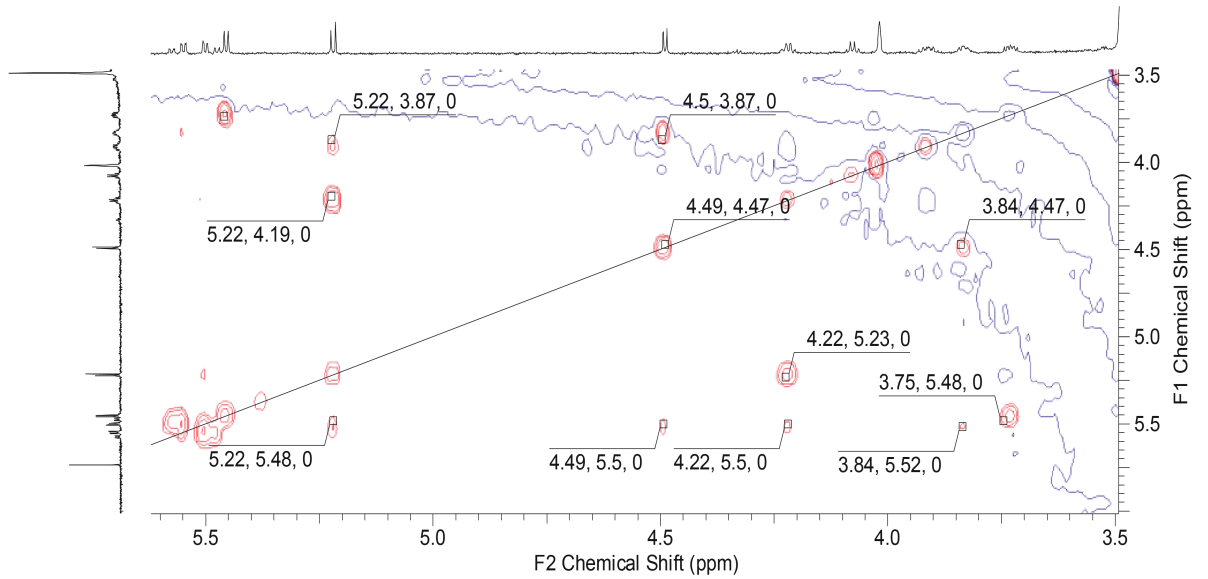


Figure S2.3 2D NMR spectra of RIF-2
(Top) TOCSY spectrum, 500 MHz, d6-DMSO (Bottom) TOCSY spectrum, 500 MHz, d6-DMSO

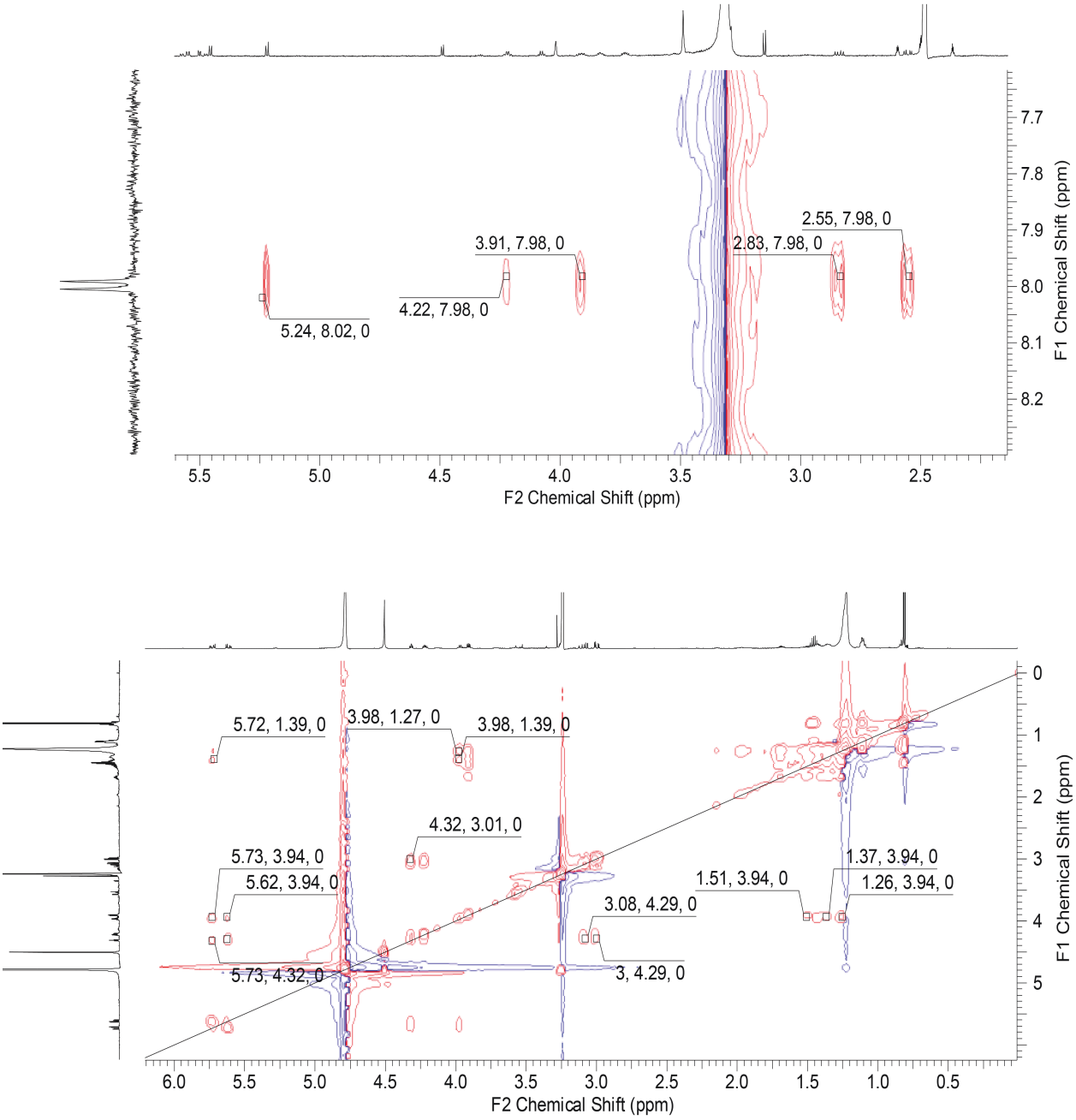


Figure S2.4 2D NMR spectra of RIF-2
(Top) TOCSY spectrum, 500 MHz, d6-DMSO. (Bottom) TOCSY spectrum (500 MHz, d4-MeOD)

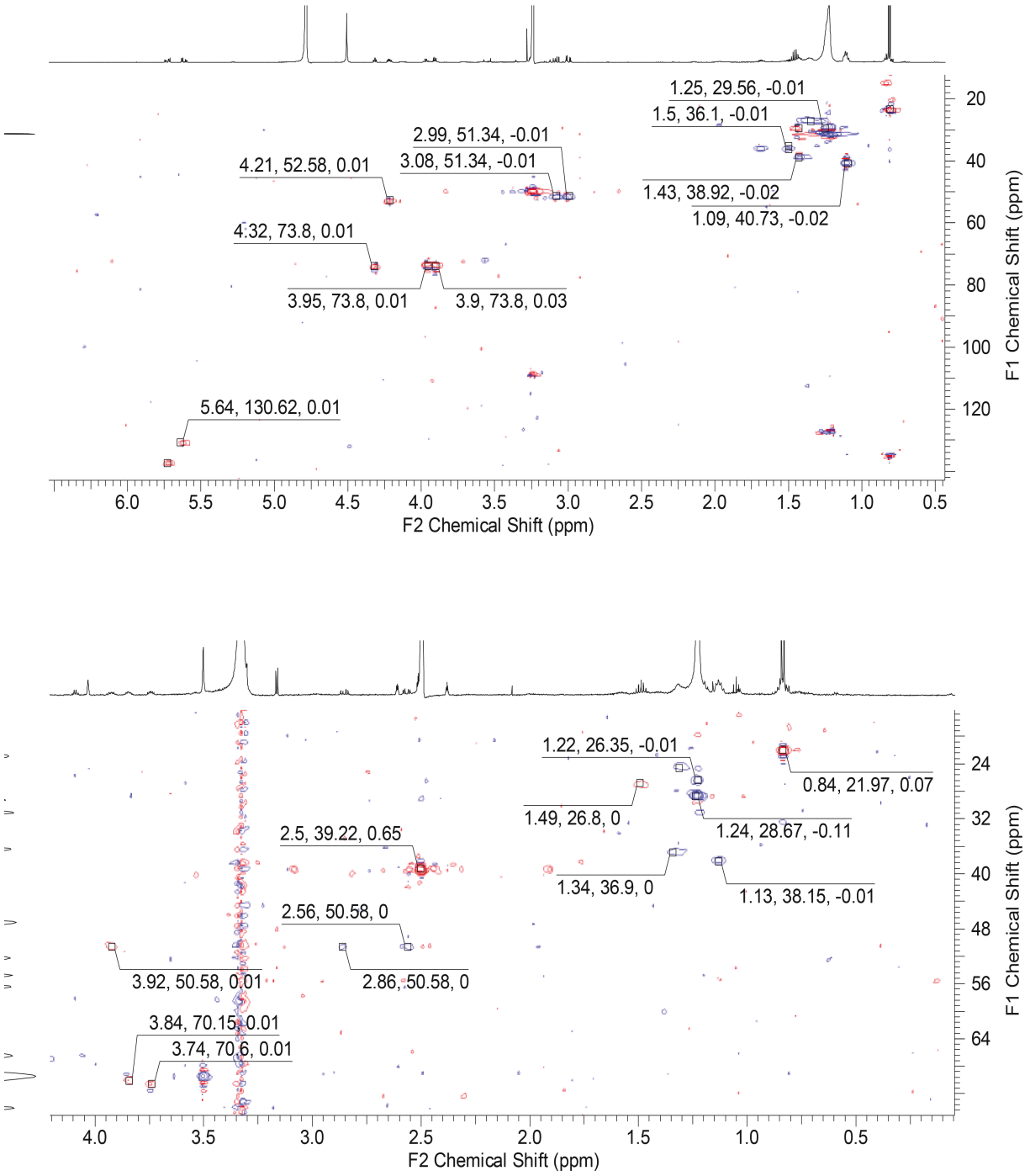


Figure S2.5 2D NMR spectra of RIF-2
(Top) HSQC spectrum, 500 MHz, d4-MeOD. (Bottom) HSQC spectrum, 500 MHz, d6-DMSO

