Title
Bacterial cues regulate multicellular development and mating in the choanoflagellate, S. rosetta

Author
Woznica, Arielle

Publication Date
2017

Peer reviewed|Thesis/dissertation

Bacterial cues regulate multicellular development and mating in the choanoflagellate, *S. rosetta*

By

Arielle Woznica

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nicole King, Chair
Professor Russell Vance
Professor Diana Bautista
Professor Brian Staskawicz

Spring 2017

Abstract

Bacterial cues regulate multicellular development and mating in the choanoflagellate, *S. rosetta*

By

Arielle Woznica

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Nicole King, Chair

Animals first diverged from their unicellular ancestors in oceans dominated by bacteria, and have lived in close association with bacteria ever since. Interactions with bacteria critically shape diverse aspects of animal biology today, including developmental processes that were long thought to be autonomous. Yet, the multicellularity of animals and the often-complex communities of bacteria with which they are associated make it challenging to characterize the mechanisms underlying many bacterial-animal interactions. Thus, developing experimentally tractable host-microbe model systems will be essential for revealing the molecules and mechanisms by which bacteria influence animal development.

The choanoflagellate *Salpingoeca rosetta*, one of the closest living relatives of animals, has emerged as an attractive model for studying host-microbe interactions. Like all choanoflagellates, *S. rosetta* feeds on bacteria; however, we have found that interactions between *S. rosetta* and bacteria extend beyond those of predator and prey. In fact, two key transitions in the life history of *S. rosetta*, multicellular "rosette" development and sexual reproduction, are regulated by environmental bacteria.

The experimental tractability of *S. rosetta* allowed us to characterize the molecules and regulatory logic underpinning the bacterial regulation of rosette development (Chapters 2 and 3). We found that the bacterium *Algoriphagus machipongonensis* produces three classes of structurally distinct lipids that are interpreted by *S. rosetta* as activators, synergistic enhancers, and inhibitors of rosette development. Although activating sulfonolipid RIFs (Rosette Inducing Factors) elicited relatively low levels of rosette development, the combined activity of the RIFs and synergizing lysophosphatidylethanolamines (LPEs; which alone had no detectable activity) was sufficient to fully recapitulate the rosette-inducing activity of *Algoriphagus* bacteria. Moreover, we identified a potent antagonist of the RIFs, IOR-1 (Inhibitor of Rosettes), but found that the synergistic activities of the RIFs and the LPEs overcame the inhibitory activities of IOR-1. We hypothesize that the integration of multiple activating, enhancing, and inhibitory bacterial cues act to ensure that rosette development is not initiated under the wrong environmental conditions.

Until recently, bacteria were not known to influence any life history transition in *S. rosetta* other than rosette development. We serendipitously discovered that the bacterium *Vibrio fischeri* produces an "aphrodisiac" that regulates sexual reproduction in *S. rosetta* (Chapter 4).

To our knowledge, the interaction between *Vibrio* and *S. rosetta* is the first known example of bacteria regulating mating in a eukaryote. After observing that *S. rosetta* cells aggregate into large swarms in response to *Vibrio* bacteria, we demonstrated that swarming, a behavior that had not been previously observed in choanoflagellates, was a prelude to sexual fusion. We next found that *Vibrio* secreted a chondroitinase aphrodisiac (EroS) that depolymerized chondroitin sulfate, a glycosaminoglycan previously thought to be restricted to animals, in the *S. rosetta* extracellular matrix. Finally, we determined mating in *S. rosetta* was triggered by low cell densities of *Vibrio* bacteria, and picomolar concentrations of EroS (as well as other bacterial chondroitinases), indicating that bacteria could plausibly trigger *S. rosetta* swarming and mating in the environment. We predict that the presence of chondroitinase-producing bacteria may indicate environmental factors that favor mating in *S. rosetta*.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix

Acknowledgements

"Above all, one must have a feeling for the organism."
- Barbara McClintock

First, I must thank Barbara McClintock for being my first science hero, and for being dead-on.

I would like to acknowledge my advisor, Nicole King, for allowing me to pursue inspiring and terrifying projects, for teaching me the value of a narrative, and for getting excited about even the most preliminary data. I have sincerely loved working with choanos.

I must thank the past and present members of the King lab. Thank you for being my mentors, collaborators, sounding boards, workout buddies, and moral support. Many of you have become my closest friends, and I am fortunate to have gotten to hang out with you every. single. day.

Finally, an enormous thank you to my family. I am so grateful to my parents, who encouraged and enabled my creativity and curiosity, and who have supported me in everything that I do. I must thank Leah, who got a real job out of college so she could afford to buy plane tickets and visit her big sister. I'd like to thank Sammy for providing unrelenting puppy love (and who deserves to be a co-author on every paper and fellowship that I have submitted).
And finally, thank you Alex, for celebrating the good-science days, for making the bad-science days not matter, and for telling me often that I'm the best scientist you know. On to the next adventure.

Chapter 1

A choanoflagellate model for host-microbe interactions

Bacterial influences on animals: past and present

The first bacteria evolved over three billion years ago, and shaped the environment in which eukaryotes would evolve some two billion years later[1,2]. It was within oceans dominated by bacteria that animals first diverged from their unicellular ancestors, and animals have lived in close association with bacteria ever since[3-6]. Though little is known about how bacteria may have influenced animal origins[7], diverse antagonistic (i.e. predator-prey and host-pathogen) and cooperative interactions with bacteria helped drive animal evolution and radiation[8,9]. The shared ancestry of bacteria and animals is evidenced by their genomes, and by the countless ways in which bacteria critically shape the biology of animals today[8].

Bacterial-animal interactions have been widely studied within the context of pathogenesis. Indeed, animals employ many mechanisms to manage their microbial environment. The production of antimicrobial peptides allows animals to protect themselves against environmental bacteria[10], and the recognition of microorganism-associated molecular patterns (MAMPs) helps animals defend against invading pathogens[11]. However, beneficial interactions with bacteria are also fundamentally important to animal health. Bacteria play an integral role in animal metabolism and nutrition by degrading otherwise indigestible dietary substances, and bacteria are essential for proper animal development and morphogenesis, processes that were long thought to be autonomous.

Only recently have we begun to fully appreciate the extent to which bacteria are partners in animal development. The symbiosis between the squid *Euprymna scolopes* and the bioluminescent bacterium *Vibrio fischeri*, in which *V. fischeri* colonizes and induces the morphogenesis of the squid's "light organ," first revealed mechanisms by which symbiotic bacteria can induce morphogenesis in animals[12-14]. Animal model systems have since yielded important insights into how commensal bacteria stimulate immune system development[15] and regulate gut morphogenesis[16] in vertebrates. Environmental bacteria that are not stably associated with animal hosts also direct developmental transitions in animals, producing exogenous molecular cues that trigger larval settlement and morphogenesis in many marine invertebrate larvae, including sponges, corals, and tubeworms[17-20].

Yet, the multicellularity of animals, and the often-complex and unseen communities of bacteria with which they are associated, make studying the molecular dialogue underlying bacterial-animal interactions challenging. Experimentally tractable host-microbe associations have proven essential for uncovering mechanisms by which bacteria influence animal development. For example, the binary nature of the *V. fischeri–E. scolopes* symbiosis (in which one microbe interacts with one animal) facilitated the isolation of bacterial factors that induce tissue morphogenesis in the squid, and germ-free animal models allow us to examine interactions between animals and a reduced number of bacterial species. Nonetheless, because relatively few tractable systems exist for studying host-microbe associations, only a handful of bacterial molecules that influence animal development have been isolated and characterized[21,22].

S. rosetta as a model for host-microbe interactions

The choanoflagellate *Salpingoeca rosetta* has recently emerged as an attractive model for studying host-microbe interactions. Choanoflagellates are a group of microbial eukaryotes that are found worldwide in fresh, brackish, marine, and even hypersaline waters[23]. As the closest living relatives of animals, choanoflagellates provide a phylogenetically relevant system for illuminating fundamental aspects of bacterial-animal interactions (Figure 1.1A). Choanoflagellates survive by eating bacteria, and have a conserved cellular architecture that is fine-tuned for bacterial prey capture: by undulating a flow-generating apical flagellum, choanoflagellates draw bacteria into a collar of actin-filled microvilli, where the bacteria are trapped and phagocytosed (Figure 1.1B,D). In this way, choanoflagellates are structurally and functionally akin to the collar cells (termed choanocytes) of sponges, specialized cells that eat bacteria. Although previously thought to be restricted to choanoflagellates and sponges, collar cells have since been described in diverse animals, where they exhibit epithelial-like apical-basal polarity and cellular structure[7]. Thus, we hypothesize that choanoflagellates likely formed the basis for the evolution of animal epithelial cells that today mediate interactions with bacteria[24,25]. Furthermore, choanoflagellate genomes have revealed that many genes once described to be animal-specific, including genes important for intercellular signaling (e.g. cadherins and receptor tyrosine kinases) and bacterial recognition (e.g. C-type lectins) in animals, are present in choanoflagellates[26-29]. Thus, choanoflagellates may offer us a unique opportunity to uncover ancestral mechanisms governing bacterial-animal associations.

It is hypothesized that the first animal was a spheroidal colony whose exterior, flagellated cells were specialized for prey capture and locomotion, and whose interior cells were specialized for digestion and mating[25]. Many species of choanoflagellates can develop, through repeated rounds of cell division, from a solitary cell into a multicellular colony, thereby recapitulating an evolutionary transition that likely occurred during the earliest stages of animal evolution (Figure 1.1B,C).

Colony formation has been best studied in the marine choanoflagellate, *Salpingoeca rosetta* (Figure 1.1B,C, Figure 2.1). In *S. rosetta*, multicellular "rosette" colonies develop from a single founding cell that undergoes serial rounds of oriented cell division, with the sister cells remaining stably attached. The orientation of the nascently divided cells around a central focus (with the flagella pointing outwards) and the production of extracellular matrix (including the secretion of the C-type lectin, Rosetteless[30]) results in the formation of a spherical, multicellular rosette[31,32]. Indeed, the process of rosette development mirrors early embryogenesis, and the resemblance of *S. rosetta* to blastula-stage animal embryos is striking.

Importantly, interactions between *S. rosetta* and bacteria extend beyond those of predator and prey. Cues produced by environmental bacteria regulate rosette development in *S. rosetta* (Figures 1.1B, 1.2A). Although bacteria from diverse genera and phyla can induce rosette development (for further discussion of this point, refer to the Appendix), the interaction between *S. rosetta* and the co-isolated, rosette-inducing bacterium *Algoriphagus machipongonensis* has served as an ecologically relevant and tractable model for studying host-microbe interactions. The ability to culture *S. rosetta* and *Algoriphagus* independently or together, along with the establishment of a simple rosette development bioassay, facilitated the isolation of the first rosette-inducing molecule, the sulfonolipid RIF-1 (Rosette Inducing Factor-1).

While the isolation of RIF-1 first demonstrated the biochemical tractability of bacterial-choanoflagellate interactions[33], RIF-1 was not sufficient to recapitulate rosette-inducing activity

of live *Algoriphagus* bacteria[34]. This observation prompted me to investigate the possibility that additional *Algoriphagus* molecules might be required to regulate rosette development (Chapter 2 and Chapter 3)[35,36].

The rich life history of *S. rosetta*

In addition to rosette development, *S. rosetta* can transition between diverse other cell types and morphologies, including: linear chain colonies, slow swimmer cells, fast swimmer cells, and substrate-attached thecate cells[31] (Figure 1.2A). *S. rosetta* also has a sexual life cycle. The *S. rosetta* genome contains a full meiotic "toolkit," and genome-wide haplotype blocks in laboratory isolates of *S. rosetta* support a history of meiotic recombination. Moreover, *S. rosetta* can partake in sexual fusion under starvation conditions[37] (Figure 1.2B), although sexual fusion occurs infrequently (in <2% of the population).

While it is sometimes possible to enrich for distinct cell types by altering laboratory culturing techniques, our understanding of the diverse life history transitions in *S. rosetta* is largely qualitative. In fact, there had been no empirical evidence for bacteria regulating a life history transition in *S. rosetta* other than rosette development until recently, when we serendipitously observed that the bacterium, *V. fischeri*, induces sexual reproduction in *S. rosetta* (Chapter 4).

Why a choanoflagellate model?

It is now widely accepted that commensal bacteria profoundly influence human health. As such, new animal systems for studying host-microbe interactions are being rapidly developed. What are the advantages of using a choanoflagellate model to study host-microbe interactions? First, there are few technical limitations to studying choanoflagellates and their associated bacteria. Choanoflagellates can be cultured quickly and continuously, and can be grown in the presence or absence of different bacteria. Moreover, most bacteria co-isolated with choanoflagellates are culturable, since the bacteria are typically planktonic and aerobic. Second, at least two key life history transitions in *S. rosetta*– rosette development and sexual reproduction– are triggered by a single bacterial species (*Algoriphagus* and *V. fischeri*, respectively). Such binary relationships are rare, yet are key to uncovering the molecules and mechanisms underlying host-microbe interactions. Finally, the breadth of interactions between bacteria and animals is remarkable, and many experimental models will be required to explore the diversity of these relationships. Nonetheless, because choanoflagellates and animals have a shared ancestry, *S. rosetta* has the potential to highlight evolutionarily conserved features of host-microbe interactions.

FIGURES

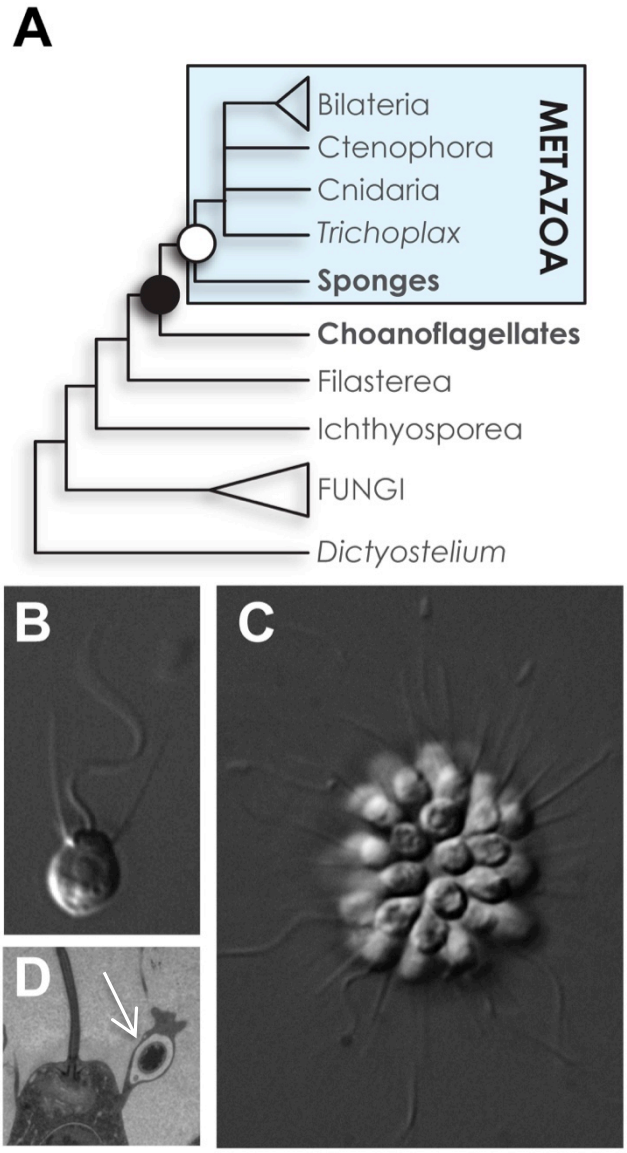


Figure 1.1. An introduction to choanoflagellates (A) Phylogenetic analyses reveal that choanoflagellates are the closest living relatives of animals. Comparisons among choanoflagellates, sponges, and eumetazoans can inform us about the last common ancestors of animals (white circle) and their last common ancestor with choanoflagellates (black circle). The choanoflagellate *S. rosetta* develops from a single cell (B) into a multicelled rosette (C) in response to bacterial cues[31]. (D) Choanoflagellates form intimate associations with bacteria. Shown is a bacterium (arrow) following engulfment by the choanoflagellate[38].

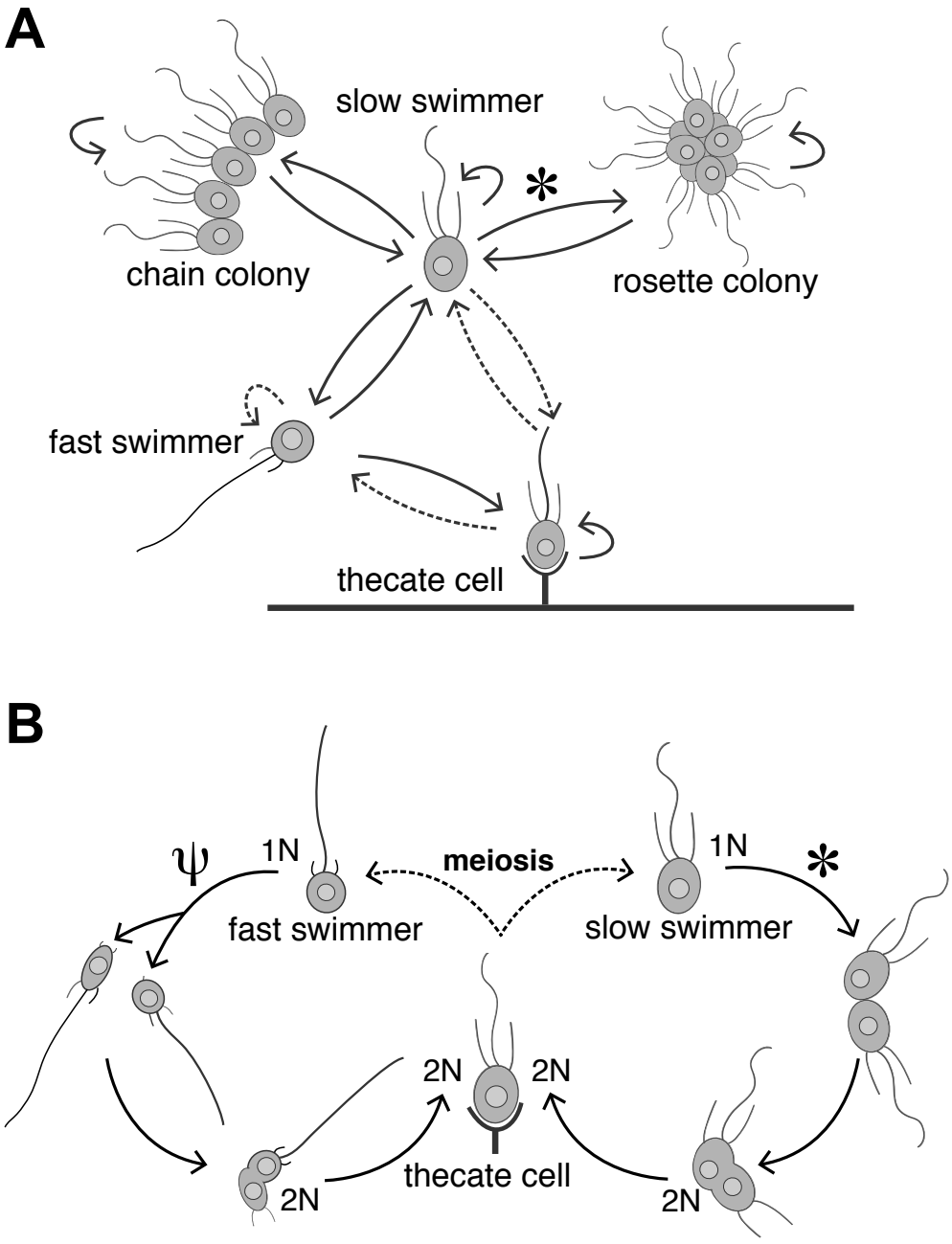


Figure 1.2. Life history of the choanoflagellate, *S. rosetta* (A) *S. rosetta* has a rich life history and differentiates into diverse cell-types. Differentiation from a single cell into a rosette is regulated by bacteria, including the co-isolated bacterium *Algoriphagus* (*). (B) Mating in *S. rosetta* is induced by two distinct environmental cues. Starvation (Ψ) triggers low frequencies of mating in fast swimmers. Some species of *Vibrio* bacteria, including *Vibrio fischeri*, trigger robust mating in slow swimmers (*).

Chapter 2

Bacteria regulate choanoflagellate development with lipid activators, inhibitors, and synergists

The results presented here were published as part of the following paper:

Woznica, A.*, Cantley, A.M.*, Beemelmanns, C., Freinkman, E., Clardy, J., King, N. (2016) Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. USA*, 113(28): 7894-7899.

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 activate, enhance, and inhibit multicellular development in a choanoflagellate.

Significance Statement

Bacterial symbionts profoundly influence the biology of their animal hosts, yet complex interactions between animals and their resident bacteria often make it challenging to characterize the molecules and mechanisms. Simple model systems can reveal fundamental processes underlying interactions between eukaryotes and their associated microbial communities, and provide insight into how bacteria regulate animal biology. In this study we isolate and characterize bacterial molecules that regulate multicellular development in the closest living relatives of animals, the choanoflagellate. We find that multiple bacterially-derived lipids converge to activate, enhance, and inhibit choanoflagellate multicellular development.

Introduction

The foundational event in animal origins, the transition to multicellularity [3,4,39], occurred in oceans filled with diverse bacteria [2,6,14,40]. 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 [8,12,13,15,18,41-48]. 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.

Choanoflagellates, a group of microbial eukaryotes that are the closest living relatives of animals [28,49-51], 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 [25,52,53]. 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 [24,25,54].

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 2.1). 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 [30-32]. Rosettes resemble morula stage embryos and the transition to multicellularity in *S. rosetta* evokes ancestral events that spawned the first animals [25,53,55].

The initiation of rosette development was recently found to be induced by a co-isolated environmental bacterium, *Algoriphagus machipongonensis* (phylum Bacteroidetes;[33,56]. The ecological relevance of the *Algoriphagus* - *S. rosetta* interaction is evidenced by the co-existence of these organisms in nature [33], and the predator-prey relationship between choanoflagellates and bacteria [38,52]. 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 [57], and prey capture studies reveal that rosettes collect more bacterial prey/cell/unit time than do single cells [58]. 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 molecular 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 (Figs. 2.2A and 2.2C), raises the possibility that additional molecules might be required [33]. 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.

Results

A newly identified 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* (see SI Materials and Methods). This culture, called 'SrEpac' (for *S. rosetta* + *E. pacifica*; [37], reproducibly yields high percentages of cells in rosettes (>80%) in response to live *Algoriphagus*, *Algoriphagus* outer membrane vesicles (OMVs; see SI Text) isolated from conditioned medium, and *Algoriphagus* bulk lipid extracts (Figure 2.2A; see SI Appendix, Fig. 2S1). 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; Fig. 2C), consistent with previous results using a different *S. rosetta* culture [33].

Because *Algoriphagus* bulk lipid extracts elicit the same rosette development response as live bacteria (Figure 2.2A), 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 (see SI Materials and Methods). Only fraction 11 was sufficient to induce rosette development, whereas all other lipid fractions lacked rosette-inducing activity at all concentrations tested (Figure 2.2B). 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 2.2C; see SI Appendix, Fig. 2S2). Structural analysis by NMR, high resolution mass spectrometry (HRMS), and tandem mass spectrometry (MSMS) revealed that the RIF mix contained RIF-1 and two structurally related but previously uncharacterized sulfonolipids with approximate molecular weights of 605 Da and 593 Da (see SI Appendix, Figs. 2S2 -17). 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 [59]. Although sulfonolipids have been reported to contribute to the gliding motility of Bacteroidetes bacteria [60,61], almost nothing is known regarding their potential roles as signaling molecules.

Additional activity-guided fractionation by HPLC allowed us to isolate pure samples of RIF-1 (35, 43) and of the 605 Da sulfonolipid. Purified RIF-1 induced maximal (~1.5%) rosette development at femtomolar to nanomolar concentrations (Figure 2.2C, inset). In contrast, the purified 605 Da sulfonolipid (hereafter "RIF-2") elicited 7-fold higher levels of rosette development (10.5% of cells in rosettes; Figure 2.2C) than RIF-1, although at micromolar concentrations. The planar structure of RIF-2 (Figure 2.3A) was determined by one and two-dimensional NMR (see SI Appendix Table S1, Figs. 2S3-17), and was found to closely resemble RIF-1, with the exception of slight structural variations of the capnoid base, which contains a double bond at C-4 and a hydroxyl group at C-6.

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. Although HRMS and HRMSMS data suggest 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 characterizing the activity of the 593 Da sulfonolipid (see SI Appendix, Figs. 2S2, S6). However, because the combination of RIF-2 and the 593 Da sulfonolipid induced rosettes at levels

indistinguishable from those of RIF-2 alone (see SI Appendix, Fig. 2S18), 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 2.2C) 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.

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 2.2B; see SI Appendix, Materials and Methods). 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 (see SI Appendix, Table S2). Therefore, we used bioactivity-guided fractionation in the presence of RIF-2 to determine the molecular basis for inhibition. HRMS and NMR experiments, together with total synthesis [36], allowed us to propose the absolute structure for the 351 Da molecule (hereafter referred to as Inhibitor of Rosettes-1, "IOR-1"; Figure 2.3C). Comprehensive methods detailing 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 [36].

Nanomolar concentrations of IOR-1 completely inhibits the ability of RIF-2 to induce rosette development, and reduces rosette development in the presence of the RIF mix (see SI Appendix, Fig. 2S19). IOR-1 is a capnine lipid that resembles the capnoid backbone of *Algoriphagus* RIFs (Figs. 2.3A and C). 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, we infer that the combined effects of the RIFs are sufficient to partially overcome the presence IOR-1.

In contrast to 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 2.2B). 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 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 2.4A).

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 2.4A; see SI Appendix, Figs. 2S20-27). We performed an olefin metathesis on the most active LPE fractions [62] 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 (see SI Appendix, Fig. 2S28). 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. Importantly, no commercially available LPEs tested in combination with the RIF mix either activated or enhanced rosette development (see SI Appendix, Table S2). 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 [34] and IOR-1 [36], 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 [63,64]. While LPEs are found in most bacterial and eukaryotic cell membranes and present in somewhat elevated concentrations in many marine and estuarine bacteria [65], little is known about how and in what contexts LPEs might act as signaling molecules [64,66].

To characterize how LPEs regulate rosette development, we started by investigating the concentrations at which the LPE mix displayed maximal enhancing activity. In contrast with the 10.5% of cells in rosettes induced by 2 M RIF-2 alone, treatment of SrEpac with 2 M RIF-2 and micromolar concentrations of the LPE mix increased rosette development five-fold to 53% (Figure 2.4B; see SI Appendix, Fig. 2S19). Furthermore, maximal levels of rosette development elicited by the RIF mix + the LPE mix matched those induced by the *Algoriphagus* lipid extract (Figure 2.2A; Figure 2.4B).

Finally, we observed that LPEs also influence RIF potency. In bioassays in which the concentration of the LPE mix was held stable at 2 M 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 2.4B). These results 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.

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 (see SI Materials and Methods), 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 2.5A). 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 2.5A). 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 2.5A).

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 (Figs. 2.5B-E). Cells in OMV-induced rosettes were tightly packed and properly localized a specific marker of rosette development, the C-type lectin protein Rosetteless (32), to the extracellular matrix-rich center of the rosette (Figure 2.5B). 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 2.5D). 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 2.6).

Discussion

Animals rely on bacteria for everything from proper metabolism to the stimulation of immune system development to the regulation of gut morphogenesis [8,67]. 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 appropriate environmental bacteria [17]. In one of the most dramatic examples of cross-talk between bacteria and an animal, *Vibrio fischeri* bacteria are recruited into crypts in the juvenile Hawaiian bobtail squid, where the bacteria then trigger post-embryonic morphogenesis of the "light organ" [12]. 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 [7,8].

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 2.6). 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* (Figs. 2.2 and 2.4). Moreover, while the *Algoriphagus* capnine IOR-1 is a potent antagonist of the RIFs [36], 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 [68] 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 [69], with the diversity and density of bacteria dramatically 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 [70-73], and the progressive specification of cell fates [74-76]. 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 trigger light organ maturation [13], and in mice, several bacterial cell wall molecules (LPS, PGN, and polysaccharide A) together shape the development of the immune system of the gut [42,43,77]. 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.

Materials and Methods

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 [31]. 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 [33]. 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; [33].

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' [37], in which *S. rosetta* is grown in the presence of the non-rosette inducing Bacteroidetes bacterium *Echinicola pacifica* [78]. In contrast with RCA, SrEpac supplemented with *Algoriphagus* forms rosettes consistently and at uniformly high levels (approaching 90%; Figure 2.2A), 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 (0.1 g/mL) elicit levels of rosette induction that recapitulate live bacteria (Figure 2.2A).

Choanoflagellate husbandry

SrEpac [37] 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, 150 L/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 H post-induction. In untreated cultures, *S. rosetta* can sometimes form small clumps of cells that break apart upon pipetting. Therefore, to quantify rosette development, 100 L 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 M 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 $1:1e^6$).

Isolation and purification of Algoriphagus lipids

General information**:** 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 (Table S1). 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 described in [33,34]. 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 (for Figure 2.2C; Figure 2.3A and B)**:** Fractions containing sulfonolipids were purified as described in [33,34]. 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 (for Figure 2.4A and B)**:** 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 $CHCl_3$ (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.

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 in [79]. 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 (Figure 2.2B)

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 g/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 g/mL. Rosette development was quantified as described above. Heat map was generated using Plotly.

Activity profile of purified lipids

The potency of purified lipids (for Figure 2.2C; Figure 2.4B; Fig. 2S18-19) 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 (Table S2) 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 50 g/mL. To test for enhancing activity, samples were tested in combination with 2 M RIF mix at concentrations of 5 and 20 g/mL. Rosette development was quantified as described above.

Immunofluorescence microscopy (Figure 2.1; Figs. 2.5B-E)

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 (Levin et al., 2014) at 1.25 ng/ L (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.

Acknowledgements

We thank Monika Abedin, David Booth, Brad Davidson, and John Rawls 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.

FIGURES

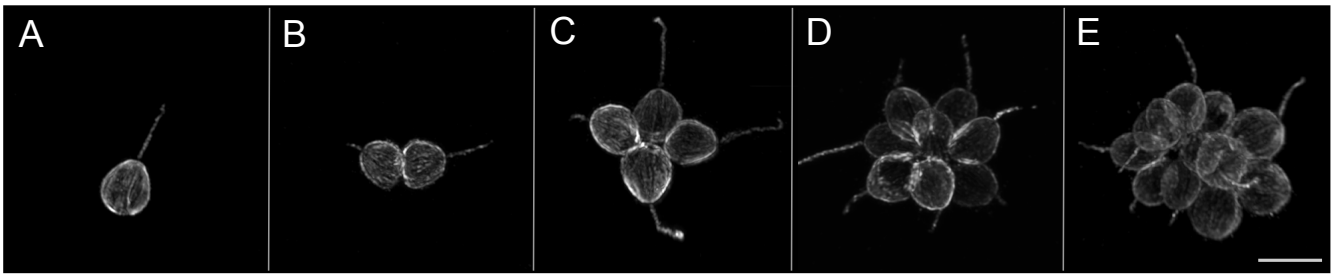


Figure 2.1. Stages of rosette development in *S. rosetta*. During rosette development, a single founding cell undergoes serial rounds of cell division, resulting in the formation of a structurally integrated rosette. Importantly, rosette development does not involve cell aggregation. Shown are a single cell (A), a pair of cells (B), a four-cell rosette (C), an eight-cell rosette (D) and a 16-cell rosette (E).

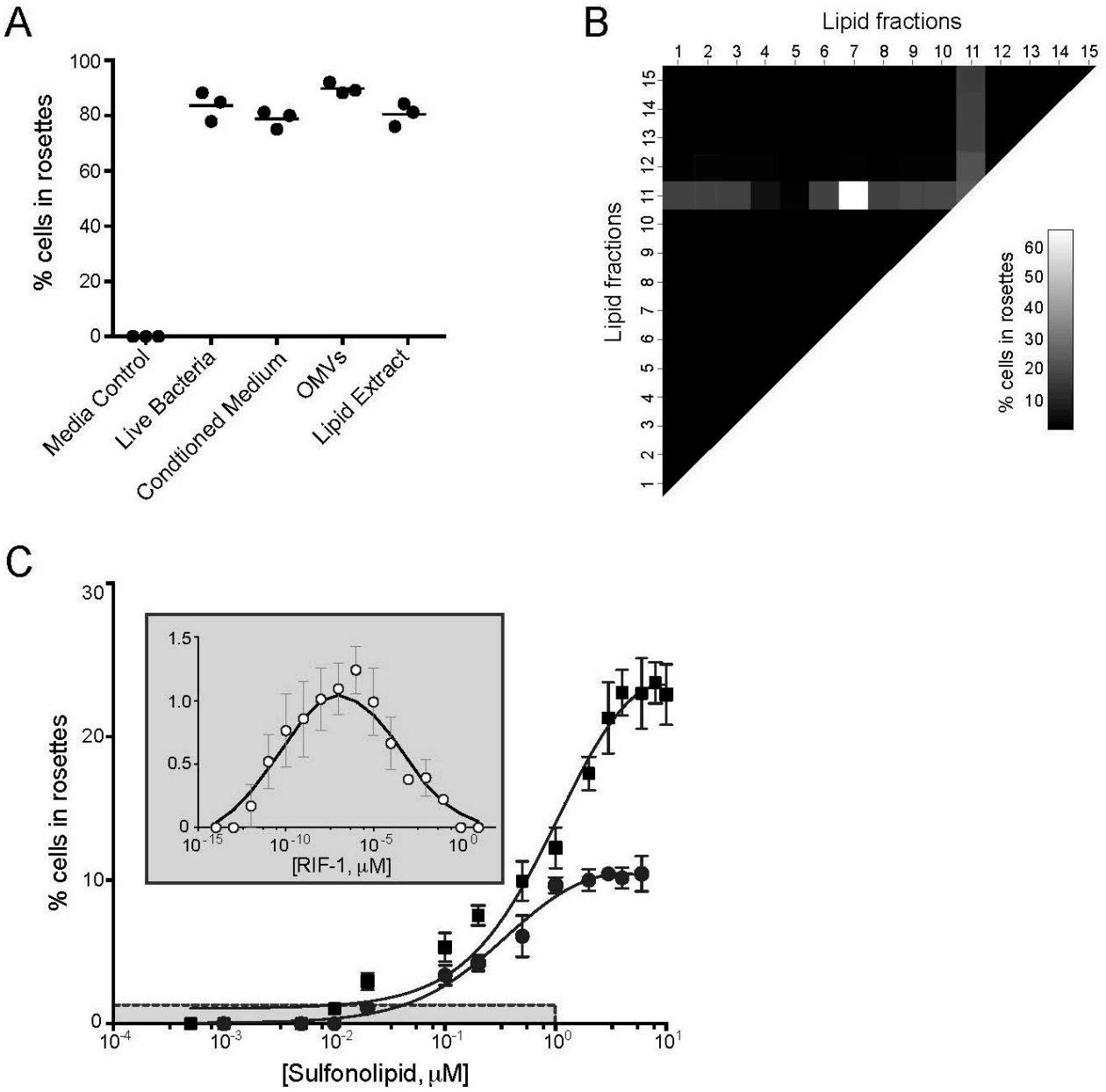


Figure 2.2. Maximal rosette development requires lipid co-factor interactions. (A) *S. rosetta* does not form rosettes when treated with media that lacks additional bacterial signals (Media Control). Maximal (~90% of cells in rosettes) or near-maximal levels of rosette development are induced by live *Algoriphagus*, *Algoriphagus* conditioned media, *Algoriphagus* OMVs, and *Algoriphagus* bulk lipid extract. **(B)** 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 g/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 of 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%. **(C)** 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 RIF-1 rosette-inducing activity. Rosette development was quantified 24 hours after induction. Minor ticks on X-axis are log spaced.

