SYMBIOSIS DRIVEN VARIATION AND SUPPLY OF NATURAL PRODUCTS

by

Ma. Diarey B. Tianero

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

Doctor of Philosophy

Department of Medicinal Chemistry

The University of Utah

December 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Ma. Diarey B. Tianero

has been approved by the following supervisory committee members:

Eric Schmidt	, Chair	06/18/15 Date Approved
Chris Ireland	, Member	06/15/15 Date Approved
Baldomero Olivera	, Member	06/15/15 Date Approved
Darrell Davis	, Member	06/15/15 Date Approved
Grzegorz Bulaj	, Member	06/15/15 Date Approved

and by Darrell Davis , Chair of

the Department of Medicinal Chemistry

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

ABSTRACT

Drug discovery and development from marine invertebrates has been fraught with two key problems, namely, the variability of occurrence and limited supply. Bacteria in symbiosis with marine invertebrates have been shown to produce most bioactive natural products isolated from these organisms, and thus are central to addressing questions of occurrence and issues of supply. Specifically, the factors that influence symbiosis influence the distribution and supply of natural products. This dissertation sought to address these two problems through studies in symbiosis and supply of symbiotic natural products.

First, the global patterns of chemical symbiosis in marine ascidians, a group of highly prolific producers of natural products, were examined. Symbiosis in ascidians is shown to be host-specific (meaning that similar species of invertebrates contain similar bacterial symbionts); further, microbiomes are shown to be equally diverse regardless of location. Secondary metabolism was also found to be host-specific, but is more sensitive to biogeographical factors as evidenced by the increase in the potency of the secondary metabolites in tropical regions. To address the supply of rare natural products, heterologous expression was used to produce useful quantities of a group of symbiotic natural products, cyanobactins. Using metabolic engineering, a platform was developed to supply cyanobactins in high-titer, and its usefulness showcased in the discovery of

novel activities of these natural products. Another facet of the supply problem is the substantial difficulty involved in synthesizing derivatives of natural products, which generally requires total chemical synthesis. On this aspect of the supply problem, the capacity of the cyanobactin pathway to generate unprecendented structural diversity by the incorporation of non-proteinogenic amino acids into this multistep, substrate-tolerant biosynthetic pathway was demonstrated.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not be possible without the help and support of the following people and funding institutions. First, I would like to thank my advisor, Eric Schmidt, for his guidance, training, and support throughout my graduate research career. I would also like to thank my committee members, Chris Ireland, Baldomero Olivera, Darrell Davis, and Grzegorz Bulaj, for their guidance and interest in my work. I would like to acknowledge NIHGMS for funding all the work in this thesis.

I would also like to thank all my collaborators and coauthors. In the ascidian microbiomes and metabolomes project, I would like to thank Tim Bugni and Tom Wyche for performing the LC/MS and PCA analyses of the ascidian extracts, Angela Presson for the assistance in statistical analyses, Mike Koch for performing the bioactivity assays, and Jason Kwan for help in microbiome and bioinformatics analyses. In the heterologous expression of cyanobactins, I would like to thank Elizabeth Pierce for performing the prenyltransferase knock out experiments and assembly of the *pat* pathway, Duane Ruffner of Symbion Discovery for the *tru* pathway vector that I used in all optimization and heterologous production experiments, John McIntosh for the initial experiments on the combination of *tru* and *mev* pathways, and Shrinivasan Raghuraman for performing all the calcium imaging experiments. In the engineering of cyanobactin pathway, I would like to thank Mohamed Donia for conceiving the project and cloning the derivatives of

cyanobactins, as well as Peter Schultz and Travis Young for providing the vectors for unnatural amino acid incorporation.

Lastly, I would like to thank my family and friends, especially my husband, John McIntosh, for all the support, patience, and confidence in the making and finishing of this dissertation.

CHAPTER 1

INTRODUCTION

Natural products variability and supply

Natural products (NPs) are an excellent source of bioactive lead compounds for drug discovery.[1] Among NP producers, marine invertebrates are known for harboring diverse, biologically potent, and structurally complex metabolites that possess drug-like qualities.[2] For example, of 36 marine natural products that were in the preclinical/clinical pipeline from 2004-2013, 30 were originally sourced from marine invertebrates.[3] However, drug discovery from marine invertebrates has been fraught with two problems, namely, the variability in isolation and limited supply. The isolation of a natural product from an organism, especially from a sessile invertebrate such as a sponge or an ascidian, does not guarantee that one will find the same natural products in subsequent collections.[4] Central to these challenges is the role of symbiotic bacteria in the production of marine invertebrate-derived natural products. [5-7]Specifically, symbiotic bacteria are central to chemical symbiosis, NP variability, and, as a result, to the supply of NPs. This dissertation addresses these two problems, the distribution and supply of symbiotic natural products.