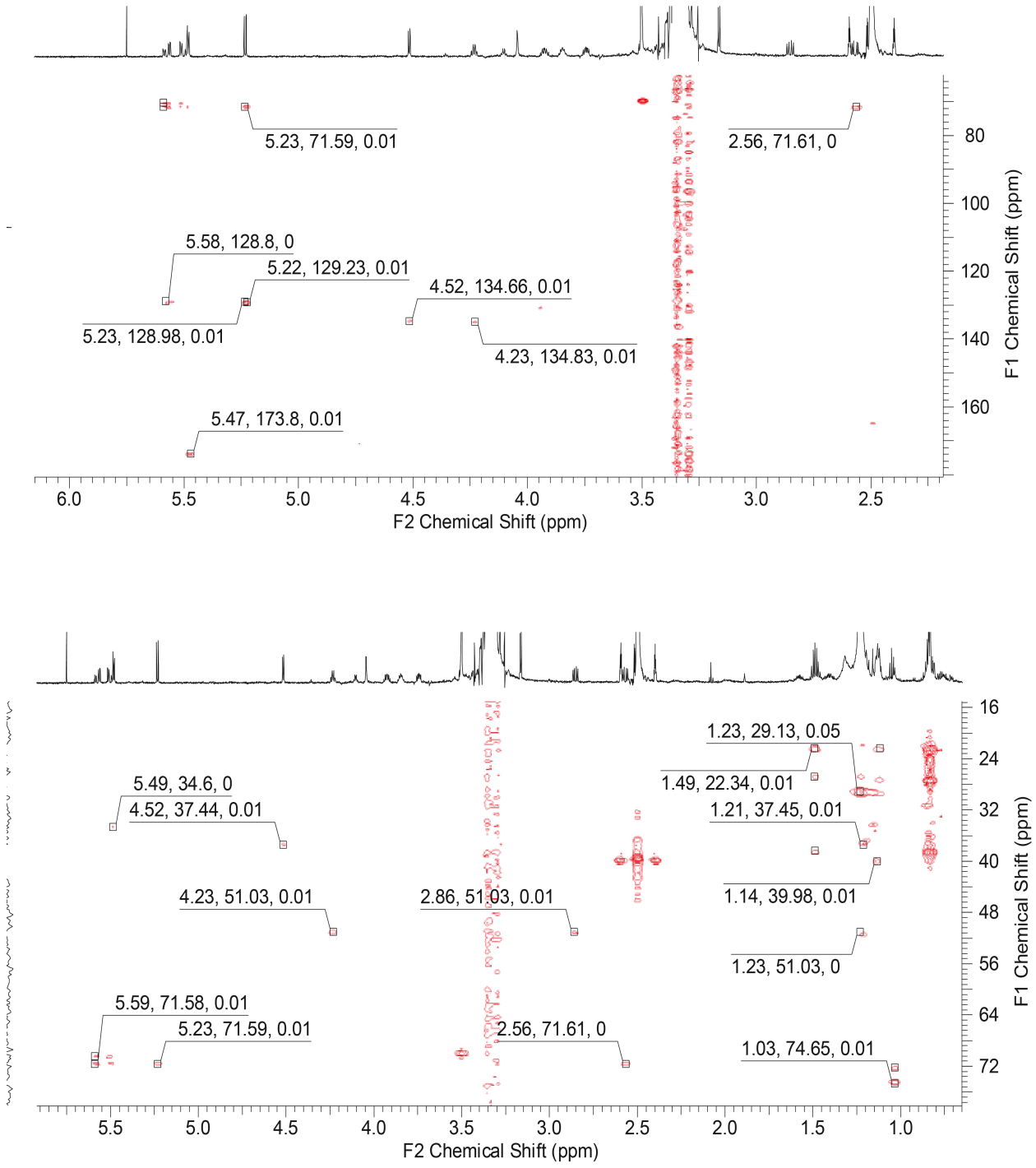


Figure S2.6 2D NMR spectra of RIF-2
(Top) HMBC spectrum, 700 MHz, d6-DMSO. (Bottom) HMBC spectrum, 700 MHz, d6-DMSO

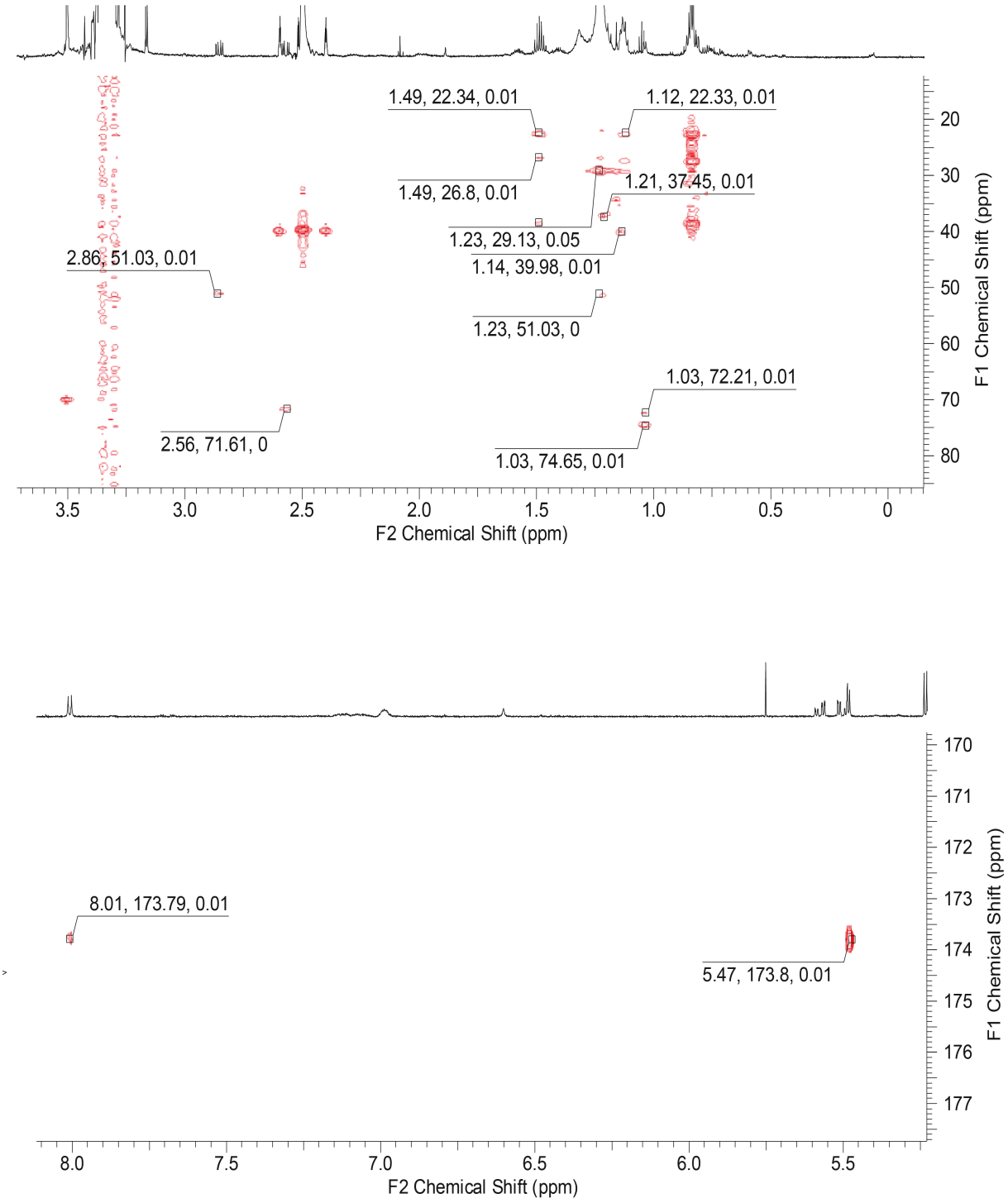


Figure S2.7 2D NMR spectra of RIF-2
(Top) HMBC spectrum, 700 MHz, d6-DMSO. (Bottom) HMBC spectrum, 700 MHz, d6-DMSO

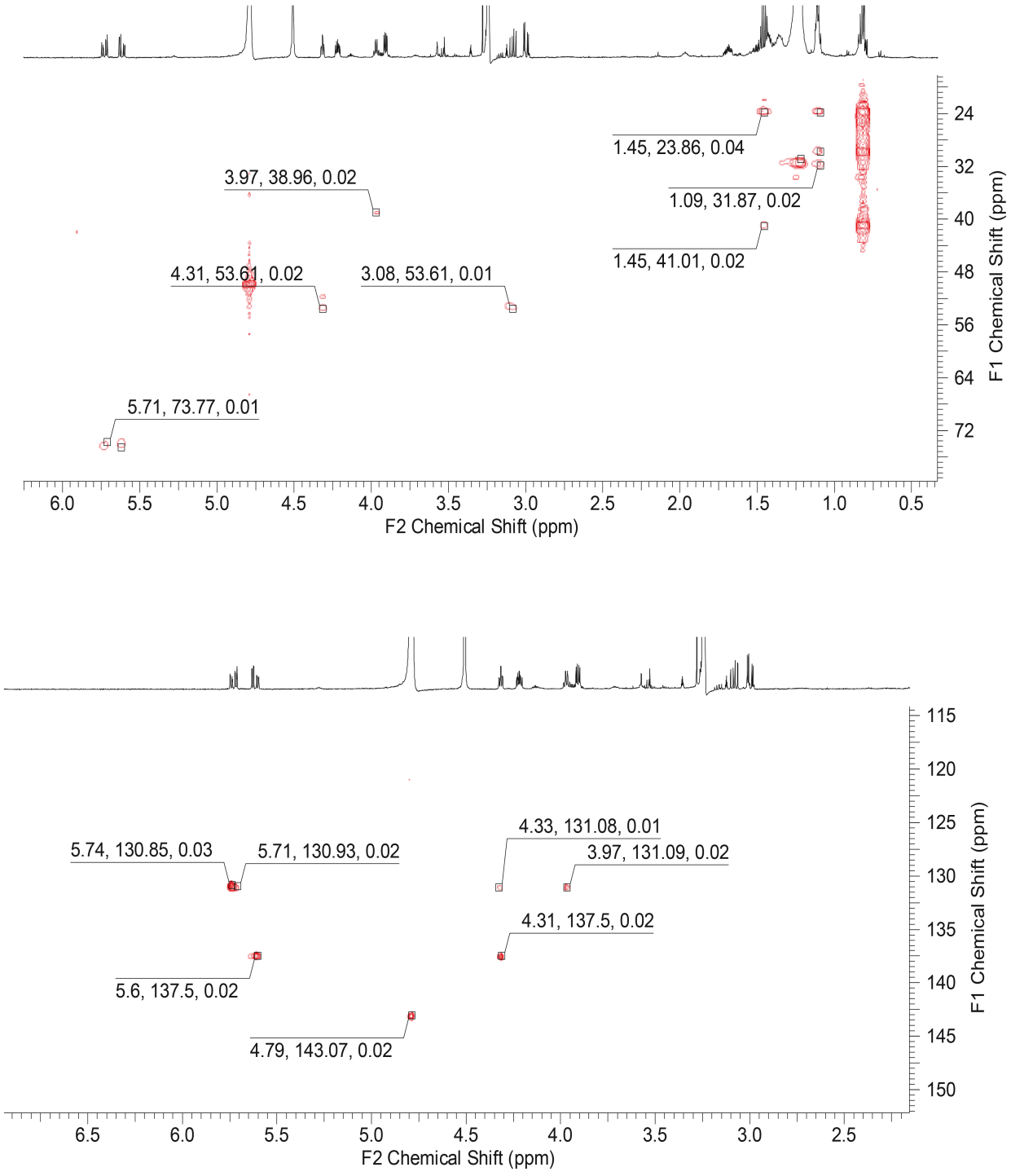
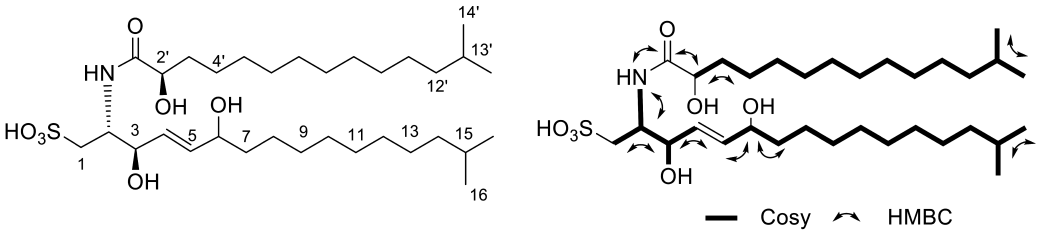
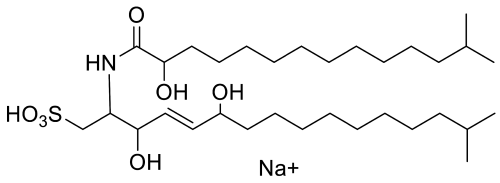


Figure S2.8 2D NMR spectra of RIF-2
(Top) HMBC spectrum, 500 <Hz, d4-MeOD. (Bottom) HMBC spectrum, 500 MHz, d4-MeOD.