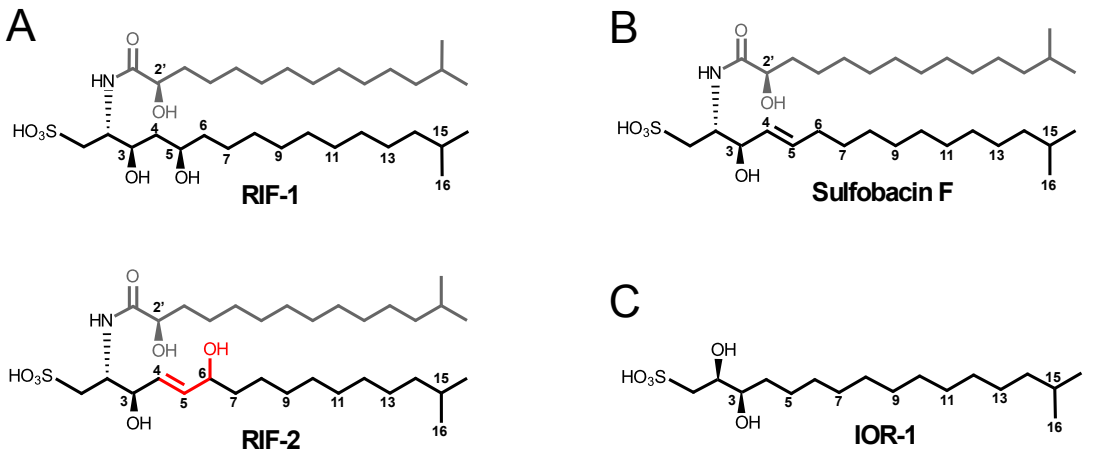


Figure 2.3. Structural similarities and differences among RIFs, an inactive sulfonolipid, and the inhibitory capnine IOR-1. (A) The three dimensional structure of RIF-1 [determined by total synthesis in [34]], compared to the proposed molecular structure of RIF-2, and **(B)** the structure of an inactive *Algoriphagus* sulfonolipid, Sulfobacin F [34]. 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.

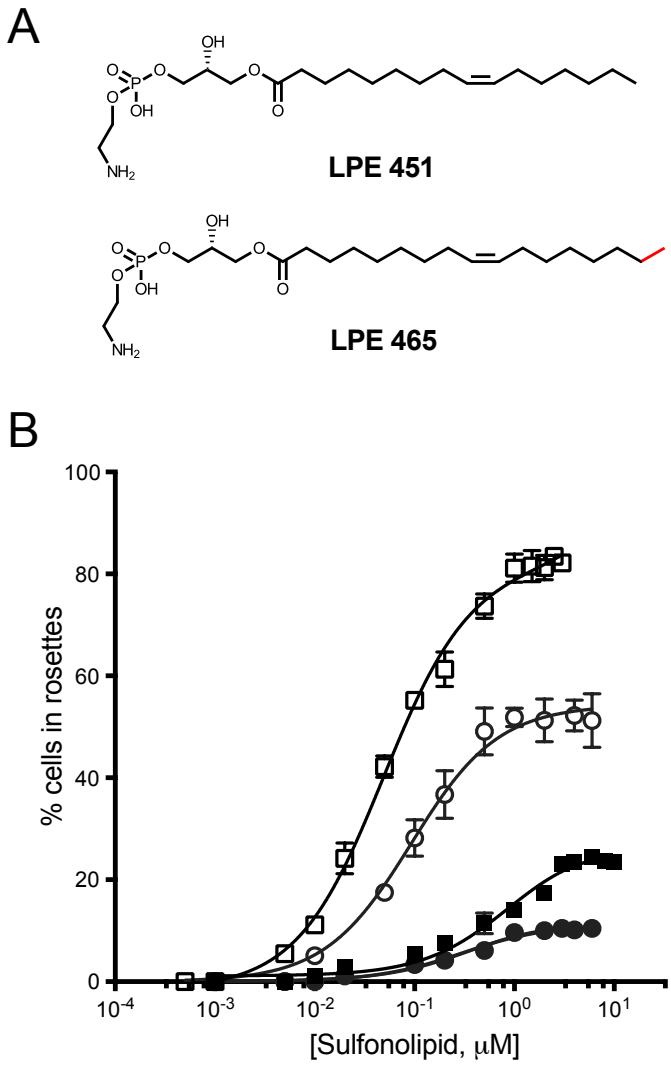


Figure 2.4. **LPEs synergize with RIFs to enhance rosette development. (A)** The structures of LPE 451 and LPE 465 as determined by NMR and tandem mass spectrometry. LPE 451 and LPE 465 differ from each other by only one methyl group along the fatty acid chain (highlighted in red). (**B**) The addition of 2 M 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.

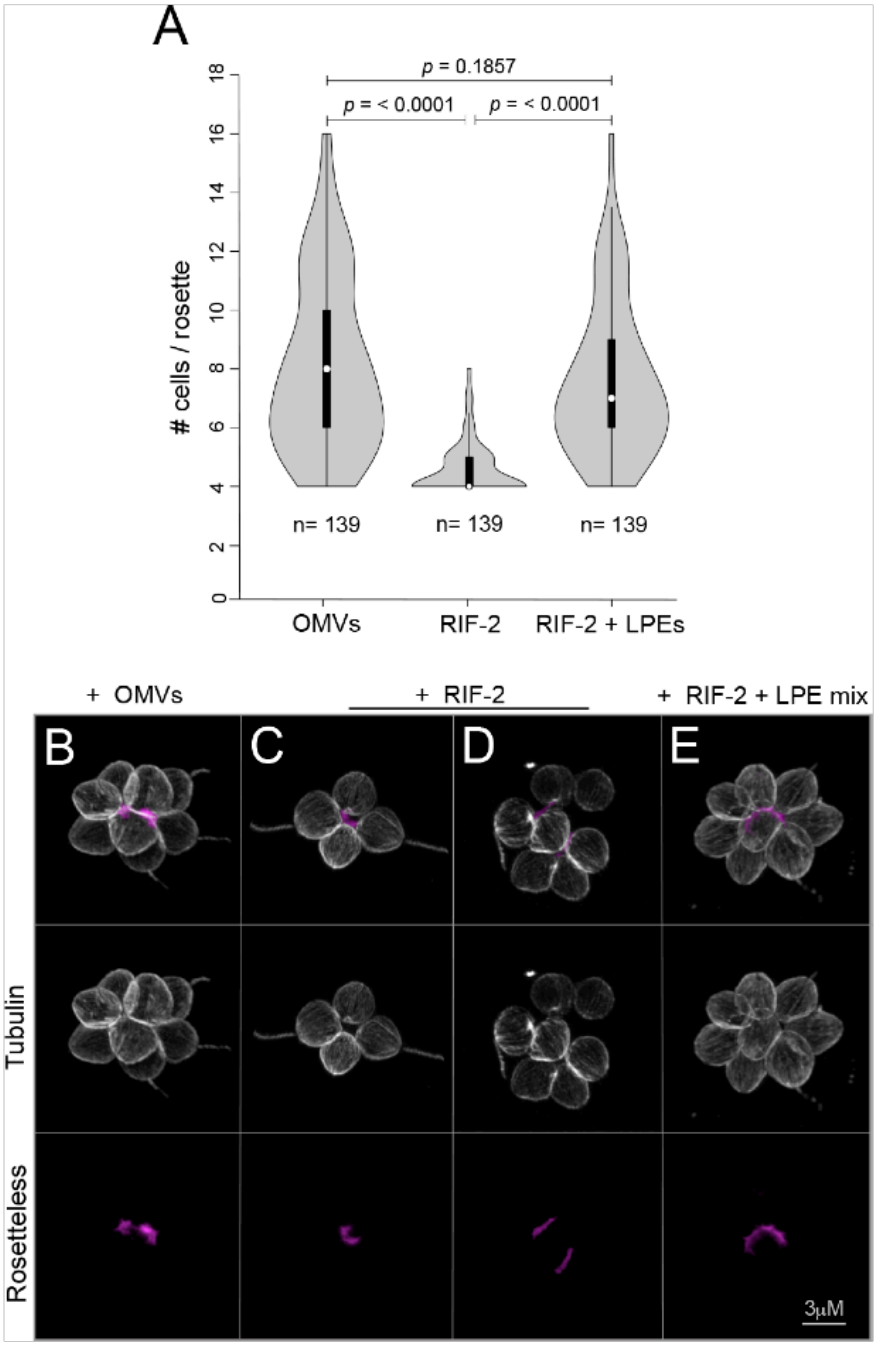


Figure 2.5. LPEs promote proper rosette development and maturation. (**A**) Frequency distribution of rosette size in SrEpac incubated with OMVs, RIF-2, and RIF-2 + LPE mix after exposure to shear stress by pipetting. Rosettes induced with RIF-2 alone contained fewer cells on average and reached a smaller 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 version 6 for Mac, GraphPad Software, La Jolla California USA. **(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 M RIF-2, 1 M RIF-2 + 2uM LPEs, or OMVs.

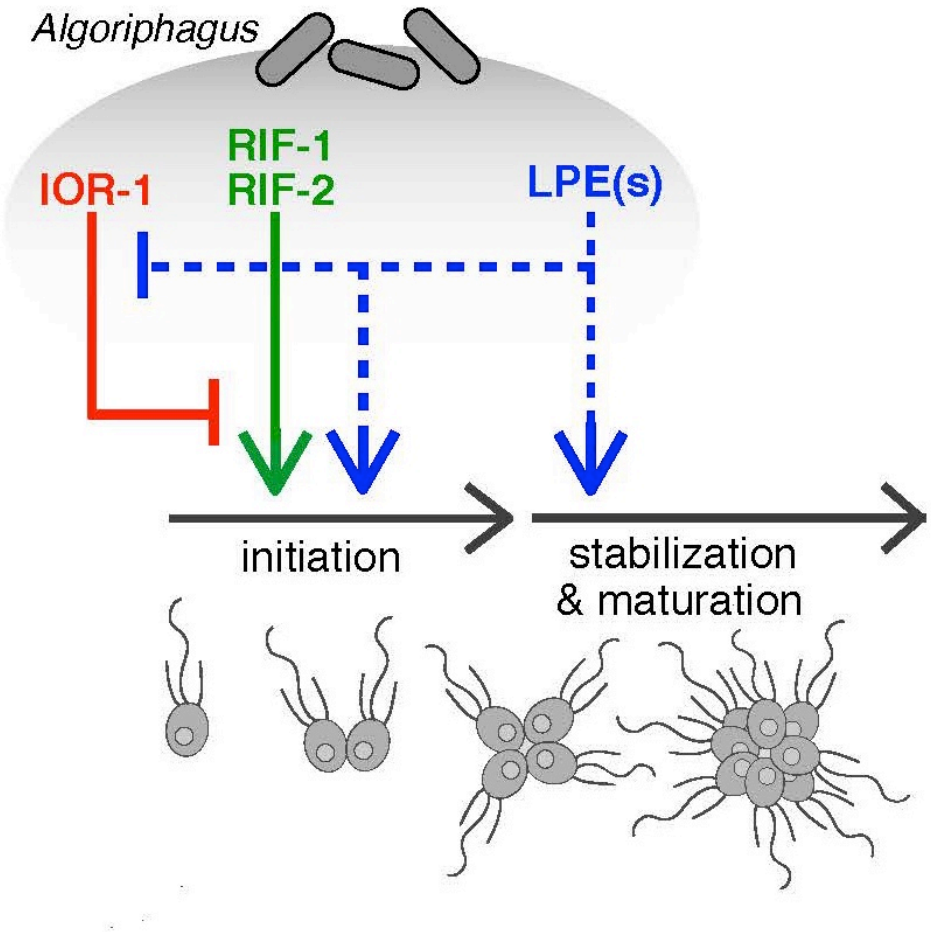


Figure 2.6. Multiple bacterial inputs regulate rosette development in *S. rosetta*.
Algoriphagus produces three chemically distinct classes of lipids – sulfonolipids, LPEs, and a capnine – that interact to alternately induce, enhance, or inhibit rosette development in *S. rosetta.* The sulfonolipids 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 [36]. Importantly, when *S. rosetta* is exposed simultaneously to RIFs and the synergistic LPEs, mature rosettes develop even in the presence of IOR-1.

Table 2S1: ^{1}H NMR proton shifts for RIF-2.

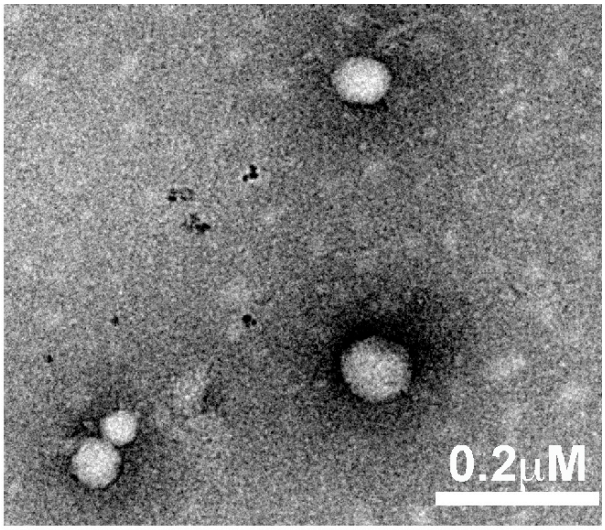
Position	RIF-2 **(600 MHz, d6-DMSO) m, H *J* in Hz)**	**(150 MHz) ^{13}C, m**	RIF-2 **(600 MHz, d4-MeOD) m, H *J* in Hz)**	**(150 MHz) ^{13}C, m**
1	2.57, dd, 1H (14.1, 4.6) 2.86, dd, 1H, (14.1, 6.1)	50.7, t	2.99, dd, 1H (14.4, 3.8) 3.07, dd, 1H (14.4, 8.2)	51.3, t
1'		173.8, s		n.d.
2	3.94, m, 1H	50.4, 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 628.422

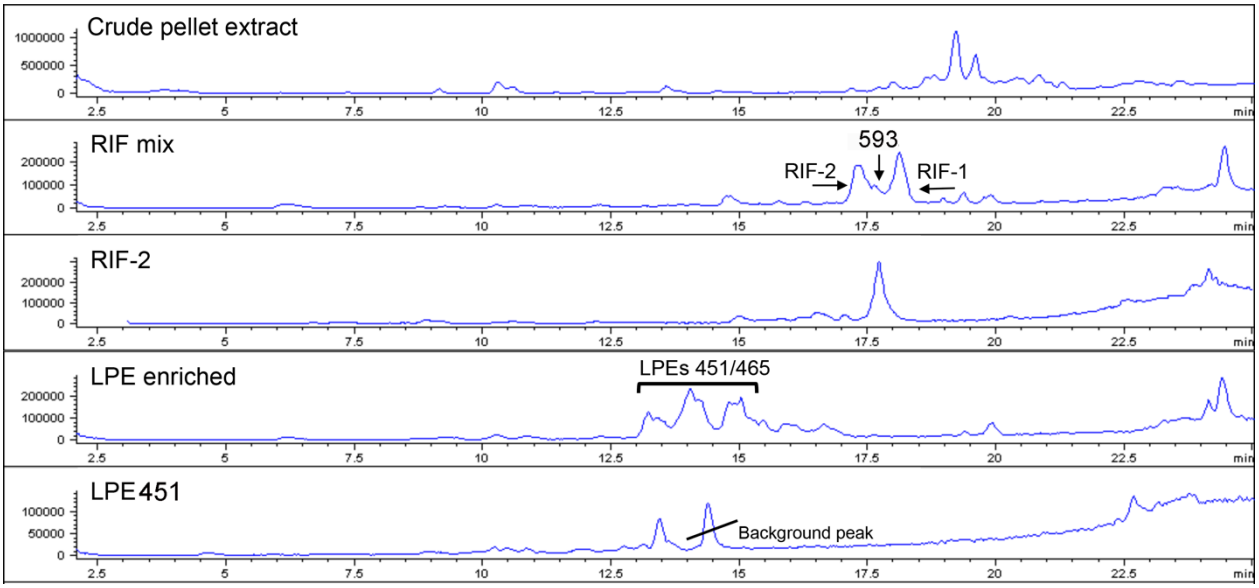
Table 2S2. Commercial lipids tested for bioactivity.

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 **	Sphignolipid	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)

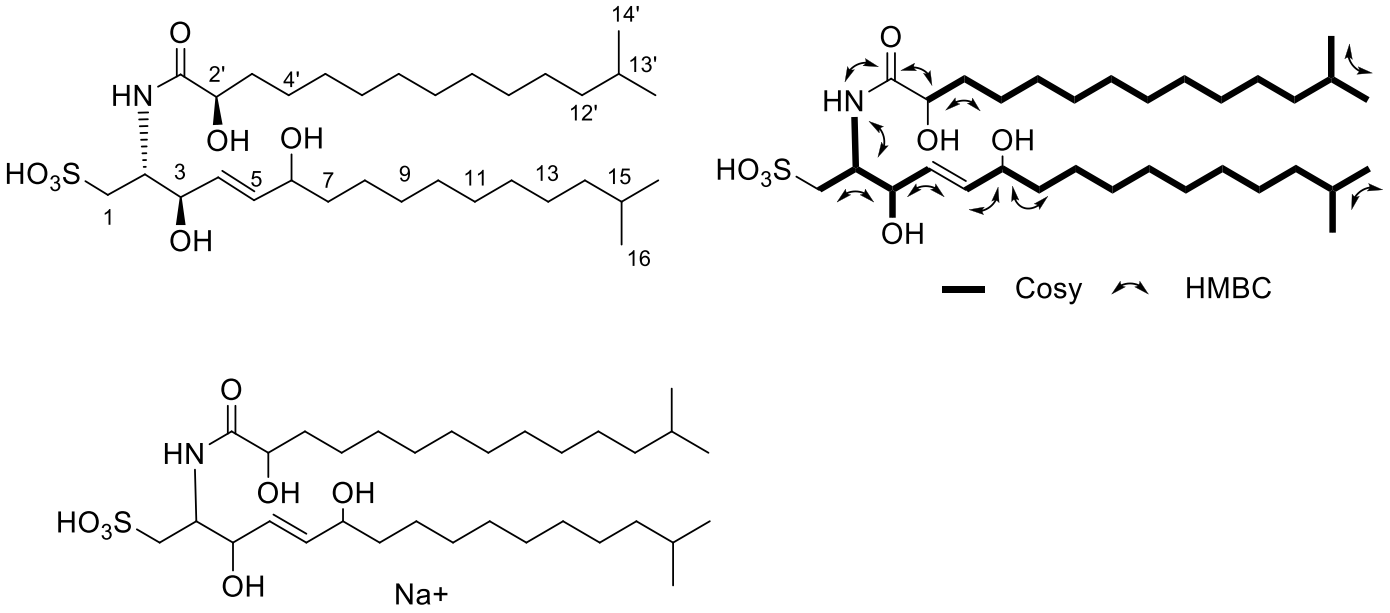
** : Also found to lack rosette inducing activity in (Alegado et al. 2012); n.d. : none detected



2S1. TEM of *Algoriphagus* OMVs. Transmission electron microscopy (TEM) images (negative staining with uranyl acetate) of OMVs isolated form *Algoriphagus* conditioned media (see Materials and Methods for OMV isolation details).

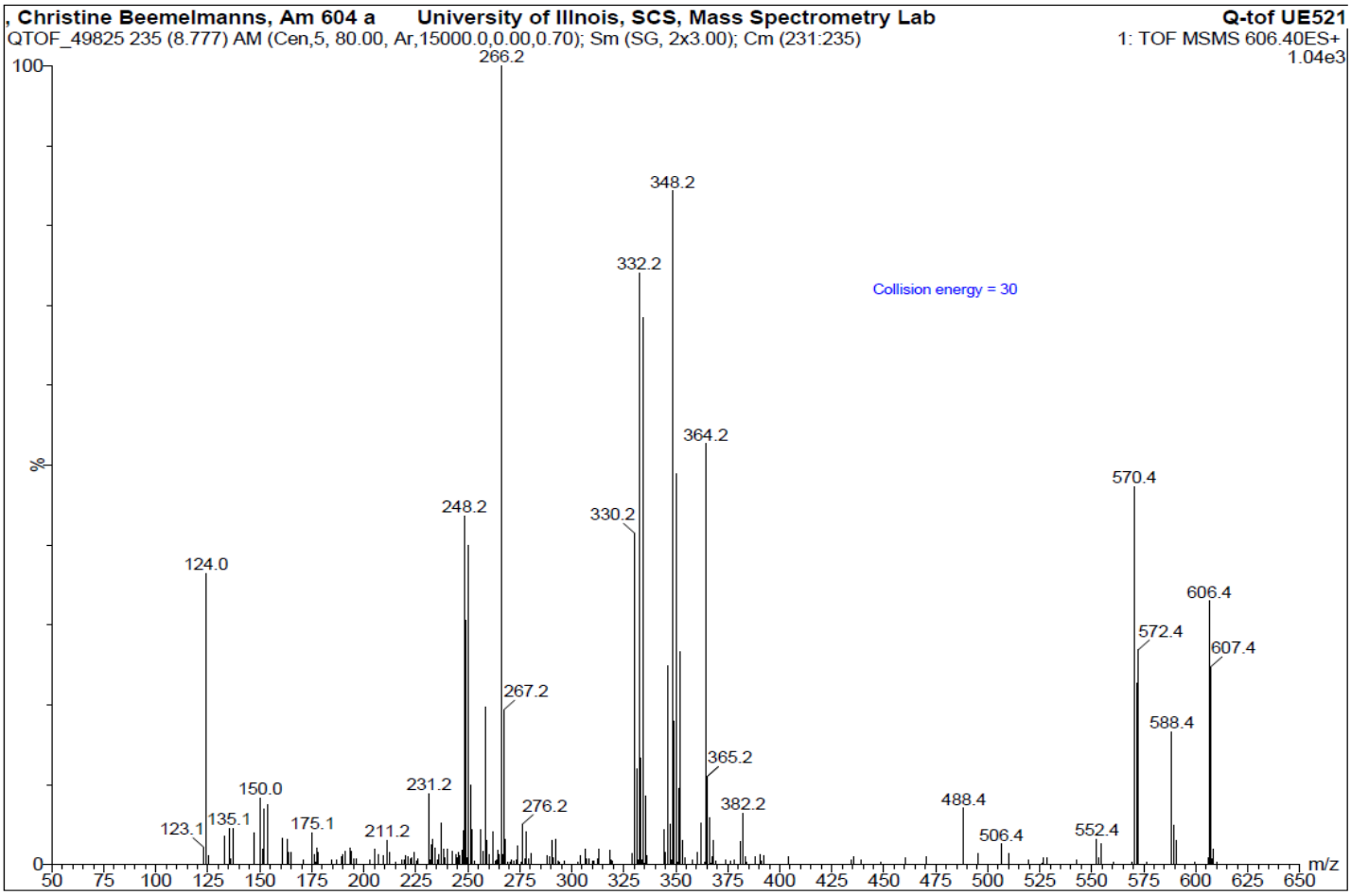


2S2. LCMS spectra showing general composition of the lipid extract, rosette inducing fractions (RIF mix and RIF-2), and synergistic fractions (LPE enriched and LPE 451). Stearic acid (depicted as "background peak" in the LPE 451 trace), a common LC-MS contaminant, was tested for activity by applying it to SrEpac both alone and in combination with the RIF mix and determined to not be bioactive (also see Table 2S2).

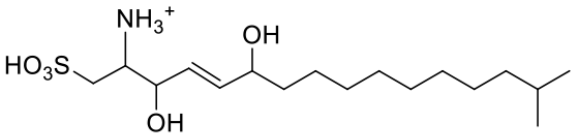


Chemical Formula: $C_{32}H_{63}NNaO_7S^+$
Exact Mass: 628.42175

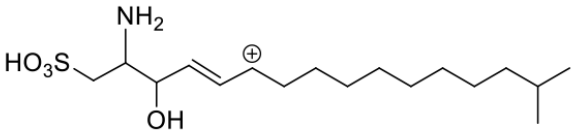
2S3. COSY and HMBC correlations for RIF-2**.**



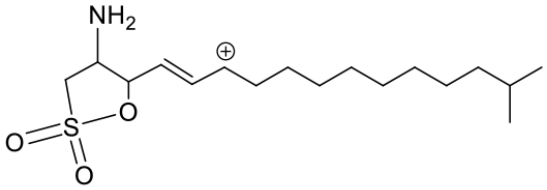
2S4. MSMS spectrum for sulfonolipid RIF-2 m/z = 606.4 $(M+H)^+$ Q-tof, pos.)



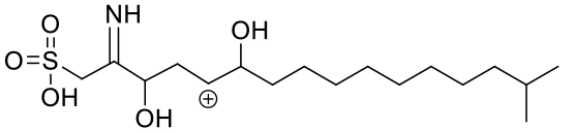
Chemical Formula: $C_{17}H_{36}NO_5S^+$
Exact Mass: 366.23087



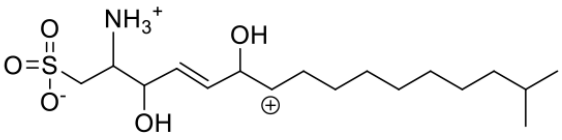
Chemical Formula: $C_{17}H_{35}NO_4S$
Exact Mass: 349.22868



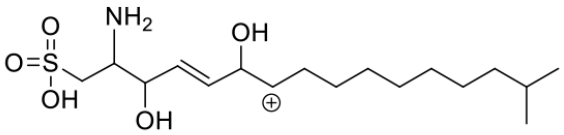
Chemical Formula: $C_{17}H_{33}NO_3S$
Exact Mass: 331.21811



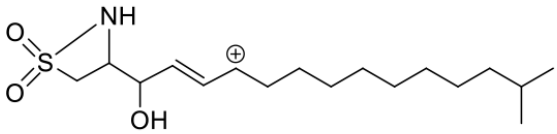
Chemical Formula: $C_{17}H_{34}NO_5S^+$
Exact Mass: 364.21522



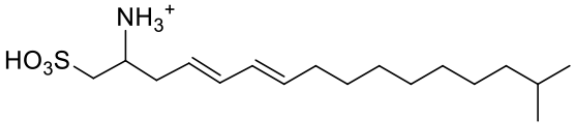
Chemical Formula: $C_{17}H_{34}NO_5S^+$
Exact Mass: 364.21522



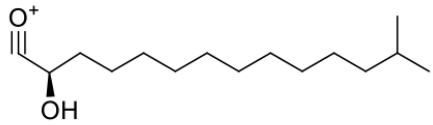
Chemical Formula: $C_{17}H_{34}NO_5S^+$
Exact Mass: 364.21522



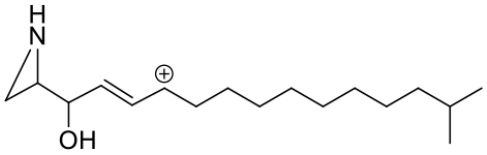
Chemical Formula: $C_{17}H_{32}NO_3S^+$
Exact Mass: 330.20974



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

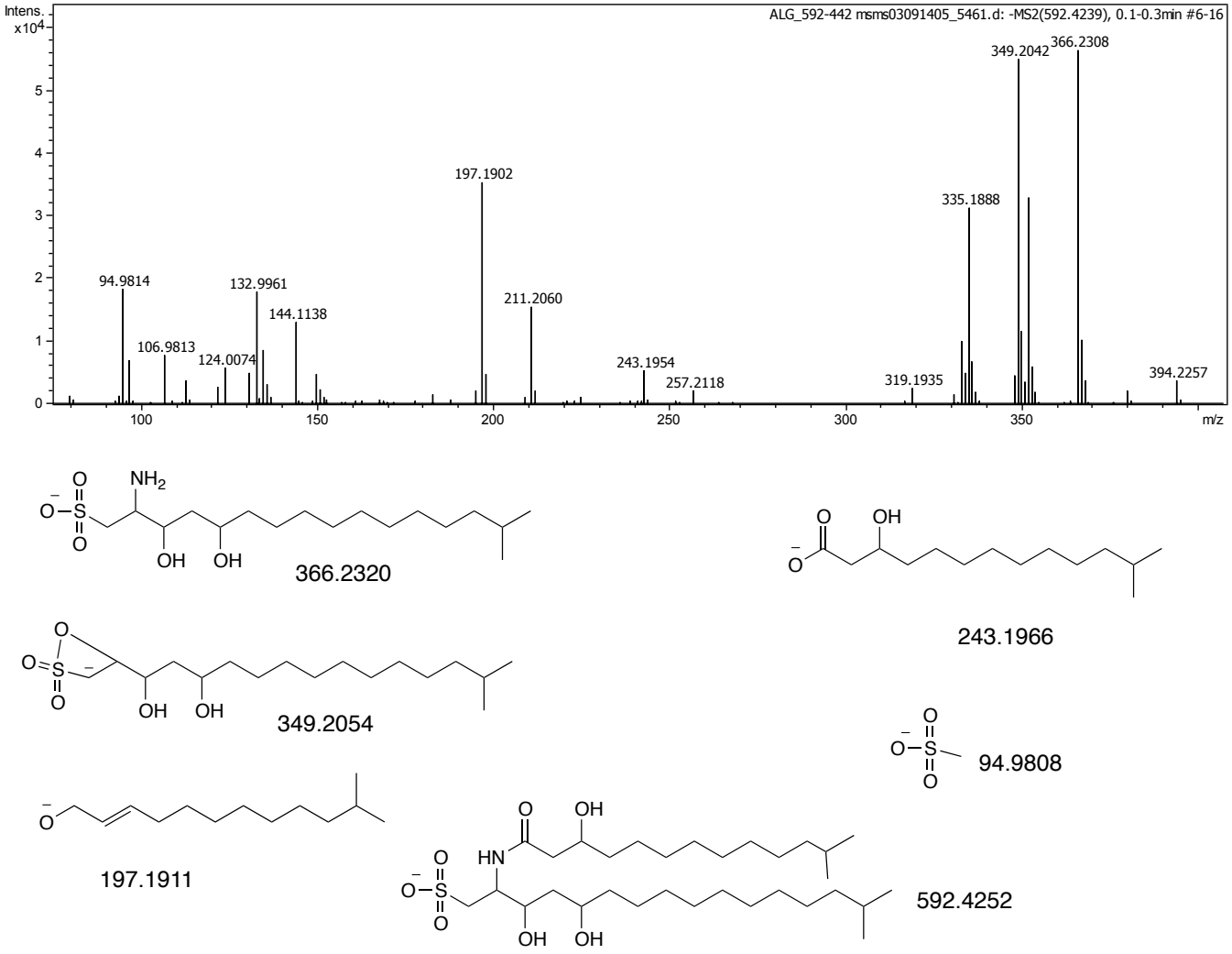


Chemical Formula: $C_{15}H_{29}O_2^+$
Exact Mass: 241.21621

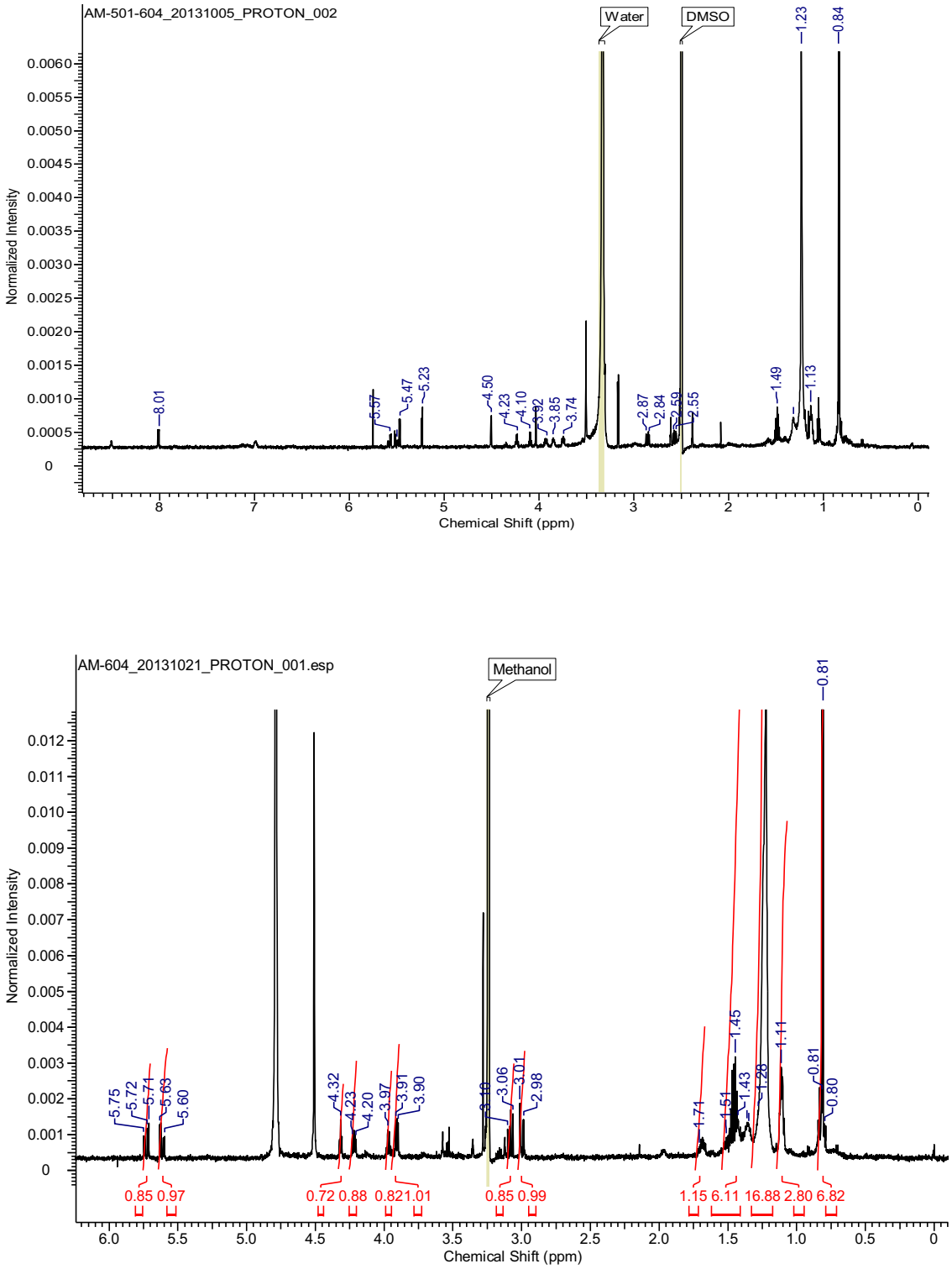


Chemical Formula: $C_{17}H_{32}NO^+$
Exact Mass: 266.24784

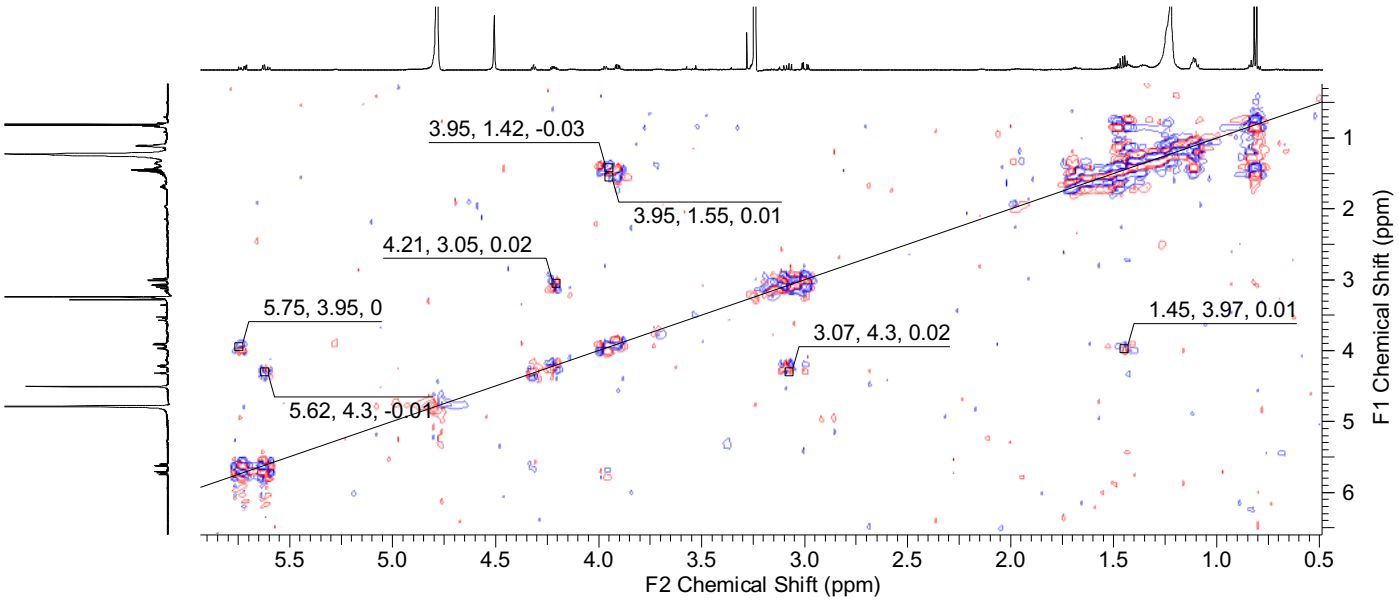
2S5. Structures of predicted fragment structures of RIF-2