Symbiosis driven variability in natural products

The following examples illustrate the influence of symbiosis on the distribution and variability of natural products from three invertebrate phyla: bryozoans, sponges, and ascidians. The identification of the symbiont producers is described in each case, as well as the factors that influence invertebrate-microbe symbiosis resulting in the sporadic occurrence of natural products. These studies emphasize the unappreciated role of cryptic host speciation, symbiont genetics, and geography in the production of compounds.

Lastly, these individual cases, among others, prompted us to examine the global distribution of ascidian microbial symbionts, which is the topic of the first part of my research.

Candidatus Endobugula sertula and the bryostatins

One of the most well-rounded studies in this field is the investigation of the chemical symbiosis between the bryozoan *Bugula neritina* and its symbiont *Candidatus* Endobugula sertula. The bryozoan *B. neritina* is a cosmopolitan species that is widely distributed in the temperate and tropical ocean regions. [8,9] It is the source of the bryostatins, powerful protein kinase C (PKC) modulators that have been extensively investigated for their anticancer, antiviral, and neuroactive properties. [9] Limited and variable isolation from the bryozoans prompted a series of studies on the biosynthetic origin and ecology of bryostatins, with the goal of finding alternative routes of supplying these high-demand molecules.

The structure of bryostatins suggests a microbial origin, specifically via polyketide synthase (PKS) machinery followed by multiple tailoring steps. Indeed, an uncultured gamma-proteobacterium associated with larvae of the bryostatin-producing *B. neritina* was identified and named *Candidatus* Endobugula sertula. [10] [11] The same bacterium was found to harbor a set of biosynthetic genes consistent with biosynthetic hypotheses as to the origin of of bryostatins. [12,13] Subsequent studies strongly supported the hypothesis that *Ca.* E. sertula produces bryostatins. [14] Final proof will await attempts to express the bryostatin gene cluster in a heterologous host, which have not been successful to date. Meanwhile, efforts continue to further our understanding of the

genetics of the *Ca.* E. sertula symbiont. Early studies on the isolation of bryostatins indicated a variable distribution in *B. neritina* species. [15] Identifying the genetic basis of bryostatins was critical to further progress in the understanding of the distribution of these molecules. Haygood and Davidson (1999) first showed that the *B. neritina* populations along the coast of California are composed of sibling species. [16] Type D (found in deep waters initially) and type S (found in shallow waters) differed by 8% in their mitochondrial CO1 sequences. Both harbored *Ca.* E. sertula symbionts that differed in their 16S rRNA gene sequences, and the bryostatin chemistry differed in each type of *B. neritina*. Subsequently, cryptic species of *B. neritina* were also identified from the western Atlantic range, [8] and a series of studies have highlighted the influence of geography in the distribution of *B. neritina* and bryostatin production. In these studies, Cape Haterras, a known biogeographical divide, was used as the reference point in the sampling of *B. neritina.* Another cryptic species of *B. neritina,* type N*,* generally inhabits the northern Atlantic ranges and does not harbor *Ca.* E. sertula symbionts. [8] This species lacks bryostatins, consistent with a producer role for *Ca*. E. sertula. Subspecies Type S of *B. neritina*, similar to the California species, is concentrated in southern areas relative to Cape Haterras and contains both *Ca.* E. sertula symbionts as well as bryostatins. The presence or absence of symbionts was therefore thought to explain completely the separation between species. Recently, however, more detailed collections have shown that the boundary of N and S species is less distinct than previously thought. [17,18] Although there is a heavy distribution of N type north of Cape Haterras, some N type species were also found south of the Cape and vice versa. Despite this apparent mixing of host species, the *Ca.* E. sertula symbionts remained geographically restricted to the south,

regardless of the species of *B. neritina*. [18] This implies that bryostatin production might be strongly influenced by geography. Higher selection pressure in the southern warmer waters may lead to acquisition of *Ca.* E. sertula symbionts that can provide chemical defense. These studies demonstrate the interplay between different factors controlling the production of bryostatins. Cryptic speciation in *B. neritina* partly explains the presence and absence of symbionts and compounds, while the host-specificity of symbionts to these cryptic species of *B. neritina* appears to explain the variation between different bryostatin derivatives. Finally, the local environment may also play a role, presumably via natural selection, in determining the presence or absence of symbiotic natural product producers. Further widespread studies of the bryostatins, *B. neritina*, and *Ca.* E. sertula will be needed to attain a complete picture of bryostatin occurence and variation.