A





Chemical Formula: $C_{32}H_{63}NNaO_7S^+$
Exact Mass: 628.42175

B

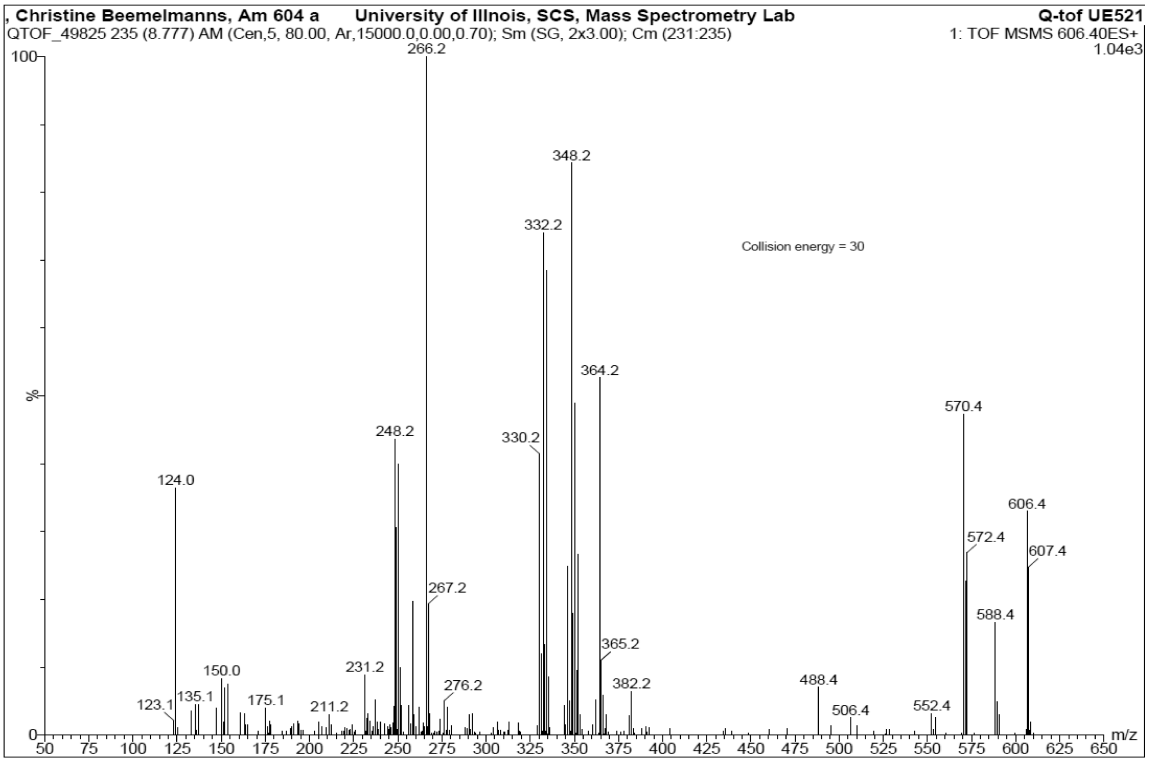
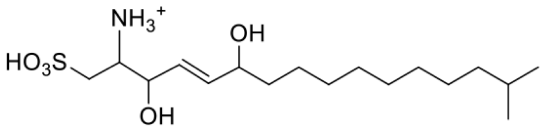
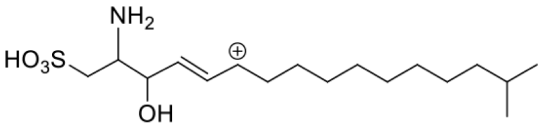


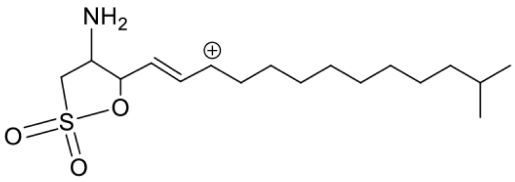
Figure S2. 9 RIF-2 structure determination
(A) COSY and HMBC correlations for RIF-2. (B) MSMS spectrum for RIF-2 m/z = 606.4 $(M+H)^+$ Q-tof, pos.



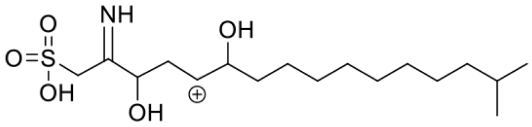
Chemical Formula: $C_{17}H_{36}NO_5S^+$
Exact Mass: 366.23087



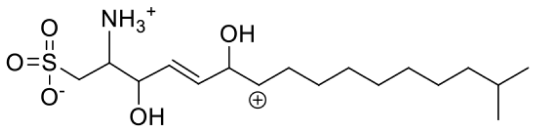
Chemical Formula: $C_{17}H_{35}NO_4S$
Exact Mass: 349.22868



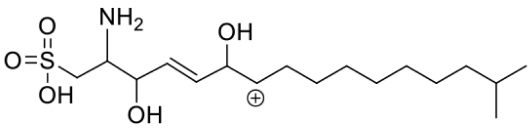
Chemical Formula: $C_{17}H_{33}NO_3S$
Exact Mass: 331.21811



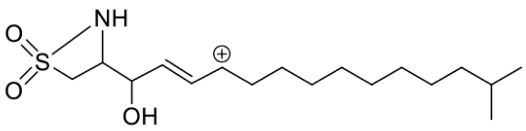
Chemical Formula: $C_{17}H_{34}NO_5S^+$
Exact Mass: 364.21522



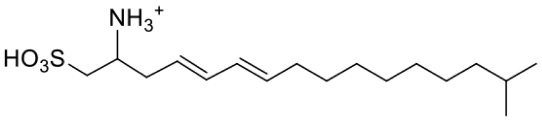
Chemical Formula: $C_{17}H_{34}NO_5S^+$
Exact Mass: 364.21522



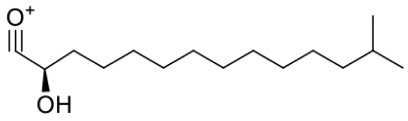
Chemical Formula: $C_{17}H_{34}NO_5S^+$
Exact Mass: 364.21522



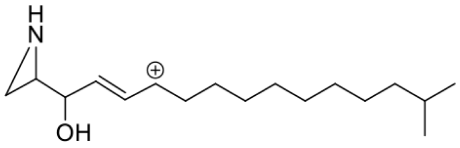
Chemical Formula: $C_{17}H_{32}NO_3S^+$
Exact Mass: 330.20974



Chemical Formula: $C_{17}H_{34}NO_3S^+$
Exact Mass: 332.22539



Chemical Formula: $C_{15}H_{29}O_2^+$
Exact Mass: 241.21621



Chemical Formula: $C_{17}H_{32}NO^+$
Exact Mass: 266.24784

Figure S2.10 Predicted fragment structures for RIF-2

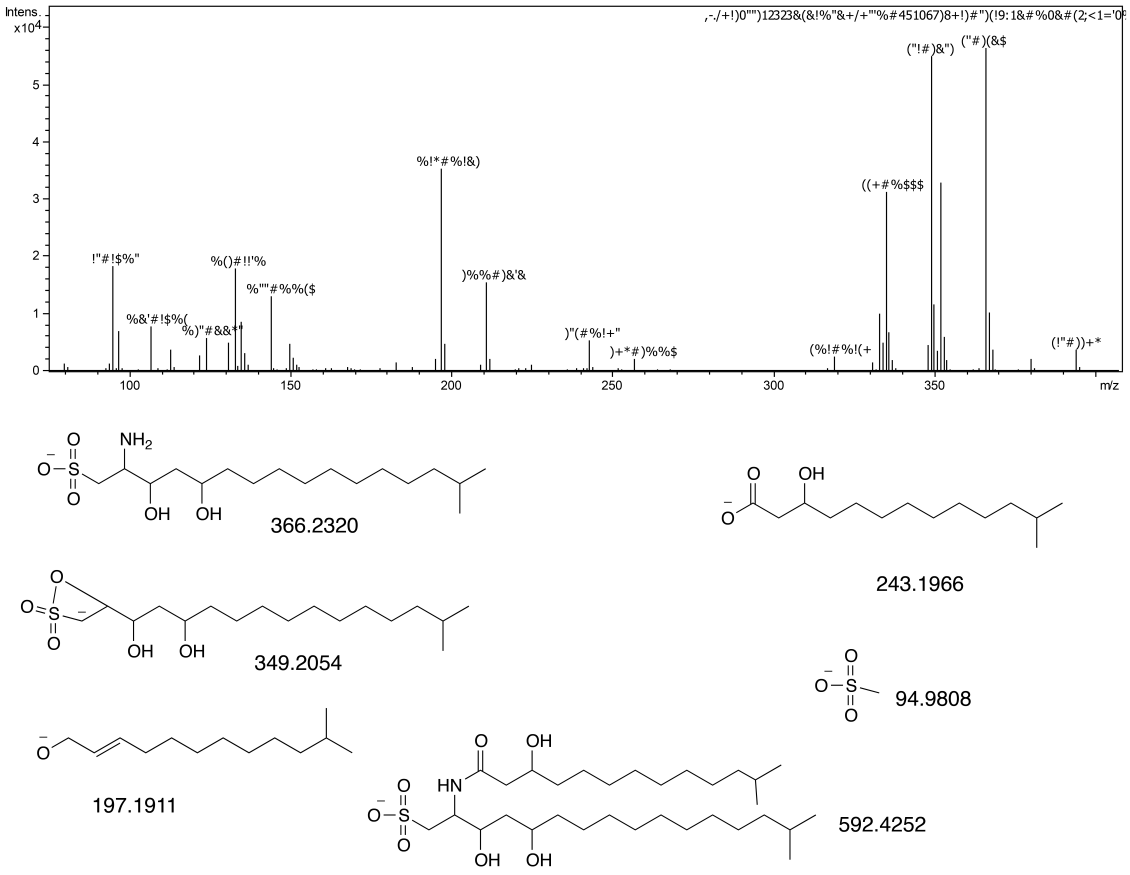


Figure S2.11 HR-MSMS sulfonolipid 592 (negative mode).