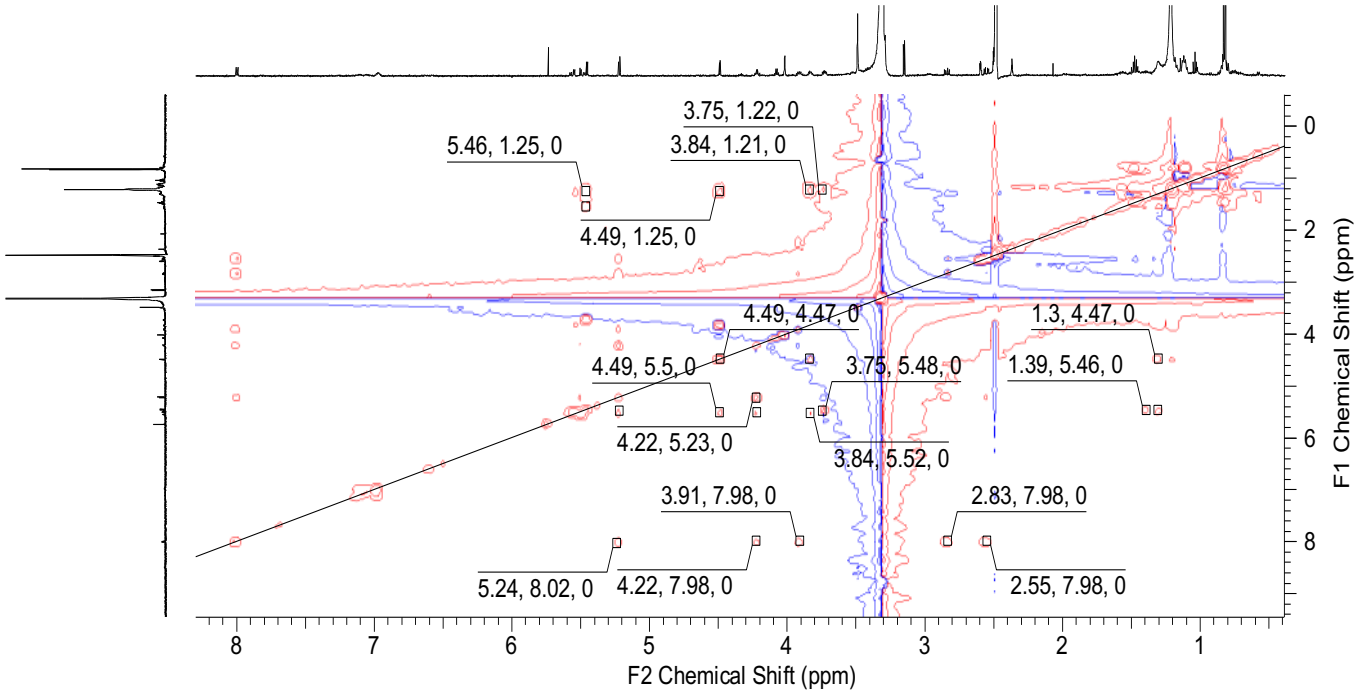
2S6. HR-MSMS sulfonolipid 593 (negative mode). Predicted fragmentation of sulfonolipid 592 (M-H). The major peak 366 (M-H) is typical of sulfonolipids (see Alegado et al. 2012; Beemelmanns et al. 2014).



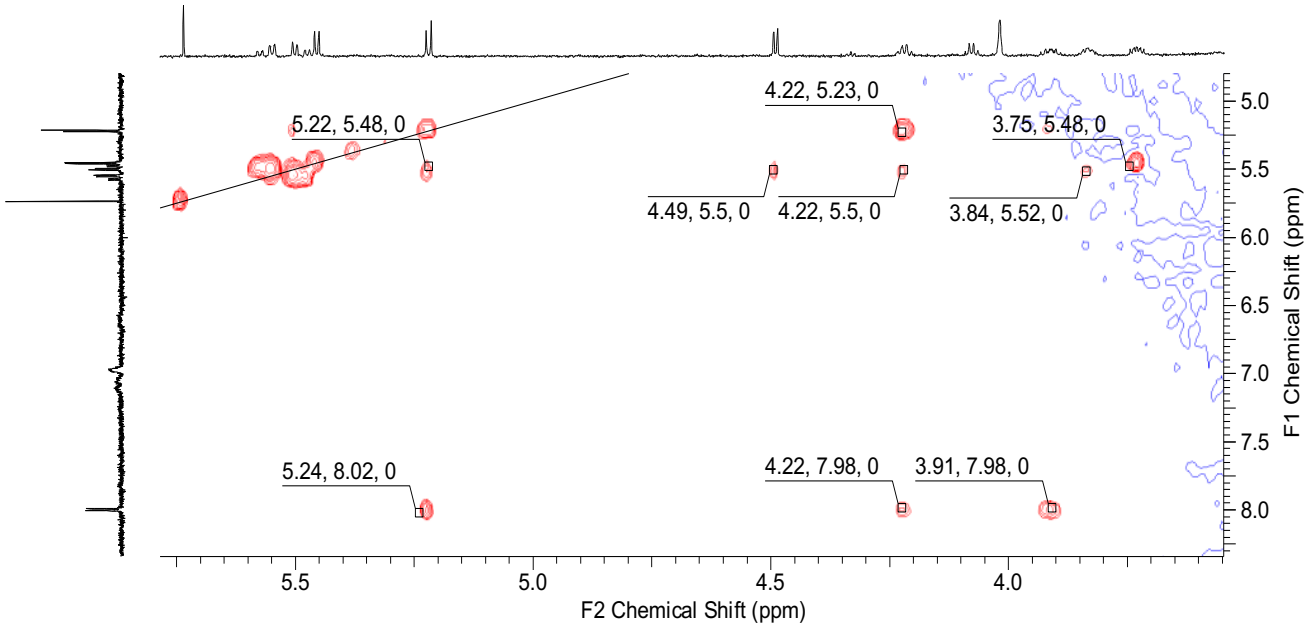
2S7. (Top) ^{1}H-NMR spectrum of RIF-2 (500 MHz, d6-DMSO), (Bottom) ^{1}H-NMR spectrum of RIF-2 (500 MHz, d4-MeOD)



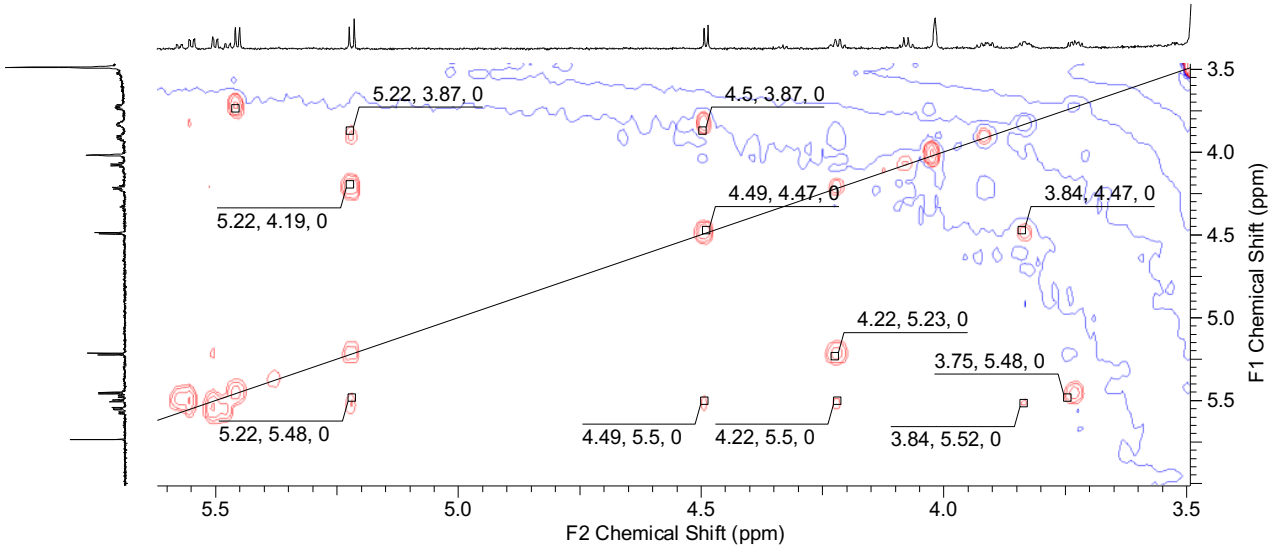
2S8. COSY spectrum of RIF-2 (500 MHz, d4-MeOD)



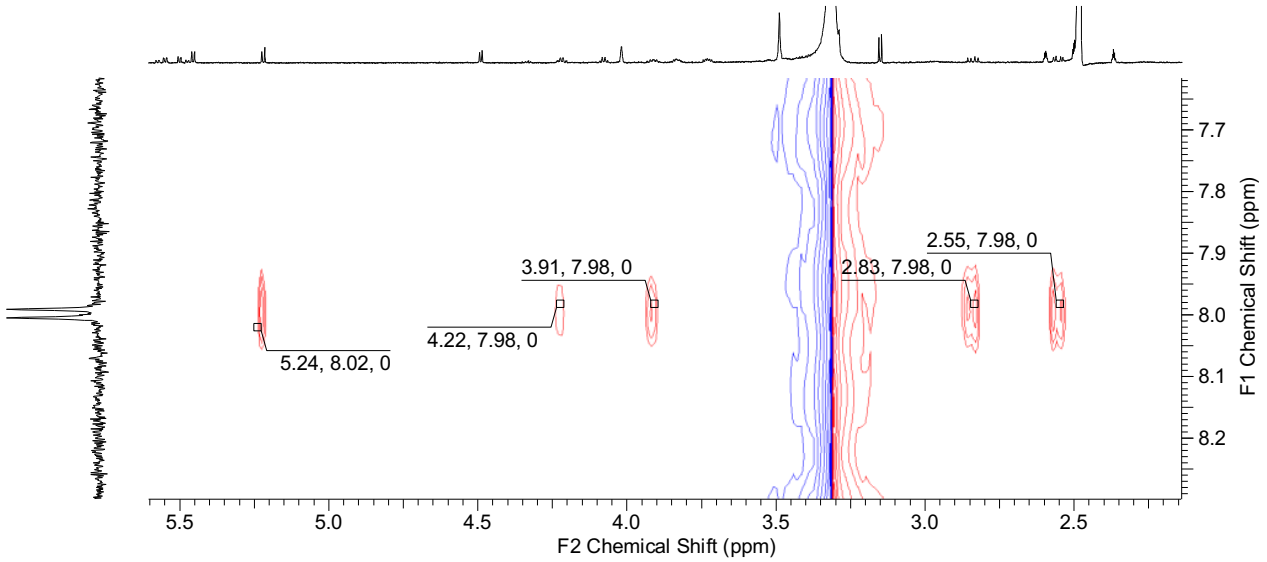
2S9. TOCSY spectrum of RIF-2 (500 MHz, d6-DMSO)



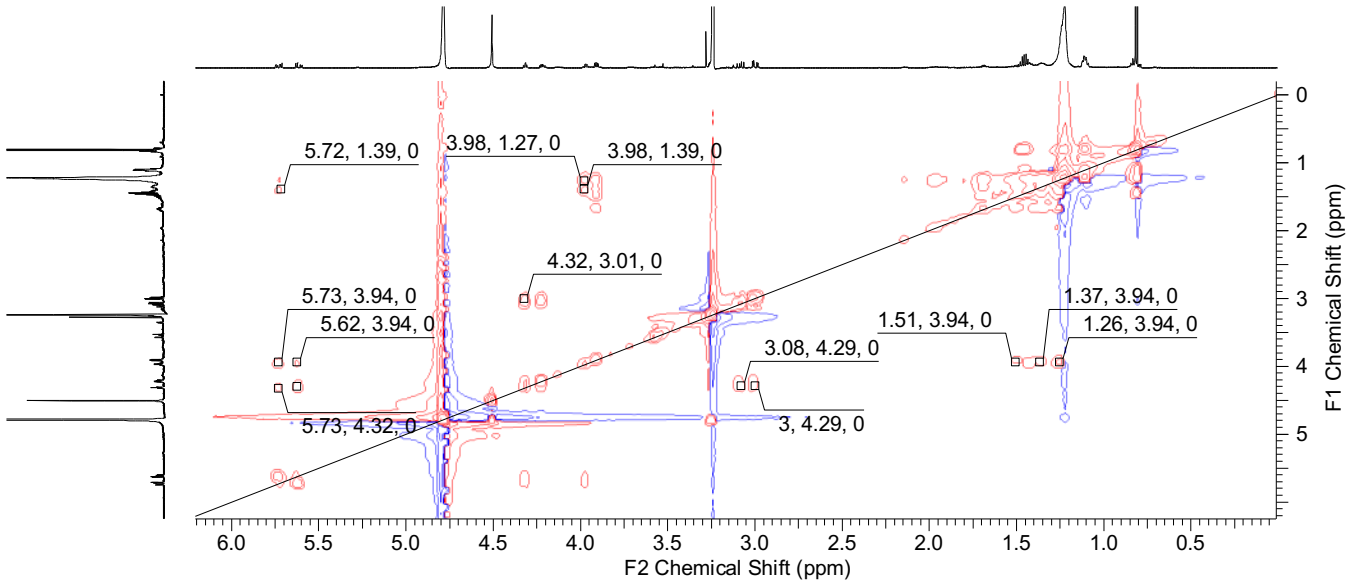
2S10. TOCSY spectrum of RIF-2 (500 MHz, d6-DMSO)



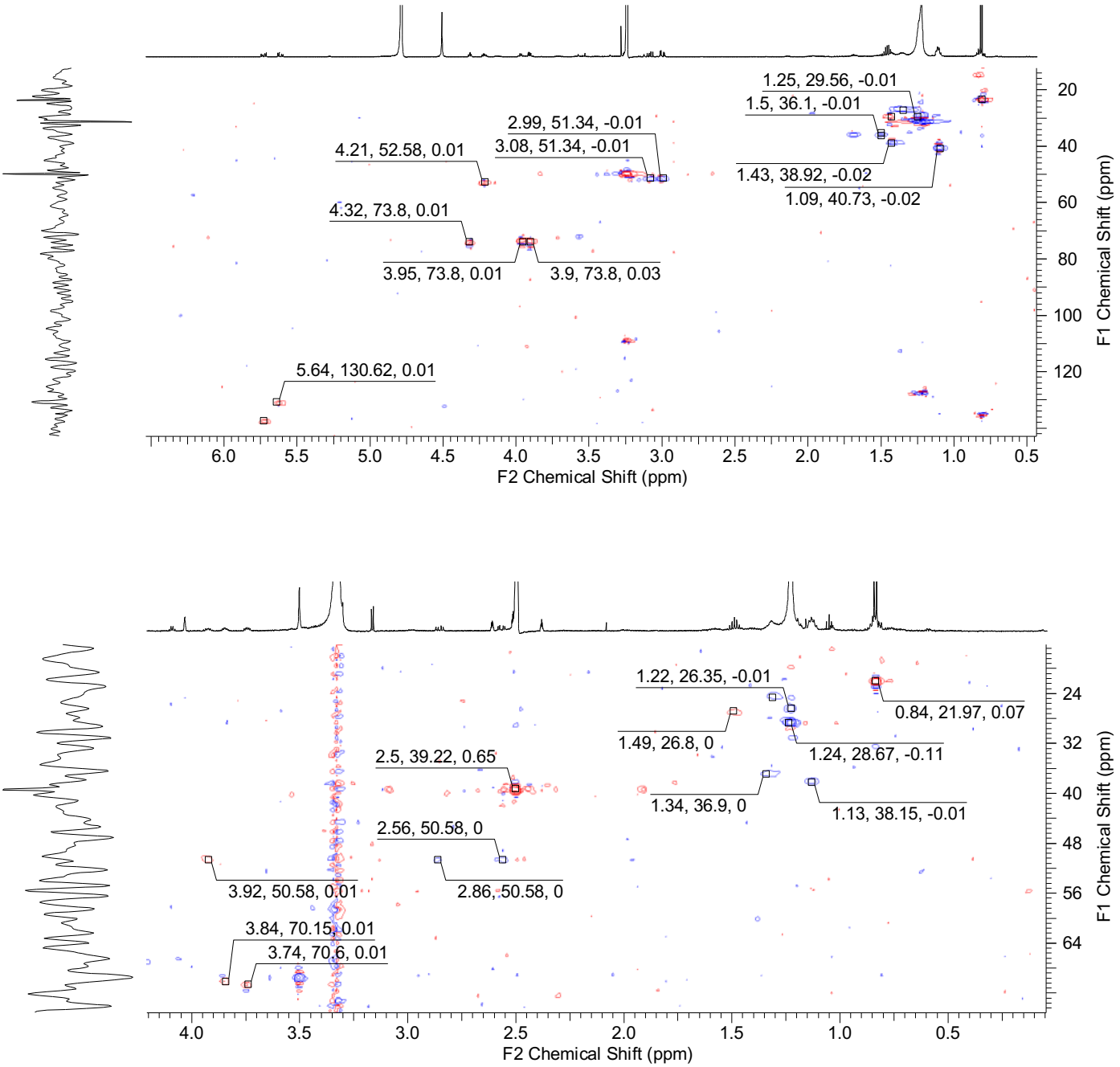
2S11. TOCSY spectrum of RIF-2 (500 MHz, d6-DMSO)



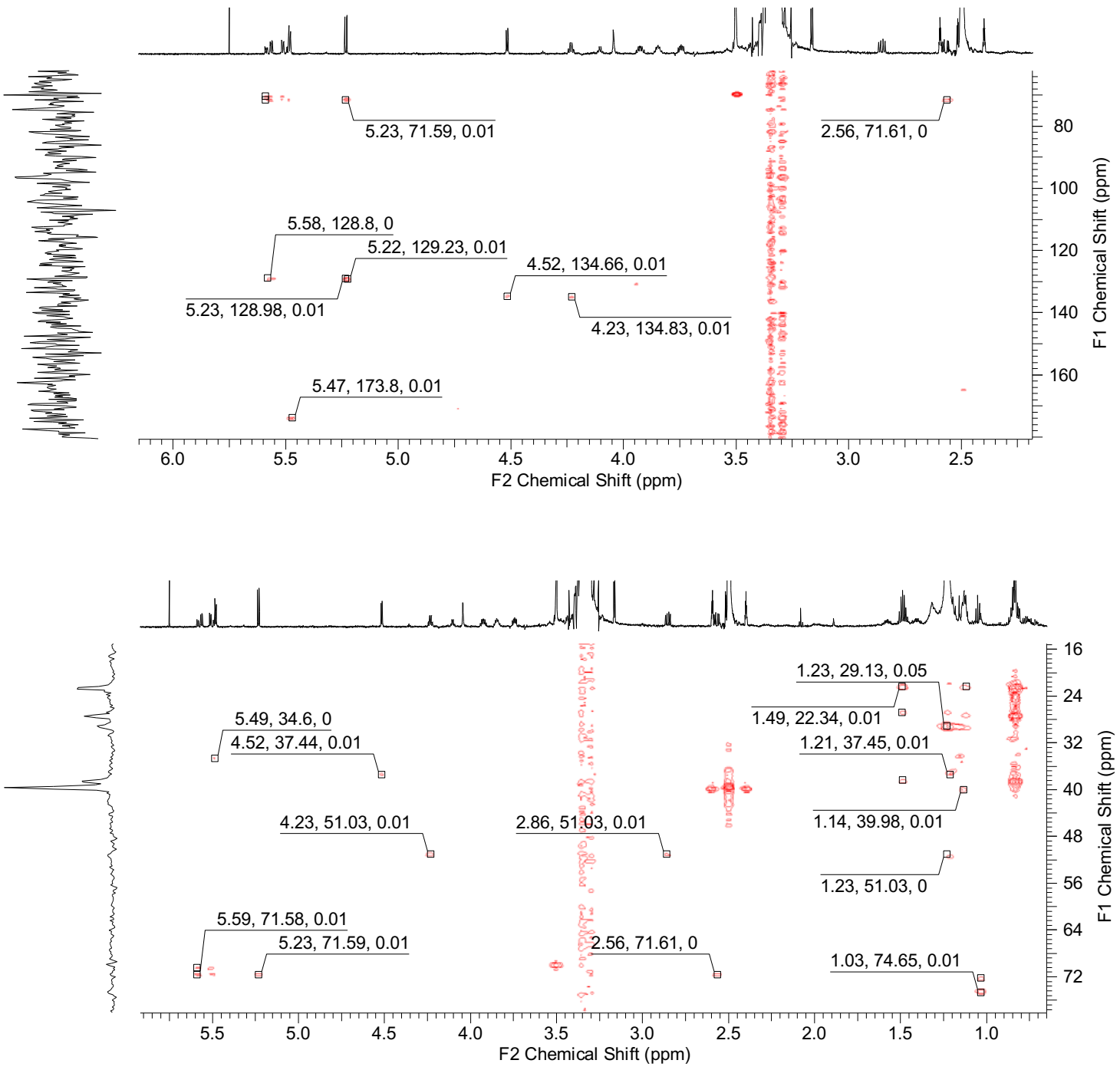
2S12. TOCSY spectrum of RIF-2 (500 MHz, d6-DMSO)



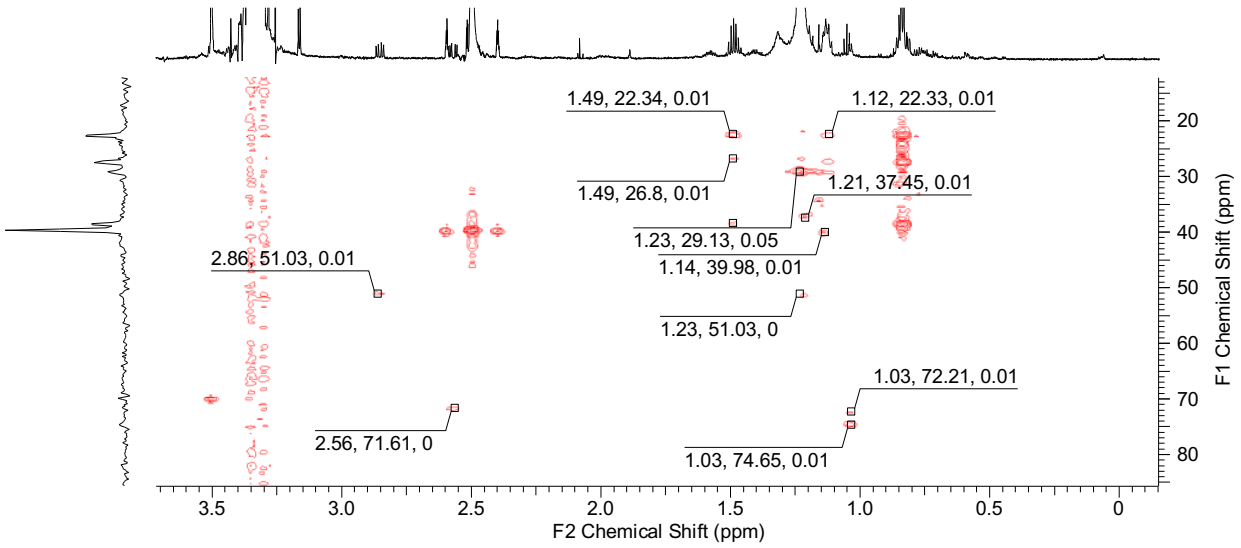
2S13. TOCSY spectrum of RIF-2 (500 MHz, d4-MeOD)



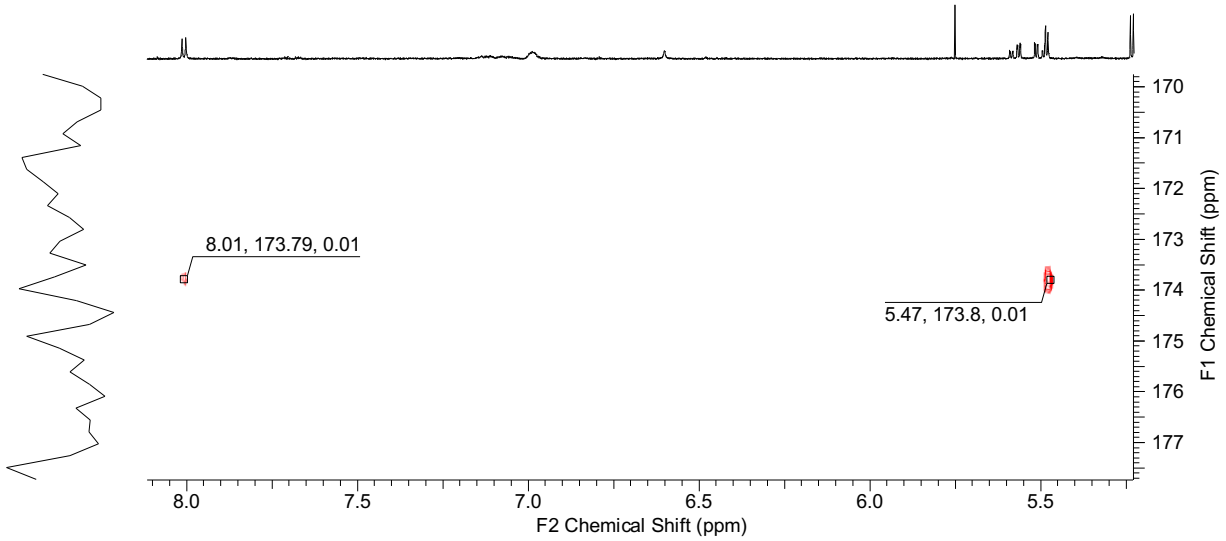
2S14. (Top) HSQC spectrum of RIF-2 (500 MHz, d4-MeOD) (Bottom) HSQC spectrum of RIF-2 (500 MHz, d6-DMSO)



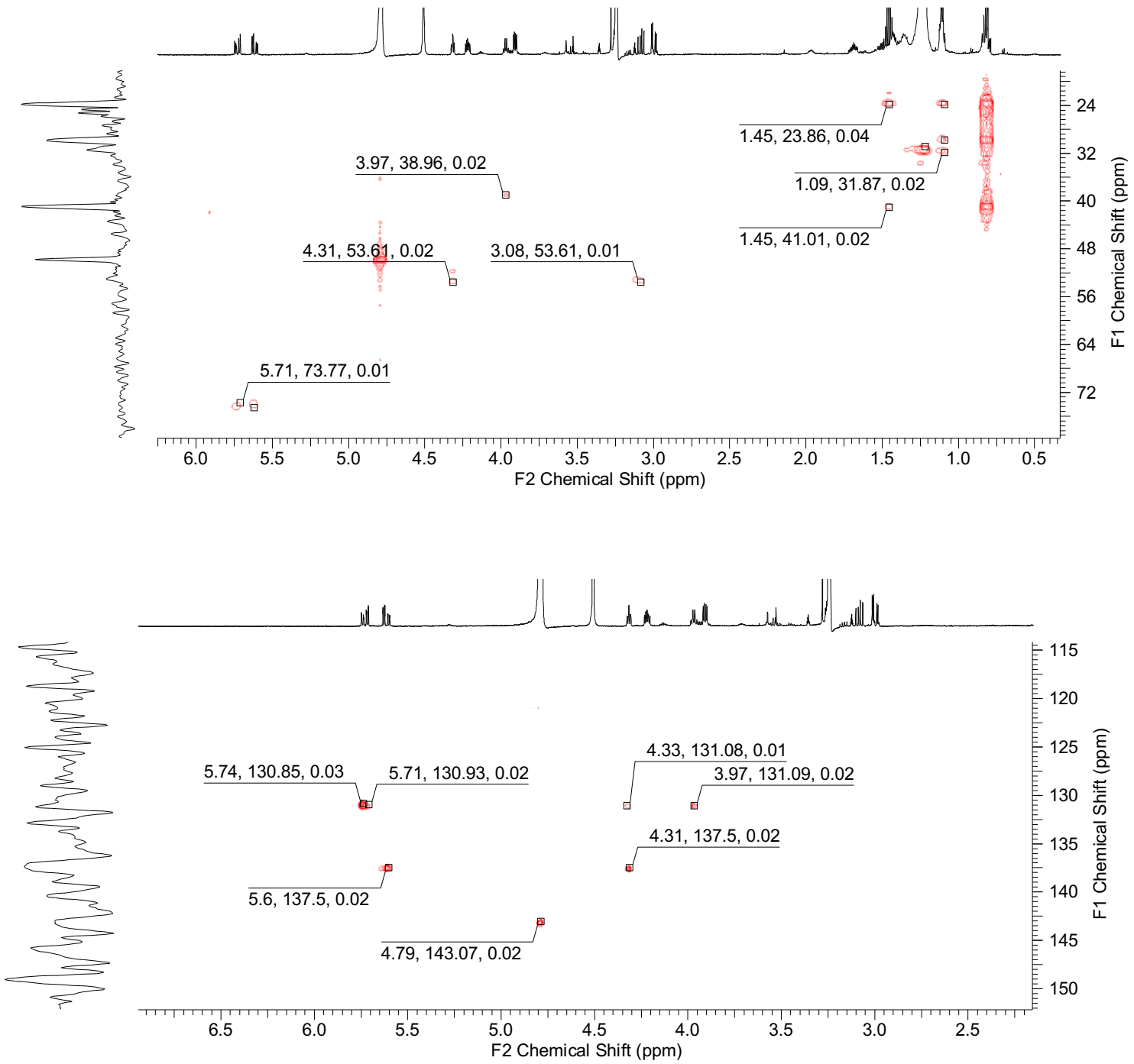
2S15. (Top) HMBC spectrum of RIF-2 (700 MHz, d6-DMSO), (Bottom) HMBC spectrum of RIF-2 (700 MHz, d6-DMSO)



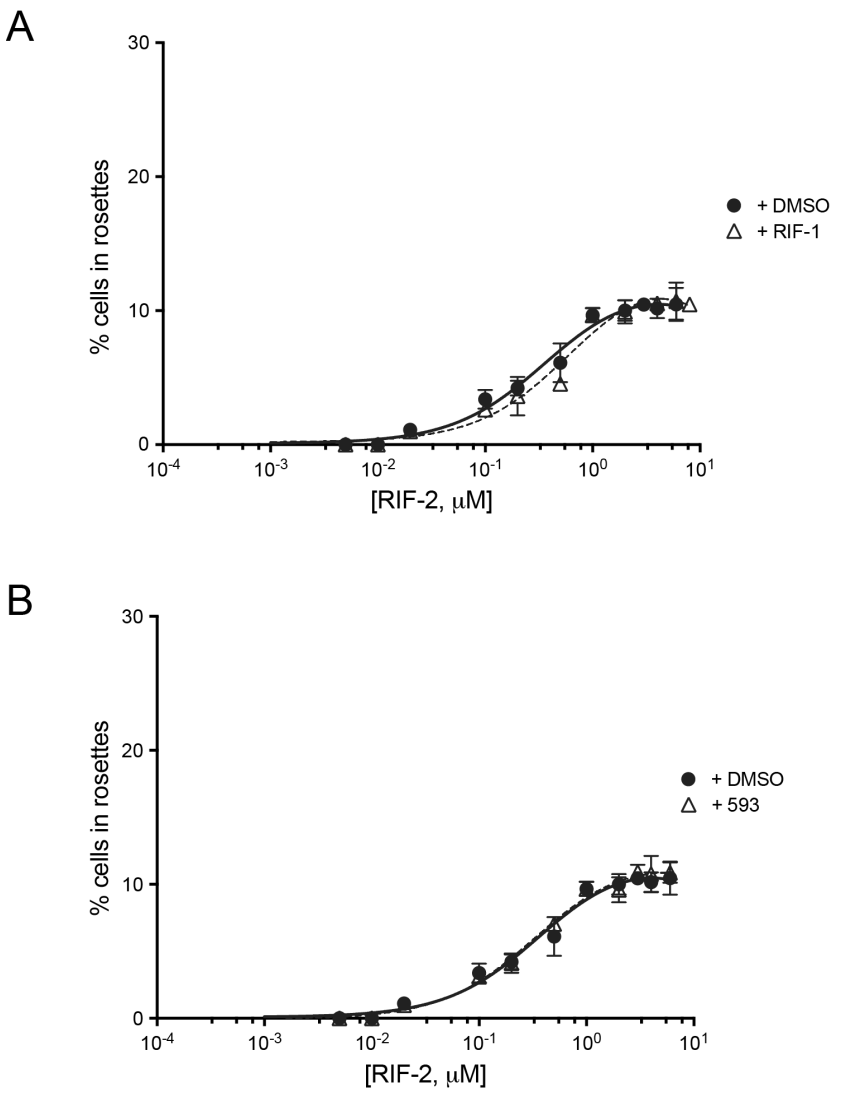
HMBC spectrum of RIF-2 (700 MHz, d6-DMSO)