Prochloron didemni and the cyanobactins

Ascidians from the family Didemnidae are among the most chemically prolific groups of marine invertebrates. [2,7] Among other natural products, they produce a family of small cyclic peptides, known as cyanobactins. [19] Genome analysis of the obligate cyanobacterial symbiont of the ascidian *Lissoclinum patella*, *Prochloron didemni*, identified the biosynthetic cluster for cyanobactins, which was found to involve a ribosomal mode of biosynthesis, in which a precursor peptide is posttranslationally modified to give bioactive structures. [20] This mode of biosynthesis is now recognized as widespread, and a large class of natural products has been assigned to the growing family of ribosomally synthesized and posttranslationally modified peptides (RiPPs). [21]

The cyanobactins are to date the most biotechnologically developed group among

natural products from uncultured symbiotic bacteria. One important milestone in the development of cyanobactins as a platform for drug discovery was the demonstration, coincident with the discovery of their biosynthetic origins, that they could be heterologously expressed in *E. coli.* [20] Also important was the discovery of the tremendous flexibility of the cyanobactin pathway. [22] This latter discovery was largely enabled by an in-depth understanding of their distribution and variation in Nature. [22,23] Diverse families of cyanobactins had been isolated from different species of didemnid ascidians since the 1980s. [19,24-27] Structural overlaps between these families led to the hypothesis that much of cyanobactin chemical diversity can be traced to hypervariability in precursor peptides. [22] Indeed, analyses of cyanobactin pathways from didemnid ascidians across the tropical Pacific showed that the sequences are virtually identical except for regions that directly encode the cyanobactin backbone.[22] In some cases, recombination of pathways leads to altered patters of posttranslational modifications, but even in such cases, precursor peptides from different pathways can be mixed and matched *in vitro* and *in vivo*. [23] Producing artificial libraries of cyanobactins would therefore only require minimal manipulation of precursor peptides, without any engineering of modifying enzymes. This in-depth knowledge of the conserved genetic architecture of cyanobactin pathways also accelerated the discovery of cyanobactin derivatives by purely genetic approaches. [28]

Cyanobactin families are also variably distributed in didemnid ascidians. For example, trunkamides, patellins, patellamides, and lissoclinamides are often found in different combinations in *L. patella* ascidians. [22,29,30] To examine this question, three genomes of *Prochloron didemni* were sequenced and compared.[29] This analysis revealed

that the *Prochloron didemni* strains were largely identical except in regions that encoded secondary metabolites. Subsequently, 24 *Prochloron didemni* strains from different *L. patella* samples across the tropical Pacific were analyzed at specific loci as well as for secondary metabolite composition. No apparent distribution patterns could be discerned however. It was hypothesized that the sporadic variation in the pathways and metabolites is ecologically relevant for *Prochloron-L. patella* symbiosis. Specifically, the presence of different pathways in different ascidians allows access to the suite of cyanobactins as needed.

A finer look at the *L. patella* phylogeny later revealed a previously unrecognized distribution. [31] Specifically, *L. patella* samples collected from the vast tropical Pacific were identified to derive from cryptic speciation events, varying by 5-10% in their mtCO1 sequences. From this analysis, the phylogeny of *L. patella* could be used to predict the distribution of cyanobactins, while the *Prochloron* symbiont was 99% identical in all cases. Thus, in an echo of the bryostatin work described above, host cryptic speciation was found to be a key driving force of the distribution of cyanobactins. A broader sampling of *L. patella* samples will be needed to show if this 'host control' phenomenon is widespread and what other factors determine this symbiosis, for example host-factors that allow for recognition of specific *Prochloron* strains, or perhaps vertical transmission of symbionts. An improved understanding of the distribution of metabolites in *L. patella* across time and space might help improve our understanding of the ecological role of cyanobactins.