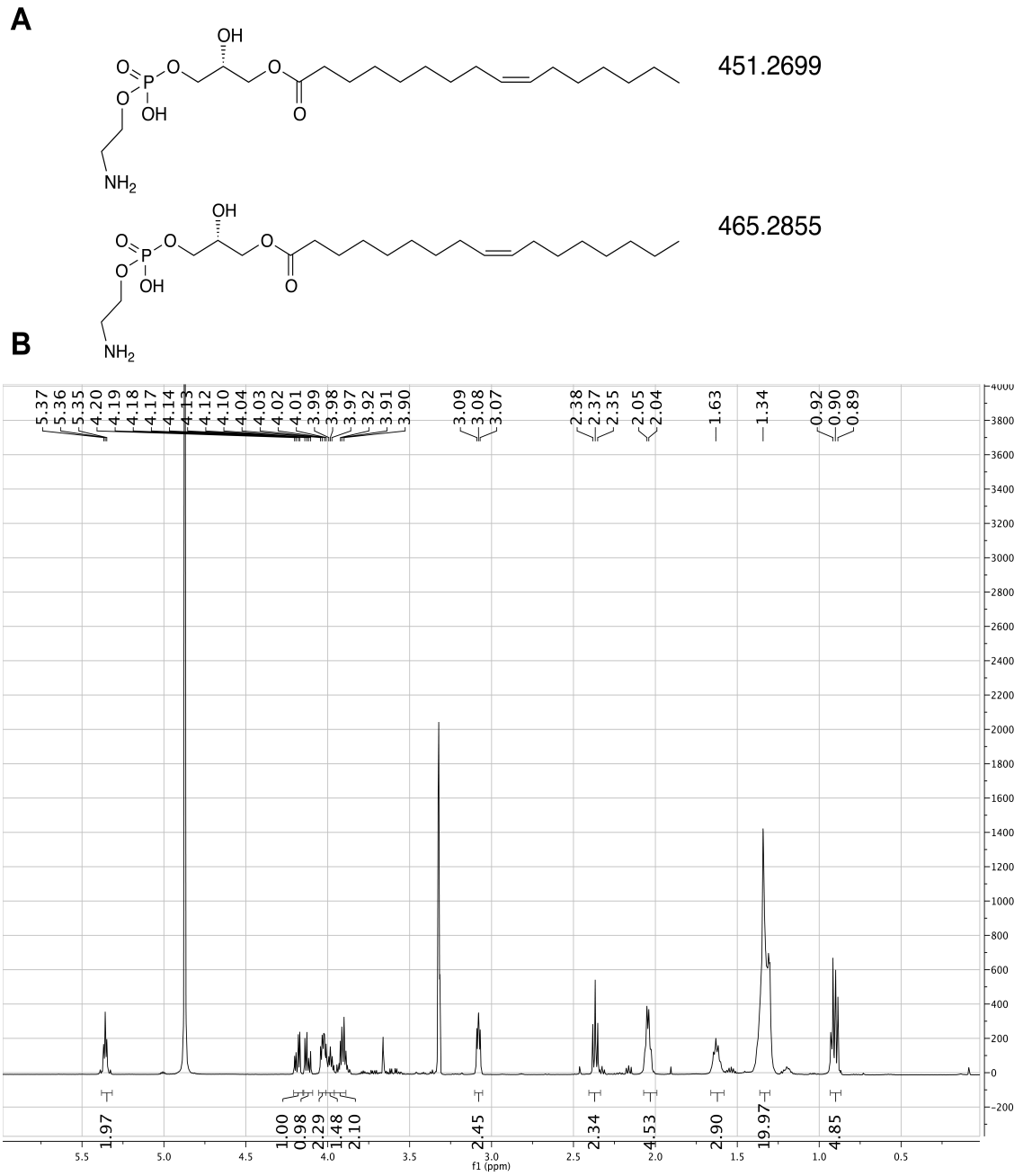


Figure S2.12 Characterization of lysophosphatidylethanolamines (LPEs)
(A) Structures and molecular weights of the two major LPEs present in the synergistic fractions **(B)** ^{1}H-NMR spectrum of the synergistic LPE-enriched fraction. (600 MHz, CD_3OD)

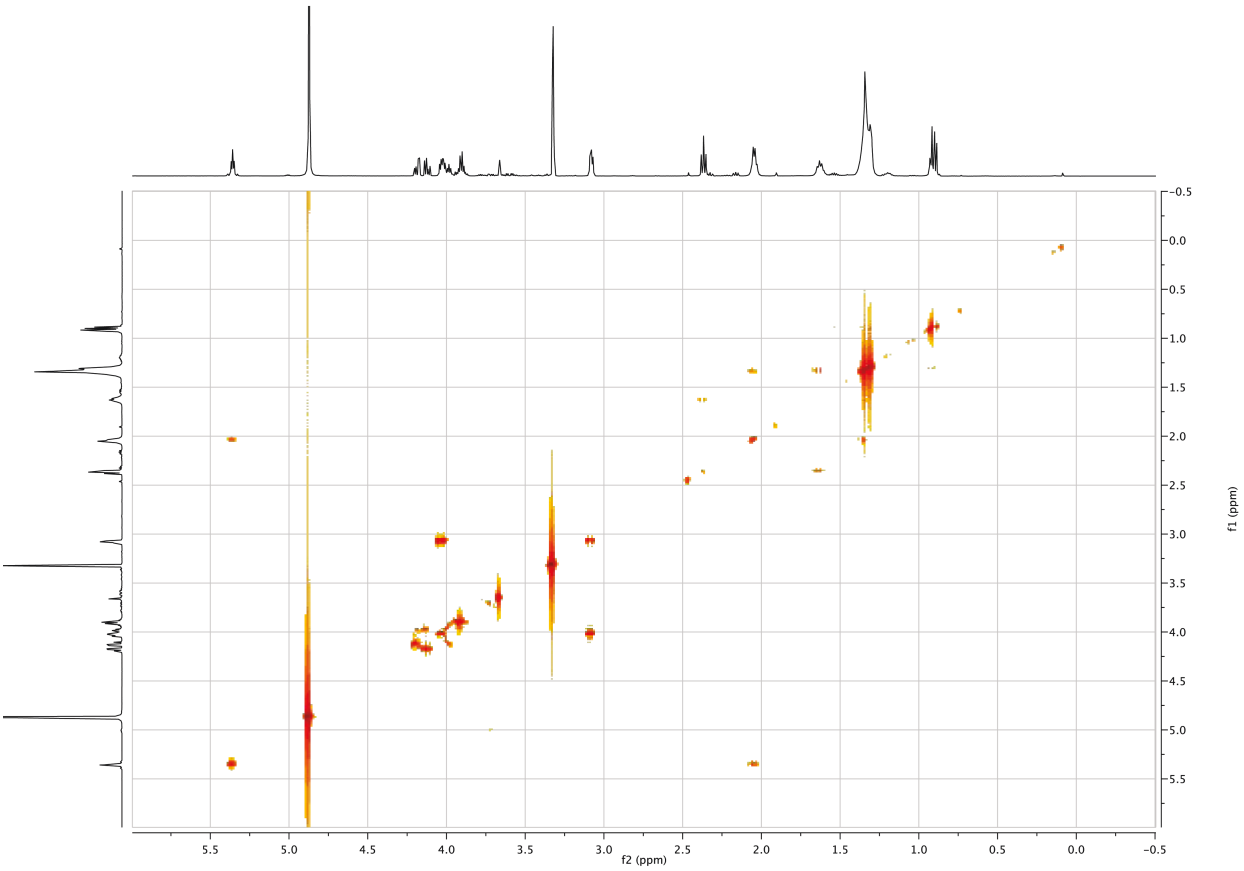


Figure S2. 13 LPE gCOSY spectrum (600 MHz, CD_3OD)

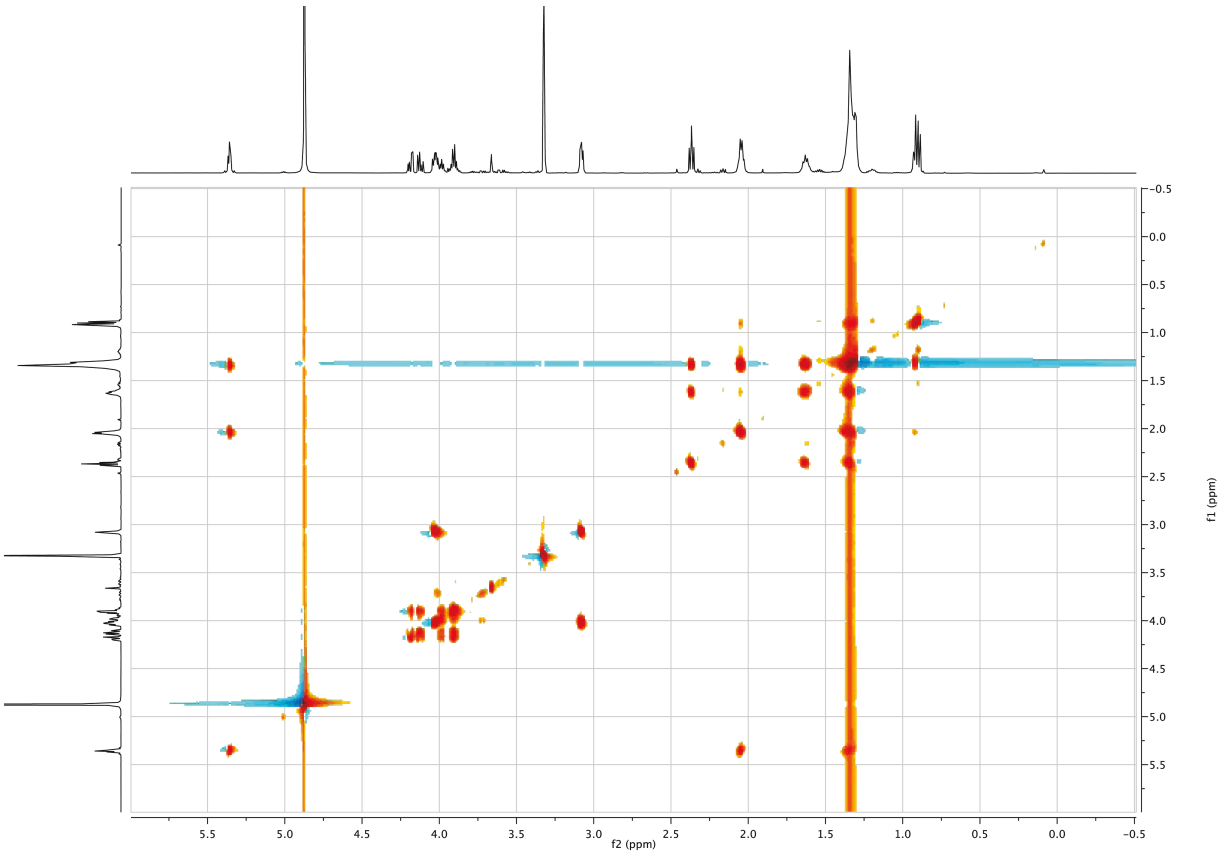


Figure S2.14 LPE TOCSY spectrum (600 MHz, CD_3OD)