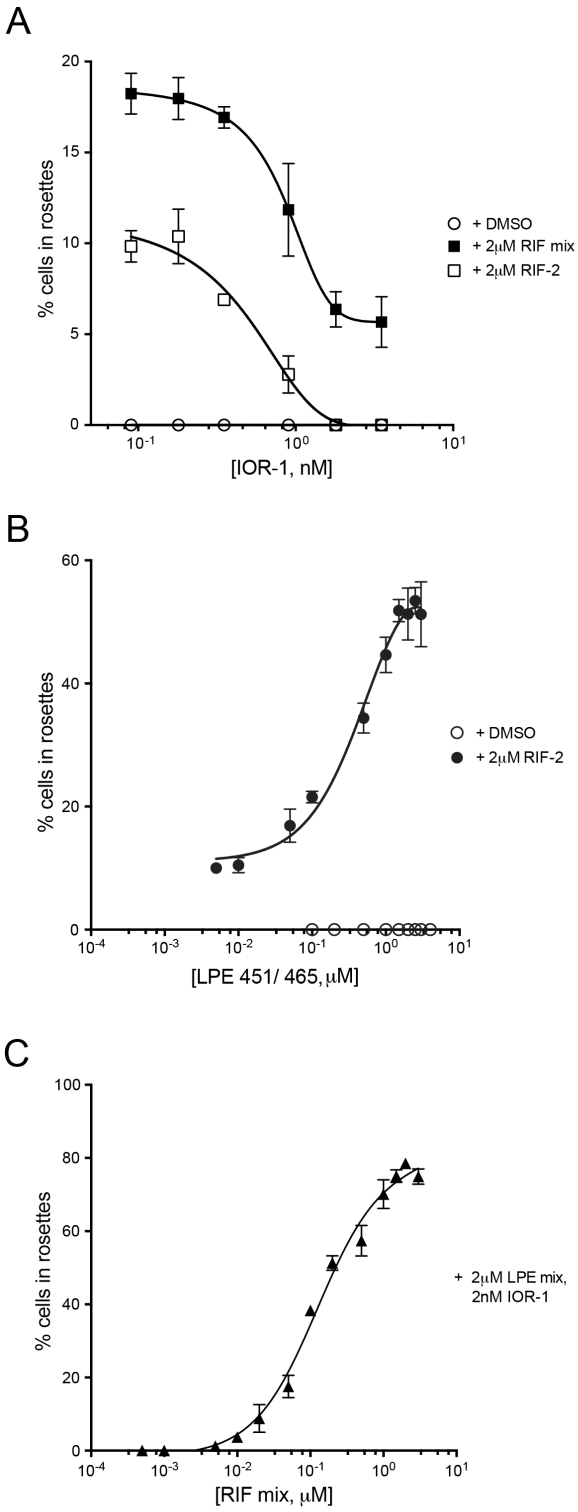
2S16. (Top) HMBC spectrum of RIF-2 (700 MHz, d6-DMSO), (Bottom) HMBC spectrum of RIF-2 (700 MHz, d6-DMSO)



2S17. (Top) HMBC spectra of RIF-2 (500 MHz, d4-MeOD), (Bottom) HMBC spectra of RIF-2 (500 MHz, d4-MeOD)

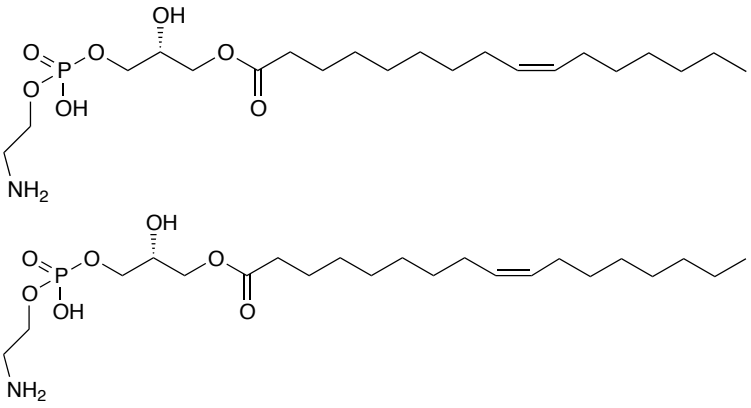


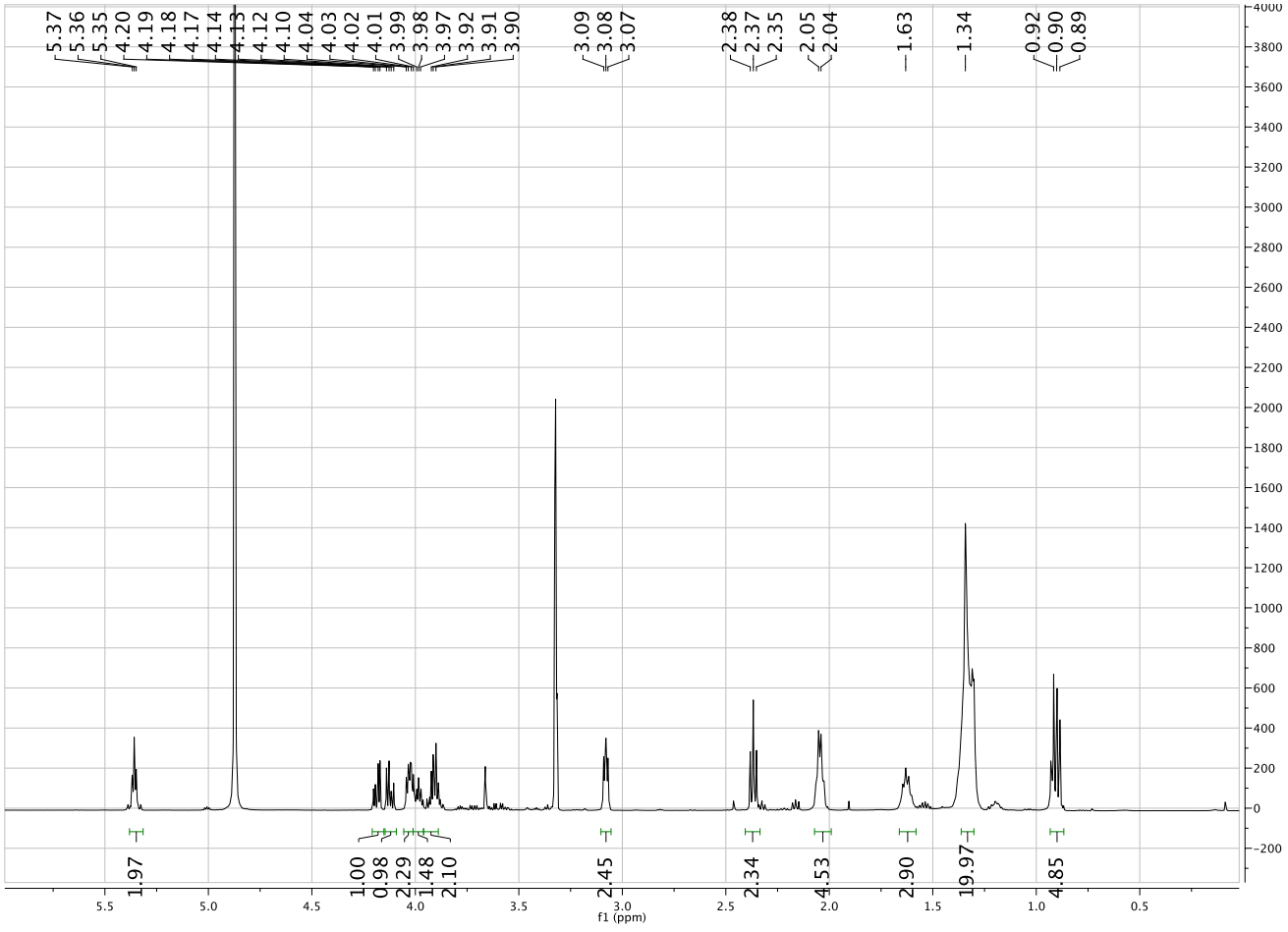
2S18. The combined activity of RIF-1 + RIF-2 does not recapitulate that of the RIF mix. **(A)** Addition of 10^{-7} M RIF-1, the concentration of RIF-1 at which maximal inducing activity was observed, to RIF-2 (triangle) does not increase levels of rosette induction compared to RIF-2 alone (circle). 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.



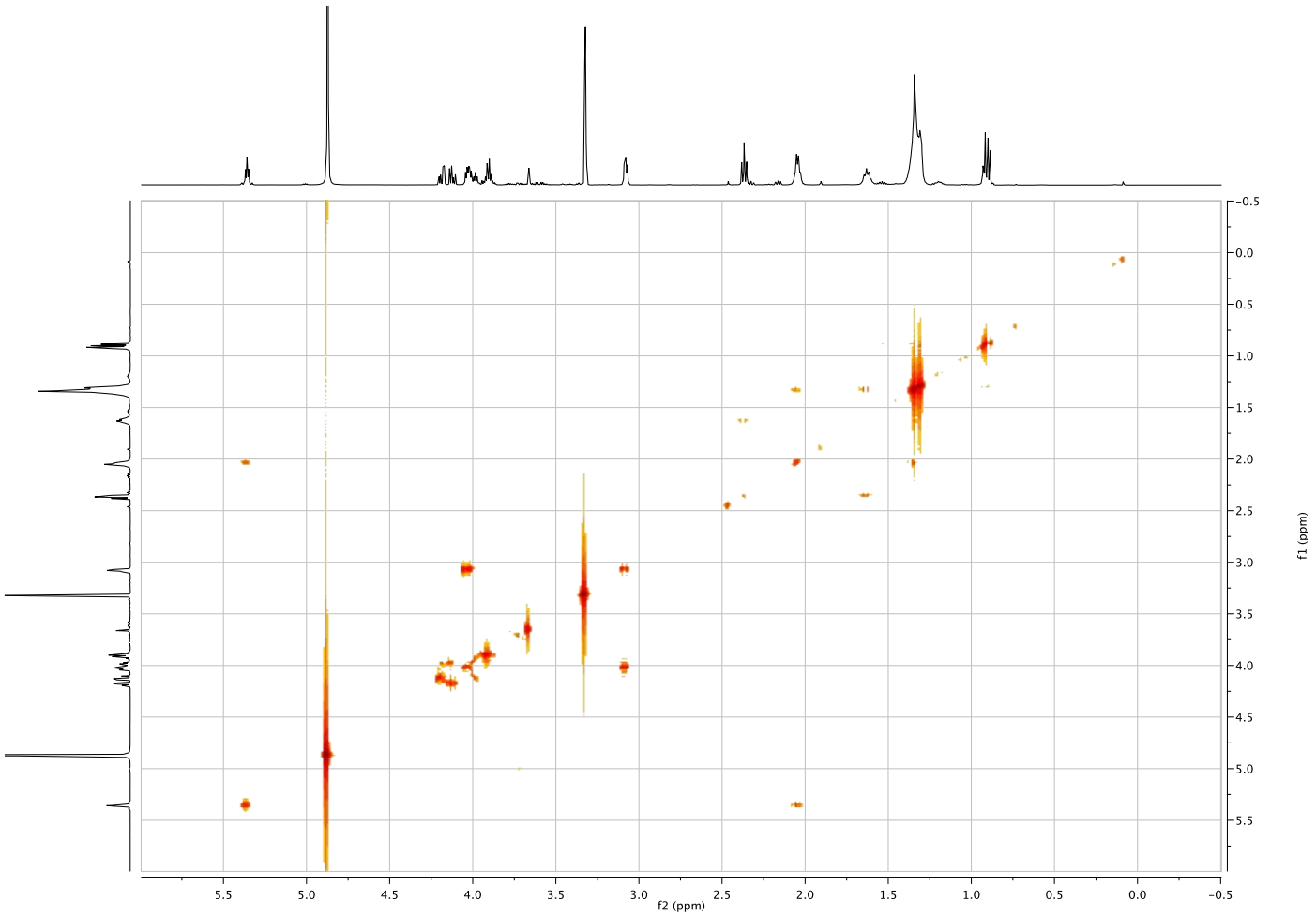
2S19. Lipid interactions inhibit and enhance RIF rosette-inducing activity.
(**A**) To determine the range of concentrations at which IOR-1 displayed inhibitory activity, IOR-1 was titrated and applied to SrEpac alone (circle), or SrEpac + 2 M RIF-2 (square). Rosette development was quantified 25 hours after induction. (**B**) 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 M RIF-2 (square). All concentrations of the LPE mix ≥ 2 M similarly enhanced rosette development in the presence of RIF-2. Rosette development was quantified 25 hours after induction. **(C)** 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 and IOR-1 were kept constant, and added to SrEpac alongside titrating concentrations of the RIF mix.

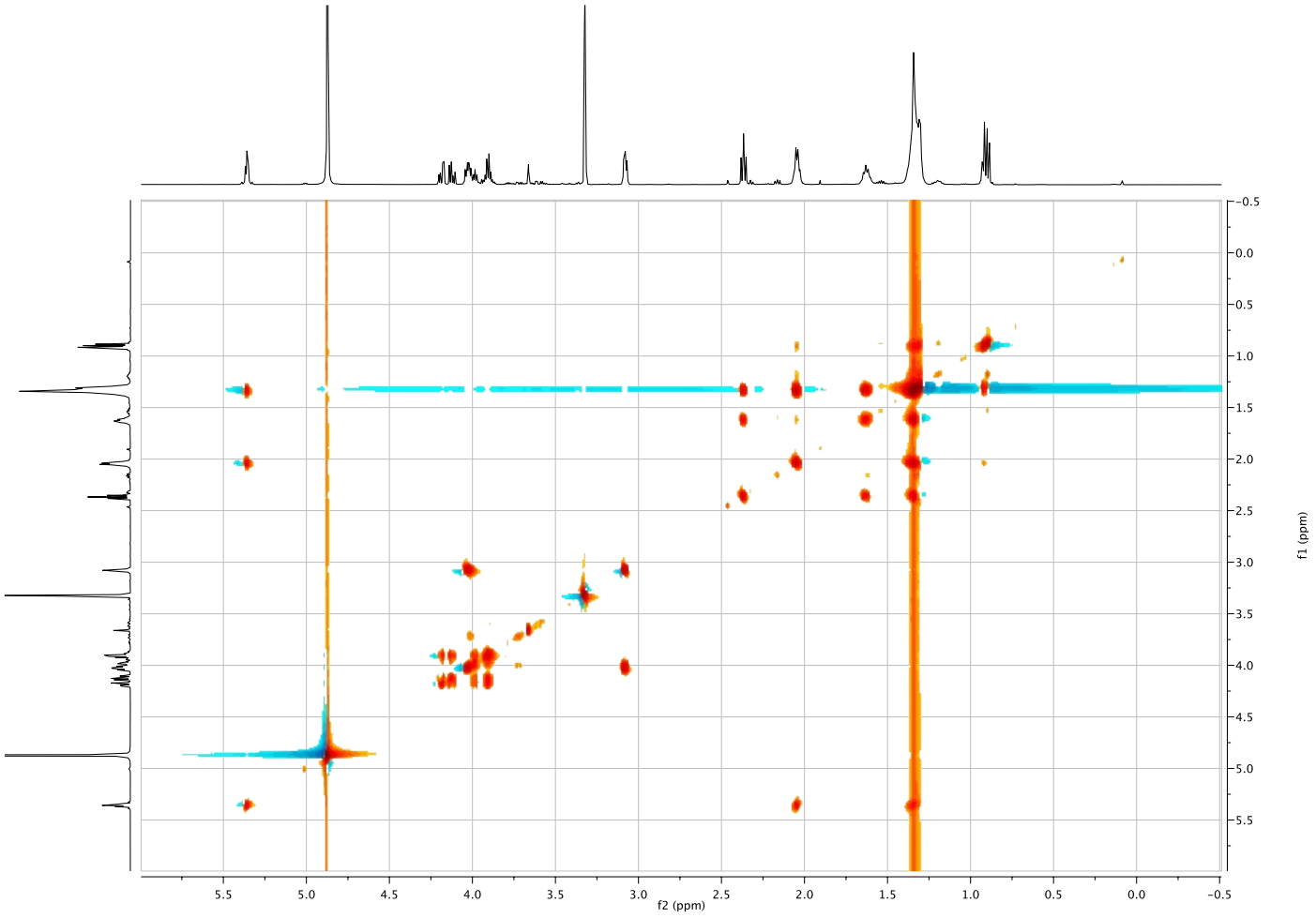




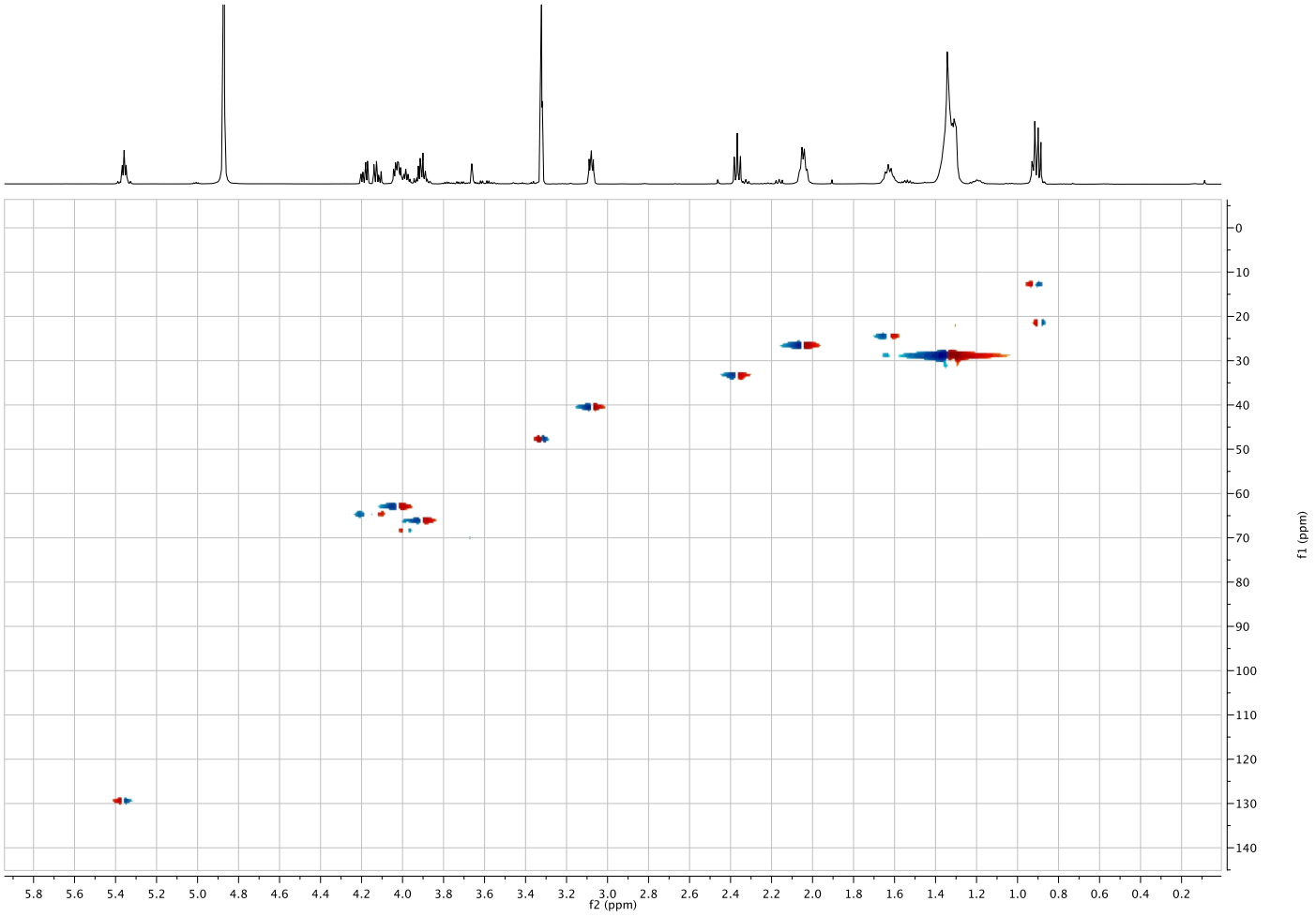
2S20. Isolated LPEs. ^{1}H NMR spectrum, 500 MHz, CD_3OD



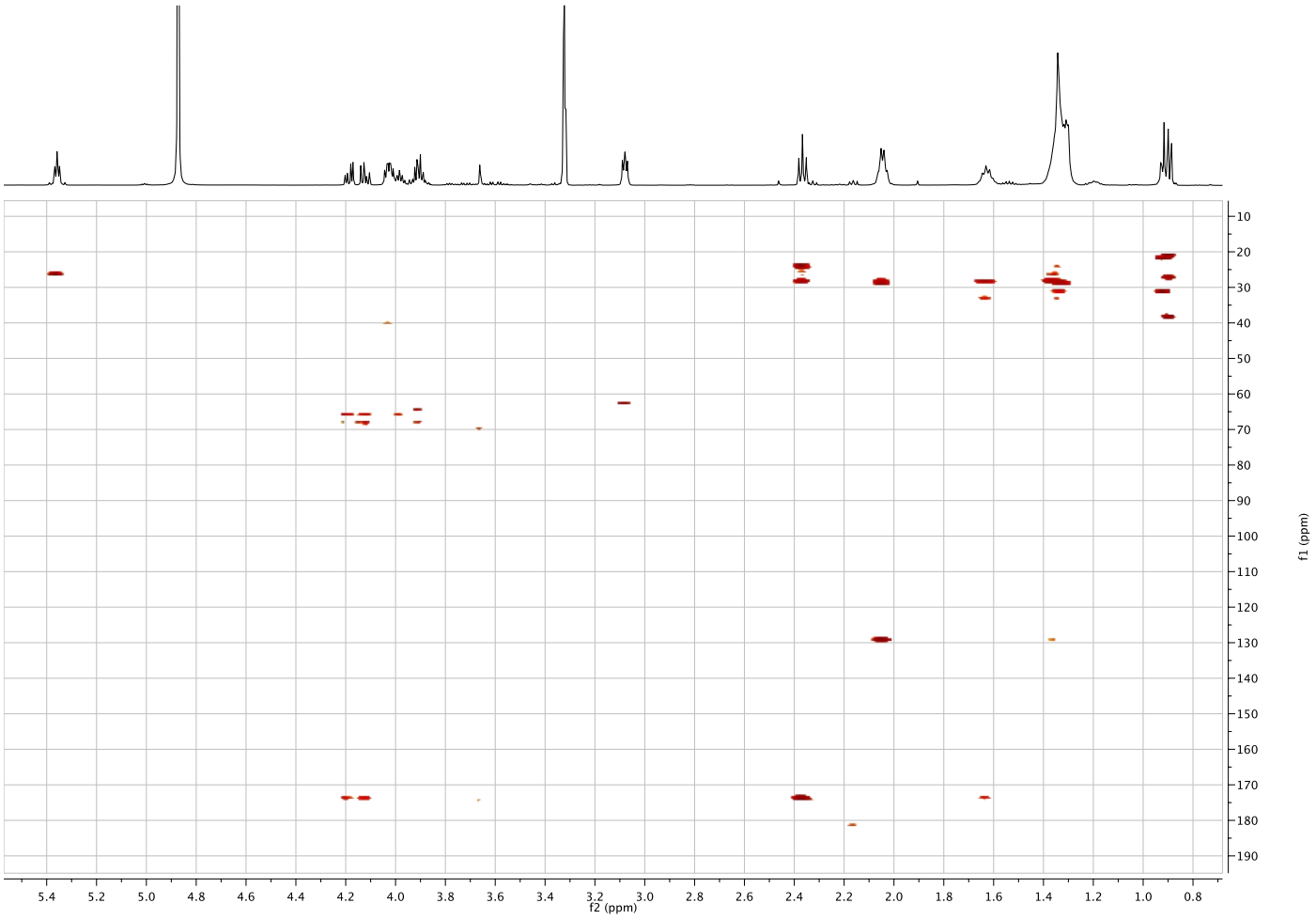
2S21. Isolated LPEs COSY spectrum, 500 MHz, CD_3OD



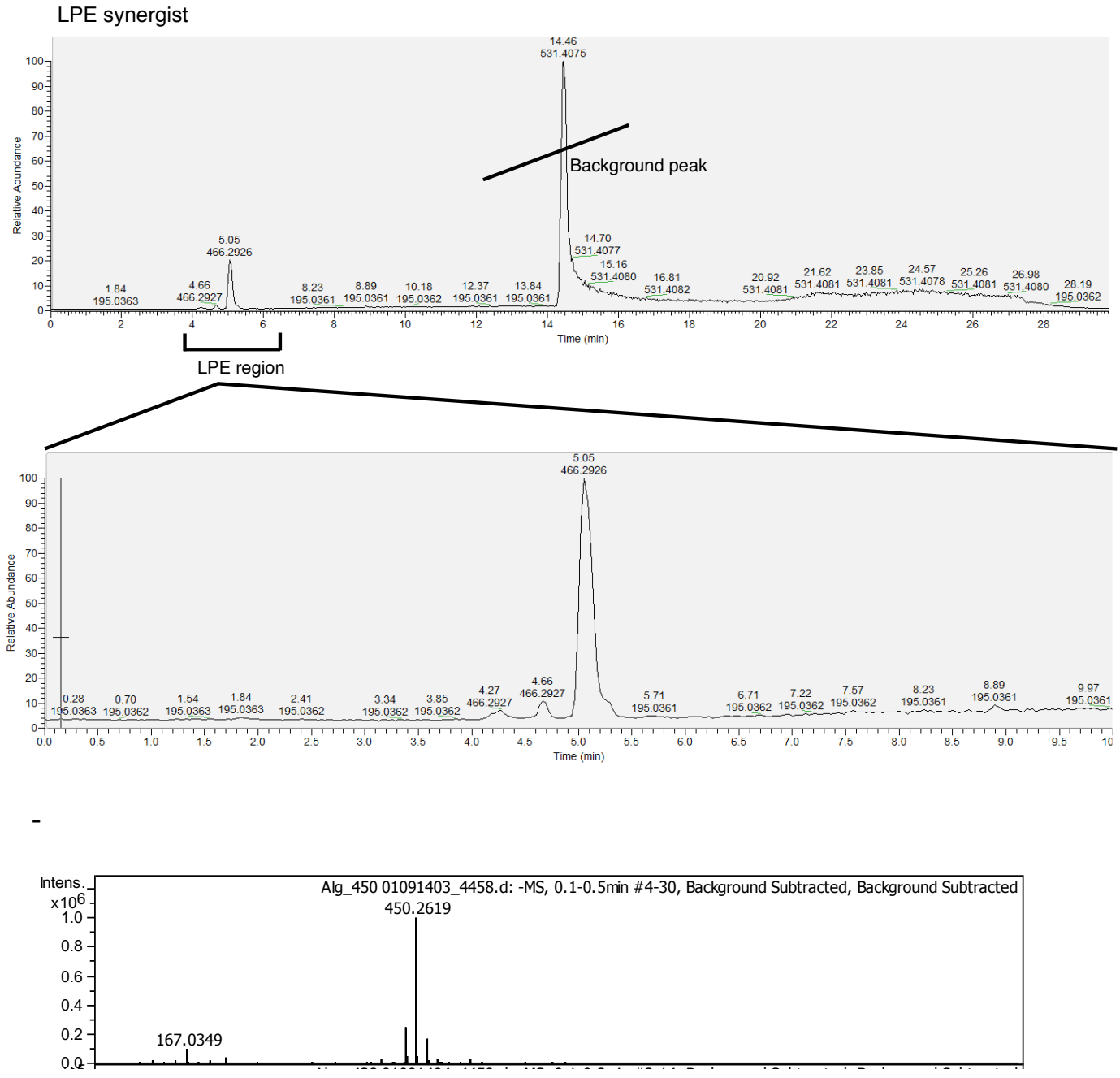
2S22. Isolated LPEs TOCSY spectrum, 500 MHz, CD_3OD



2S23. Isolated LPEs, gHSQCAD spectrum, ^{1}H: 500 MHz, ^{13}C: 126 MHz, CD_3OD

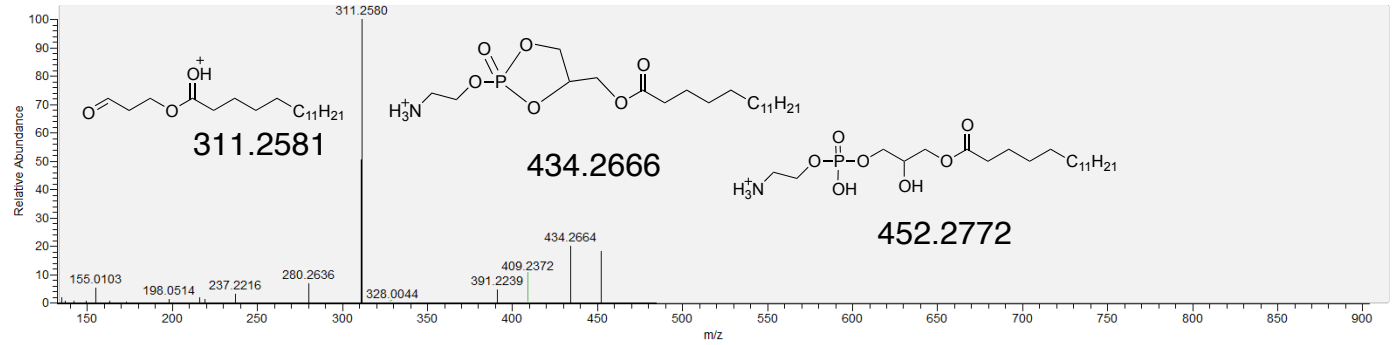


2S24. Isolated LPEs, gHMBCAD spectrum, ^{1}H: 500 MHz, ^{13}C: 126 MHz, CD_3OD

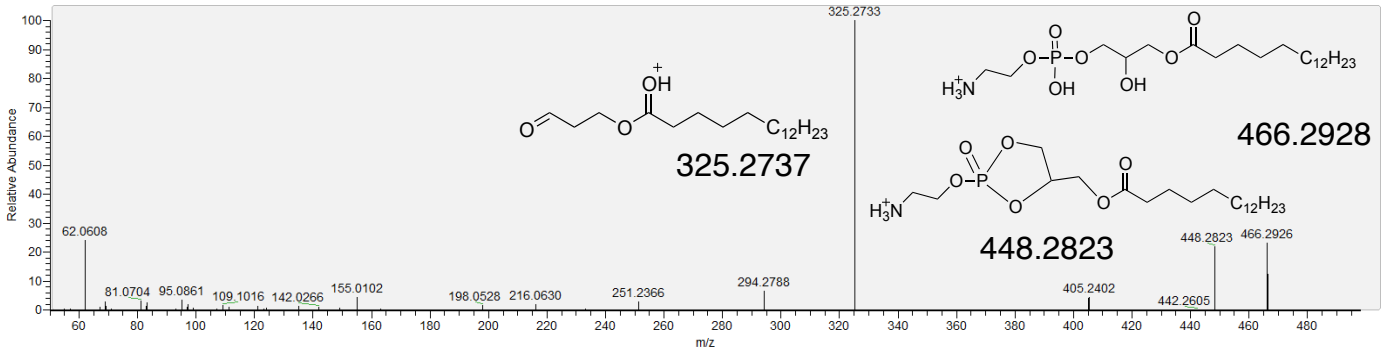


-

2S25. HRMS of LPEs. HR LC-MS of synergistic fraction 465: (M+H) = 466 (top and middle panel), and HRMS of synergistic fraction 451: (M-H) = 450 (bottom panel).



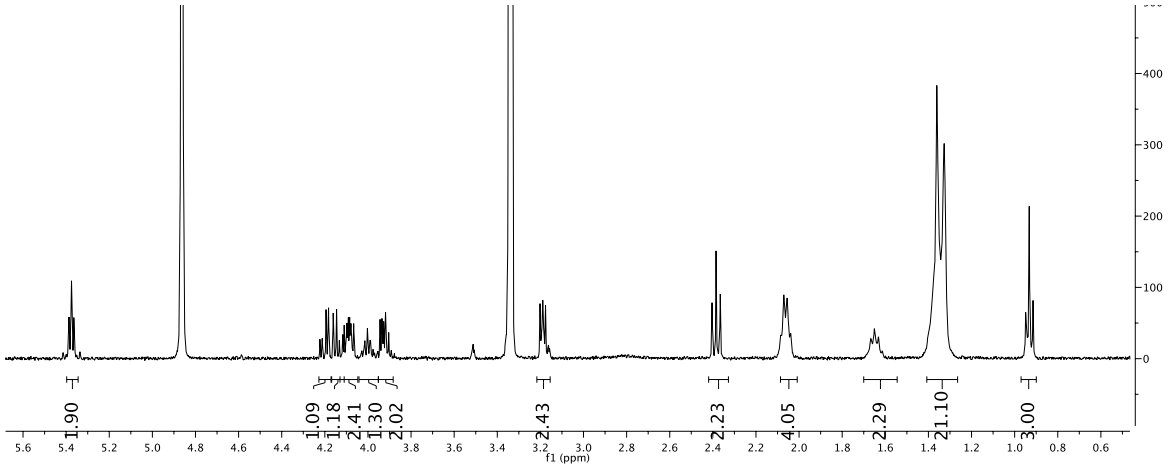
MS/MS pos. 452.2772



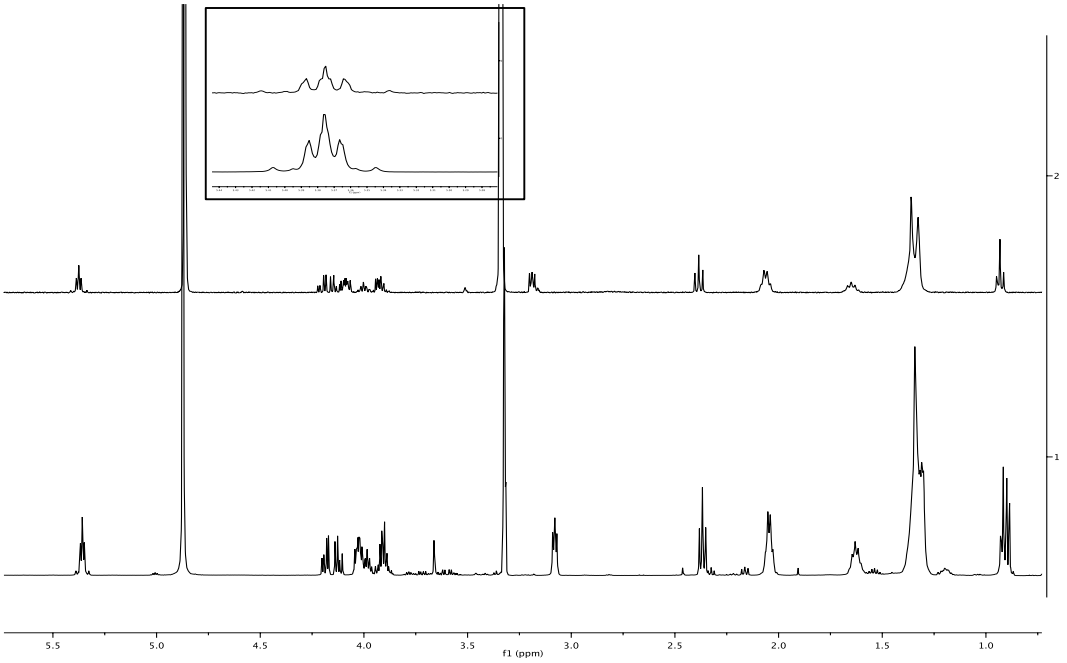
MS/MS pos. 466.2928

2S26. LPE-465 MSMS and predicted fragments (positive mode). Fragmentation of synergist LPE 451 (top panel), fragmentation of synergist LPE 465 (bottom panel).