Candidatus Entotheonella sp. and a suite of peptides and polyketides

The marine sponges from the family Theonellidae are another group of marine invertebrates that have been studied for their structurally complex and bioactive NPs. [32] Most of these molecules are complex polyketides including the swinholides and discodermolides, as well as peptides such as keramamides, polytheonamides, theopalau - amides, or hybrid PKS-NRPS-derived molecules such as the calyculins, onnamides, and theopederins. [32-35] The microbial origin of these metabolites has been proposed early on. Indeed, microscopy [36,37] and later biosynthetic studies confirmed this hypothesis. [38,39]

Perhaps the earliest attempt to implicate symbiotic bacteria in the biosynthesis of sponge-derived metabolites was from the work of Bewley *et al.* in the mid-1990s. [36] Separation of microbial cells by differential centrifugation and subsequent chemical analysis showed that different microbial fractions contained different metabolites. For example, a fraction enriched in unicellular bacteria contained swinholides, while the fraction containing filamentous bacteria contained cyclic peptides of the theopalauamide family. In 2000, Schmidt *et al.* described the filamentous bacteria as a novel δ-proteobacteria, named *Candidatus* Entotheonella palauensis. [40] In this key study, the relationship of chemistry to the symbiont genotype was first proposed. More than a decade later, using a combination of single cell genomics and metagenomics, Wilson *et al.* confirmed that the *Ca.* Entotheonella species previously identified by Schmidt *et al.* are in fact the producers of most *Theonella swinhoei*-derived natural products. [41] Different strains of *Ca.* Entotheonella, produce different types of NPs present in the same *T. swinhoei* sample. *Ca.* Entotheonella species have also been described from other Theonellid sponges, where they are responsible for making large PKS-NRPS

compounds.[42]

It is becoming apparent that *Ca*. Entotheonella strains are diverse and widespread among sponges. In particular, a PCR survey showed that *Ca.* Entotheonella strains are present in different sponges outside the Theonillidae and also collected from different locations. [41] No apparent patterns of distribution have been identified so far. It is clear, however, that symbiosis with different strains lead to the isolation of different compounds in sponges. Studies on the distribution, host-specificity, and chemistry of *Ca*. Entoheonella will be critical to advance our understanding of their distribution and to facilitate exploiting their potential in drug discovery.

A global view of chemical symbiosis

The specific cases above emphasize the roles of symbiotic bacteria in the variability and distribution of natural products. However, they only represent a small subset of marine natural products. The first part of my thesis addresses this problem by examining global patterns in the distribution of chemical symbiosis that could guide natural products discovery. We focused on ascidians, examining the specificity, diversity, and stability of the microbiomes and secondary metabolism across hosts, geography, and time. We hypothesized that in the tropical regions where there is a higher degree of predation and competition, the selective pressure to survive leads to acquisition of more symbionts that could aid with the host' fitness (e.g., chemical defense) and therefore leads to higher microbial diversity. The results are discussed in Chapter 2.

Supply of symbiotic natural products

Natural products that are produced within the context of a symbiotic relationship are often potent because they are evolutionarily refined to interact with relevant biological targets. Consequently, they are often produced in limited amounts in the animals. The bryostatins described above are an excellent example of such compounds. The amount of bryostatins that can be isolated ranges from 10^{-5} to 10^{-7} wt% of the animal colony (18 grams from 14 tons of *B. neritina* in one example). [9] The same is true for *L. patella* metabolites, the patellazoles, which are products of the endosymbiont, *Ca.* endolissoclinum patella. [43] The patellazoles are highly cytotoxic, with an IC50 of 332 pM in cell assays. They are isolated at 0.08% dry weight of the ascidian and are sparsely found, severely hampering their development.

Another barrier in the supply of NPs from symbiotic bacteria is the inability to cultivate the producing symbionts. Symbiotic bacteria in established relationships with their host often lose functions that are required for independent lifestyles outside the hosts. In extreme cases, the genome of symbiotic bacteria is degraded over time, retaining only the genes that are relevant to the symbiosis such as the secondary metabolic genes, while depending on the host for the rest of its metabolic needs. [44]

Chemical synthesis is often the next step in supplying the natural products, especially those that are highly sought for their bioactivities. However, there are obvious drawbacks to total chemical synthesis. Natural products are structurally complex, often containing multiple chiral centers, and often require lengthy synthetic routes, thus posing a significant barrier to achieving a practical supply. [45] The total synthesis of bryostatins, for example, was completed in 79 steps in 1990s, and it was not until recently (2011) that

several research groups achieved the total synthesis with 36 to 43 steps. [46,47]

Heterologous expression of biosynthetic pathways in genetically tractable hosts such as *E. coli* is one of the most promising alternatives to symbiotic NP supply. [48 49] Although not without challenges, heterologous expression offers many advantages over total chemical synthesis and other complementary approaches such as aquaculture. [50] Once optimized, natural products can be supplied from scaled fermentations without the need for multiple reactions and in between purification steps required of total synthesis. In the second part of my dissertation, we used heterologous expression to address the supply problem in the symbiotic NPs, the cyanobactins.