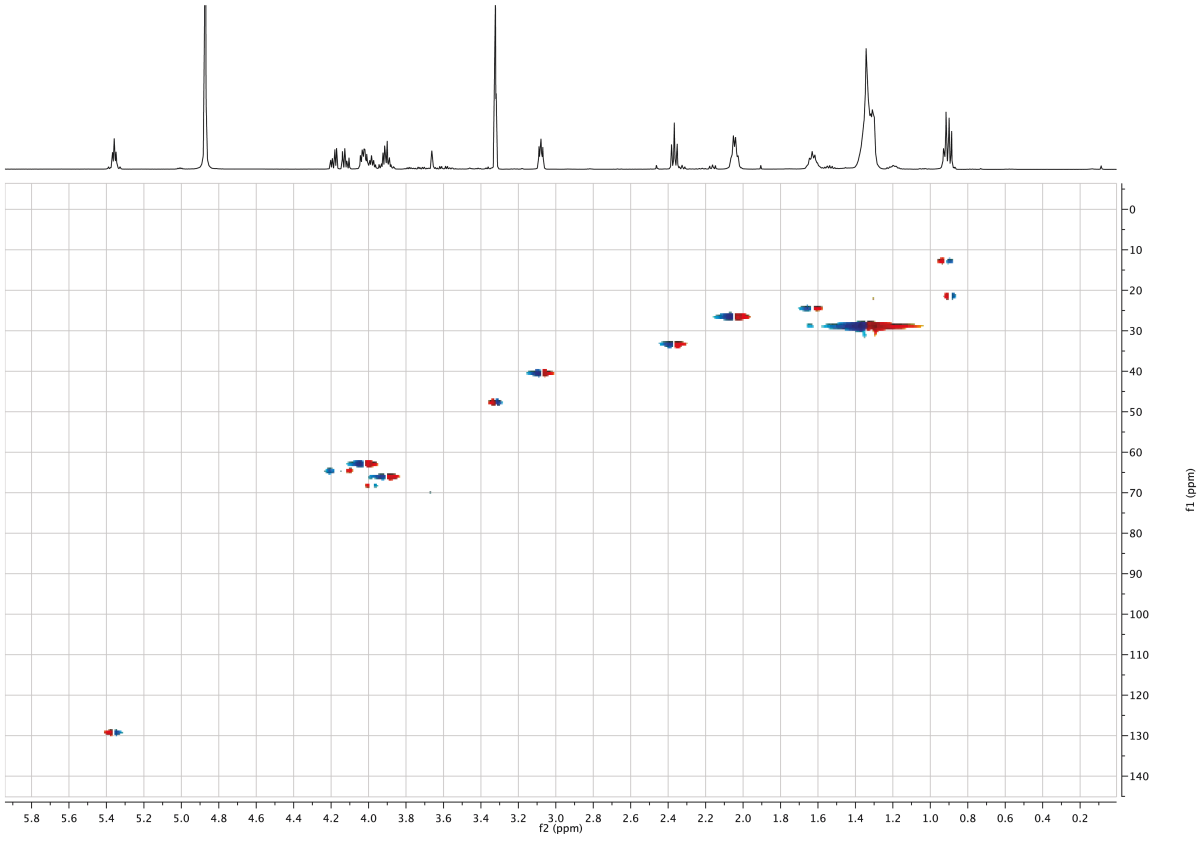


Figure S2. 15 LPE HSQCAD spectrum (600 MHz, CD_3OD)

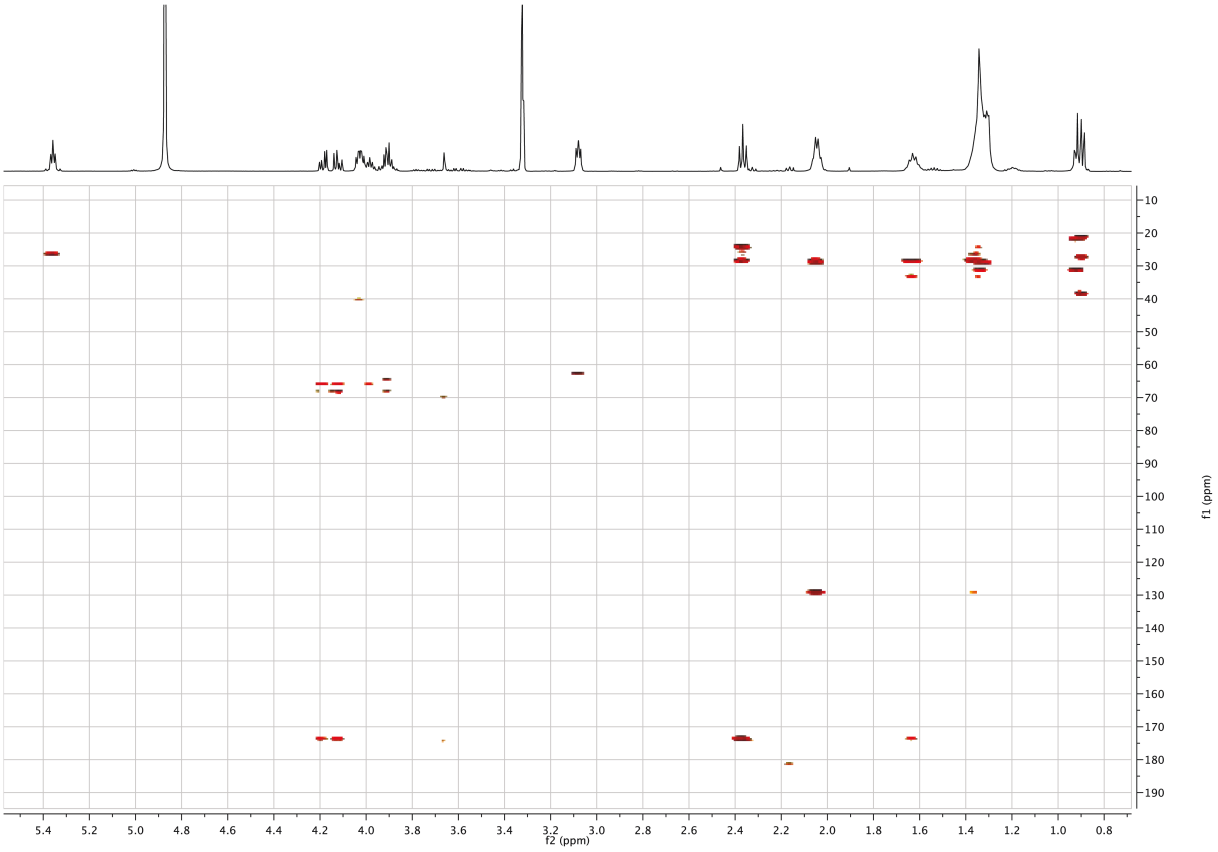


Figure S2.16 LPE HMBC spectrum (600 MHz, CD_3OD)

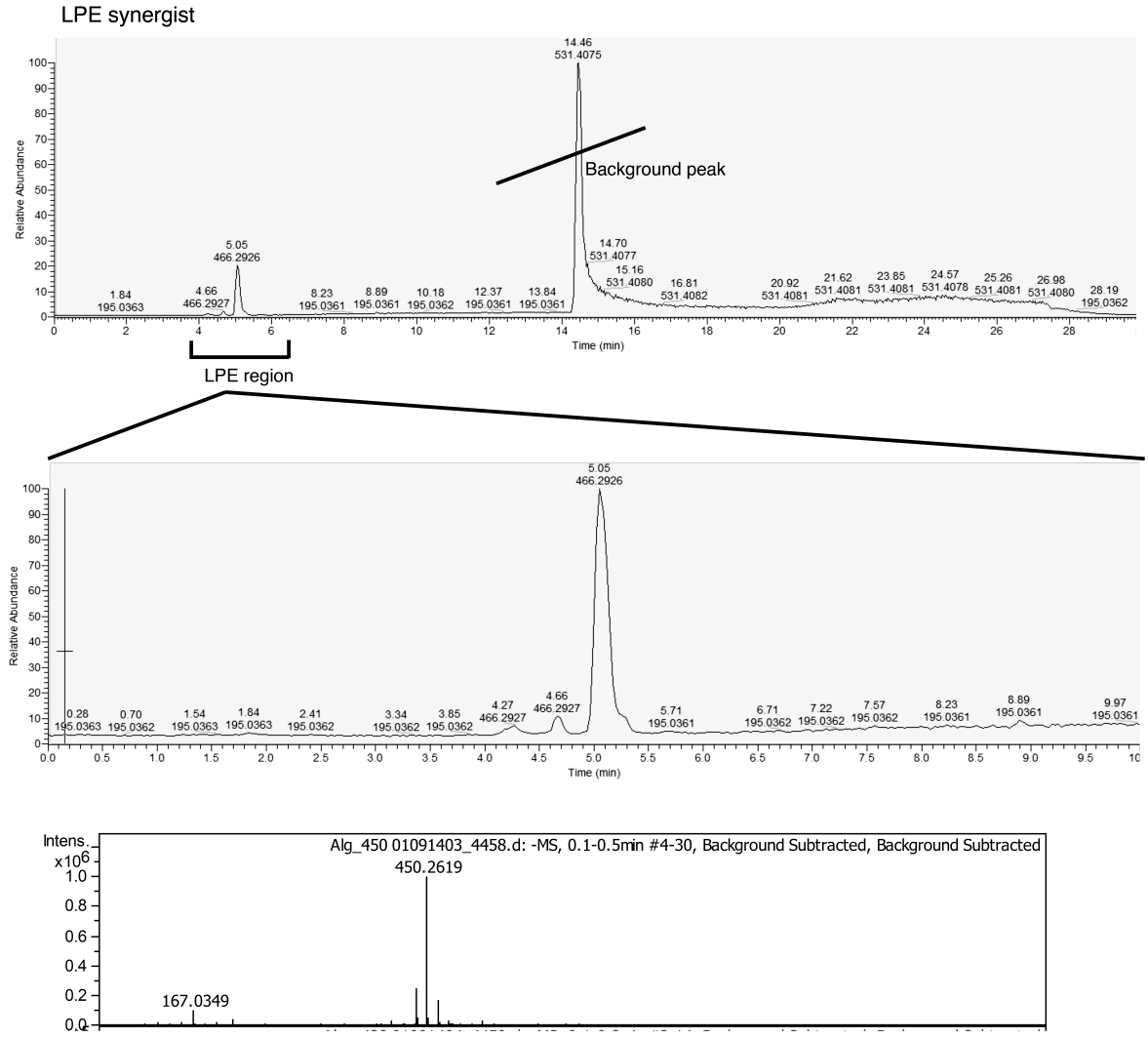
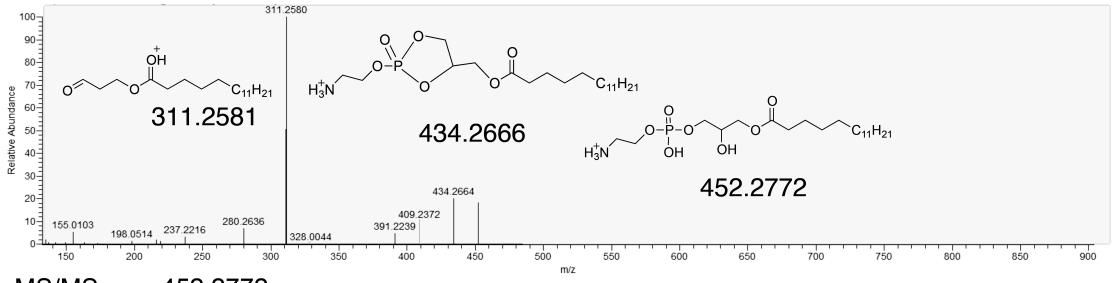
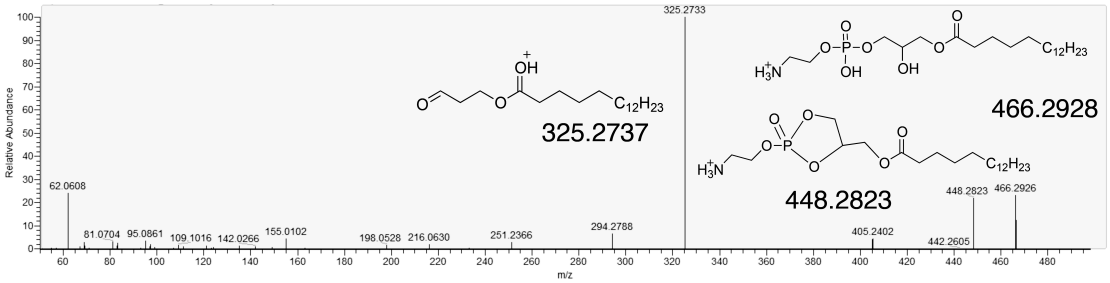