A

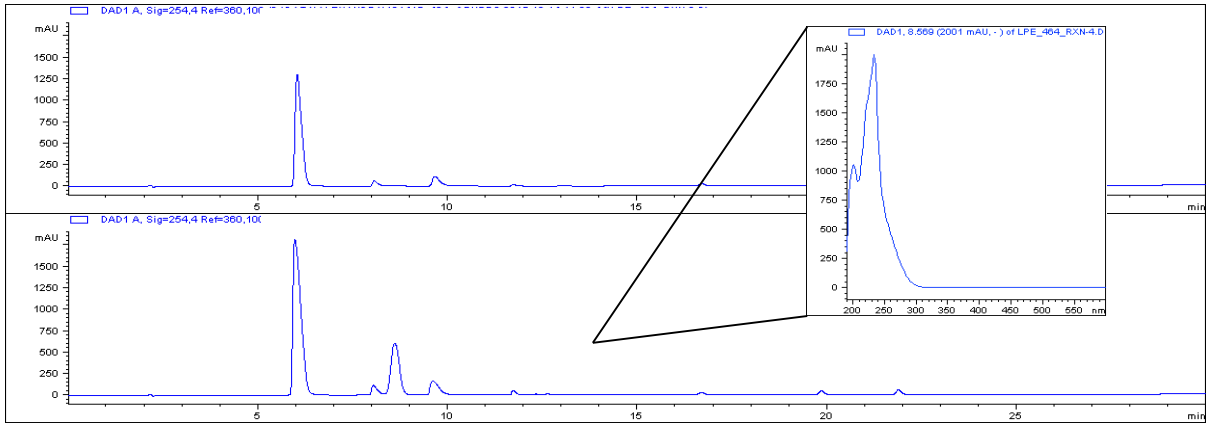


B

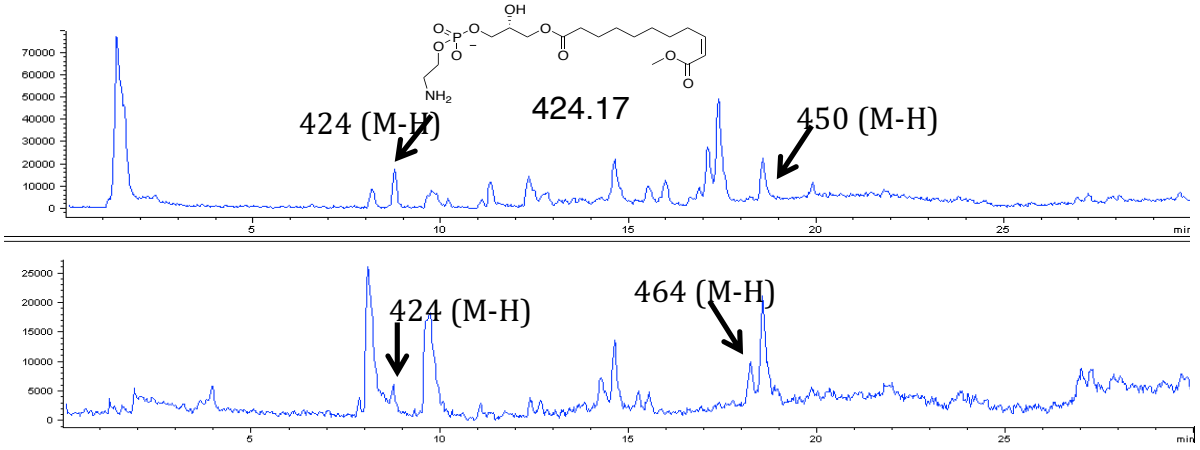


2S27. **NMR spectrum of *Algoriphagus* LPEs compared to a commercial LPE standard. A)** Standard LPE (18:1(9Z)). Avanti Lipids cat no. 846725. **B)** Overlay of commercial LPE (top) vs. LPEs isolated from *Algoriphagus.* Inset: *cis* vinylic protons of commercial LPE (18:1(9Z)): J = 5.0 Hz (top). Vinylic protons of isolated LPE-450/464: J = 4.9 Hz (bottom)

A



B



2S28. Grubbs reaction progress and Grubbs reaction. (A) Top: UV trace (254 nm) immediately after catalyst addition. Bottom: UV trace 3 hours after Grubb's catalyst addition. UV trace of indicated peak ~220 nm – indicative of α,β-unsaturated ester.
(B) Grubbs reaction 10- 450 and 464. Top: MS (-) trace of 450 Grubbs reaction products (crude). Relevant species indicated. Bottom: MS (-) trace of 464 Grubbs reaction products (crude). Both show major acrylate species with mass 424 (M-H) indicating a **Δ**-9 acyl chain is the major species.

Chapter 3

Isolation and synthesis of a bacterially-produced inhibitor of rosette development

The results presented here were published as part of the following paper:

Cantley, A.M.*, Woznica, A.*, Beemelmanns, C., King, N., Clardy, J. (2016) Isolation and synthesis of a bacterially-produced inhibitor of rosette development in choanoflagellates. *J. Am. Chem. Soc*., 138(13): 4326-4329.

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 laboratories 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 the 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.

Results

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 [80]. 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 [30-32,51]. While predominately unicellular, several species of choanoflagellate, including *Salpingoeca rosetta,* alternate between unicellular and multicellular states. 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. 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.

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 coisolated with *S. rosetta* that serves as prey for the choanoflagellate[33,34]. Subsequent synthesis of the inducing

molecule, termed *r*osette-*i*nducing *f*actor-1 (RIF-1), revealed the absolute configuration of the molecule as well as the strict stereochemical requirements for activity (Figure 3.1) [61].

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*.

We performed a chloroform/methanol extraction on the cell pellet of *Algoriphagus* and fractionated the extract by reversed-phase (C-18) HPLC using a broad elution range in order to expand our search beyond sulfonolipids[33]. We then tested each fraction in combination with inducers of rosette development to determine whether any of the 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 (Figure 3.1). We identified two adjacent fractions that reduced rosette formation when treated in combination with either 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 $C_{17}H_{35}O_5S$. One- and two-dimensional NMR experiments permitted us to propose a planar structure for this molecule, which we have named *i*nhibitor *o*f *r*osettes (IOR-1) (Figure 3.1). IOR-1 is optically active ($[\alpha]_D^{22}$ = +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 concentration of 2.5 nM (Figure 3.2), which corresponds with our observation of IOR-1's single-digit-nanomolar concentration in *Algoriphagus*-conditioned medium (Supporting Information, Methods).

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 [81-83]the sulfonic acid headgroup 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 an $-NH_2$ 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[60,61]. 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. The structural similarity between IOR-1 and these signaling molecules suggests that they may also share functional similarities.

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 with 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 C2 and C3, we reasoned that we could reduce an alkyl chain ending in a propargylic alcohol to either the corresponding *cis*- or *trans*-alkene and then perform Sharpless asymmetric dihydroxylations 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.

To reach 15-methylhexadec-2-yn-1-ol **(3)**, we started with commercially available 10-undecyn-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 that described previously. Propargylic alcohol **3** was obtained through acetylide formation and subsequent nucleophilic addition to paraformaldehyde. At this stage our synthetic strategy diverged to obtain both the *cis*- and *trans*-alkene. 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*-alkene *E*-4. These reductions were achieved in acceptable yields of 74% and 70% respectively. 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. This stage proved suitable to purify the diols via chiral chromatography, yielding enantiopure 6A, 6B, 6C, and 6D (Scheme 1).

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 was 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 the final product (14–20%), allowed a much simpler purification process and higher overall conversion.

The ^{1}H NMR spectra of compounds IOR-1A and IOR-1B were identical to that of isolated IOR-1, whereas compounds IOR-1C and IOR-1D exhibited different chemical shifts of protons at positions C1, C2, and C3. Determination of the optical rotations for these molecules revealed matching signs and values for IOR-1A and IOR-1 (Supporting Information, Methods), suggesting that 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. Full dose–response curves revealed almost identical activity for IOR-1A compared to the isolated inhibitor, whereas IOR-1B displayed no activity (Figure 3.2); unsurprisingly, IOR-1C and IOR-1D were also inactive. Given both the spectroscopic and biological data, we were able to determine the absolute configuration of IOR-1 as 2*S*,3*R*. Significantly, only one stereoisomer of the inhibitor is active, reprising the theme that these

molecules interact in a highly specific manner with their target. We further validated this specificity by testing a handful of commercially available IOR-1 analogues, and none were active at concentrations ranging from 0.1 ng/mL to 1 μg/mL.

Furthermore, this specificity suggests that 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 nonspecific interactions even if the probes were of somewhat lower potency. Our synthetic route also supplies us with inactive stereoisomers of IOR-1, which can serve as 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 2-amine and 3-hydroxyl groups observed in RIF-1, which is by far the most common stereochemistry for sulfonolipids and sphingolipids[81]. While not unprecendented, it is quite rare for capnine bases to exhibit the syn configuration, and biosynthesis of the *syn*-diol has not been reported. 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 C2 during conversion from an amino group. 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 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–bacterium 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 this is true, production of factors that attenuate rosette-colony formation would confer an apparent benefit to the producing bacteria[57]. 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 the 2*S*,3*R* 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.

FIGURES

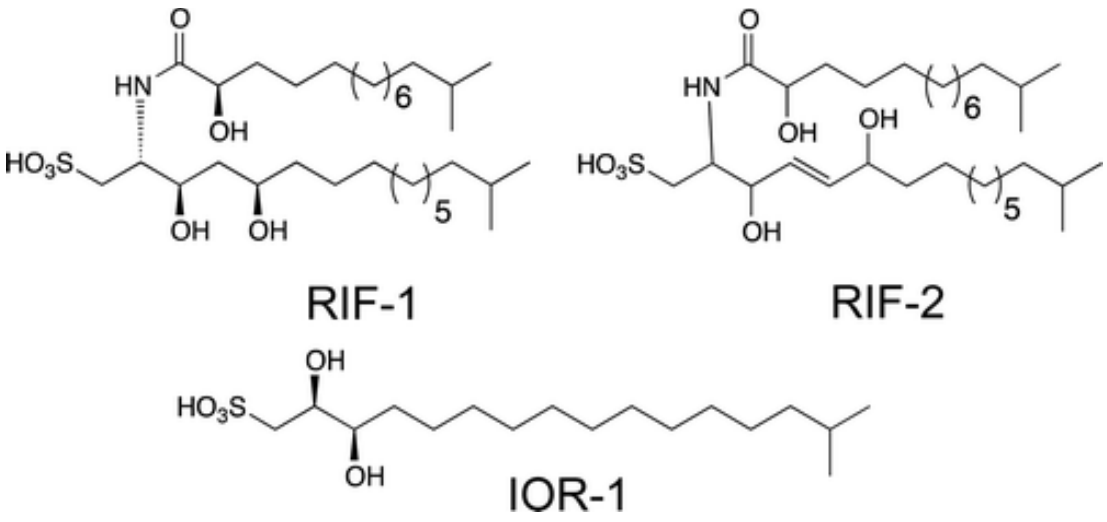


Figure 3.1. Previously isolated rosette-inducing molecules RIF-1 and RIF-2 and inhibitor of rosettes (IOR-1).

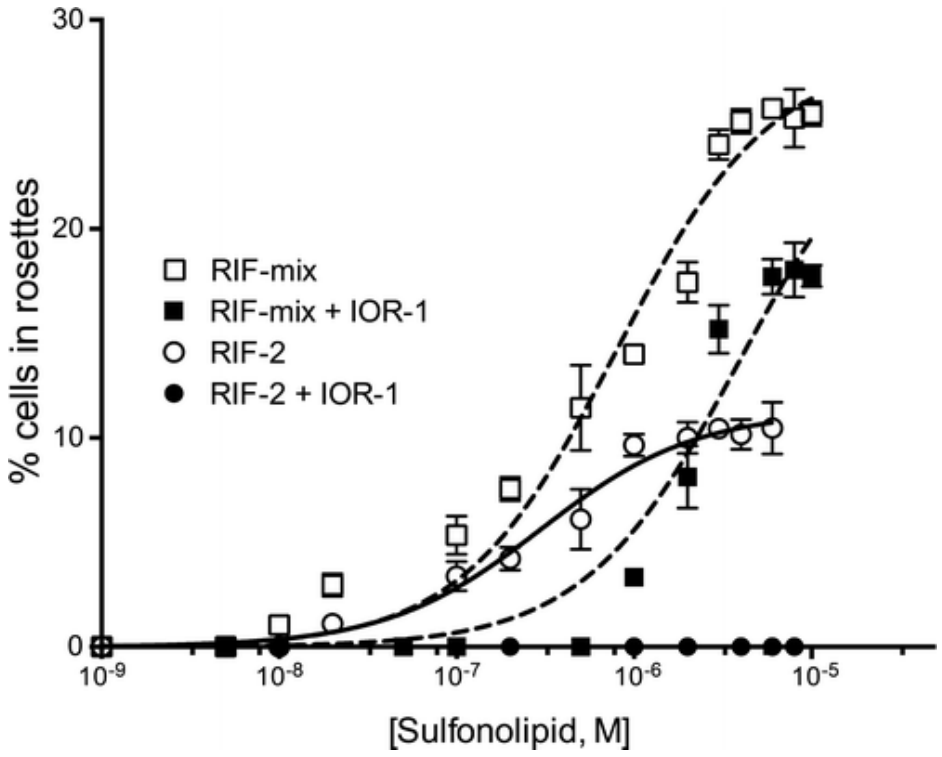
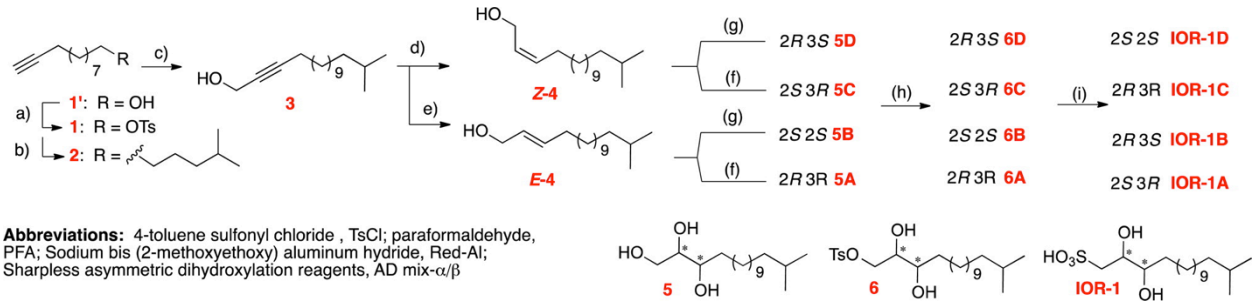


Figure 3.2. Co-treatment of IOR-1 (2.5 nM) with RIF-2 and RIF-mix. Graphs were generated using GraphPad Prism 6 statistical software. The rosette induction data were analyzed using a one-site (specific binding) model.



Scheme 1. Synthesis of IOR-1 Stereoisomers A–D

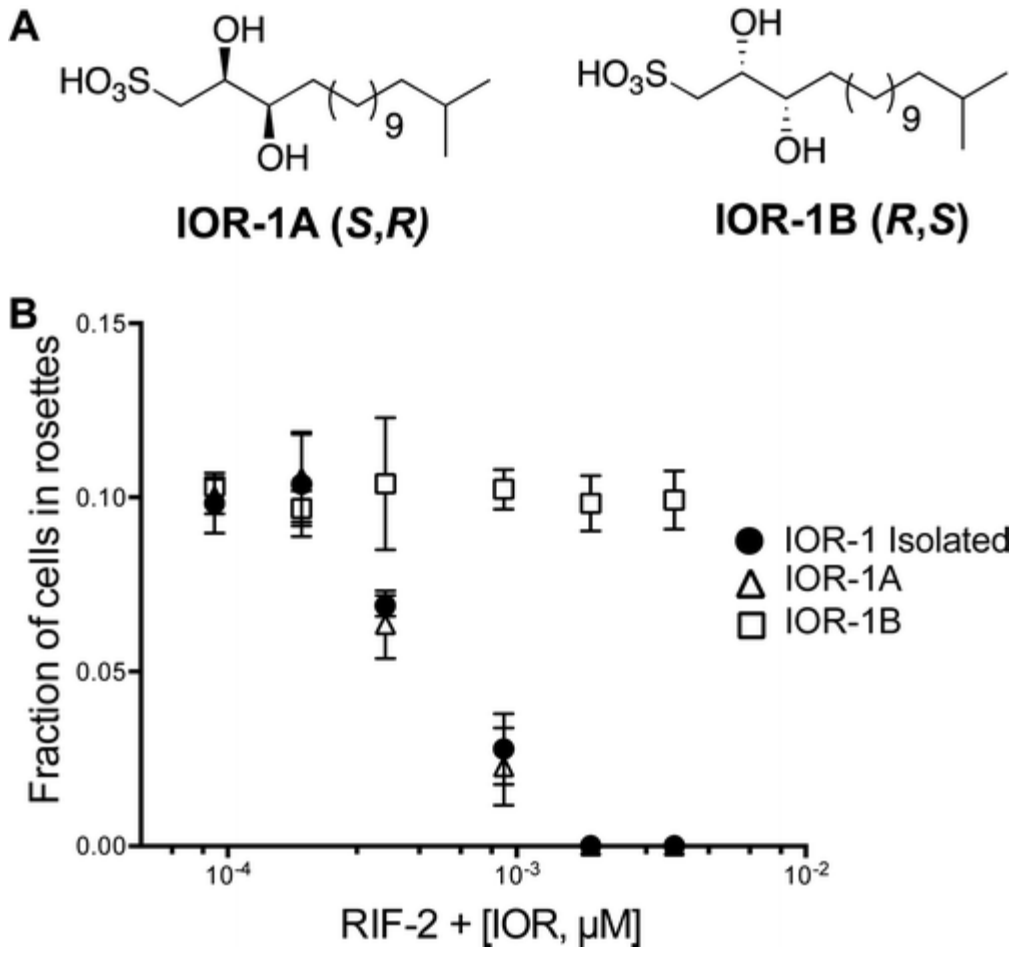


Figure 3.3. IOR-1A recapitulates isolated IOR-1. (A) Structures of the synthetic compounds IOR-1A and IOR-1B. (B) Comparison of dose–response curves of IOR-1A, IOR-1B, and isolated IOR-1. RIF-2 treated at 2 µM.

Chapter 4

An aphrodisiac produced by *Vibrio fischeri* stimulates mating in the closest living relatives of animals

Introduction

Bacterial–eukaryotic interactions are ubiquitous, and the influences of bacteria on eukaryotes vary from subtle to profound. Yet, because eukaryotes are often associated with complex and unseen communities of bacteria, the breadth of eukaryotic biological processes regulated by bacteria and the underlying molecular dialogue often remain mysterious. Nonetheless, studies of experimentally tractable host-microbe pairs have revealed a growing number of examples in which bacteria regulate eukaryotic cell biology and development, in some cases using molecular cues that mediate pathogenesis in other contexts[13].

One of the closest living relatives of animals, the marine choanoflagellate *S. rosetta*, has emerged as an attractive model for uncovering bacterial cues that regulate eukaryotic development. Like all choanoflagellates, *S. rosetta* survives by eating bacteria[23,38]. However, interactions between *S. rosetta* and bacteria extend far beyond those of predator and prey. In prior work, we demonstrated that the developmental switch triggering the formation of multicellular "rosettes" from a single founding cell[32] is regulated by specific lipids produced by the environmental bacterium *Algoriphagus machipongonensis*[33,35,84]. Rosette development is one of at least six different developmental switches in the sexual and asexual phases of *S. rosetta*'s dynamic life history[31,37], but until now was the only choanoflagellate process known to be regulated by bacterial cues.

We report here on our serendipitous discovery that sexual reproduction in *S. rosetta* is regulated by a secreted cue from the marine bacterium *Vibrio fischeri.*

Results

S. rosetta *forms mating swarms upon exposure to* V. fischeri

V. fischeri is perhaps best understood as a model for bacterial quorum sensing and as a symbiont required for the induction of light organ development in the squid, *Euprymna scolopes*[85]. Although *Vibrio* spp. are known symbionts, commensals, and pathogens of animals[86], *V. fischeri* does not induce rosette development[33] and was not previously known to influence any aspect of *S. rosetta* biology. We were therefore surprised to observe that the addition of live *V. fischeri* bacteria to a culture of single-celled, motile *S. rosetta* induced cells to gather rapidly into loose aggregates or "swarms," each composed of between 2-50 cells (Figure 4.1A,B; Figure 4.4A). This dynamic and previously unobserved swarming behavior began as early as 15 minutes after induction with *V. fischeri*, with individual *S. rosetta* cells often moving between swarms that periodically broke apart or merged with other swarms. In its timescale, mechanism, and outcome, swarming was clearly unrelated to the *Algoriphagus*-induced developmental process by which a single *S. rosetta* cell divides repeatedly to form a rosette[32].

Although swarming has not previously been reported in choanoflagellates and the biological significance of swarming in *S. rosetta* was not immediately obvious, swarming is

associated with mating in diverse motile eukaryotes, including amoebae, ciliates, crustaceans, insects, fish, birds, and bats[87-93]. Therefore, we hypothesized that swarming in *S. rosetta* might indicate mating. To investigate whether *V. fischeri*-induced swarming is a prelude to mating, we sought to determine whether the hallmarks of mating in microbial eukaryotes (cell fusion, nuclear fusion, and meiotic recombination[37,94,95]) occur in *S. rosetta* following treatment with *V. fischeri.*

Our lab previously found that starved *S. rosetta* cells mate at low frequencies (<2% of the population) after starvation for 11 days[30,37]. In contrast, as early as 30 minutes after induction with live *V. fischeri* or conditioned medium isolated from a pure *V. fischeri* culture, *S. rosetta* cells formed swarms and then began to pair up and fuse (Figure 4.1C). Once paired, cell fusion took as little as three minutes and all observed cell pairs were oriented with their basal poles (opposite the flagellum) touching. Paired cells subsequently fused along the basal pole, resulting in the formation of a binucleate cell harboring two flagella (Figure 4.1D). After cell fusion, the two nuclei congressed and fused, and one of the two parental flagella eventually retracted (Figure 4.4B), resulting in a diploid cell.

While cell fusion and nuclear fusion are consistent with mating, parasexual processes can occur in the absence of meiotic recombination[96]. Therefore, to test whether swarming was associated with the initiation of a true sexual cycle, we used *V. fischeri* to induce the formation of heterozygous diploids in *S. rosetta* cultures and then examined their offspring for evidence of meiosis and recombination (Figure 4.1E). To produce heterozygous diploid cells, two haploid *S. rosetta* strains (R+ and R–) containing previously characterized polymorphisms[30] were mixed either in the presence of *V. fischeri* conditioned medium (VFCM), or in conditioned medium from *Echinicola pacifica* (EPCM), a prey bacterium[30] that does not induce swarming, as a negative control (Figure 4.1E). After VFCM or EPCM exposure, 48 clones were isolated from each culture condition. Although we cannot directly measure the ploidy of live *S. rosetta* cells, heterozygous diploids can be readily identified by genotyping. While all clones (48/48) reared from the EPCM-treated culture contained un-recombined parental genotypes, 10/48 clones isolated following VFCM treatment were shown by genotyping to be heterozygous diploids. We surmised that the heterozygous diploids were the result of outcrossed mating, and found that further passaging of these clones yielded motile haploid progeny. 147 haploid progeny from three different heterozygous diploids were clonally isolated and genotyped at polymorphic sites across the genome, providing evidence for independent assortment and meiotic recombination (Figure 4.1E). Taken together, these results demonstrate that *V. fischeri* produces an aphrodisiac that induces swarming and mating in *S. rosetta*.

Bioactivity-guided fractionation revealed that the* V. fischeri *aphrodisiac is a protein

The formation of large swarms by *S. rosetta* single cells in response to VFCM combined with automated image analysis provided the basis for a quantitative bioassay (Figure 4.2A,B; Methods). As a baseline, we found that 30 minutes after treatment with VFCM, *S. rosetta* cells consistently formed swarms containing between 5-35 cells each, whereas cells did not form clusters in response to EPCM. Using this bioassay, we first tested whether *V. fischeri* cues involved in quorum sensing (e.g. homoserine lactones)[97,98] or those required for its symbiosis with the squid *Euprymna scolopes* (e.g. lipopolysaccharide (LPS) and peptidoglycan (PGN))[13,99] might contribute to swarming induction in *S. rosetta*. A set of five different *V. fischeri* mutant strains that are deficient in quorum sensing were all wild type for swarming induction[97], as were seven mutants in polysaccharide export pathways required for symbiosis with *E. scolopes*[33]

(Table 4.1). Moreover, treatment of *S. rosetta* with purified quorum sensing molecules (Table 4.2) and *V. fischeri* outer membrane vesicles (OMVs) containing LPS and PGN[34] (Figure 4.5A) also failed to elicit mating, suggesting that the cue(s) required for *S. rosetta* mating induction likely differ from factors required either for quorum sensing or squid colonization.

We next turned to an unbiased, activity-guided fractionation and found that the aphrodisiac was enriched in VFCM, even after depletion of OMVs (Figure 4.5A). The aphrodisiac produced by *V. fischeri* could be recovered from VFCM by ammonium sulfate precipitation, and the activity of the ammonium sulfate fraction was sensitive to both heat and protease treatment, suggesting that the activity might be proteinaceous (Figure 4.2C). We therefore separated all proteins precipitated from VFCM by size exclusion and anion exchange chromatography, and tested the protein fractions in the swarming bioassay (Figure 4.2A-C, Figure 4.5). The swarming activity tracked with a single ~90kD protein, which was revealed by mass spectrometry to be the uncharacterized *V. fischeri* protein VF_A0994, hereafter referred to as EroS (Figure 4.2C; GenPept Accession YP_206952). To test whether EroS was sufficient to induce swarming in *S. rosetta,* we heterologously expressed EroS in *E. coli* and found that purified EroS recapitulated the swarm-inducing activity of live *V. fischeri* and VFCM (Figure 4.2D). Purified EroS was also sufficient to induce mating between two *S. rosetta* strains (R+ and R–;Table 4.3), demonstrating that a single protein secreted by *V. fischeri* is sufficient to induce both swarming and mating in *S. rosetta.*

The S. rosetta *aphrodisiac is a chondroitinase*

To understand the mechanism by which *V. fischeri* induces choanoflagellate mating, we set out to determine the biochemical function of EroS. The EroS protein sequence contains a predicted glycosaminoglycan (GAG) lyase domain (supported by the detection of PFAM[100] domains PF08124, PF02278, PF02884). GAGs are linear polysaccharides that are integral components of the animal extracellular matrix (ECM). GAG lyases depolymerize GAGs through an elimination mechanism that distinguished them hydrolases and are produced, and often secreted, by a subset of primarily pathogenic bacteria and fungi[101]. GAG lyases are also produced by human commensals, including by gut bacteria in the genus *Bacteroides*[102,103]. Through the alignment of EroS with amino acid sequences from bacterial GAG lyases with solved structures, we found that EroS harbors conserved residues (His-278 and Tyr-287) at sites required for catalytic activity (Figure 4.3A)[104-107].

GAGs are diverse and the substrate specificities of GAG lyases cannot be deduced from sequence alone[101]. Moreover, GAGs are thought to be eumetazoan-specific innovations[‡‡] that facilitated animal multicellularity[109,110], and are not known to exist in choanoflagellates. Therefore, we next set out to answer three questions: (1) does EroS exhibit GAG-degrading activity, (2) is the enzymatic activity of EroS required for its function as an aphrodisiac, and (3) what are its substrates in *S. rosetta*?

We found that purified EroS degrades GAG substrates *in vitro,* and is thus a *bona fide* GAG lyase. GAGs are classified based on their disaccharide units: heparan sulfate, chondroitin sulfate, dermatan sulfate, hyaluronic acid, and keratan sulfate[101]. EroS showed strong lyase activity toward purified chondroitin sulfate and hyaluronan, but not heparan sulfate or dermatan

[‡‡] Although some pathogenic bacteria produce extracellular GAGs to evade host immune response, bacterial GAGs are the result of an independently evolved biosynthetic pathway and remain unsulfated[108]

sulfate (Figure 4.3B; Figure 4.6). We did not test keratan sulfate because it does not contain uronic acid and therefore cannot be depolymerized by GAG lyases[111].

We next asked whether the enzymatic activity of EroS is important to its function as an aphrodisiac. Well-characterized chondroitin lyases from other bacteria (ABC chondroitinase from *Proteus vulgaris* and AC chondroitinase from *Flavobacterium heparinum*) induced swarming and mating in *S. rosetta* at levels resembling those induced by EroS (Figure 4.3C, Table 4.3), indicating that the GAG lyase activity of EroS is both necessary and sufficient for its function as an aphrodisiac.

Although sulfated GAGs were previously thought to be restricted to animals, key heparan biosynthetic enzymes have been detected in the genome of the choanoflagellate *Monosiga brevicollis*[112], and we have further found that the *S. rosetta* genome encodes homologs of enzymes required for heparan and chondroitin biosynthesis (Figure 4.3D, Figure 4.7A)[28,29]. To test whether chondroitin is produced by *S. rosetta,* we treated polysaccharides isolated from planktonic *S. rosetta* with the broad specificity ABC chondroitinase from the bacterium *P. vulgaris* (Figure 4.3E, Figure 4.7B,C). The *P. vulgaris* ABC chondroitinase liberated chondroitin disaccharides, demonstrating that *S. rosetta* indeed produces chondroitin. Finally, to test whether EroS can degrade *S. rosetta* chondroitin we treated *S. rosetta* polysaccharides with EroS and found that it released unsulfated chondroitin and chondroitin-6-sulfate disaccharides, indicating that *S. rosetta* chondroitin is a target of EroS (Figure 4.3E, Figure 4.7B,C).

Chondroitin sulfate in animal ECM can be found covalently linked to core proteins, thus forming proteoglycans. To determine whether chondroitin disaccharides released from proteoglycans play a role in stimulating mating, we tested various products of EroS digestion for aphrodisiac activity (Figure 4.9). Conditioned media from EroS-induced *S. rosetta* cells did not trigger swarming in naïve *S. rosetta*, nor did the digested products of commercial chondroitin sulfate treated with EroS. Moreover, swarming was not induced by any combination or concentration of unsulfated and 6-sulfated chondroitin disaccharides tested (Figure 4.9). These results lead us to hypothesize that the structural modification of proteoglycans by EroS, rather than the chondroitin disaccharides products of EroS digestion, are important for activating the swarming and mating pathway in *S. rosetta*.

Finally, because swarming has not been previously described in choanoflagellates, we investigated whether *V. fischeri* might induce swarming under plausible environmental conditions. We found that EroS is secreted constitutively by *V. fischeri* when grown under either high or low nutrient conditions (Table 4.4), and observed that cultures of *S. rosetta* swarm in response to *V. fischeri* densities as low as $4x10^2$ cells/mL– the equivalent of one *V. fischeri* cell per 500 *S. rosetta* cells – within 30 minutes of exposure (Figure 4.10A). Moreover, EroS was sufficient to trigger robust swarming in *S. rosetta* at concentrations as low as 5 pM (Figure 4.10B), making EroS as potent as the sex pheromones produced by volvocine algae[113] and by marine invertebrates[114,115]. Together, these data suggest that *V. fischeri*, which ranges in density in the oligotrophic oceans from $6x10^2$ cells/mL to $>1x10^4$ cells/mL during blooms[116], or other chondroitinase-producing bacteria could plausibly trigger *S. rosetta* swarming and mating *in natura*.

Discussion

We have discovered that a secreted bacterial chondroitinase induces mating in *S. rosetta,* one of the closest living relatives of animals. To our knowledge, the interaction between *V.*

fischeri and *S. rosetta* is the first known example of bacteria regulating mating in a eukaryote. Moreover, through the study of this novel interkingdom interaction, we found that mating in *S. rosetta* is initiated in response to the degradation of chondroitin sulfate, a glycosaminoglycan previously thought to be restricted to animals.

The first hint that *V. fischeri* might induce mating came from the observation of *S. rosetta* swarms following exposure to the bacterium. By increasing local population density, swarming has previously been found to facilitate mating in diverse amoebae, flagellates, crustaceans, cnidarians, polychaetes, insects, fish, and birds[87-91,117-119]. As in other organisms that swarm, the connection of swarming to mating may be critical, since their aquatic, pelagic lifestyle can make it challenging to find mates. Indeed, under the starvation conditions that trigger *S. rosetta* mating in the absence of swarming, mating takes >500X longer (~11 days) and occurs in only 2% of the population[30].

Most previously characterized examples of coordinated mating behaviors are regulated by pheromonal cues. Conspecific swarming is initiated by diverse aggregation pheromones (for example, ester and isoprenoid pheromones in beetles[120,121] and peptide pheromones in sea slugs[122] and polychaetes[123]), and free-spawning marine animals produce pheromones to synchronize gamete release and enhance fertilization success[124]. Biotic and abiotic cues from the environment can also help coordinate mating behavior in animals. For example, spawning in marine invertebrates is correlated with phytoplankton blooms, and some sea urchins, mussels, and polychaetes spawn after exposure to small molecules produced by environmental phytoplankton[125-127], although these cues remain structurally elusive. Just as phytoplankton blooms are hypothesized to signify a nutrient-rich environment for spawning, the presence of chondroitinase-producing bacteria may indicate an environmental condition, or the convergence of multiple environmental factors, that favor mating in *S. rosetta*. Although *V. fischeri* was the first bacterium observed to regulate mating in *S. rosetta*, we have since identified other bacteria that similarly induce swarming and mating (Table 4.1). Therefore, we predict that mating in *S. rosetta* might be regulated by diverse species of bacteria in nature, and hypothesize that swarming is a common occurrence within the natural life history of *S. rosetta*.

Our discovery that *V. fischeri* produces a chondroitinase that functions as an aphrodisiac also revealed that *S. rosetta* produces chondroitin sulfate, providing the first biochemical evidence for this important GAG in a non-animal and extending its evolutionary history to the premetazoan era. In an interesting parallel to the induction of *S. rosetta* mating by a chondroitinase, GAGs and sulfated polysaccharides mediate mating in diverse internally and externally fertilizing animals where they provide a protective coating around oocytes[128]. In the case of the mammalian oocyte, which is surrounded by the GAG hyaluronan, sperm penetrate the protective coating by secreting hyaluronidase, a glycoside hydrolase[129]. Of course, GAGs like hyaluronan and chondroitin sulfate are also an essential part of the ECM in somatic cells of animals, where they contribute to a range of functions that include the maintenance of cell adhesion through interactions with ECM molecules, the integration of signals from the extracellular milieu, and the stabilization of collagen fibers[31]. In the future, it will be interesting to explore whether *S. rosetta* GAGs similarly function to mediate cell recognition in the context of fertilization[§§] and cell adhesion in the context of multicellular rosette development.