Metabolic engineering and fermentation strategies for NP supply

Tools and strategies for the metabolic engineering of natural products expressions in heterologous hosts are becoming increasingly available. [51-53] Some rational techniques involve increasing precursor supply, overexpression of bottleneck enzymes in the pathways, improving regulation of gene expression, and channeling intermediates, precursors into desired pathways. [53] These, in tandem with empirical approaches such as optimization of fermentation conditions, media, additives, cofactors, and other primary metabolites, have been used for the improved supply of natural products *in vivo*. [54,55]

Increasing the precursor supply has been employed in several pathways, including polyketides, terpenoids, and shikimate pathways. As an example, in the production of actinorhodin, the supply of malonyl-CoA from acetyl-CoA was increased through the overexpression of acetyl-CoA carboxylase resulting in a 6-fold increase. [56] Similarly, the overexpression of the first enzyme in the shikimate pathway, DAHP, resulted in an

increase in the expression of Balhimycin, which required the aromatic amino acids produced through the shikimate pathway. [57] The terpenoid pathways are among the extensively studied tools in metabolic engineering. Terpenoid pathways use both precursors DMAPP and IPP as building blocks for longer terpene compounds. [58] In production of NPs in *E. coli*, these precursors are usually supplied by improving the efficiency of the native MEP pathway or via the heterologous expression of a mevalonate pathway derived from higher organisms. [59-61]

Empirical methods have also been used in the optimization of fermentation conditions for production of NPs. One example is in the optimization of 6-deoxyerythronolide B production in *E. coli*, where the effect of media, and subsequently its individual components were studied. In this example, tryptone, and subsequently, tryptone components, glutamine, and threonine amino acids were found to increase compound production by a further 16-fold. [55]

The cyanobactin pathway is one of the few marine NP pathways that have been successfully expressed in a heterologous host. [5,20,28] Since the first expression of cyanobactins in *E. coli*, there have been continuous efforts to improve and optimize the expression system in order to establish a robust platform that can supply practical amounts of cyanobactins. [28,62]However, despite the stability in the detection of compounds from the extracts, the best yield that we were able to obtain previously was 300 μg/L.

To improve the production of cyanobactins in *E. coli*, we used a combination of simple media and fermentation optimization, as well as co-expression of a heterologous mevalonate pathway to supply isoprene precursors. In Chapter 3, we show an overall

improvement in yield of about 100-fold, which is the highest titer for the expression of cyanobactins so far. With this ability to supply cyanobactins, we started to explore their bioactivities, which resulted in the discovery of a new activity for patellin 2 [63] in dorsal root ganglia neurons.

Engineering for expansion of cyanobactin diversity

One key advantage that synthetic compound possess over natural products is that it is vastly easier to generate derivatives of synthetic compound in order to optimize the bioactivity of a given compound. In contrast, it is much more difficult to discover or engineer natural product derivatives, and natural product chemists generally lack the ability to make deterministic changes in compounds, but are limited to what few derivatives are naturally occurring, or those that can be generated by disabling late-stage biosynthetic enzymes.

Within this context, the cyanobactin pathway is particularly notable, in that it can be readily engineered to allow the production of derivative compounds due to the high substrate tolerance of the modifying enzymes. In the first proof of this capacity, the precursor peptide of the patellamide pathway, patE2, encoding the natural product ulithiacyclamide was modified to encode a peptide mimic of the rattlesnake venom peptide, eptifibatide. [22] The artificial cyanobactin was referred to as eptidemnamide, with the replacement of the original disulfide with an amide bond. Successful production of eptidemnamide demonstrated that cyanobactin enzymes are substrate-tolerant beyond just the natural cyanobactin sequences. Eptidemnamide contained charged amino acid residues, arginine, and aspartic acid, and others that were not found in natural

cyanobactins, specifically, glycine, tryptophan, and glutamine. [22] Successive experiments in the engineering of cyanobactins further supported this finding, including swapping of precursor peptides in the trunkamide pathway, [23] and producing a new cyanobactin called minimide in the *tru* pathway that was discovered by genome mining. [28]

The capacity of the cyanobactin pathway for generating chemical diversity was subjected to a quantitative test with the construction of random mutant libraries based on the sequence of trunkamide. Double and quadruple mutant subsets of a projected library of 3.2 million derivatives were expressed and chemically analyzed to examine the amino acid selectivity in the different positions of the compound. From these subsets, 58% of the unique mutant genes were produced *in vivo*. Libraries of 10^8 compounds were made, and by extrapolation from these experiments, a minimum of 10^6 compounds can be produced from a single cyanobactin pathway. [62]