Figure S2.17 HRMS of LPEs

HRMS depicting the two major LPE components of the synergistic fraction.



MS/MS pos. 452.2772



MS/MS pos. 466.2928

Figure S2.18 Tandem mass spectrometry of LPEs

High resolution MSMS of two major LPEs (452/466) and predicted fragmentation patterns.

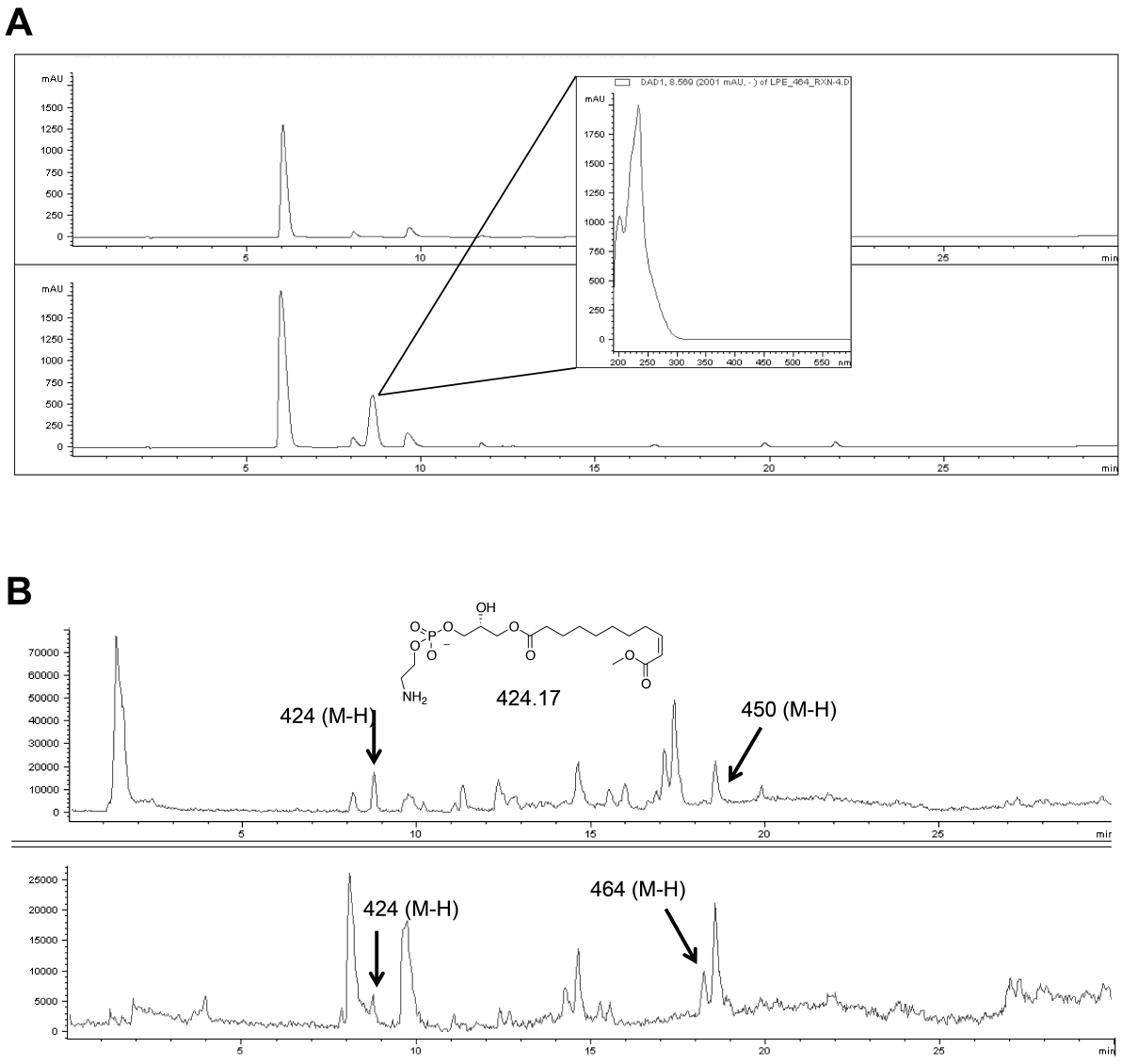


Figure S2. 19 Grubb's metathesis

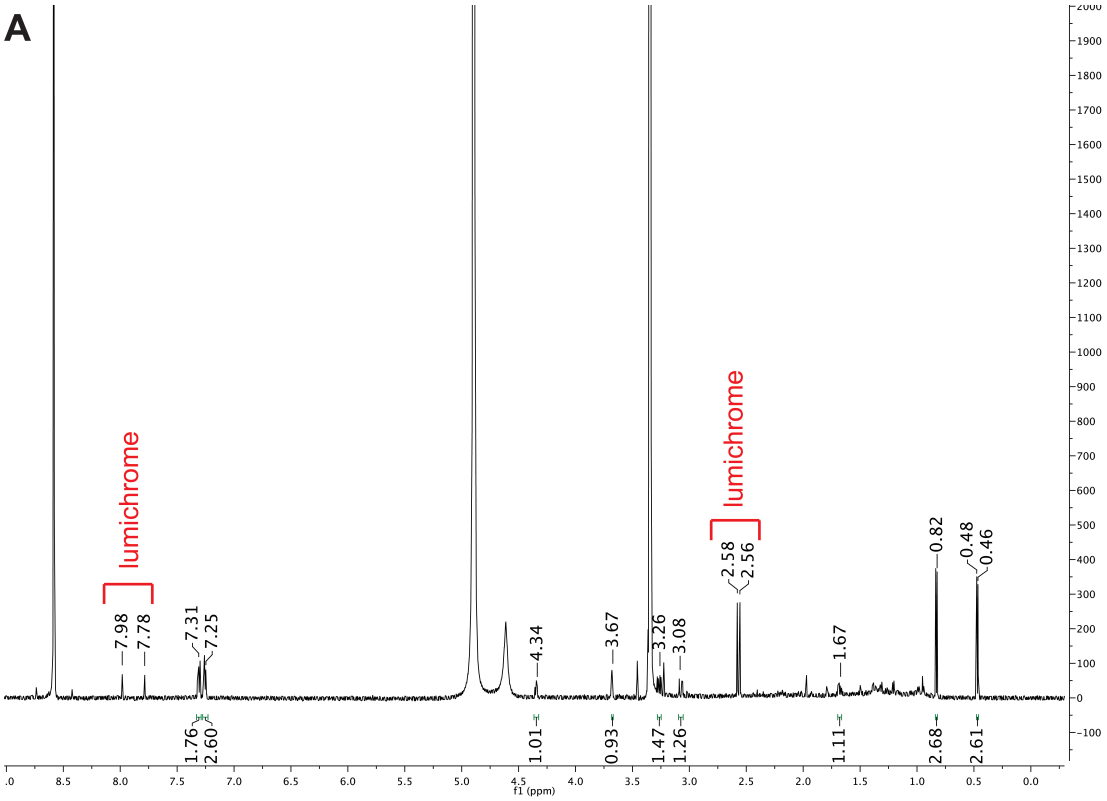
(A) Liquid chromatography UV trace of Grubb's reaction. Top spectrum shows the starting material and the bottom spectrum depicts the completed reaction. Inset indicates the UV of the α/β-unsaturated ester of the methyl acrylate product **(B)** LCMS (negative mode) shows both the starting material (indicated as 450 or 464) and the methyl acrylate product, indicated as 424.