[§§] Indeed, the *S. rosetta* genome encodes homologs of GAG lyases (EGD79853.1; EGD79387.1) although neither are expressed under laboratory conditions and both have thus far been intransigent to heterologous expression or biochemical analysis

Methods

Culture media

Artificial seawater (ASW), cereal grass media (CG media), and Sea Water Complete media (SWC) were prepared as described previously [35,37]. Artificial sea water (ASW) 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. SWC media was made by adding 250 mg/L peptone, 150 mg/L yeast extract, 150 L/L glycerol in artificial sea water. CG media was made by infusing ASW with cereal grass pellets (Basic Science Supplies, Rochester NY).

Choanoflagellate husbandry

SrEpac[37] (*S. rosetta* grown in the presence of *Echinicola pacifica* bacteria, ATCC PRA-390) was propagated in 5% Sea Water Complete media (SWC diluted to 5% vol/vol in ASW) at 22°C. SrEpac was passaged 1:20 into 19mL fresh 5% SWC every other day to obtain stationary growth phase cultures (cells were grown in 25cm^2 Corning cell culture flask). Prior to all induction bioassays, unless otherwise indicated, cells were diluted to approximately 1x10^5 cells/mL in ASW at the time of induction.

Immunofluorescence microscopy

Stationary-phase cells were induced with *V. fischeri* conditioned media or *E. pacifica* conditioned media and fixed at intervals of 10 minutes, 30 minutes, 1 hour, 2 hours, and 4 hours after induction. After vortexing, cells were fixed for 5 min in 6% acetone followed by 10 minutes in 4% formaldehyde. Cells were then allowed to settle for 30 min onto poly-L-lysine coated coverslips (BD Bioscience). Cells were stained with E7 anti-tubulin antibody (1:200; Developmental Studies Hybridoma Bank), Alexa Fluor 488 anti-mouse (1:1000; 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.

Mating stages were assigned based on the following criteria: orientation of paired cells, fusion of cell bodies, localization and number nuclei, number of flagella. Cell fusion could be clearly distinguished from cell division for several reasons, including (1) fusing cells are paired basally, whereas recently divided sister cells are paired laterally, (2) flagella remain elongated during the fusion process, but are retracted throughout cell division and (3) DNA remains uncondensed throughout cell fusion, but is condensed during cell division.

Isolation of conditioned media (including VFCM and EPCM)

Vibrio fischeri ES114 (ATCC 700601) and all other *Vibrio* species (Table 4.2) were grown by shaking in 200mL 100% SWC media for 30H at 20°C, and pelleted by centrifugation. *E. pacifica* was grown by shaking in 200mL 100% SWC for 30H at 30°C, and pelleted by centrifugation. Cell-free supernatant was then vacuum filtered twice through a 0.22 M filter (EMD Millipore Stericup) to obtain 100% CM. Concentrated conditioned media was obtained using 30kD and 50kD molecular weight cut off centrifugal filter units (Amicon).

Inducing mating and meiosis

<u>*S. rosetta* strains:</u> All crosses were performed between two *S. rosetta* strains with previously verified single nucleotide polymorphisms (SNPs), R- (previously referred to as Rosetteless) and R+ (previously referred to as Isolate B)[30]. Prior to inducing mating, stationary phase cultures

were obtained by passaging Rosetteless 1:20 into 19mLs fresh 5% SWC media every other day, and Isolate B 1:10 into 20mLs fresh 25% CG media every two days.

Inducing mating: Stationary phase R+ and R- cultures were counted and diluted to the same cell density ($1x10^6$ cells/mL). R+ and R- cultures were mixed in equal proportions, pelleted, and resuspended in fresh 25% CG media to obtain a final cell density of $1x10^6$ cells/mL. Mating crosses were performed in 2mL total volumes under the following induction conditions: 5% (V/V) *E. pacifica* conditioned media (EPCM), 5% (V/V) *Vibrio fischeri* conditioned media (VFCM), 0.5nM VF_rGAG lyase, 5 "units" Chondroitinase ABC (Sigma C3667), and 5 "units" Chondroitinase AC (Sigma C2780). Cells were allowed to mate for 16 hours, after which the induced culture was pelleted and washed twice in 25% CG media to prevent further mating prior to limiting dilution.

Isolating diploids by limiting dilution: Mated cells were clonally isolated by limiting dilution into 96-well plates containing 25% CG media. For all crosses performed, the probability of clonal isolation at this step was between 0.85 and 0.92. A Although we cannot directly measure the ploidy of live *S. rosetta* cells, the differentiation of planktonic motile cells into substrate-attached "thecate" cells correlates with the transition to diploidy[30]. After five days of growth, isolates were phenotyped and then divided into two populations. For each isolate, one population was rapidly passaged to induce meiosis (see below), and the other population was used for DNA extraction. For DNA extraction, isolates were expanded into 1mL of 5% CG media to prevent meiosis, and grown for three days in 24-well plates. DNA was extracted from each isolate using the following method: 500 L of cells were pelleted and resuspened in 20 L base solution (25mM NaOH, 2mM EDTA). Base solutions from the isolates were transferred to a PCR plate, boiled at 100°C for 20 min, and cooled at 4°C for 5 min. 20 L Tris solution (40mM Tris-HCl, pH 7.5) was then added to each sample. 1 L of this sample was used as the DNA template for genotyping reactions. To identify which isolates were the result of outcrossed mating, isolates were genotyped at two unlinked microsatellite markers that are polymorphic between the R+ and R-parental strain[30]. All outcrossed diploids isolated were phenotypically thecate as opposed to motile planktonic. No thecate isolates were observed in control EPCM treated cultures.

Isolation of haploid meiotic progeny: Immediately after phenotyping, clones isolated by limiting dilution were passaged 1:10 into 1mL fresh 25% CG media to induce meiosis. Thecate clones that were outcrossed diploids typically gave rise to a clear mixture of haploid chains and rosettes after two days. Haploids were clonally isolated by limiting dilution into 96-well plates containing 25% CG media, and phenotyped after five days. Meiosis was confirmed either by 1) genotyping at two unlinked microsatellite markers, or 2) by genotyping at 38 markers using KASP technology (LGC Genomics, Beverly, MA)[30].

Genotyping meiotic progeny: To confirm genome-wide recombination, haploid progeny isolated from the 5% VFCM-induced cross were genotyped at 38 markers[30]. Briefly, three outcrossed diploids (named A2, A3, and H2) were rapidly passaged to induce meiosis, and clones from each outcrossed diploid were isolated by limiting dilution. The probability of clonal isolation at this step was .94 for A2, .93 for A3, and .91 for H2. A total of 147 haploid isolates from the three outcrossed diploids were phenotyped and expanded for subsequent DNA extraction and genotyping.

Quantifying mating swarms

Inductions were set up in 100 L volumes in 96-well glass bottom plates (Ibidi 89626). Assays were imaged at 10X magnification using transmitted light (bright field) on the Zeiss

Observer Z.1 platform using a Hamamatsu C11440 camera. An automated sequence was set up such that each sample was imaged at 4 distinct locations throughout the well.

Images were batch processed in ImageJ to ensure consistency. After applying the 'Smooth' command to reduce background bacterial signal, the 'Find Edges' command was applied to further highlight the phase-bright choanoflagellate cells. Images were then converted to black and white using the 'Make Binary' command, followed by the "Close" command to fill in small holes.

Finally, images were analyzed using the 'Analyze Particles' command to calculate the area of each cell or swarm (the white space in Figure 4.2A',B') within an image.

Isolating the *Vibrio fischeri* mating induction factor

Preparation of >30kD-enriched VCM: Eight 1L cultures of *V. fischeri* ES114 were grown shaking for in 100% SWC for 24 h at 25ºC. Cultures were pelleted at 16,000 × g, and the supernatants were concentrated to 120 mL using a tangential flow filtration device with a 30 kDa centramate filter (Pall #OS030T12). The supernatant was then further clarified by pelleting 39,000 × g.

Ammonium sulfate precipitation: >30kD-enriched VCM was treated with 1 M Tris-HCl (pH 7.6) and fractionally precipitated with increasing (40%-75%) concentrations of ammonium sulfate. Precipitates were resuspended in water, and tested in the swarming bioassay.

Size exclusion chromatography: Active ammonium sulfate precipitation fractions were combined and concentrated to 1 mL. 0.85 mL was injected on a HiPrep^{TM} 16/60 Sephacryl^{TM} S-200 High Resolution column (GE Healthcare Life Sciences #17-1166-01) using an AKTA Explorer FPLC instrument. Proteins were eluted with 30 mM Tris-HCl (pH 7.7, 4 ºC) at 0.5 mL/min for 120 mL, and 2 mL fractions were collected. Adjacent fractions were paired and tested in the swarming bioassay, as well as analyzed by PAGE.

Anion exchange chromatography:

Active SEC fractions were combined and concentrated to 1 mL in Solvent A (20 mM L-histidine, pH 6.0) and injected at 2.5 mL/min into a HiPrep^{TM} 16/10 Q XL column (GE Healthcare Life Sciences #17-5092-01). Proteins were eluted in 2 mL fractions over a 300 mL linear gradient (0-100%) of Solvent A to Solvent B (1 M NaCl in 20 mM L-histidine, pH 6.0). Fractions were tested in the swarming bioassay and analyzed by PAGE.

Renaturing proteins after PAGE: Proteins from highly active AEX fractions were concentrated and mixed with 15 µg/mL β-lactoglobulin carrier protein, and run in adjacent lanes through a NuPAGE^{TM} 4-12% Bis-Tris polyacrylamide gel. Evenly spaced bands were excised from one lane, and the remaining gel was stained with Coomassie blue R-250 and retained for mass spectrometry. The excised slices were crushed, and then extracted for 6 hours with 300 µL of elution buffer (50 mM Tris-HCl pH 7.7, 100 µM EDTA, 1 mM DDT, 150 mM NaCl, 0.1% sodium dodecyl sulfate [SDS], 100 µg/mL bovine serum albumin [BSA], pH 7.7). Proteins were precipitated with 1200 µL cold acetone, incubated on dry ice for 30 minutes, and pelleted at 16,000 × g for 10 minutes at 4 ºC. Air-dried pellets were dissolved in 10 µL of solubilization buffer (50 mM Tris-HCl pH 7.7, 100 µM EDTA, 1 mM DDT, 150 mM NaCl, 20% glycerol, 6 M guanidine hydrochloride) for 20 minutes, and then diluted with 500 µL of dilution buffer (50 mM Tris-HCl pH 7.7, 100 µM EDTA, 1 mM DDT, 150 mM NaCl, 20% glycerol). Proteins further renatured for 3.5 H at room temperature, and were concentrated to 20 µL. Proteins excised from bands were then tested in the swarming bioassay.

Mass Spectrometry: Because only one excised protein band displayed bioactivity, the corresponding slice retained for mass spectrometry was subjected to trypsin digestion and LC-MS/MS at the Proteomics/Mass Spectrometry Laboratory at UC Berkeley.

Heterologous expression and purification of EroS

eroS was amplified by PCR (Phusion® DNA polymerase) from *V. fischeri* ES114 genomic DNA (Forward primer: 5'-GCCTCTGTCGACGCAAAAAATACCCAAACACCAC; Reverse primer: 5'-AATTAAGCGGCCGCCGTCTTGAATTGTTACTTGGAAAGAATAAG). After digestion with *SalI*-HF and *NotI*-HF (New England Biolabs), the *eroS* gene was ligated into a pET6xHN-N vector (Clontech) and transformed into OneShot BL21(DE3) cells (Invitrogen) for expression. *E. coli* were grown at 37 ºC, 200 rpm shaking in LB media supplemented with 100 µg/mL ampicillin. After growth to OD 1.0, the temperature was decreased to 16 ºC, and protein expression was induced by addition of 1 mM IPTG. After 24 hours, cells were pelleted and lysed with xTractor™ buffer (Clontech #635651), purified with HisTALON™ gravity columns (Clontech #635655), and further purified by enterokinase cleavage (Millipore enterokinase cleave capture kit #69067).

Amino acid sequence alignments

Amino acid sequences from *V. fischeri* (VF_A0994, GenBank: AAW88064.1) and characterized bacterial GAG lyases [*A. aurescens* (AC lyase, PDB: 1RWG_A), *S. coelicolor* (AC lyase, PDB: 2WDA_A), *S. agalactiae* (hyaluronate lyase, PDB: 1LXM_A), *F. heparinum* (AC lyase, PDB: 1CB8_A), and *P. vulgaris* (ABC chondroitinase, GenBank: ALL74069.1)] were aligned using Clustal Omega multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalo/). Conserved amino acid residues were highlighted using BoxShade (http://www.ch.embnet.org/software/BOX_form.html).

EroS GAG lyase activity *in vitro*

The glycosaminoglycan cleavage activity of purified Eros was determined *in vitro* as previously described [130]. Briefly, GAG standards [hyaluronic acid (Sigma #H5388), chondroitin sulfate (Sigma #C4384), dermatan sulfate (Sigma #C3788), and heparin (Sigma #H3393)] were dissolved to a concentration of 1 mg/mL in buffer solution (50 mM NaH_2PO_4/Na_2HPO_4, 0.5 M NaCl, pH 8.0). 1 mL of GAG standard was added to 5 µL of enzyme solution. The assays were performed in quartz cuvettes (1 cm pathlength) at 23 ºC, and UV absorbance measurements (232 nm) were taken directly on the reacting mixture.

AQUA quantification of EroS

Absolute Quantification (AQUA) peptide[131] was used to accurately quantify the concentration of purified EroS and EroS in *Vibrio fischeri* conditioned media. Briefly, *V. fischeri* was grown for 8 H in 100% SWC media and 5% SWC media. Purified EroS and concentrated *Vibrio fischeri* culture supernatants were loaded onto a PAGE gel. The gel was stained with Coomassie blue R-250, and bands containing VF_A0994 were excised and sent to the Taplin Mass Spectrometry facility (Harvard Medical School) for further analysis. Synthetic AQUA peptide (TQITDDTYQNFFD[KC13N15], Sigma-Aldrich) and trypsin were added to each excised band, and LC-MS/MS was performed on the digested peptides using a Thermo Scientific Orbitrap. The amount of EroS present in each gel slice was calculated by comparing MS2 peak intensities of the native peptide with the internal AQUA synthetic peptide standard.

***S. rosetta* polysaccharide isolation and GAG disaccharide analysis**

6 x 500 mL cultures of SrEpac were grown in 5% SWC until mid-stationary phase, and washed 3x to reduce bacterial load before being pelleted, flash frozen, and lyophilized. 125 mg

of lyophilized *S. rosetta* sample was sent to the Complex Carbohydrate Research Center for GAG isolation, digestion, and SAX-HPLC.

Polysaccharides were isolated from the *S. rosetta* sample and digested with either EroS or chondroitinase ABC (Sigma C3667) for chondroitin disaccharide analysis, and heparinases I, II, and III (Dextra Laboratories) for heparan disaccharide analysis. Briefly, a ratio of 10 L *S. rosetta* polysaccahrides to 1uL of enzyme was incubated for 24 hours. Samples were heated to 100°C for 5 minutes to inactivate the enzyme, and centrifuged at 14,000 rpm for 30 minutes prior to SAX-HPLC.

SAX-HPLC was carried out on an Agilent system using a 4.6x250 mm Waters Spherisorb analytical column with 5μm particle size at 25°C. Detection was performed by post-column derivatization. Briefly, the eluent from the column was combined with a 1:1 mixture of 0.25 M NaOH and 1% 2-cyanoacetamide pumped at a flow rate of 0.5 mL/min from a post-column reactor. The eluent was heated to 130°C in a 10m reaction coil, then cooled in a 50-cm cooling coil and directed into a Shimadzu fluorescence detector (λex = 346 nm, λem = 410). Commercial standard disaccharides (Dextra Laboratories) were used for identification of each disaccharide based on elution time, as well as calibration.

Testing bioactivity of chondroitin disaccharides

Chondroitin disaccharides and chondroitin sulfate were tested for bioactivity by treating *S. rosetta* with unsulfated chondroitin disaccharides, chondroitin-6-sulfate disaccharides, unsulfated chondroitin + chondroitin-6-sulfate disaccharides, and chondroitin sulfate (from shark cartilage) at concentrations ranging from 0.0001M-0.1M. Cells were imaged and quantified after 30 minutes, 1 hour, and 3 hours.

Degradation products of chondroitin sulfate were generated by incubating 100 g of chondroitin sulfate with 1 L of either EroS or ABC chondroitinase (*P. vulgaris*) overnight. Enzymatic activity was killed by incubating samples at 80°C for 5 minutes. The resulting degradation products were tested for bioactivity at concentrations ranging from 0.0001M-0.1M. Cells were imaged and quantified after 30 minutes, 1 hour, and 3 hours.

FIGURES

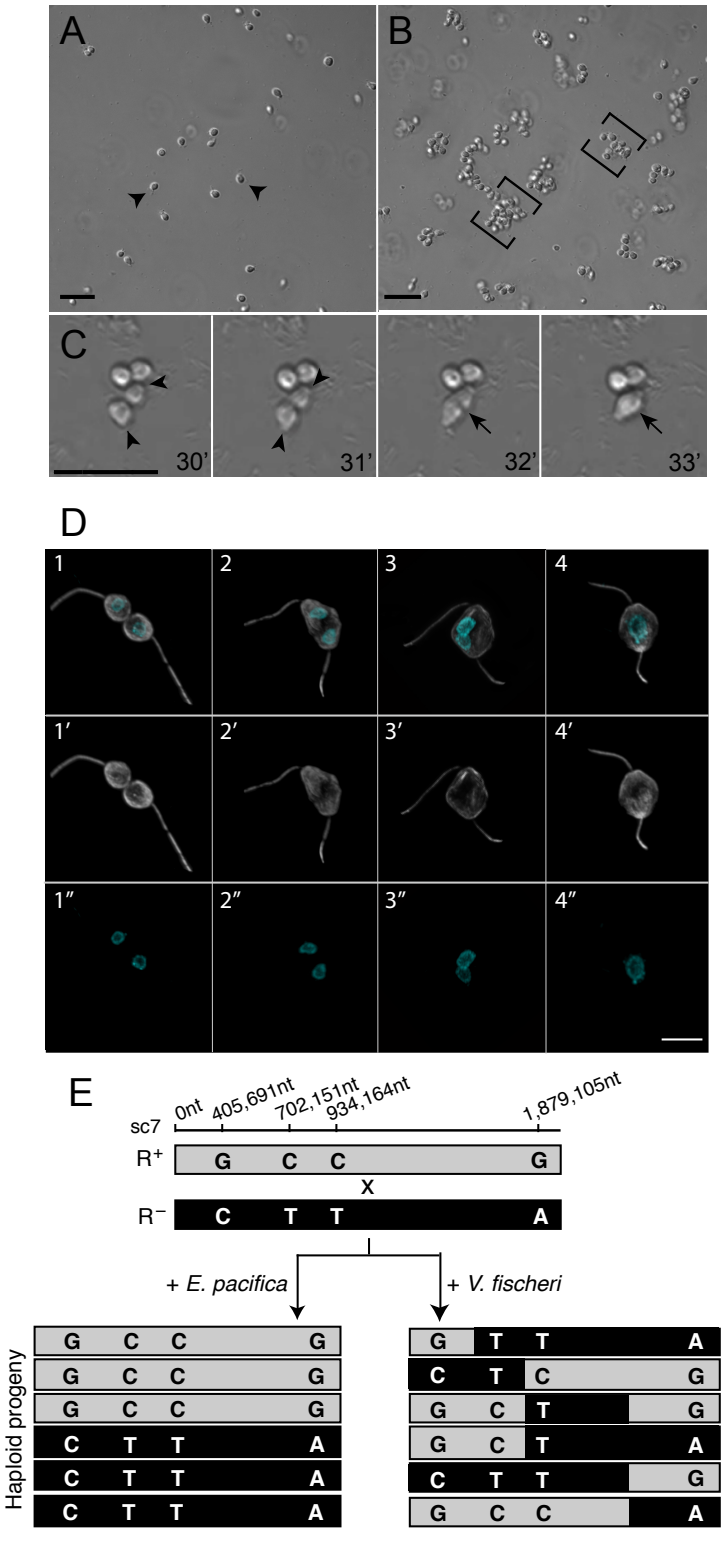

A
B
C
30'
31'
32'
33'
D
1
2
3
4
1′
2′
3′
4′
1″
2″
3″
4″
E
sc7
0nt
405,691nt
702,151nt
934,164nt
1,879,105nt
R+
G C C G
X
R−
C T T A
+ E. pacifica
+ V. fischeri
Haploid progeny
G C C G
G C C G
G C C G
C T T A
C T T A
C T T A
G T T A
C T C G
G C T G
G C T A
C T T G
G C C A

Figure 4.1. *V. fischeri* bacteria induce swarming and mating in the choanoflagellate, *S. rosetta*. (A) In the absence of *V. fischeri,* motile *S. rosetta* cells (arrowheads) are evenly dispersed. **(B)** Within 30 minutes of exposure to *V. fischeri, S. rosetta* motile cells aggregate into large swarms (brackets). Scale bar = 20 m. **(C)** *S. rosetta* cells within a swarm pair and fuse. Prior to fusion, cells reposition themselves such that their basal membranes are adjacent and their apical flagella point away (31'; arrowheads mark apical pole of unfused cells). Cell fusion takes only minutes, and occurs along the basal membrane (32'; indicated by arrow), resulting in a single, elongated cell (33'; indicated by arrow). Scale bar = 20 m. **(D)** Stages of cell and nuclear fusion in *S. rosetta* mating pairs. Haploid mating pairs are oriented with their basal poles (opposite the flagellum) touching (D1), and cell fusion proceeds along the basal membrane, resulting in a binucleated cell with two flagella (D2). Nuclei then congress towards the midline (D3), where the nuclei undergo nuclear fusion, resulting in a diploid cell (D4). Anti-tubulin antibody (D1'-4'; white) highlights the cell body and flagellum, and Hoechst (D1''-4''; cyan) highlights the nucleus. Scale bar = 5 m **(E)** Evidence for meiotic recombination in *S. rosetta* following exposure to *V. fischeri*. Two haploid, genotypically distinct *S. rosetta* strains [R+(grey shading) and R- (black shading)] were mixed in the presence of either *E. pacifica* conditioned media or *V. fischeri* conditioned media for 16 hours. Haploid progeny were clonally isolated and genotyped at polymorphic markers across the genome, (Supplemental Data). We show here genotyping results for four representative loci along supercontig 7 (sc7). All clones isolated from *E. pacifica*-treated cultures contained unrecombined parental genotypes, while haploid clones isolated from *V. fischeri*-treated cultures showed clear evidence of recombination. Top numbers show marker genomic positions along sc7.

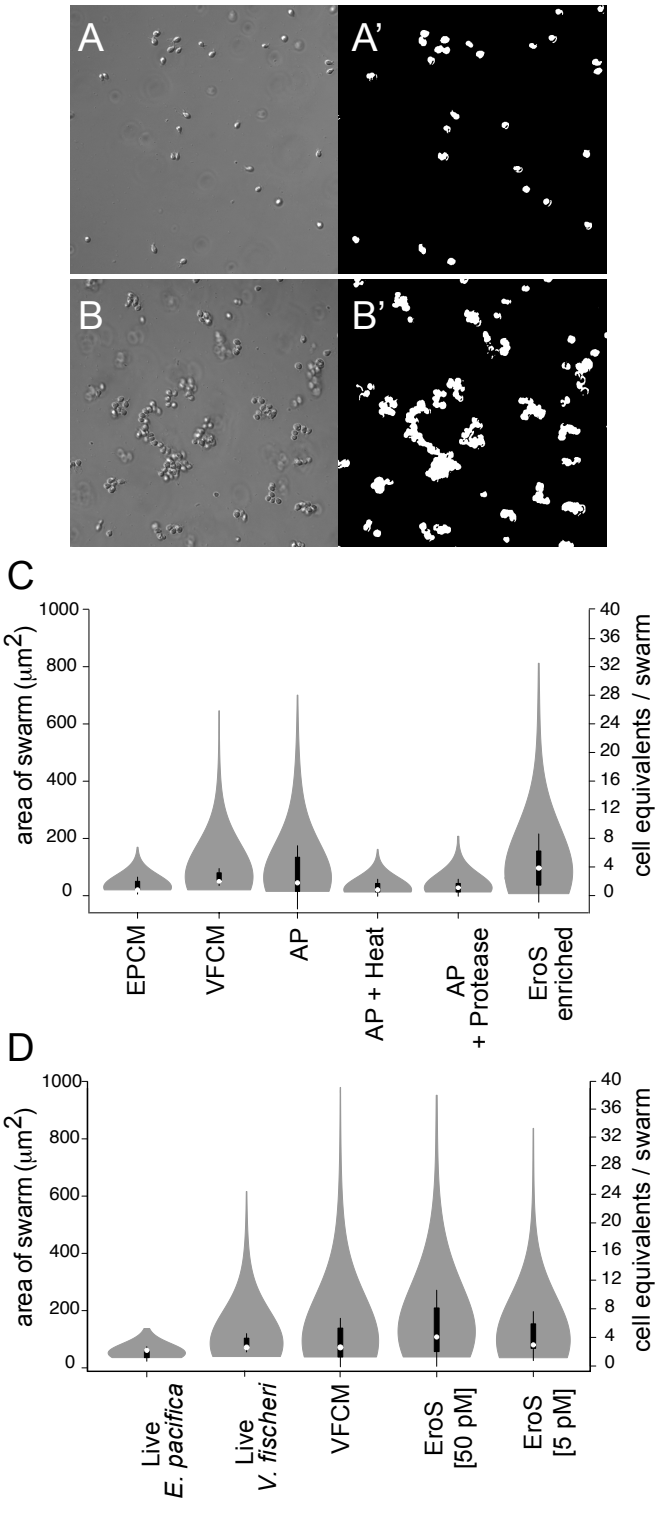

A
A'
B
B'
C
area of swarm (μm2)
cell equivalents / swarm
1000
800
600
400
200
0
40
36
32
28
24
20
16
12
8
4
0
EPCM
VFCM
AP
AP + Heat
AP + Protease
EroS enriched
D
area of swarm (μm2)
cell equivalents / swarm
Live E. pacifica
Live V. fischeri
VFCM
EroS [50 pM]
EroS [5 pM]

Figure 4.2. Bioactivity-guided isolation of the *V. fischeri* aphrodisiac. (A, B) Automated image analysis allowed quantification of *S. rosetta* swarming in response to *V. fischeri*-derived activity. Pictured are *S. rosetta* cells 30 minutes after treatment with *E. pacifica* conditioned media (A) or *V. fischeri* conditioned media (B). By generating a binary mask (A', B') we could measure the area of the swarm, and estimate the number of cells ("cell equivilants") per swarm. **(C)** Swarming in *S. rosetta* is induced by compounds in the ammonium sulfate precipitation of *V. fischeri* culture supernatant (AP), but not by AP exposed to heat (80°C for 10 minutes; AP + Heat) or proteases (AP + Protease). The aphrodisiac activity tracked with a ~90kD protein band that was revealed by mass spectrometry to be the *V. fischeri* EroS protein (VF_A0994). **(D)** EroS triggers mating at plausible environmental concentrations. Purified EroS induces swarming in *S. rosetta* at concentrations as low as 5 pM, and is sufficient to fully recapitulate the aphrodisiac activity of live *V. fischeri* bacteria and VFCM.

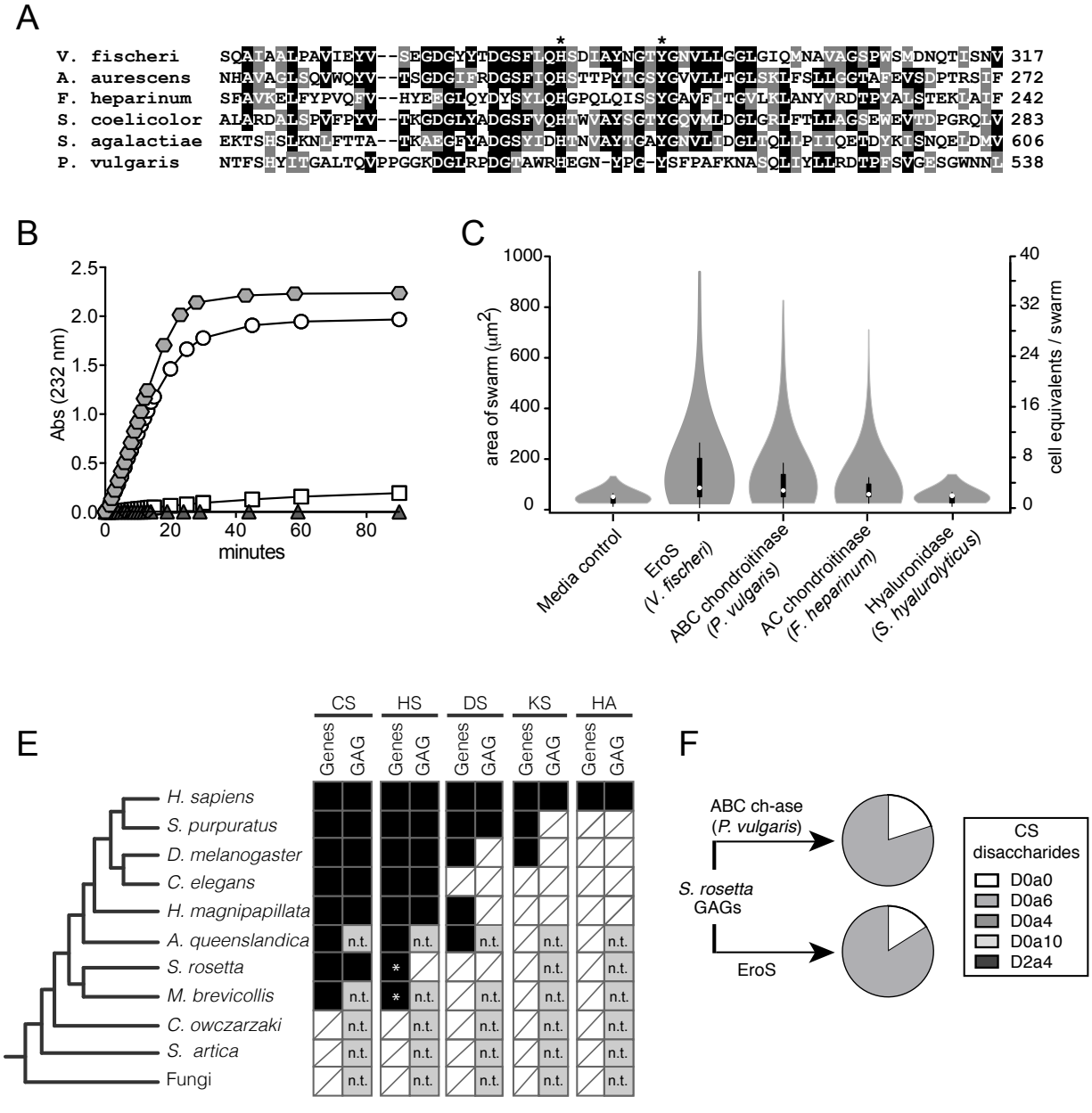

A
V. fischeri
A. aurescens
F. heparinum
S. coelicolor
S. agalactiae
P. vulgaris
B
Abs (232 nm)
minutes
C
area of swarm (μm2)
cell equivalents / swarm
Media control
EroS (V. fischeri)
ABC chondroitinase (P. vulgaris)
AC chondroitinase (F. heparinum)
Hyaluronidase (S. hyalurolyticus)
E
CS
HS
DS
KS
HA
Genes
GAG
H. sapiens
S. purpuratus
D. melanogaster
C. elegans
H. magnipapillata
A. queenslandica
S. rosetta
M. brevicollis
C. owczarzaki
S. artica
Fungi
n.t.
F
ABC ch-ase (P. vulgaris)
S. rosetta GAGs
EroS
CS disaccharides
D0a0
D0a6
D0a4
D0a10
D2a4

Figure 4.3. The *V. fischeri* aphrodisiac is a GAG lyase that degrades *S. rosetta* chondroitin. **(A)** Alignment of the *V. fischeri* EroS amino acid sequence to diverse bacterial GAG lyases reveals that *V. fischeri* harbors conserved His and Tyr residues (indicated by *) at sites required for catalytic activity in characterized GAG lyases. Amino acids with >50% conservation between sequences are shaded (black shading for identical amino acids and grey shading for similar amino acids. **(B)** Purified EroS degrades chondroitin sulfate and hyaluronan. EroS was incubated with purified chondroitin sulfate (open circle), hyaluronan (grey hexagon), dermatan sulfate (open square), and heparan sulfate (grey triangle), and lyase activity of EroS was measured by monitoring the abundance of unsaturated oligosaccharide products with an absorbance at 232nm. Chondroitin sulfate and hyaluronan oligosaccharides accumulated rapidly in the presence of EroS, indicating depolymerization, whereas heparan sulfate and dermatan sulfate were not depolymerized by EroS. **(C)** The chondroitinase activity of EroS is necessary and sufficient for its function as an aphrodisiac. EroS protein with mutations in predicted catalytic resides fail to induce swarming in *S. rosetta. P. vulgaris* ABC chondroitinase and *F. heparinum* AC chondroitinase are sufficient to induce swarming at levels similar to EroS, whereas *S. hyalurolyticus* hyaluronidase fails to induce swarming, indicating that chondroitinase activity is necessary and sufficient for aphrodisiac activity. **(D)** Phylogenetic distribution of diverse GAGs [CS= chondroitin sulfate; HS= heparan sulfate; DS= dermatan sulfate; KS= keratan sulfate; HA= hyaluronan], and their biosynthetic genes. The presence (black box) and absence (white box with slash) of genes required for the biosynthesis of GAGs (Gene) and biochemical evidence for GAGs (Expt) in select opisthokonts. *; Ori et al. (2011) identified putative HS biosynthetic orthologs in the *M. brevicollis* genome, and similar levels of conservation are observed in *S. rosetta.* n.t.; not tested (experiments have not been performed to biochemically profile GAGs). **(E)** *S. rosetta* produces chondroitin that can be degraded by EroS. Polysaccharides isolated from *S. rosetta* were treated with either *P. vulgaris* ABC chondroitinase, an enzyme that can degrade many modifications of chondroitin disaccharide (CS disaccharides), or EroS. Both ABC chondroitinase and EroS yielded similar degradation products of unsulfated chondroitin (D0a0) and chondroitin-6-sulfate (D0a6). This indicates that unsulfated chondroitin and chondroitin-6-sulfate, but not chondroitin-4-sulfate (D0a4), chondroitin-4,6-sulfate (D0a10), or chondroitin-2,4-sulfate (D2a4), are present in *S. rosetta*.