Although these experiments demonstrated that the cyanobactin pathway has a high substrate tolerance for proteinogenic amino acids, this does impose a limitation on the chemical diversity on cyanobactins. Thus, we posed the question of whether it might be possible to incorporate functionalities beyond the 20 proteinogenic amino acids. Incorporation of non-proteinogenic amino acids are often hallmarks of non-ribosomal peptides, among which are some of the most bioactive natural products known. [64] NRPS pathways are widespread, but are extraordinarily difficult to engineer. [65,66]Combining multistep, substrate-tolerant, RiPP pathways with the structural complexity of NRPS therefore would be an unprecedented and important advance in the development of natural products as drugs. In Chapter 4, we describe the successful incorporation of unnatural amino acids (UAAs) into cyanobactins. Prior to this report, the incorporation of

UAAs has only been done in proteins, [67] as well as in lantipeptides *in vitro*. [68]

References

1 Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta* **1830**, 3670-3695, doi:10.1016/j.bbagen.2013.02.008 (2013).

2 Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. & Prinsep, M. R. Marine natural products. *Nat Prod Rep* **30**, 237-323, doi:10.1039/c2np20112g (2013).

3 Martins, A., Vieira, H., Gaspar, H. & Santos, S. Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Mar Drugs* **12**, 1066-1101, doi:10.3390/md12021066 (2014).

4 Molinski, T. F., Dalisay, D. S., Lievens, S. L. & Saludes, J. P. Drug development from marine natural products. *Nat Rev Drug Discov* **8**, 69-85, doi:10.1038/nrd2487 (2009).

5 Piel, J. Metabolites from symbiotic bacteria. *Nat Prod Rep* **26**, 338-362, doi:10.1039/b703499g (2009).

6 Florez, L. V., Biedermann, P. H., Engl, T. & Kaltenpoth, M. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat Prod Rep*, doi:10.1039/c5np00010f (2015).

7 Schmidt, E. W. The secret to a successful relationship: lasting chemistry between ascidians and their symbiotic bacteria. *Invertebrate Biology : A Quarterly Journal of the American Microscopical Society and the Division of Invertebrate Zoology/ASZ* **134**, 88-102, doi:10.1111/ivb.12071 (2015).

8 McGovern, T. M. & Hellberg, M. E. Cryptic species, cryptic endosymbionts, and geographical variation in chemical defences in the bryozoan Bugula neritina. *Mol Ecol* **12**, 1207-1215 (2003).

9 Trindade-Silva, A. E., Lim-Fong, G. E., Sharp, K. H. & Haygood, M. G. Bryostatins: biological context and biotechnological prospects. *Curr Opin Biotech* **21**, 834-842, doi:10.1016/j.copbio.2010.09.018 (2010).

10 Woollacott, R. M. Association of bacteria with bryozoan larvae. *Mar Biol* **65**, 155-158, doi:10.1007/BF00397080 (1981).

11 Haygood, M. G. & Davidson, S. K. Small-subunit rRNA genes and in situ hybridization with oligonucleotides specific for the bacterial symbionts in the

larvae of the bryozoan Bugula neritina and proposal of "Candidatus endobugula sertula". *Appl Environ Microbiol* **63**, 4612-4616 (1997).

12 Davidson, S. K., Allen, S. W., Lim, G. E., Anderson, C. M. & Haygood, M. G. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "Candidatus Endobugula sertula" of the bryozoan Bugula neritina. *Appl Environ Microbiol* **67**, 4531-4537 (2001).

13 Sudek, S. *et al.* Identification of the putative bryostatin polyketide synthase gene cluster from "Candidatus Endobugula sertula", the uncultivated microbial symbiont of the marine bryozoan Bugula neritina. *J Nat Prod* **70**, 67-74, doi:10.1021/np060361d (2007).

14 Lopanik, N. B. *et al.* In vivo and in vitro trans-acylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. *Chem Biol* **15**, 1175-1186, doi:10.1016/j.chembiol.2008.09.013 (2008).

15 Pettit, G. R. in *Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products* Vol. 57 *Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products* (eds W. Herz, G. W. Kirby, W. Steglich, & Ch Tamm) Ch. 3, 153-195 (Springer Vienna, 1991).