Table S2.2 Commercially available lipids tested for rosette induction or inhibition

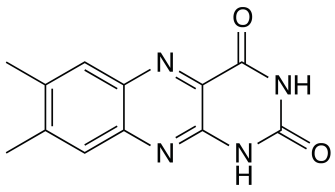
Name	Description	Inducing/Enhancing Activity	Inhibitory Activity	Company
Oleoylethanolamine	Ethanolamide	n.d.	n.d.	Sigma (O0383)
Oleic acid	Fatty acid	n.d.	n.d.	Avanti (861809)
Monosialoganglioside **	Glycosphingolipid	n.d.	n.d.	Sigma (G7641)
1-myristoyl-2-hydroxy-glycero-3-phosphoethanolamine	Lysophosphatidylethanolamine	n.d.	n.d.	Avanti (856735P)
1-octadecanoyl-sn-glycero-3-phosphoethanolamine	Lysophosphatidylethanolamine	n.d.	n.d.	Avanti (110700)
1-Oleoyl-2-hydroxy-sn-glycero-3-phosphoeethanolamine	Lysophosphatidylethanolamine	n.d.	n.d.	Avanti (846725P)
1-palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	Lysophosphatidylethanolamine	n.d.	n.d.	Avanti (856705P)
1-tetradecanoyl-sn-glycero-3-phosphoethanolamine	Lysophosphatidylethanolamine	n.d.	n.d.	Avanti (110697)
L-alpha-lysophosphatidylethanolamine	Lysophosphatidylethanolamine	n.d.	n.d.	Avanti (850095P)
Sphingosylphosphorylcholine	Lysosphingolipid	n.d.	n.d.	Sigma (S4257)
Sphingomyelin **	Sphingolipid	n.d.	n.d.	Avanti (860062)
Ceramide-1-phosphate	Sphingolipid	n.d.	n.d.	Avanti (860652P)
Sphingosine-1-phosphate	Sphingolipid	n.d.	n.d.	Sigma (S9666)
Ceramide phosphorylethanolamine	Sphingolipid	n.d.	n.d.	Sigma (C4987)
Dihydrosphingosine	Sphingosine precursor	n.d.	n.d.	Avanti (110758)

** : Also found to lack rosette inducing activity in (Alegado et al. 2012); n.d. : none detected

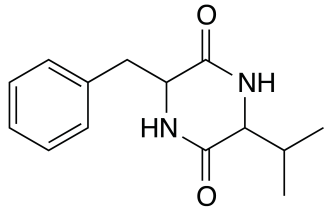
Appendix 3: Supplementary materials for Chapter 4



B



lumichrome
$C_{12}H_{10}N_4O_2$
HRMS (M+H)
measured: 243.0878
expected: 243.0877



val-phe-diketopiperazine
$C_{14}H_{18}N_2O_2$
HRMS (M+H)
measured: 247.1442
expected: 247.1441

Figure S3.1 Structure elucidation of lumichrome and val-phe-dkp isolated from Bx11
(A) Proton NMR spectrum (600 MHz, CD_3OD) of most active fraction from Bx11 ethyl acetate extract. Fraction contains both lumichrome and val-phe-DKP. **(B)** Lumichrome and val-phe-DKP structural information. High resolution mass spectrometry measured on QTOF LCMS (ESI).

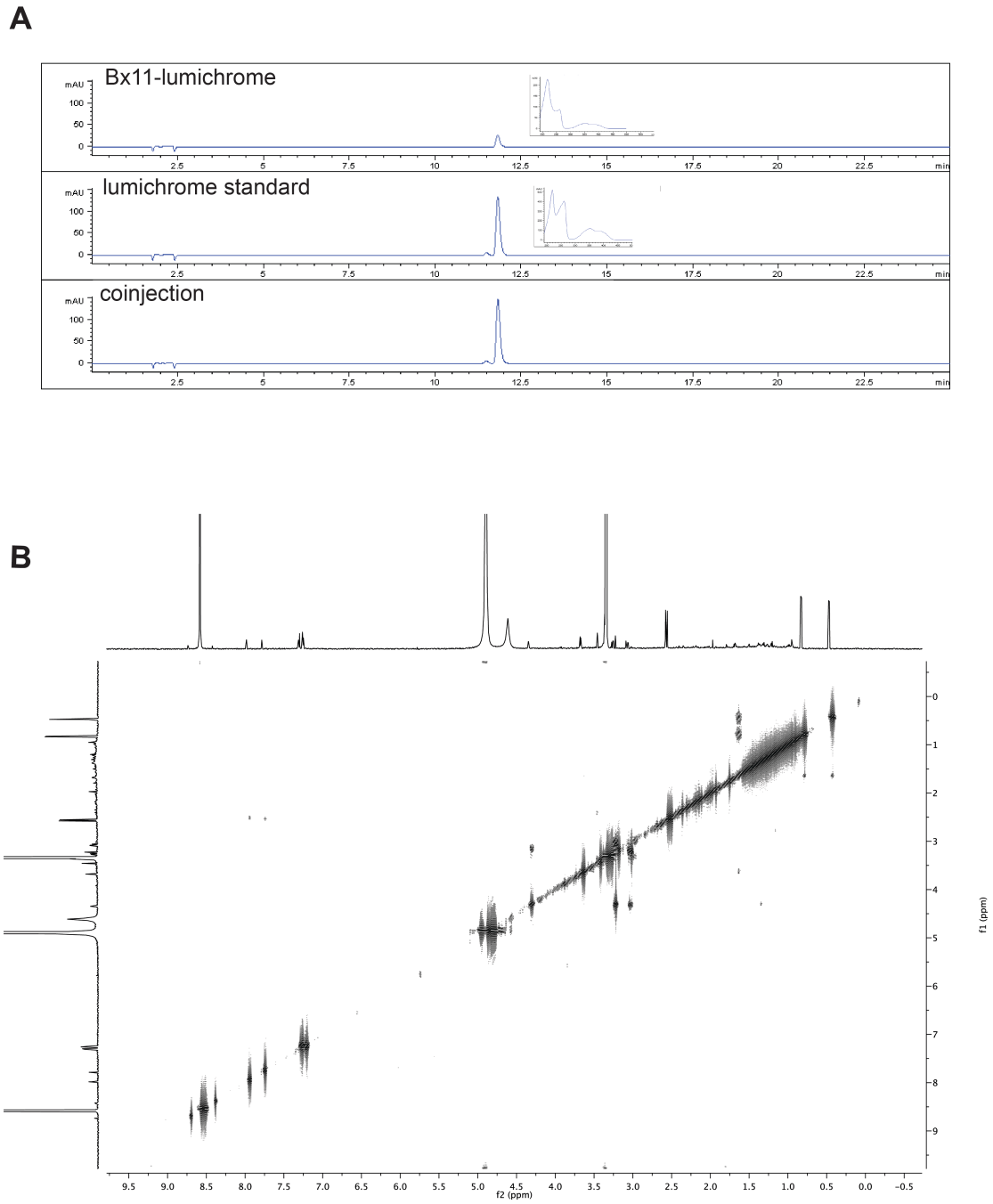


Figure S3. 2 Structural characterization of lumichrome and val-phe-DKP

(A) Co-injection onto LCMS (C18 column) of isolated lumichrome and commercially available lumichrome (Sigma Aldrich 103217). Gradient of 10%-100% acetonitrile + 0.1% formic acid. **(B)** COSY (600 MHz, CD_3OD) spectrum of Bx11 active fraction

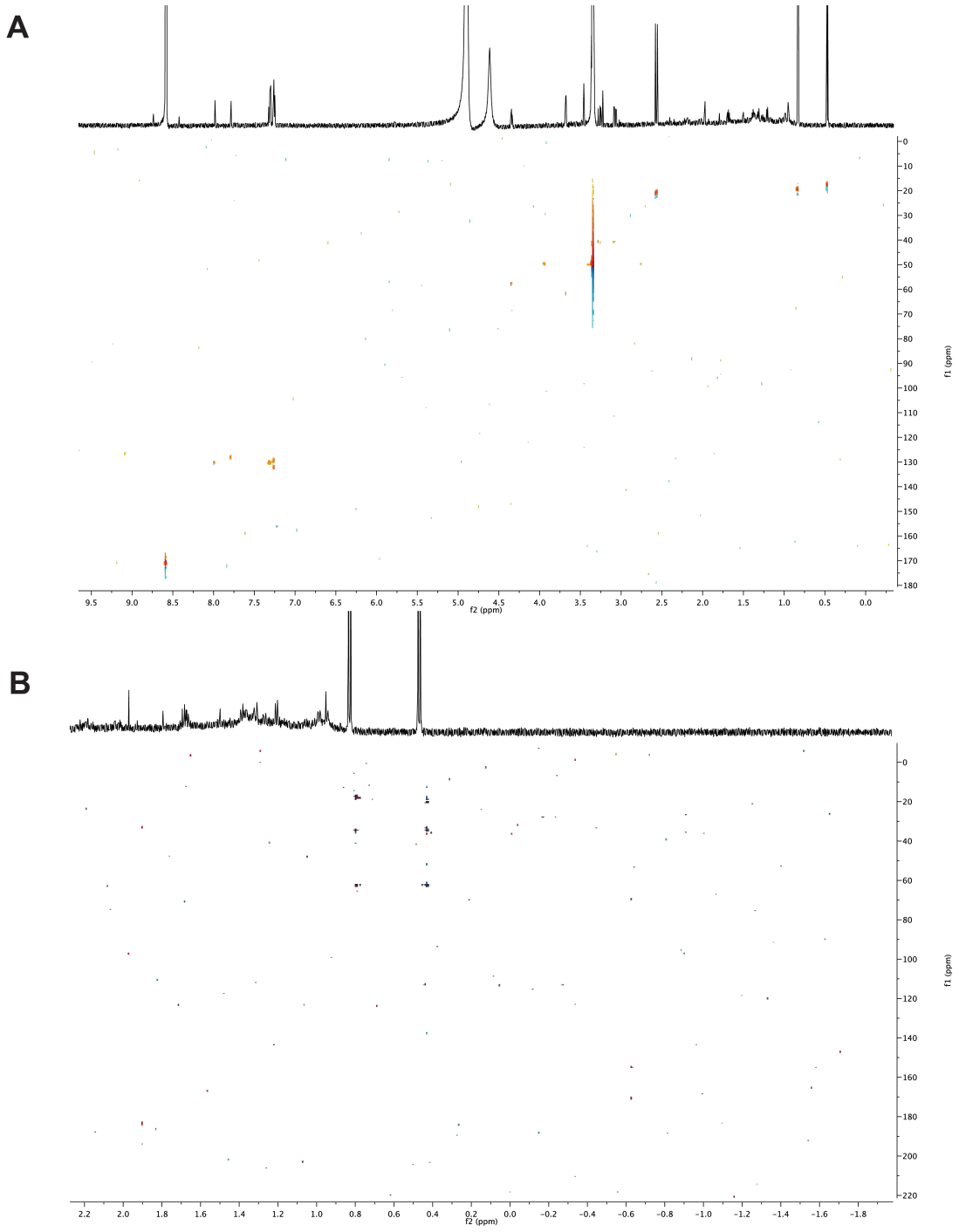
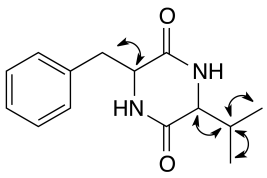
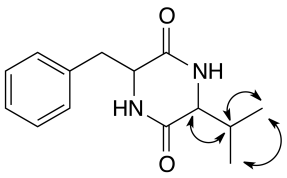


Figure S3. 3 2D NMRs of Bx11 active fraction
(A) HSQC spectrum (600 MHz, CD_3OD) **(B)** HMBC spectrum (600 MHz, CD_3OD)

A

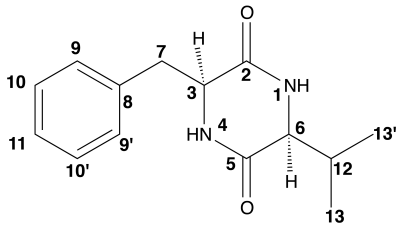


COSY key correlations



HMBC key correlations

B



Position	L-Val-L-Phe DKP (400 MHz, CD_3OD) ^{1}H, m, H (*J* in Hz)	D-Val-L-Phe DKP (400 MHz, CD_3OD) ^{1}H, m, H (*J* in Hz)	Isolated Val-Phe DKP (400 MHz, CD_3OD) ^{1}H, m, H (*J* in Hz)
1-NH	-	-	-
2	-	-	-
3	4.34, td, 1H (5.2, 1.6)	4.31, t, 1H (4.3)	4.35, td, 1H (5.2, 1.5)
4-NH	-	-	-
5	-	-	-
6	3.67, dd, 1H (4.6, 1.6)	2.97, dd, 1H (2.9, 0.9)	3.68, dd, 1H (4.6, 1.6)
7	3.26, dd, 1H (13.8, 5.4) 3.07, dd, 1H (13.8, 4.7)	3.29, dd, 1H (13.8, 4.1) 3.05, dd, 1H (13.8, 4.6)	3.27, dd, 1H (13.8, 5.3) 3.07, dd, 1H (13.8, 4.7)
8	-	-	-
9, 9', 11	7.27-7.23, m, 3H	7.34-7.29, m, 3H	7.27-7.24, m, 3H
10, 10'	7.33-7.28, m, 2H	7.28-7.24, m, 2H	7.33 – 7.29, m, 2H
12	1.68, m, 1H	2.17, m, 1H	1.68, m, 1H
13	0.83, d, 3H (7.1)	0.93, d, 3H (7.2)	0.83, d, 3H (7.0)
13'	0.47, d, 3H (6.8)	0.87, d, 3H (6.8)	0.47, d, 3H (6.7)

Figure S3.4 val-phe structural information

(A) Key correlations from COSY and HMBC spectra **(B)** List of NMR shifts comparing isolated and synthetized val-phe diketopiperazines.