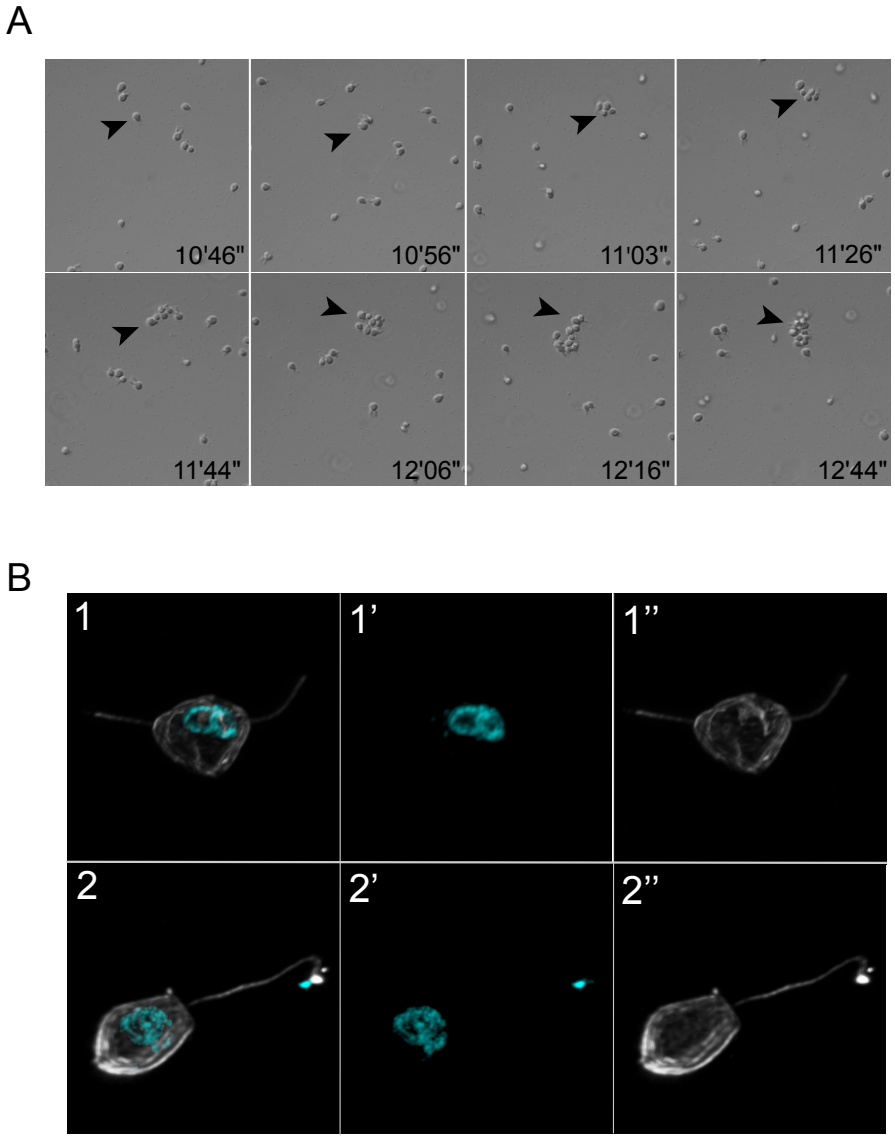


Figure 4.4. *V. fischeri* induces swarming and mating in *S. rosetta*. (A) Stills of swarm formation after induction with *V. fischeri* bacteria. Arrowhead tracks the formation and movement of a single swarm over time. **(B)** Nuclear fusion in mating pairs of *S. rosetta* following treatment with *V. fischeri*. Pictured are late stages of nuclear congression and fusion. Following cell fusion, the nuclei congress towards the center of the bi-flagellated cell (B1-1''), and fuse (B2-2''). The final result of nuclear fusion is a diploid cell, harboring a single flagellum (B2-2''). Hoechst (B1',2'; cyan) highlights the nucleus, and anti-tubulin antibody (B1'',2''; white) highlights the cell body and flagellum.

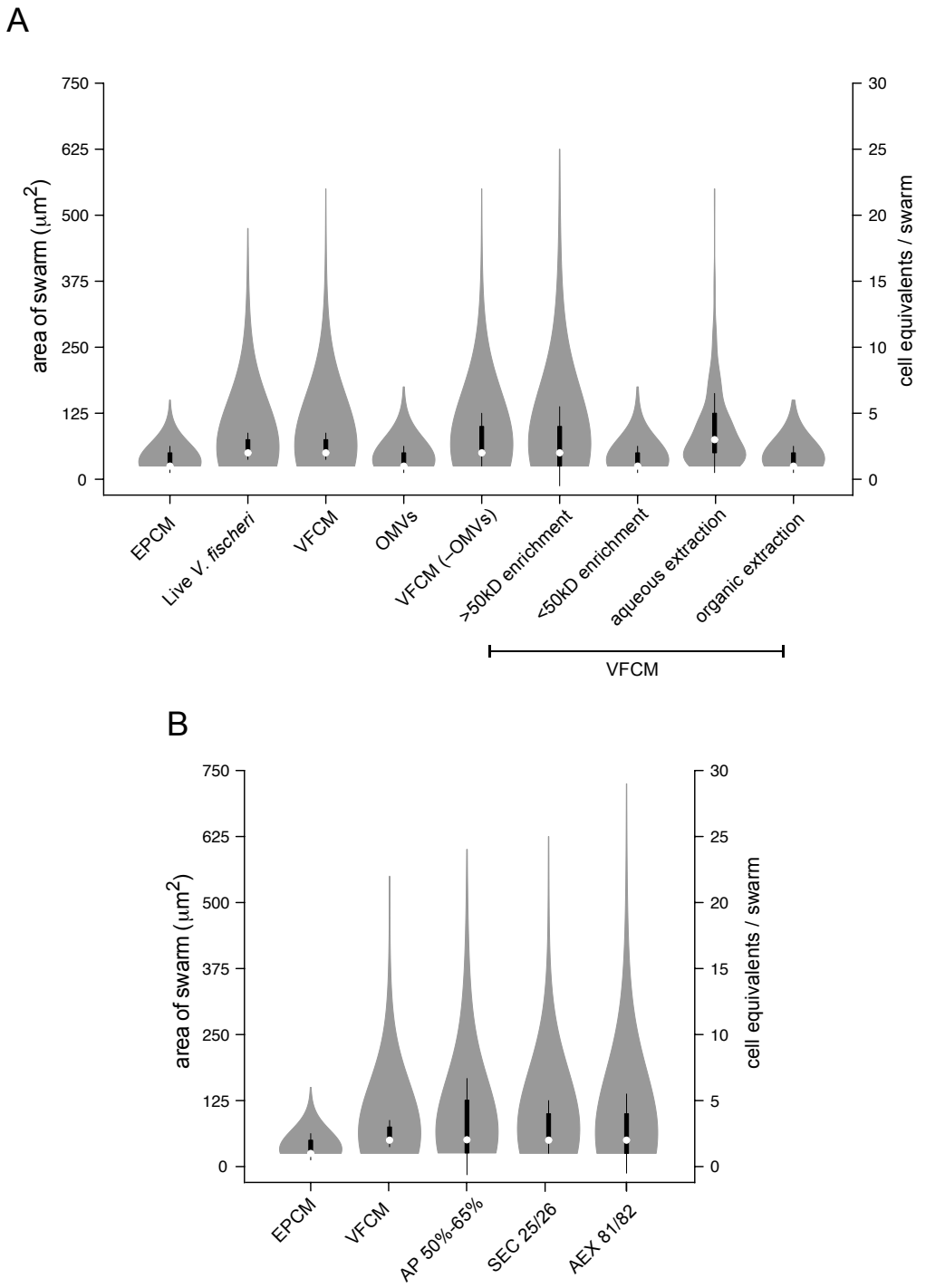


Figure 4.5 Bioactivity-guided isolation of EroS. (A) Swarming in *S. rosetta* is induced by large (>50kD), water-soluble factors present in *V. fischeri* conditioned media (VFCM). **(B)** Isolation of EroS from VFCM. To identify the source of the aphrodisiac activity, proteins were precipitated from VFCM (AP 50%-65%) and separated by size exclusion (SEC) and anion exchange (AEX) chromatography. A protein band of ~90kD, later determined to be EroS (VF_A0994), was abundant in the bioactive SEC (SEC 25/26) and AEX (AEX 81/82) fractions.

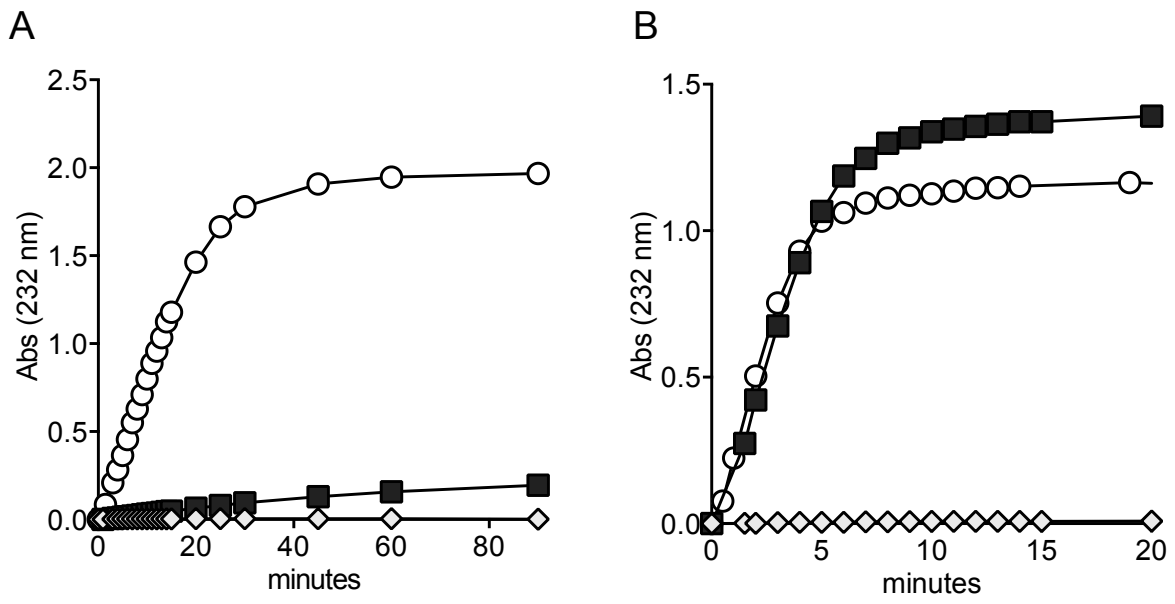
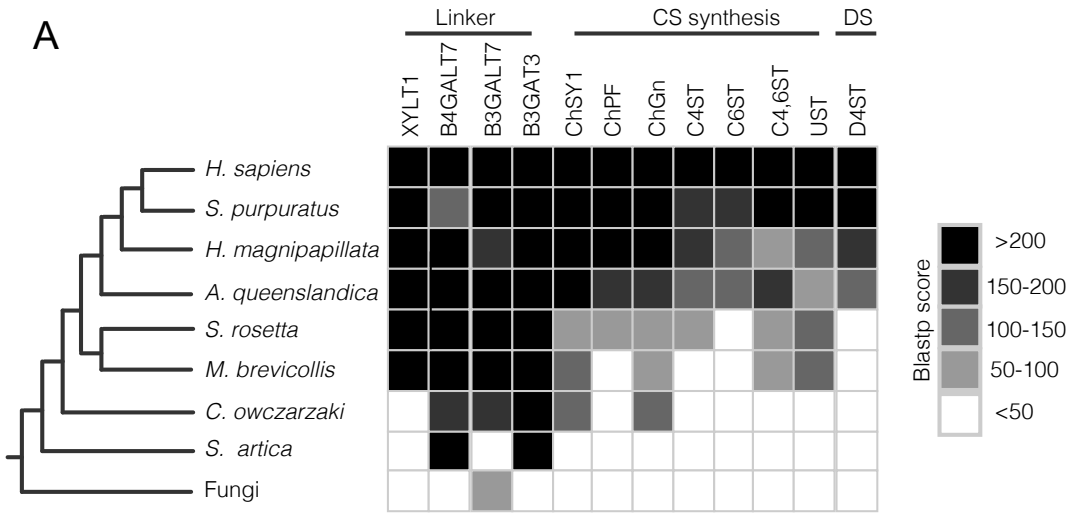
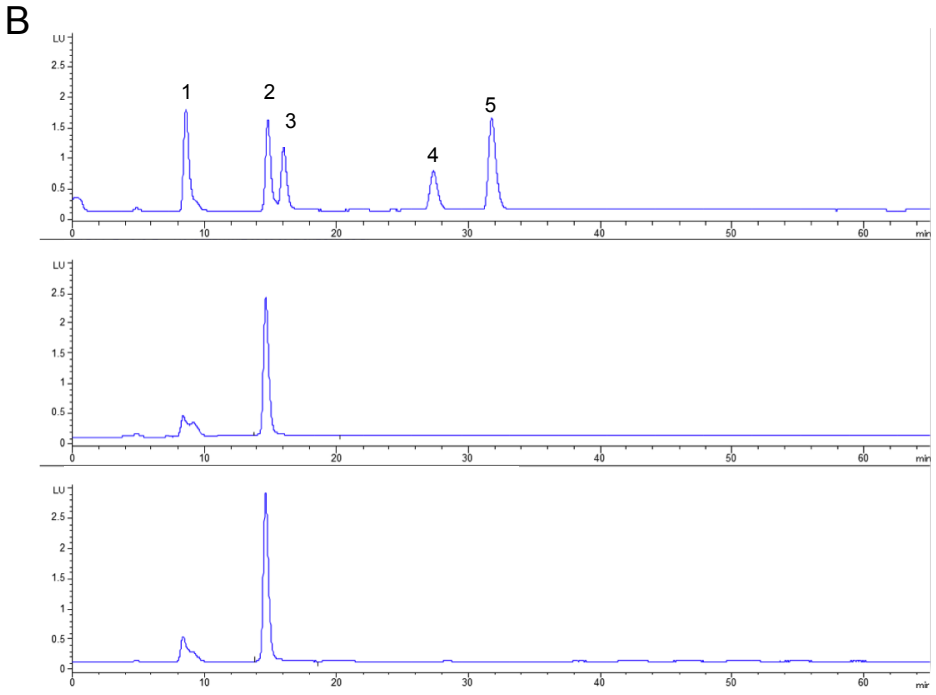


Figure 4.6. EroS is a chondroitin AC lyase. (A) EroS degrades chondroitin sulfate AC, but not chondroitin sulfate B *in vitro*. EroS was incubated with purified chondroitin sulfate AC (open circle) and chondroitin sulfate B (grey square), and lyase activity of EroS was measured by monitoring the abundance of unsaturated oligosaccharide products with an absorbance at 232nm. Chondroitin sulfate AC, but not chondroitin sulfate B, accumulated rapidly in the presence of EroS. White diamond represents a no enzyme control. **(B)** Chondroitinase ABC (*P. vulgaris*), a positive control for *in vitro* chondroitin degradation assays, rapidly depolymerizes both chondroitin sulfate AC (open circle) as well as chondroitin sulfate B (grey square). White diamond represents a no enzyme control.





C

	Chondroitinase ABC		EroS	
CS	µg	*% total*	µg	*% total*
D0a0	12.25	*20*	11.57	*16*
D0a6	50.45	*80*	60.96	*84*
D0a4	n.d.		n.d.	
D0a10	n.d.		n.d.	
D2a4	n.d.		n.d.	
Total CS	62.69	*100*	72.50	*100*

Figure 4.7. Chondroitin sulfate produced by *S. rosetta* can be degraded by EroS. (A) Orthologs of chondroitin sulfate (CS) synthesis, but not dermatan sulfate (DS) synthesis, are

present in the genomes of choanoflagellates *S. rosetta* and *M. brevicollis*. Genes identified as "linker" synthesize the proteoglycan linker tetrasaccharide and are important for the biosynthesis of multiple types of GAG, whereas the genes identified as "CS synthesis" are specific to CS biosynthesis. The gene identified as "DS" is required for the formation of dermatan sulfate. All query sequences used were human orthologs. If multiple subject sequences were hits for a single query sequence, the ortholog with the highest Blastp score was chosen. **(B)** *S. rosetta* produces chondroitin that can be degraded by ABC chondroitinase and EroS. Polysaccharides isolated from *S. rosetta* were treated with either ABC chondroitinase from *P. vulgaris* (center plot) or EroS (bottom plot). Degradation products from samples treated with ABC chondroitinase and EroS were separated by SAX-HPLC (X-axis indicates time, Y-axis indicates abundance) and compared to the following chondroitin disaccharide standards (top plot): (1) D0a0, unsulfated chondroitin; (2) D0a6, chondroitin-6-sulfate; (3) D0a4, chondroitin-4-sulfate; (4) D0a10, chondroitin-4,6-sulfate; (5) D2a4, chondroitin-2,4-sulfate. Unsulfated and 6-sulfated chondroitin disaccharides were present at similar abundance in both the ABC chondroitinase and EroS – treated samples, whereas all other chondroitin disaccharides were below the limit of detection. **(C)** Quantification of chondroitin disaccharide products produced by *ABC* chondroitinase (*P. vulgaris*) and EroS treatment of *S. rosetta* polysaccharides. Disaccharide abbreviations: D0a0=unsulfated chondroitin; D0a6=chondroitin-6-sulfate; D0a4= chondroitin-4-sulfate; D0a10=chondroitin-4,6-sulfate; D2a4=chondroitin-2,4-sulfate.

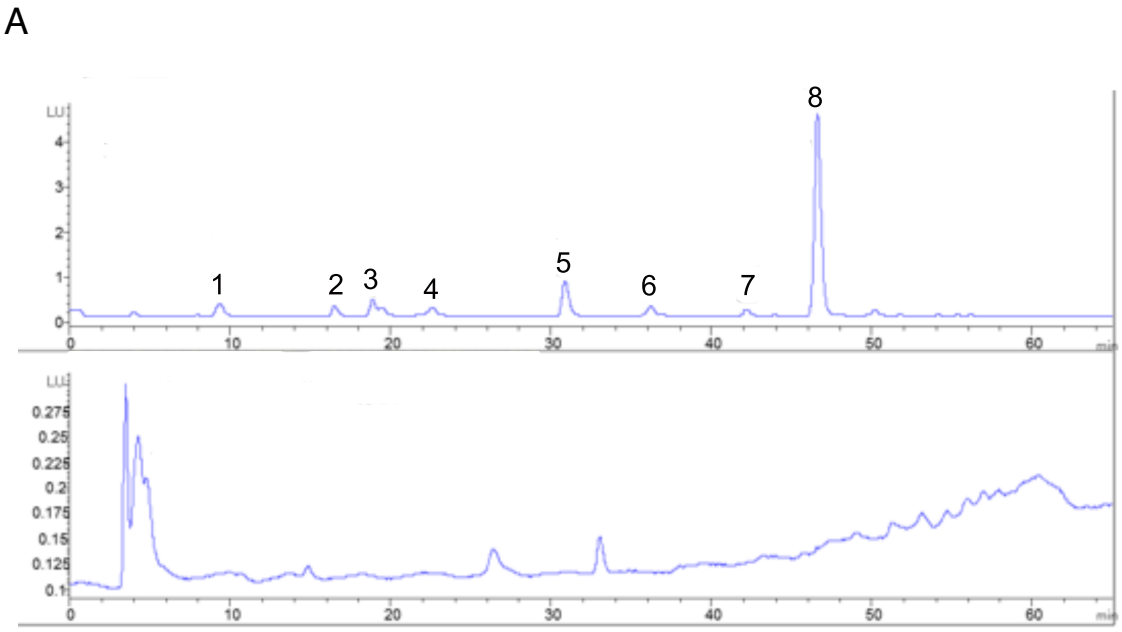


Figure 4.8. *S. rosetta* does not produce heparan sulfate. Polysaccharides isolated from *S.*

rosetta (bottom plot) were treated with Heparinase I, Heparinse II, and Heparinase III (Dextra Laboratories) separated by SAX-HPLC (X-axis indicates time, Y-axis indicates abundance) and compared to the following heparan sulfate disaccharide standards (top plot): (1) D0A0; (2) D0S0; (3) D0A6; (4) D2A0; (5) D0S6; (6) D2S0; (7) D2A6; (8) D2S6. No heparan disaccharides were present above the limit of detection in the *S. rosetta* polysaccharide sample.

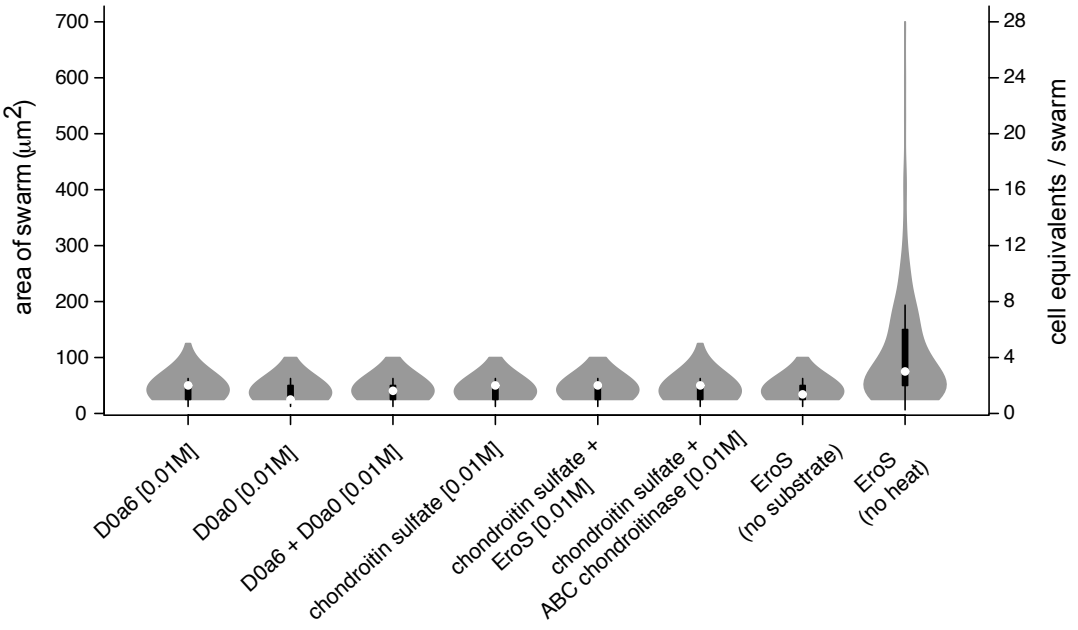


Figure 4.9. Swarming in *S. rosetta* is not induced by chondroitin sulfate or chondroitin disaccharides. Neither commercial chondroitin disaccharides (6SCS and 0SCS), or chondroitin disaccharides generated via the depolymerization of chondroitin sulfate by EroS or ABC chondroitinase (*P. vulgaris*), are sufficient to induce swarming in *S. rosetta*.

A

V. fischeri density (cells/mL)	*S. rosetta* density (cells/mL)	# *Vibrio* : # *S. rosetta*	Time to swarm
$2.0x10^3$	$2.0x10^6$	1:1000	30 minutes
$4.0x10^2$	$2.0x10^5$	1:500	30 minutes
$2.0x10^2$	$2.0x10^5$	1:1000	60 minutes
$4.0x10^2$	$2.0x10^4$	1:50	30 minutes
$1.0x10^2$	$2.0x10^4$	1:200	90 minutes

B

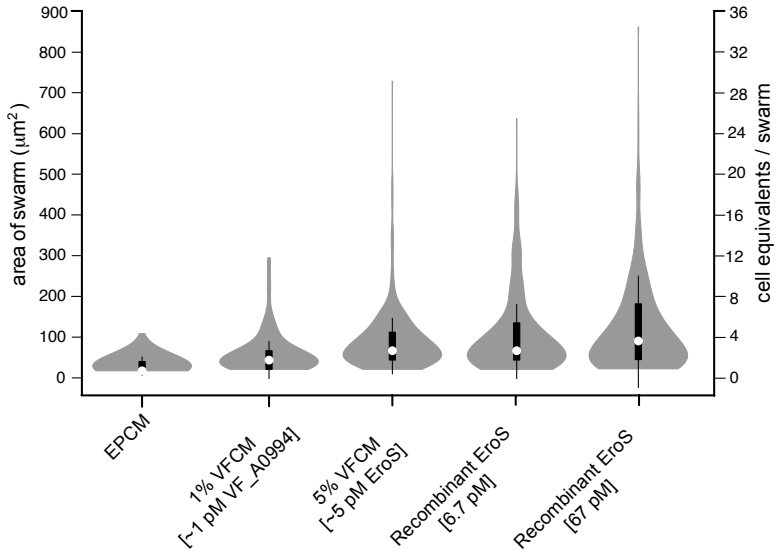


Figure 4.10. *V. fischeri* induces mating in *S. rosetta* under plausible environmental conditions. (A) *S. rosetta* swarms in response to low numbers of *V. fischeri* bacteria in a cell density-dependent manner. *S. rosetta* at high cell densities ($2.0x10^6$ cells/mL) swarms in response to as few as one *V. fischeri* cell per 1000 *S. rosetta* cells within 30 minutes of exposure, whereas swarming in *S. rosetta* at lower cell densities ($2.0x10^5$ cells/mL) within a similar time frame requires at least one *V. fischeri* cell per 500 *S. rosetta* cells. **(B)** Picomolar concentrations of secreted (5% VFCM) and purified EroS are sufficient to induce swarming in *S. rosetta*.

Table 4.1. Bacteria tested in swarming bioassay

Species	Strain	Genotype	Live bacteria	Conditioned media	Concentrated conditioned media	Accession information
Vibrio fischeri	ES114	WT	+	+	+	ATCC 700601
		ΔLuxO	+	+	+	Lupp et al. 2003
		ΔAinS	+	+	+	Lupp et al. 2003
		ΔLuxI	+	+	+	Lupp et al. 2003
		ΔAinSΔLuxS	+	+	+	Lupp and Ruby 2004
		ΔLuxR	+	+	+	Lupp and Ruby 2004
		ΔSypC	+	+	+	Shibata et al. 2012
		ΔSypH	+	+	+	Shibata et al. 2012
		ΔSypI	+	+	+	Shibata et al. 2012
		ΔSypK	+	+	+	Shibata et al. 2012
		ΔSypM	+	+	+	Shibata et al. 2012
		ΔSypO	+	+	+	Shibata et al. 2012
		ΔSypQ	+	+	+	Shibata et al. 2012
Vibrio fischeri	MJ11		+	+	+	BAA-1741
Vibrio tubiashii			+	+	+	ATCC 19105
Vibrio orientalis			+	+	+	ATCC 33934
Vibrio harveyi			–	–	–	ATCC 14126
Vibrio natriegens			–	–	–	ATCC 8110
Vibrio parahaemolyticus			Ψ	Ψ	Ψ	ATCC 17802
Vibrio alginolyticus			Ψ	Ψ	Ψ	PRJNA13571
Vibrio anguillarum			–	–	–	ATCC 19181
Vibrio ordalii			–	–	–	ATCC 33509
Vibrio cholera	YB1A01		n.t.	–	–	PRJNA281423
Vibrio metoecus	YB4D01		–	–	–	PRJNA281423
Vibrio mimicus			–	–	–	ATCC 33653
Echinicola pacifica			–	–	–	DSM 19836
Echinicola pacifica	From starved *S. rosetta* co-culture		n.t.	–	–	ATCC PRA-390

+: swarming observed; –: no swarming observed; Ψ: settling observed; n.t.: not tested

Table 4.2. Purified molecules tested in swarming bioassay

Molecule	Induces Swarming?	Source
3-O-C_6-(L)-HSL	–	Cayman Chemical (10011207)
C8-HSL	–	Cayman Chemical (10011199)
Cyclic di-GMP	–	Sigma SML1228
Chondroitinase ABC	+	Proteus vulgaris (Sigma C3667)
Chondroitinase AC	+	Flavobacterium heparinum (Sigma C2780)
Chondroitinase B	–	Flavobacterium heparinum (Sigma C8058)
Heparinase I and III	–	Flavobacterium heparinum (Sigma H3917)
Hyaluronate lyase	–	Staphylococcus hyalurolyticus (Sigma H1136)
Chitinase	–	Streptomyces griseus (Sigma C9830)
O-Glycosidase	–	Streptococcus pneumonia (Sigma G1163)
Lysozyme	–	Sigma L67876
Collagenase	–	Clostridium histolyticum (Sigma C0130)
Unsulfated chondroitin	–	Sigma C3920
Chondroitin-6-sulfate	–	Sigma C4170
Chondroitin sulfate (shark cartilage)	–	Sigma C4384

Table 4.3. Chondroitinase-induced mating in *S. rosetta*

Inducing factor	Clonal isolates (#)	% thecate isolates (presumable diploid)	% outcrossed diploid isolates
E. pacifica CM (5% Vol/Vol)	88	0	0
EroS (*V. fischeri,* 50pM)	53	17%	15%
Chondroitinase ABC (*P. vulgaris,* 5 units)	62	21%	17%
Chondroitinase AC (*F. heparinum,* 5 units)	52	13%	11%

Table 4.4. Quantification of purified Eros and Eros secreted by *V. fischeri*

	Purified EroS	Low nutrient media	High nutrient media
MS2 Peak intensity (sequence TQITDDTYQNFFDK)	1.43E+05	9.08E+02	2.77E+03
MS2 Peak intensity (1 pMol standard)	1.07E+04	2.24E+04	1.86E+04
pMol EroS in excised gel band	13.4	0.0405	0.149
Sample volume in gel band (μL)	2.0	18.0	6.0
[EroS] loaded on gel	6.7 μM	2.3 nM	25 nM
[EroS] in culture supernatant		6.8 pM	75 pM

Appendix

Future directions and other thoughts

Towards identifying a RIF receptor

Several lines of evidence indicate that *S. rosetta* perceives bacterial cues that modulate rosette development (the RIFs, LPEs, and IOR-1) via a receptor-mediated mechanism. First, the rosette-inducing activity of the RIFs saturates in a manner that is consistent with a ligand-receptor interaction. Second, because the RIFs and LPEs induce robust rosette development at micromolar concentrations, it is unlikely that the bioactivity of the lipids is the result of their incorporation into the *S. rosetta* membrane. And finally, the potency and specificity of IOR-1 suggest that it functions as a competitive inhibitor of RIF-2.

So how do we go about identifying the *S. rosetta* receptors for these molecules? I spent a great deal of time thinking about this question over the past several years, and I have explored several candidate and unbiased approaches for identifying RIF receptors. Here, I will briefly summarize select findings, and suggest an approach for identifying a RIF receptor in the future.

GPCRs as candidate receptors

At the start of my thesis work, only a handful of animal receptors had been demonstrated to interact with lipid ligands, and many of these receptors (including sphingolipid receptors) are absent from the *S. rosetta* genome. I selected a group of seven transmembrane G-protein coupled receptors (GPCRs) as the targets of my candidate approach, because several putative GPCRs are present in the *S. rosetta* genome, and because many techniques were already available for studying GPCR signal transduction.

Further annotation of *S. rosetta* GPCR domain structures revealed that all but one protein likely belonged to a class of highly diverse and largely uncharacterized adhesion GPCRs (Table A1). A classical readout of GPCR activation is the release of intracellular calcium that occurs as part of GPCR signal transduction. Thus, calcium imaging is a widely-utilized method employed to identify ligands for orphan GPCRs. However, the signal transduction pathway following adhesion GPCR activation differs from other classes of GPCRs, and calcium release cannot be used as a reliable readout of ligand binding.

Therefore, I set out to use the Transfluor Assay, a technique that monitors receptor recycling via localization of a fluorescently labeled beta-arrestin protein, to determine if *Algoriphagus* lipids activate candidate *S. rosetta* GPCRs. I heterologously expressed several *S. rosetta* GPCRs in beta-arrestin-GFP U2OS cells, and assessed the localization of beta-arrestin in the presence and absence of *Algoriphagus* lipids. If GPCR activation occurred, beta-arrestin would be translocated from a diffuse state throughout the cytoplasm, into bright puncta associated with vesicles for receptor recycling. Although one *S. rosetta* GPCR appeared to be activated in response *Algoriphagus* lipids (Figure A1), it became clear that adhesion GPCRs are challenging proteins to work with biochemically. Moreover, because there were no techniques available for perturbing gene or protein function in *S. rosetta*, I ultimately decided that it was not yet the right time to pursue this project.

A constitutive rosette-forming mutant

With evidence that a forward genetic approach could identify genes required for rosette development, I decided to explore the idea of screening/selecting for constitutive rosette-forming mutants. Although such a mutant would surely be rare, I hypothesized that if *Algoriphagus* lipids were indeed perceived by a receptor, a constitutively active form of this receptor or downstream signaling components should result in a mutant that remains a rosette even in the absence of inducing molecules.

I spent a short amount of time performing and optimizing mutagenesis and size-selections for constitutive mutants. However, during these experiments, I was concurrently collecting data that discouraged me from wholeheartedly pursing the constitutive selection. I had found that RIFs alone only induced the development of small, fragile rosettes, and that the synergistic activities of two distinct classes of *Algoriphagus* lipids, the RIFs and the LPEs, were required to induce robust rosette development[35]. Moreover, my result that the IOR-1 specifically inhibited RIF-2, but not the LPEs, led me to hypothesize that multiple *S. rosetta* receptors are required to perceive distinct *Algoriphagus* cues. The size selection I had implemented meant that I was not recovering small, "RIF only" rosettes, and that the large rosettes I had been looking for would have likely required constitutive mutations in multiple receptors or downstream effectors. And so, the very rare mutant I initially set out to isolate became impossibly rare in my eyes, and I surrendered to the constitutive selection.

IOR-1: our savior?

Although my (admittedly insufficient) attempts at identifying *S. rosetta* receptors for *Algoriphagus* lipids were unsuccessful, I am still optimistic that these receptors will be uncovered. The rosette development inhibitor, IOR-1, is a tractable biochemical probe for identifying a RIF receptor, and should not be overlooked[84]. Not only is IOR-1 incredibly potent and specific to the RIFs, unlike the inducing sulfonolipids, it is small and easy to synthesize (and the inactive diastereomers of IOR-1 are ideal controls for biochemical assays). Although calling IOR-1 our savior may be hyperbolic, it is a tool that I would have been thrilled to exploit as a new graduate student.

The diversity of rosette-inducing bacteria and molecules

At the start of my thesis work, it was already known that species from several genera of Bacteroidetes bacteria, including *Algoriphagus*, induce rosette development in *S. rosetta*[33]. Although the exact nature of the molecular cues produced by Bacteroidetes inducing bacteria were uncharacterized, the phylogenetic distribution of the bacteria led us to hypothesize that the cues were structurally related to those produced by *Algoriphagus*.

We have since found that diverse bacteria from phyla other than Bacteroidetes, including Actinobacteria, Verrucomicrobia, and Firmicutes, also regulate rosette development (Figure A2). Moreover, preliminary findings show that the molecular profiles of non-Bacteroidetes inducing bacteria differ from *Algoriphagus*, suggesting that the inducing cues produced by these bacteria are structurally unrelated to the RIF sulfonolipids.

How are these structurally diverse cues perceived by *S. rosetta*? We have already demonstrated that the RIF sulfonolipids have exceedingly specific structure-activity relationships[34,35]. I therefore find it unlikely that a single *S. rosetta* receptor is sufficient to interpret structurally diverse molecules. Instead, I hypothesize that *S. rosetta* expresses several

receptors that each perceive a subset of structurally-related inducing cues, and converge onto a common downstream rosette-development pathway.

The diverse nature of rosette-inducing bacteria is, in my opinion, very exciting because it shows that *S. rosetta* has much room to grow as a model for bacterial-eukaryotic interactions. Eukaryotes are surrounded by complex communities of bacteria that produce both general microbial signatures, as well as distinct molecular cues. The ability of *S. rosetta* to perceive diverse rosette-inducing bacteria provides us with the opportunity to study mechanisms by which eukaryotes interpret and respond to complex communities of environmental bacteria.