16 Davidson, S. K. & Haygood, M. G. Identification of sibling species of the bryozoan Bugula neritina that produce different anticancer bryostatins and harbor distinct strains of the bacterial symbiont "Candidatus Endobugula sertula". *Biol Bull* **196**, 273-280 (1999).

17 Fehlauer-Ale, K. H. *et al.* Cryptic species in the cosmopolitan Bugula neritina complex (Bryozoa, Cheilostomata). *Zool Scr* **43**, 193-205, doi:10.1111/zsc.12042 (2014).

18 Linneman, J., Paulus, D., Lim-Fong, G. & Lopanik, N. B. Latitudinal variation of a defensive symbiosis in the Bugula neritina (Bryozoa) sibling species complex. *PLoS One* **9**, e108783, doi:10.1371/journal.pone.0108783 (2014).

19 Donia, M. S. & Schmidt, E. W. in *Comprehensive Natural Products II* (eds Mander Lew, & Liu Hung-Wen) Vol. 2, 539-558 (Elsevier, 2010).

20 Schmidt, E. W. *et al.* Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proc Natl Acad Sci U S A* **102**, 7315-7320, doi:10.1073/pnas.0501424102 (2005).

21 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108-160, doi:10.1039/c2np20085f (2013).

22 Donia, M. S. *et al.* Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* **2**, 729-735, doi:10.1038/nchembio829 (2006).

23 Donia, M. S., Ravel, J. & Schmidt, E. W. A global assembly line for cyanobactins. *Nat Chem Biol* **4**, 341-343, doi:10.1038/nchembio.84 (2008).

24 Ireland, C., Scheuer,P.J. Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate. *J Am Chem Soc* **102**, 5688-5691 (1980).

25 Ireland, C. M., Durso, A.R., Newman, R.A, Hacker, M.P. Antineoplastic cyclic peptides from the marine tunicate *Lissoclinum patella*. *J Org Chem* **47**, 1807-1811 (1982).

26 Carroll, A. R. *et al.* Patellins 1-6 and trunkamide A: novel cyclic hexa-, hepta- and octa-peptides from colonial ascidians, Lissoclinum sp. *Aust J Chem* **49**, 659-667 (1996).

27 Degnan, B. M. *et al.* New cyclic peptides with cytotoxic activity from the ascidian Lissoclinum patella. *J Med Chem* **32**, 1349-1354 (1989).

28 Donia, M. S., Ruffner, D. E., Cao, S. & Schmidt, E. W. Accessing the hidden majority of marine natural products through metagenomics. *Chembiochem* **12**, 1230-1236, doi:10.1002/cbic.201000780 (2011).

29 Donia, M. S., Fricke, W. F., Ravel, J. & Schmidt, E. W. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS One* **6**, e17897, doi:10.1371/journal.pone.0017897 (2011).

30 Schmidt, E. W., Donia, M. S., McIntosh, J. A., Fricke, W. F. & Ravel, J. Origin and variation of tunicate secondary metabolites. *J Nat Prod* **75**, 295-304, doi:10.1021/np200665k (2012).

31 Kwan, J. C. *et al.* Host control of symbiont natural product chemistry in cryptic populations of the tunicate Lissoclinum patella. *PLoS One* **9**, e95850, doi:10.1371/journal.pone.0095850 (2014).

32 Winder, P. L., Pomponi, S. A. & Wright, A. E. Natural products from the Lithistida: a review of the literature since 2000. *Mar Drugs* **9**, 2643-2682, doi:10.3390/md9122643 (2011).

33 Fusetani, N. & Matsunaga, S. Bioactive sponge peptides. *Chem Rev* **93**, 1793-1806, doi:10.1021/cr00021a007 (1993).

34 Bewley, C. A. & Faulkner, D. J. Lithistid sponges: star performers or hosts to the stars. *Angew Chem Int Edit* **37**, 2162-2178, doi:10.1002/(SICI)1521-3773(19980904)37:16<2162::AID-ANIE2162>3.0.CO;2-2 (1998).

35 Schmidt, E. W., Bewley, C. A. & Faulkner, D. J. Theopalauamide, a bicyclic glycopeptide from filamentous bacterial symbionts of the lithistid sponge Theonella swinhoei from Palau and Mozambique. *J Org Chem* **63**, 1254-1258, doi:10.1021/jo9718455 (1998).

36 Bewley, C. A., Holland, N. D. & Faulkner, D. J. Two classes of metabolites fromTheonella swinhoei are localized in distinct populations of bacterial symbionts. *Experientia* **52**, 716-722, doi:10.1007/bf01925581 (1996).