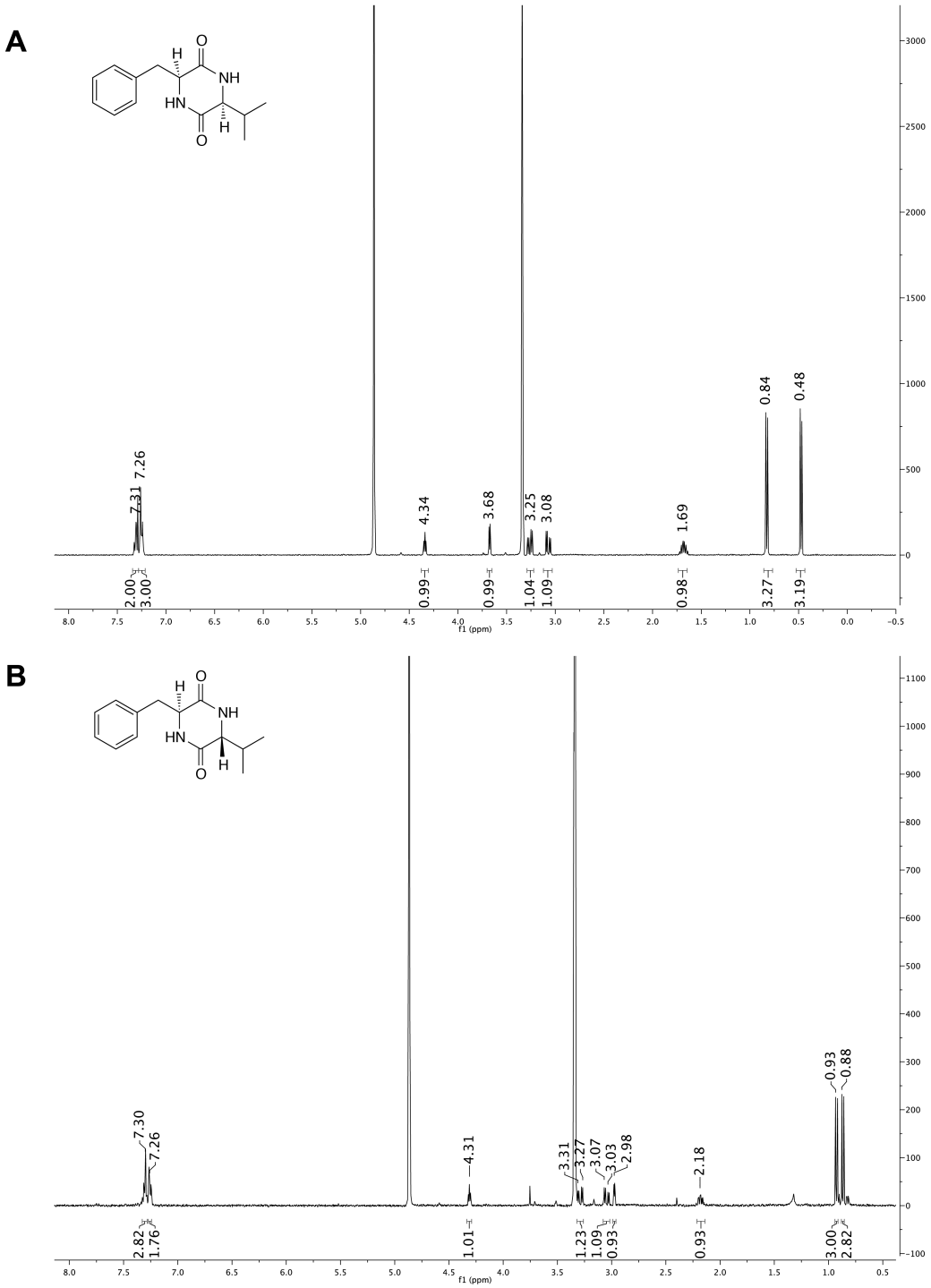


Figure S3. 5 ^{1}H NMR spectra for synthetic diketopiperazines
(A) ^{1}H NMR spectrum (400 MHz, CD_3OD) for L-val-L-phe DKP (B) ^{1}H NMR spectrum (400 MHz, CD_3OD) for D-val-L-phe DKP

Table S3. 1 Biosynthesis and virulence-related genes from the Bx11 genome

Gene name/type	Start (bp)	End (bp)
terpene	Chr1: 1,466, 257	1,487,096
NRPS-t1pks cluster	Chr1: 2,061,178	2,123,404
phosphonate cluster	Chr2: 606,670	648,439
terpene	Chr2: 49,851	82,384
terpene	Chr2: 107,235	139,986
hserlactone-NRPS	Chr2: 823,935	44, 896
Type III SS (Ysc, Hrp, FliI, Flh, Lcr)	Chr1: 389,035	406,163
Type VI SS (Rhs, VgrG)	Chr1: 543,303	555,468
Type VI SS (Rhs, VgrG, ClpB, Vas)	Chr1: 2,252,754	2,299,408
Type II/IV SS (Flp/Tad pilus)	Chr1: 2,902,982	2,917,994
Type III SS (FliI, Flh, Ysc)	Chr1: 2964458	2977530
Type I SS (Lap, Lss, RTX)	Chr1: 3,158,153	3,167,624
Toxin(PIN domain)	Chr1: 2,568,028	2,568,399
Antitoxin (VapB)	Chr1: 2,567,834	2,568,031
Type VI SS (Vas, Lmp, ClpB, VgrG)	Chr2: 147, 458	163,416
Type II/IV SS (Flp/Tad pilus)	Chr2: 213,055	219,564

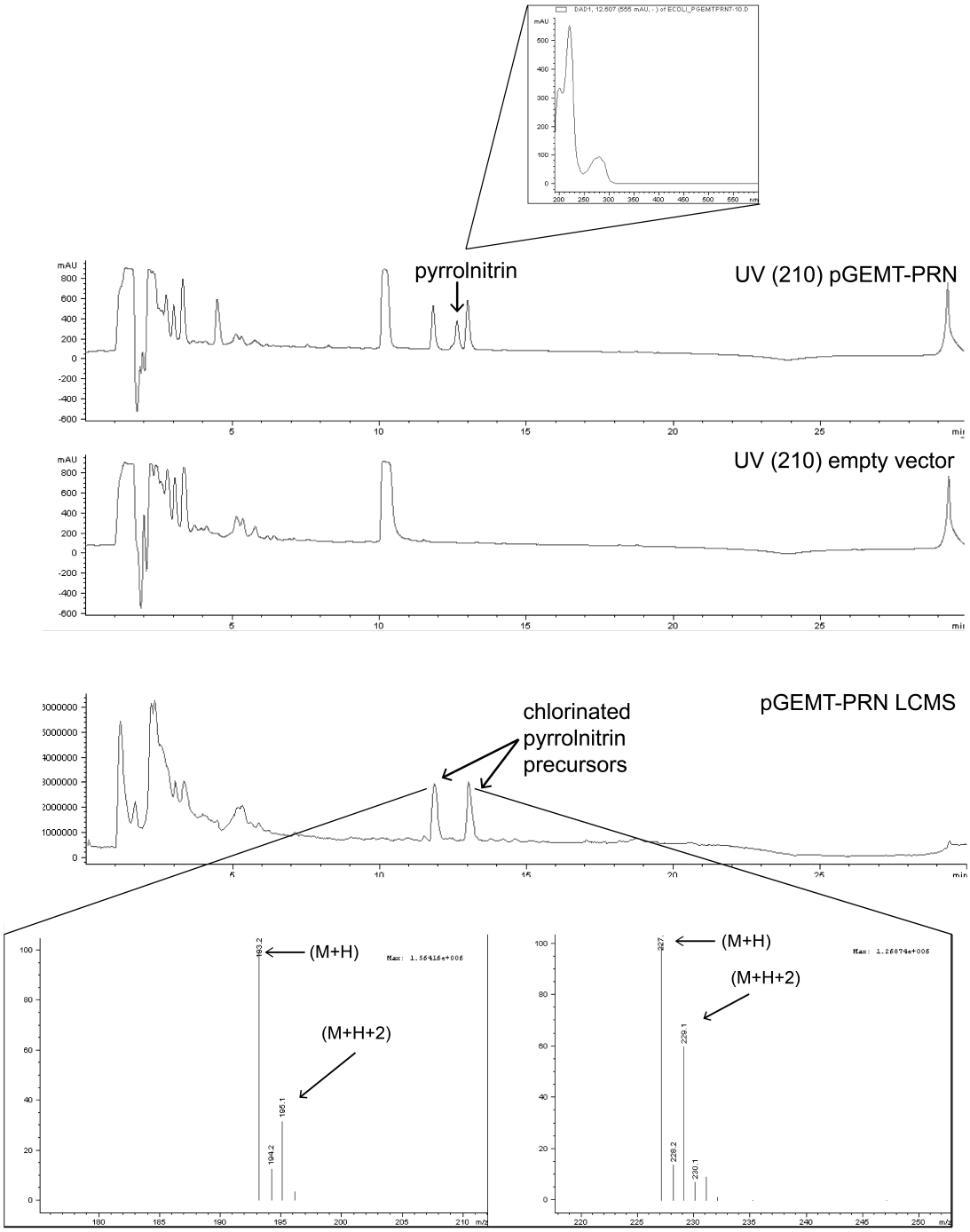


Figure S3.6 Pyrrolnitrin expression from PRN operon cloned into pGEMT vector

Pyrrolnitrin producton was detected by LCMS. (**Top**) UV (210) traces reveal peaks associated with the clone containing PRN and were not detected in the control. UV vis of the indicated peak resembles the known UV for pyrrolnitrin. (**Bottom**) LCMS confirmed the presence of precursors to pyrrolnitrin with isotope patterns consistent with chlorination. Pyrrolnitrin does not ionize well and could not be detected by MS under these conditions