Characterizing the molecular basis of mating

This section is particularly challenging to write because there are so many unanswered questions surrounding mating in *S. rosetta*. Although the discovery that a chondroitinase (*Vfis*HCLase) secreted by *Vibrio fischeri* induces mating in *S. rosetta* will undoubtedly facilitate experimental approaches, characterizing the molecular basis of mating in *S. rosetta* will hardly be trivial.

Putative GAG-lyases in choanoflagellate genomes

A natural question is whether mating in *S. rosetta* can also be regulated by endogenous pheromones. Under starvation conditions and in the absence of chondroitinase-producing bacteria, *S. rosetta* can mate with low efficiency (mating takes >500X longer and occurs in less than 2% of the population), suggesting that multiple mechanisms may regulate mating in *S. rosetta*. Interestingly, I identified two proteins encoding GAG lyase domains in the genome of *S. rosetta* (I also identified three proteins with similar homology in the genome of *M. brevicollis,* although I have not observed swarming in *M. brevicollis*) (Figure A3).

GAG lyases are primarily produced by bacteria and a subset of fungi, and are very rare in eukaryotes. Proteins encoding GAG lyase domains are completely uncharacterized in holozoans, which is unsurprising given that, in addition to choanoflagellates, I could only identify GAG lyase domains in the sponge *O. carmela* (although not in *A. queenslandica*), and in select cnidarians (corals *E. pallida* and *A. digitigera*, sea anemone *N. vectensis,* and polyp *H. vulgaris)* (Table A2).

Because the putative *S. rosetta* GAG lyases harbored conserved residues at sites predicted for catalytic activity (Figure A3), I hypothesized that the *S. rosetta* GAG lyases may serve as an endogenous mechanism for regulating mating. RNAseq analysis indicated that although the *S. rosetta* GAG lyases are annotated correctly, both are expressed at very low levels in all *S. rosetta* cell types analyzed. Therefore, I chose to heterologously express and purify the *S. rosetta* proteins. Each protein came with a unique set of challenges. I did not detect induction of EGD79853 under any growth condition or in any expression strain of *E. coli*, indicating that the protein was likely toxic to the bacteria. On the other hand, EGD79387 expressed marvelously, but was incredibly prone to aggregation and entirely insoluble. I found that EGD79387 contained two regions of poly-glutamine repeats, which likely accounted for its aggregative nature. Although I generated deletion constructs of EGD79387 which removed these poly-glutamine repeats, these constructs still proved to be insoluble.

The hypothesis that endogenous GAG lyases may regulate mating in *S. rosetta* under certain environmental conditions remains untested. Nonetheless, it is wildly intriguing that *S. rosetta* contains putative GAG lyases, given how infrequently these proteins are detected outside of bacteria and fungi.

***Identifying the proteoglycan target of* VfisHCLase**

Identifying the *S. rosetta* chondroitin sulfate proteoglycan (CSPG) target of *Vfis*HCLase may help reveal the mechanisms by which mating pairs interact (i.e. cell-cell recognition and fusion), or enable the identification of downstream effectors required for mating. Although a handful of CSPGs are well studied in vertebrates, these proteoglycans are primarily associated with neural development and regeneration[132], and no clear orthologues of these proteins are present in the *S. rosetta* genome. However, many proteoglycan core proteins contain characteristic small leucine-rich repeats (SLRPs), along with specific residues required for the attachment of glycosaminoglycan chains[133]. Despite the absence of CSPG orthologues in *S. rosetta,* it may be possible to identify putative *S. rosetta* CSPGs by looking for genes that contain sequence motifs, such as SLRPs and chondroitin sulfate serine attachment residues, that are conserved among characterized CSPG core proteins.

If a candidate approach is to be considered, the basement membrane proteoglycan Bamacan is rather interesting. Bamacan can occur in certain cell types as either a proteoglycan secreted at the basement membrane, or as an intracellular protein involved in sister chromatid cohesion (known as structural maintenance of chromosome 3, SMC3)[134]. Although the intracellular function of SMC3 is highly conserved in eukaryotes, little is known about the phylogenetic distribution and function of Bamacan as a basement membrane CSPG. *S. rosetta* has a clear orthologue of Bamacan/SMC3 (PTSG_07983), however, whether this protein solely functions as a cohesin, or is also present extracellularly as a CSPG, is unknown.

Alternatively, it may be possible to take a combined biochemical and proteomics approach to identify *S. rosetta* CSPGs. Novel CSPGs were successfully identified in *C. elegans* (an organism lacking orthologues of mammalian CSPGs) by cleverly exploiting the intrinsic biochemical properties of proteoglycans[135]. Because chondroitin chains are degraded during mating induction in *S. rosetta*, it may be possible to identify the target proteoglycan by simply looking for differences in the banding pattern of glycopeptides isolated from cells before and after exposure to *V. fischeri*.

Yet, identifying the *S. rosetta* targets of *V. fischeri* may still not help us answer one question critical to understanding mating in *S. rosetta*: are there mating types in *S. rosetta*? Diverse cell morphologies are observed during starvation-induced mating, and it is hypothesized that anisogamous mating occurs under starvation conditions[37]. In contrast, I have not observed morphological differentiation of *S. rosetta* during *Vibrio*-induced mating. Although cells are often of different mating types during isogamous mating, mechanisms of mating type determination are understood for very few isogamous eukaryotes. Therefore, understanding if and how mating type is regulated, and whether mating type switching occurs in *S. rosetta,* would provide important insight into the evolution of sex determination.

FIGURES

Table A1. Putative GPCRs encoded in the *S. rosetta* genome.

***S. rosetta* gene ID**	**PFAM domains**	**GPCR Classification**
PTSG_03655	7TM, EGF, laminin	Adhesion
PTSG_06041	7TM, GPS	Adhesion
PTSG_06956	7TM, GPS	Adhesion
PTSG_09376	7TM, GPS, fibronectin	Adhesion
PTSG_09542	7TM, GPS, TIG	Adhesion
PTSG_09543	7TM, GPS, TIG	Adhesion
PTSG_09717	7TM, GPS	Adhesion
PTSG_09821	7TM, GPS, collagen	Adhesion
PTSG_11558	7TM, cysteine residues	Class C (glutamate)

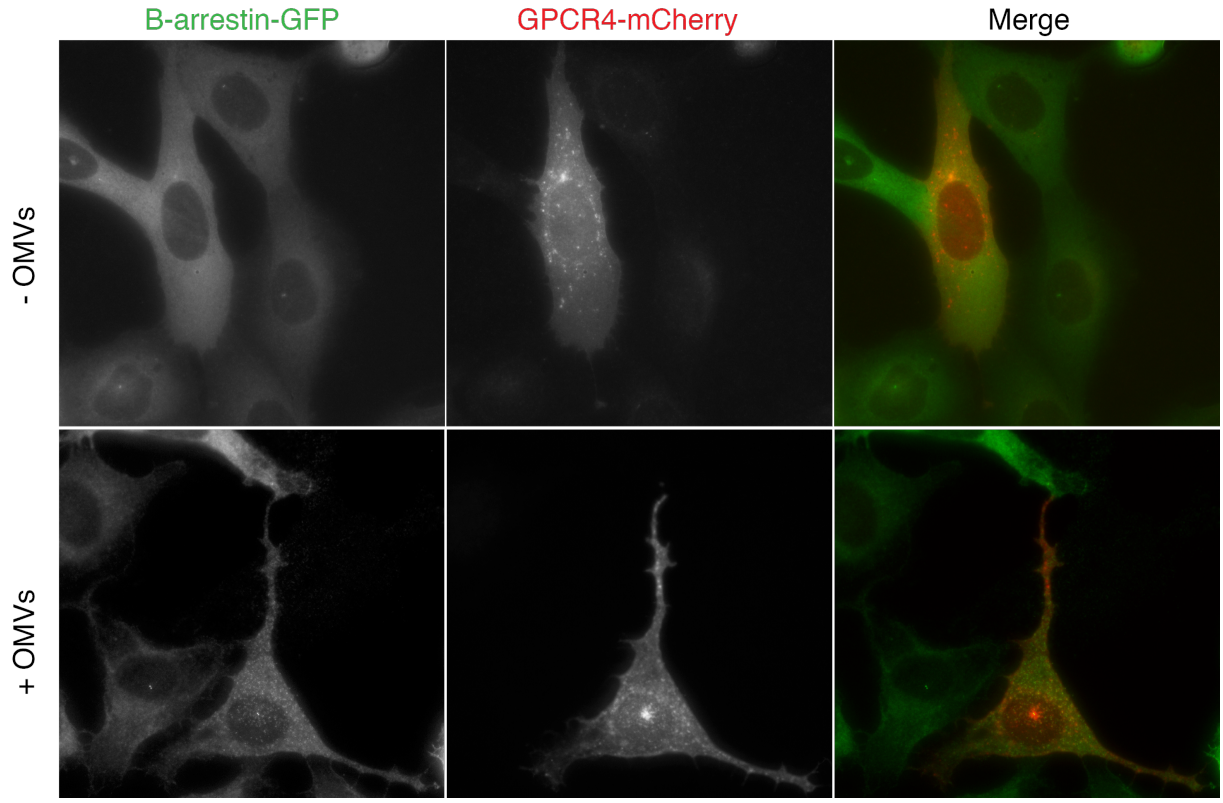


Figure A1. *S. rosetta* GPCR4 (PTSG_09376) is activated in response to *Algoriphagus* OMVs. In U2OS cells expressing GPCR4 (GPCR4-mCherry), beta-arrestin (B-arrestin-GFP) is trafficked from diffuse locations throughout the cytoplasm to distinct puncta on the membrane after exposure to *Algoriphagus* OMVs. Relocalization of beta-arrestin is consistent with GPCR receptor activation and subsequent desensitization through receptor recycling.

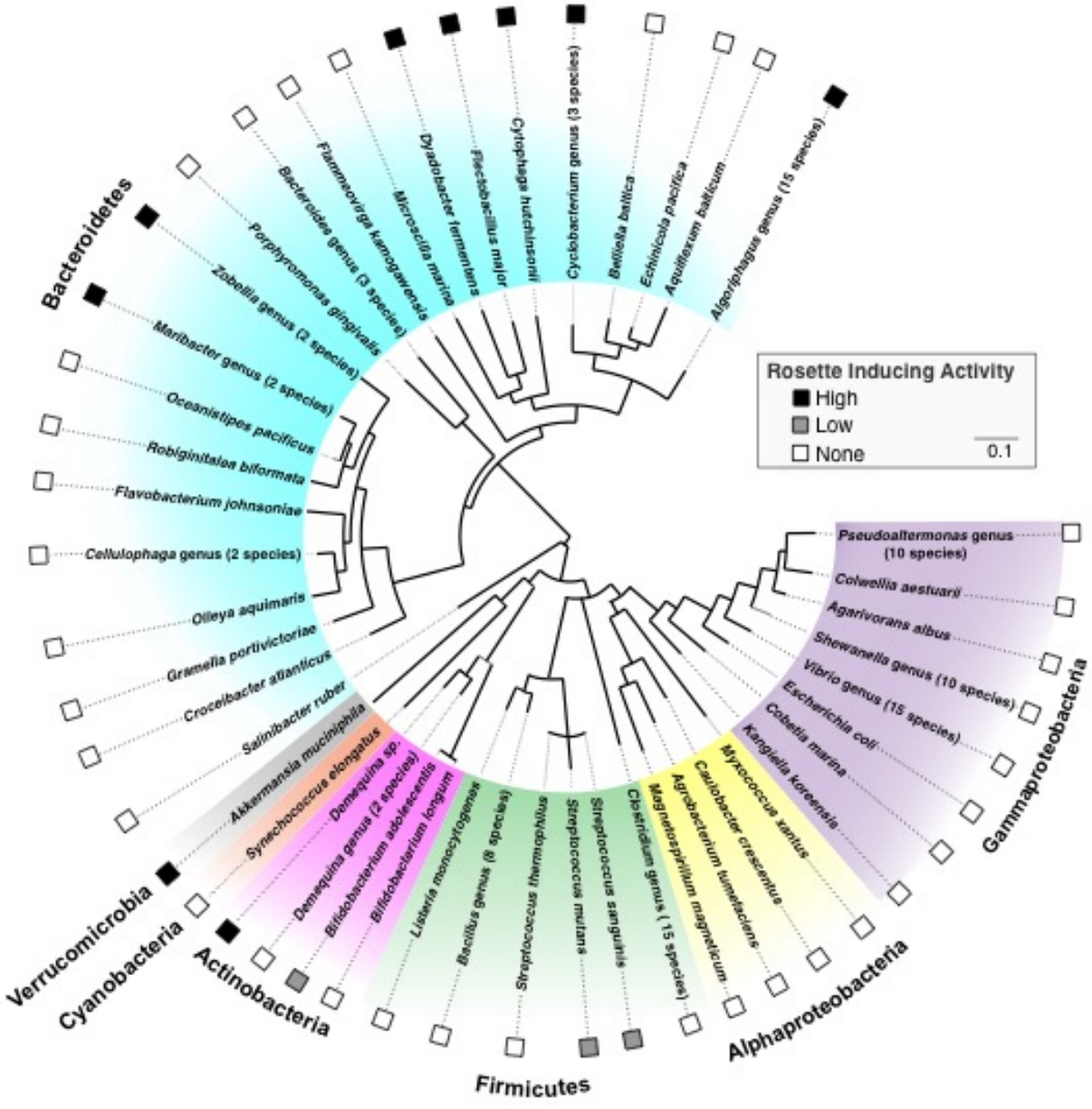


Figure A2. Phylogenetically diverse bacteria induce rosette development in *S. rosetta*. Although most bacteria found to induce rosette development belong to the Bacteroidetes phylum, rosette development also occurs in response to bacteria from Verrucomicrobia, Actinobacteria, and Firmicutes. A "High" level of rosette inducing activity indicates that live bacteria and conditioned media were sufficient to trigger robust rosette development, while a "Low" level indicates that rosette development occurred only in response to isolated bacterial lipids. "None" indicates that the level of rosette induction was below the limit of detection.

Figure A3. Choanoflagellates genomes encode putative GAG lyases. Alignment of the *V. fischeri Vfis*HCLase amino acid sequence to *S. rosetta* (Sr) and *M. brevicollis* (Mb) GAG lyase domain-containing proteins reveals that the choanoflagellate sequences harbor conserved His and Tyr residues (indicated by *) at sites required for catalytic activity in characterized GAG lyases. Amino acids with >50% conservation between sequences are shaded (black shading for identical amino acids and grey shading for similar amino acids. (Bottom) Accession information for all proteins containing GAG lyase domains that are present in sequenced holozoans.

Table A2. Putative GAG lyases are present in choanoflagellates and other holozoans.

	Species	Accession
Choanoflagellates		
	Salpingoeca rosetta	EGD79853.1
		EGD79387.1
	Monosiga brevicollis	EDQ89421.1
		EDQ87570.1
		EDQ84702.1
		EDQ84181.1
Porifera		
	Oscarella carmela	comp47422_c0_seq1:3-2225(+)
Cnidaria		
	Nematostella vectensis	EDO41162.1
		EDO30805.1
		EDO33283.1
		EDO30225.1
	Acropora digitifera	XP_015775436.1
		XP_015766608.1
		XP_015766570.1
	Exaiptasia pallida	KXJ21045.1
		KXJ05346.1
	Hydra vulgaris	XP_004212482.2

1. Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology 2004 4:1* **4,** 2 (2004).
2. Narbonne, G. M. The Ediacara Biota: Neoproterozoic Origin of Animals and Their Ecosystems. *Annu. Rev. Earth Planet. Sci.* **33,** 421–442 (2005).
3. King, N. The unicellular ancestry of animal development. *Dev Cell* **7,** 313–325 (2004).
4. Carroll, S. B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature* **409,** 1102–1109 (2001).
5. Grosberg, R. K. & Strathmann, R. R. The Evolution of Multicellularity: A Minor Major Transition? *Annu. Rev. Ecol. Evol. Syst.* **38,** 621–654 (2007).
6. Knoll, A. H. The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* **39,** 217–239 (2011).
7. Alegado, R. A. & King, N. Bacterial influences on animal origins. *Cold Spring Harbor Perspectives in Biology* **6,** a016162–a016162 (2014).
8. Mcfall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA* **110,** 3229–3236 (2013).
9. Seilacher, A. Biomat-related lifestyles in the Precambrian. *PALAIOS* **14,** 86–93 (1999).
10. Hancock, R. E. & Scott, M. G. The role of antimicrobial peptides in animal defenses. *Proceedings of the National Academy of Sciences* **97,** 8856–8861 (2000).
11. Stuart, L. M., Paquette, N. & Boyer, L. Effector-triggered versus pattern-triggered immunity: how animals sense pathogens. *Nat. Rev. Immunol.* **13,** 199–206 (2013).
12. 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).
13. Koropatnick, T. A. *et al.* Microbial factor-mediated development in a host-bacterial mutualism. *Science* **306,** 1186–1188 (2004).
14. McFall-Ngai, M. The development of cooperative associations between animals and bacteria: establishing détente among domains. *American zoologist* (1998).
15. Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci USA* **99,** 15451–15455 (2002).
16. Hooper, L. V. Bacterial contributions to mammalian gut development. *Trends in Microbiology* **12,** 129–134 (2004).
17. Hadfield, M. G. Biofilms and marine invertebrate larvae: what bacteria produce that larvae use to choose settlement sites. *Ann Rev Mar Sci* **3,** 453–470 (2011).
18. Webster, N. S. *et al.* Metamorphosis of a scleractinian coral in response to microbial biofilms. *Appl Environ Microb* **70,** 1213–1221 (2004).
19. Hay, M. E. Marine chemical ecology: chemical signals and cues structure marine populations, communities, and ecosystems. *Ann Rev Mar Sci* **1,** 193–212 (2009).
20. Tebben, J. *et al.* Induction of Larval Metamorphosis of the Coral Acropora millepora by Tetrabromopyrrole Isolated from a Pseudoalteromonas Bacterium. *PloS one* **6,** e19082–8 (2011).
21. Pacheco, A. R. & Sperandio, V. Inter-kingdom signaling: chemical language between bacteria and host. *Curr Opin Microbiol* **12,** 192–198 (2009).
22. Hughes, D. T. & Sperandio, V. Inter-kingdom signalling: communication between

bacteria and their hosts. *Nat Rev Microbiol* **6,** 111–120 (2008).
23. Leadbeater, B. S. C. *The Choanoflagellates: Evolution, Ecology, and Biology*. (Cambridge University Press, 2015).
24. Artis, D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat. Rev. Immunol.* **8,** 411–420 (2008).
25. Richter, D. J. & King, N. The genomic and cellular foundations of animal origins. *Annu. Rev. Genet.* **47,** 509–537 (2013).
26. King, N. & Carroll, S. B. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc Natl Acad Sci USA* **98,** 15032–15037 (2001).
27. Abedin, M. & King, N. The premetazoan ancestry of cadherins. *Science* **319,** 946–948 (2008).
28. King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451,** 783–788 (2008).
29. Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14,** R15 (2013).
30. Levin, T. C., Greaney, A. J., Wetzel, L. & King, N. The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *Elife* **3,** (2014).
31. Dayel, M. J. *et al.* Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Developmental biology* **357,** 73–82 (2011).
32. Fairclough, S. R., Dayel, M. J. & King, N. Multicellular development in a choanoflagellate. *Curr Biol* **20,** R875–6 (2010).
33. Alegado, R. A. *et al.* A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *Elife* **1,** e00013 (2012).
34. Beemelmanns, C. *et al.* Synthesis of the rosette-inducing factor RIF-1 and analogs. *J. Am. Chem. Soc.* **136,** 10210–10213 (2014).
35. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc Natl Acad Sci USA* **113,** 7894–7899 (2016).
36. 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.* (2016). doi:10.1021/jacs.6b01190
37. Levin, T. C. & King, N. Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. *Current Biology* **23,** 2176–2180 (2013).
38. Dayel, M. J. & King, N. Prey capture and phagocytosis in the choanoflagellate *Salpingoeca rosetta*. *PloS one* **9,** e95577 (2014).
39. Tyler, S. Epithelium--the primary building block for metazoan complexity. *Integr. Comp. Biol.* **43,** 55–63 (2003).
40. Stanley, S. M. An ecological theory for the sudden origin of multicellular life in the late precambrian. *Proceedings of the National Academy of Sciences* **70,** 1486–1489 (1973).
41. Hooper, L. V. & Gordon, J. I. Commensal host-bacterial relationships in the gut. *Science* **292,** 1115–1118 (2001).
42. 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).
43. Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* **456,** 507–510 (2008).
44. 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).
45. Weiss, B. L., Maltz, M. & Aksoy, S. Obligate symbionts activate immune system development in the tsetse fly. *J Immunol* **188,** 3395–3403 (2012).
46. Bosch, T. C. G. Cnidarian-microbe interactions and the origin of innate immunity in metazoans. *Annu Rev Microbiol* **67,** 499–518 (2013).
47. An, D. *et al.* Sphingolipids from a Symbiotic Microbe Regulate Homeostasis of Host Intestinal Natural Killer T Cells. *Cell* **156,** 123–133 (2014).
48. Shikuma, N. J. *et al.* Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* **343,** 529–533 (2014).
49. 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).
50. 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).
51. 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).
52. Leadbeater, B. S. C. & C, M. A microscopical study of a marine species of *Codosiga* James-Clark (Choanoflagellata) with special reference to the ingestion of bacteria. *Biol. J. Linn. Soc.* 337–347 (1974).
53. Nielsen, C. Six major steps in animal evolution: are we derived sponge larvae? *Evol Dev* **10,** 241–257 (2008).
54. Arendt, D., Benito-Gutierrez, E., Brunet, T. & Marlow, H. Gastric pouches and the mucociliary sole: setting the stage for nervous system evolution. *Philos Trans R Soc Lond, B, Biol Sci* **370,** 20150286–20150286 (2015).
55. Haeckel, E. The gastrea-theory, the phylogenetic classification of the animal kingdom and the homology of the germ-lamelle. *Q. J. Microscop. Sci.* **14,** 142–165 (1874).
56. Alegado, R. A. *et al.* Complete genome sequence of Algoriphagus sp. PR1, bacterial prey of a colony-forming choanoflagellate. *Journal of Bacteriology* **193,** 1485–1486 (2011).
57. 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).
58. Kreft, J. M. Effects of forming multicellular colonies or attaching to surfaces on feeding rates of the choanoflagellate Salpingoeca rosetta. (2010).
59. Bartke, N. & Hannun, Y. A. Bioactive sphingolipids: metabolism and function. *The Journal of Lipid Research* **50,** S91–S96 (2008).
60. 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).
61. Godchaux, W., III & Leadbetter, E. R. Sulfonolipids are localized in the outer membrane of the gliding bacterium *Cytophaga johnsonae*. *Arch Microbiol* 42–47 (1988).
62. 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).
63. Ishii, I., Fukushima, N., Ye, X. & Chun, J. Lysophospholipid receptors: signaling and

biology. *Annu. Rev. Biochem.* **73,** 321–354 (2004).
64. Makide, K. *et al.* Novel lysophosphoplipid receptors: their structure and function. *J Lipid Res* **55,** 1986–1995 (2014).
65. Foster, J. W., Spector, M. P. & Moat, A. G. *Microbial Physiology*. 1–25 (2003).
66. Grzelczyk, A. & Gendaszewska-Darmach, E. Novel bioactive glycerol-based lysophospholipids: new data -- new insight into their function. *Biochimie* **95,** 667–679 (2013).
67. Sommer, F. & Bäckhed, F. The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol* **11,** 227–238 (2013).
68. 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).
69. Stocker, R. Marine microbes see a sea of gradients. *Science* **338,** 628–633 (2012).
70. Kimelman, D. & Kirschner, M. Synergistic induction of mesoderm by FGF and TGF-beta and the identification of an mRNA coding for FGF in the early Xenopus embryo. *Cell* **51,** 869–877 (1987).
71. Ruiz i Altaba, A. & 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).
72. Bachiller, D. *et al.* The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature* **403,** 658–661 (2000).
73. Niehrs, C. Regionally specific induction by the Spemann-Mangold organizer. *Nat Rev Genet* **5,** 425–434 (2004).
74. 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).
75. 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).
76. 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).
77. 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).
78. Nedashkovskaya, O. I. *et al.* 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).
79. Tafesse, F. G. *et al.* Disruption of Sphingolipid Biosynthesis Blocks Phagocytosis of Candida albicans. *PLoS Pathog* **11,** e1005188 (2015).
80. Hibberd, D. J. Observations on the ultrastructure of the choanoflagellate Codosiga botrytis (Ehr.) Saville-Kent with special reference to the flagellar apparatus. *J Cell Sci* **17,** 191–219 (1975).
81. Merrill, A. H. De novo sphingolipid biosynthesis: a necessary, but dangerous, pathway. *Journal of Biological Chemistry* **277,** 25843–25846 (2002).
82. Hannun, Y. A. & Obeid, L. M. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nat. Rev. Mol. Cell Biol.* **9,** 139–150 (2008).

83. Takikawa, H., Nozawa, D., Kayo, A. & Muto, S. Synthesis of sphingosine relatives. Part 22. Synthesis of sulfobacin A, B and flavocristamide A, new sulfonolipids isolated from Chryseobacterium sp. *Journal of the Chemical …* (1999).
84. 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).
85. Mcfall-Ngai, M. Divining the Essence of Symbiosis: Insights from the Squid-Vibrio Model. *PLoS Biol* **12,** e1001783 (2014).
86. Thompson, F. L., Austin, B. & Swings, J. *The biology of vibrios.* (ASM Press, 2006).
87. Watson, G. J., Bentley, M. G., Gaudron, S. M. & Hardege, J. D. The role of chemical signals in the spawning induction of polychaete worms and other marine invertebrates. *Journal of Experimental Marine Biology and Ecology* **294,** 169–187 (2003).
88. Giese, A. C. Comparative physiology: annual reproductive cycles of marine invertebrates. *Annu Rev Physiol* **21,** 547–576 (1959).
89. Downes, J. A. The swarming and mating flight of Diptera. *Annual Review of Entomology* **14,** 271–298 (1969).
90. Buskey, E. J. Components of mating behavior in planktonic copepods. *Journal of Marine Systems* **15,** 13–21 (1998).
91. Avery, M. I. Lekking in birds: choice, competition and reproductive constraints. *Ibis* **126,** 177–187 (1984).
92. O'Day, D. H. Aggregation during sexual development in *Dictyostelium discoideum*. *Can. J. Microbiol.* **25,** 1416–1426 (1979).
93. Veith, M., Beer, N., Kiefer, A., Johannesen, J. & Seitz, A. The role of swarming sites for maintaining gene flow in the brown long-eared bat. *Heredity (Edinb)* **93,** 342–349 (2004).
94. Dini, F. & Nyberg, D. in *Advances in Microbial Ecology* **13,** 85–153 (Springer US, 1993).
95. Bell, G. *Sex and death in protozoa: the history of an obsession*. (Cambridge University Press, 1988).
96. Goodenough, U. & Heitman, J. Origins of eukaryotic sexual reproduction. *Cold Spring Harbor Perspectives in Biology* **6,** a016154–a016154 (2014).
97. Lupp, C., Urbanowski, M., Greenberg, E. P. & Ruby, E. G. The *Vibrio fischeri* quorum-sensing systems ain and lux sequentially induce luminescence gene expression and are important for persistence in the squid host. *Molecular Microbiology* **50,** 319–331 (2003).
98. Lupp, C. & Ruby, E. G. *Vibrio fischeri* uses two quorum-sensing systems for the regulation of early and late colonization factors. *Journal of Bacteriology* **187,** 3620–3629 (2005).
99. Shibata, S., Yip, E. S., Quirke, K. P., Ondrey, J. M. & Visick, K. L. Roles of the structural symbiosis polysaccharide (syp) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in *Vibrio fischeri*. *Journal of Bacteriology* **194,** 6736–6747 (2012).
100. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44,** D279–D285 (2016).
101. Zhang, F., Zhang, Z. & Linhardt, R. J. in *Handbook of Glycomics* 59–80 (Elsevier, 2010).
102. Ahn, M. Y. *et al.* Characterization of a Bacteroides species from human intestine that

degrades glycosaminoglycans. *Can. J. Microbiol.* **44,** 423–429 (2011).
103. Hong, S. W. *et al.* Purification and characterization of novel chondroitin ABC and AC lyases from Bacteroides stercoris HJ-15, a human intestinal anaerobic bacterium. *The FEBS Journal* **269,** 2934–2940 (2002).
104. Weijun Huang *et al.* Active site of chondroitin AC lyase revealed by the structure of enzyme−oligosaccharide complexes and mutagenesis. *Biochemistry* **40,** 2359–2372 (2001).
105. Han, W., Wang, W., Zhao, M., Sugahara, K. & Li, F. A novel eliminase from a marine bacterium that degrades hyaluronan and chondroitin sulfate. *J Biol Chem* **289,** 27886–27898 (2014).
106. Linhardt, R. J., Avci, F. Y., Toida, T., Kim, Y. S. & Cygler, M. in *Chondroitin sulfate: structure, role and pharmacological activity* **53,** 187–215 (Elsevier, 2006).
107. Shaya, D. *et al.* Composite active site of chondroitin lyase ABC accepting both epimers of uronic acid. *Glycobiology* **18,** 270–277 (2008).
108. DeAngelis, P. L. Microbial glycosaminoglycan glycosyltransferases. *Glycobiology* **12,** 9R–16R (2002).
109. DeAngelis, P. L. Evolution of glycosaminoglycans and their glycosyltransferases: implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *Anat Rec* **268,** 317–326 (2002).
110. Yamada, S., Sugahara, K. & Özbek, S. Evolution of glycosaminoglycans. *Communicative & Integrative Biology* **4,** 150–158 (2011).
111. Garron, M.-L. & Cygler, M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20,** 1547–1573 (2010).
112. Ori, A., Wilkinson, M. C. & Fernig, D. G. A systems biology approach for the investigation of the heparin/heparan sulfate interactome. *J Biol Chem* **286,** 19892–19904 (2011).
113. Kochert, G. Sexual Pheromones in Volvox Development. *Sexual Interactions in Eukaryotic Microbes* 73–93 (2012). doi:10.1016/B978-0-12-524160-1.50009-6
114. Bartels-Hardege, H. D. *et al.* Sex pheromones in marine polychaetes: a biologically active volatile compound from the coelomic fluid of female *Nereis* (*Neanthes*) *japonica*. *Journal of Experimental Marine Biology and Ecology* **201,** 275–284 (1996).
115. Li, W. *et al.* Bile acid secreted by male sea lamprey that acts as a sex pheromone. *Science* **296,** 138–141 (2002).
116. Jones, B. W., Maruyama, A., Ouverney, C. C. & Nishiguchi, M. K. Spatial and Temporal Distribution of the Vibrionaceae in Coastal Waters of Hawaii, Australia, and France. *Microb Ecol* **54,** 314–323 (2007).
117. Omori, M. & Hamner, W. M. Patchy distribution of zooplankton: Behavior, population assessment and sampling problems. *Marine Biology* **72,** 193–200 (1982).
118. Hamner, W. M. & Dawson, M. N. A review and synthesis on the systematics and evolution of jellyfish blooms: advantageous aggregations and adaptive assemblages. *Hydrobiologia* **616,** 161–191 (2009).
119. Sorensen, P. W. & Wisenden, B. D. *Fish Pheromones and Related Cues*. (John Wiley & Sons, 2015).
120. Kartika, T., Shimizu, N. & Yoshimura, T. Identification of Esters as Novel Aggregation Pheromone Components Produced by the Male Powder-Post Beetle, Lyctus africanus Lesne (Coleoptera: Lyctinae). *PloS one* **10,** e0141799 (2015).

121. Wertheim, B., van Baalen, E.-J. A., Dicke, M. & Vet, L. E. M. Pheromone-mediated aggregation in nonsocial arthropods: an evolutionary ecological perspective. *Annual Review of Entomology* **50,** 321–346 (2005).
122. Painter, S. D., Clough, B., Garden, R. W., Sweedler, J. V. & Nagle, G. T. Characterization of Aplysia Attractin, the First Water-borne Peptide Pheromone in Invertebrates. *The Biological Bulletin* **194,** 120–131 (2016).
123. Ram, J. L., Müller, C. T., Beckmann, M. & Hardege, J. D. The spawning pheromone cysteine-glutathione disulfide ('nereithione') arouses a multicomponent nuptial behavior and electrophysiological activity in Nereis succinea males. *FASEB J* **13,** 945–952 (1999).
124. Babcock, R., Mundy, C., Keesing, J. & Oliver, J. Predictable and unpredictable spawning events: in situ behavioural data from free-spawning coral reef invertebrates. *Invertebrate Reproduction & Development* **22,** 213–227 (2011).
125. Smith, J. R. & Strehlow, D. R. Algal-induced spawning in the marine mussel Mytilus californianus 4. *International Journal of Invertebrate Reproduction* **6,** 129–133 (1983).
126. Starr, M., Himmelman, J. H. & Therriault, J.-C. Isolation and properties of a substance from thediatom *Phaeodactylum tricornutum* whic h induces spawning in the sea urchin *Strongylocentrotus droebachiensis*. *Marine Ecology Progress Series* **79,** 275–287 (1992).
127. Starr, M., Himmelman, J. H. & Therriault, J.-C. Direct Coupling of Marine Invertebrate Spawning with Phytoplankton Blooms. *Science* **247,** 1071–1074 (1990).
128. Miller, D. J. & Ax, R. L. Carbohydrates and fertilization in animals. *Mol. Reprod. Dev.* **26,** 184–198 (1990).
129. Salustri, A., Camaioni, A., Di Giacomo, M., Fulop, C. & Hascall, V. C. Hyaluronan and proteoglycans in ovarian follicles. *Hum. Reprod. Update* **5,** 293–301 (1999).
130. Wang, W. *et al.* Cloning and characterization of a novel chondroitin sulfate/dermatan sulfate 4-O-endosulfatase from a marine bacterium. *J Biol Chem* **290,** 7823–7832 (2015).
131. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences* **100,** 6940–6945 (2003).
132. Avram, S., Shaposhnikov, S., Buiu, C. & Mernea, M. Chondroitin Sulfate Proteoglycans: Structure-Function Relationship with Implication in Neural Development and Brain Disorders. *BioMed Research International* **2014,** 1–11 (2014).
133. Iozzo, R. V. The Family of the Small Leucine-Rich Proteoglycans: Key Regulators of Matrix Assembly and Cellular Growth. *Crit. Rev. Biochem. Mol. Biol.* (2008). doi:10.3109/10409239709108551
134. Wu, R.-R. & Couchman, J. R. cDNA Cloning of the Basement Membrane Chondroitin Sulfate Proteoglycan Core Protein, Bamacan: A Five Domain Structure Including Coiled-Coil Motifs. *J Cell Biol* **136,** 433–444 (1997).
135. Olson, S. K., Bishop, J. R., Yates, J. R., Oegema, K. & Esko, J. D. Identification of novel chondroitin proteoglycans in Caenorhabditis elegans: embryonic cell division depends on CPG-1 and CPG-2. *J Cell Biol* **173,** 985–994 (2006).