37 Schirmer, A. *et al.* Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge Discodermia dissoluta. *Appl Environ Microbiol* **71**, 4840-4849, doi:10.1128/AEM.71.8.4840-4849.2005 (2005).

38 Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proc Natl Acad Sci U S A* **101**, 16222-16227, doi:10.1073/pnas.0405976101 (2004).

39 Freeman, M. F. *et al.* Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* **338**, 387-390, doi:10.1126/science.1226121 (2012).

40 Schmidt, E. W., Obraztsova, A. Y., Davidson, S. K., Faulkner, D. J. & Haygood, M. G. Identification of the antifungal peptide-containing symbiont of the marine sponge Theonella swinhoei as a novel δ-proteobacterium, Candidatus Entotheonella palauensis. *Mar Biol* **136**, 969-977, doi:10.1007/s002270000273 (2000).

41 Wilson, M. C. *et al.* An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58-62, doi:10.1038/nature12959 (2014).

42 Wakimoto, T. *et al.* Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat Chem Biol* **10**, 648-655, doi:10.1038/nchembio.1573 (2014).

43 Kwan, J. C. *et al.* Genome streamlining and chemical defense in a coral reef symbiosis. *Proc Natl Acad Sci U S A* **109**, 20655-20660, doi:10.1073/pnas.1213820109 (2012).

44 McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**, 13-26, doi:10.1038/nrmicro2670 (2012).

45 Morrison, K. C. & Hergenrother, P. J. Natural products as starting points for the synthesis of complex and diverse compounds. *Nat Prod Rep* **31**, 6-14, doi:10.1039/c3np70063a (2014).

46 Wender, P. A. & Schrier, A. J. Total synthesis of Bryostatin 9. *J Am Chem Soc* **133**, 9228-9231, doi:10.1021/ja203034k (2011).

47 Lu, Y., Woo, S. K. & Krische, M. J. Total Synthesis of Bryostatin 7 via C-C Bond Forming Hydrogenation: Merged Redox-Construction Events for Synthetic Efficiency. *J Am Chem Soc* **133**, 13876-13879, doi:10.1021/ja205673e (2011).

48 Ongley, S. E., Bian, X., Neilan, B. A. & Muller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep* **30**, 1121-1138, doi:10.1039/c3np70034h (2013).

49 Luo, Y., Cobb, R. E. & Zhao, H. Recent advances in natural product discovery. *Curr Opin Biotech* **30**, 230-237, doi:10.1016/j.copbio.2014.09.002 (2014).

50 Leal, M. C. *et al.* Marine microorganism-invertebrate assemblages: perspectives to solve the "supply problem" in the initial steps of drug discovery. *Mar Drugs* **12**, 3929-3952, doi:10.3390/md12073929 (2014).

51 Zhang, H., Boghigian, B. A., Armando, J. & Pfeifer, B. A. Methods and options for the heterologous production of complex natural products. *Nat Prod Rep* **28**, 125-151, doi:10.1039/c0np00037j (2011).

52 Challis, G. L. Engineering Escherichia coli to produce nonribosomal peptide antibiotics. *Nat Chem Biol* **2**, 398-400, doi:10.1038/nchembio0806-398 (2006).

53 Pickens, L. B., Tang, Y. & Chooi, Y. H. Metabolic engineering for the production of natural products. *Ann Rev Chem Biomol Eng* **2**, 211-236, doi:10.1146/annurev-chembioeng-061010-114209 (2011).

54 Sun, L., Zeng, J., Zhang, S., Gladwin, T. & Zhan, J. Effects of exogenous nutrients on polyketide biosynthesis in Escherichia coli. *App Microbiol Biot* **99**, 693-701, doi:10.1007/s00253-014-6212-7 (2015).

55 Pistorino, M. & Pfeifer, B. A. Efficient experimental design and micro-scale medium enhancement of 6-deoxyerythronolide B production through Escherichia coli. *Biotechnol Progr* **25**, 1364-1371, doi:10.1002/btpr.250 (2009).

56 Ryu, Y. G., Butler, M. J., Chater, K. F. & Lee, K. J. Engineering of primary carbohydrate metabolism for increased production of actinorhodin in

Streptomyces coelicolor. *Appl Environ Microbiol* **72**, 7132-7139, doi:10.1128/AEM.01308-06 (2006).

57 Thykaer, J. *et al.* Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster. *Metab Eng* **12**, 455-461, doi:10.1016/j.ymben.2010.05.001 (2010).

58 Kuzuyama, T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotech Bioch* **66**, 1619-1627 (2002).

59 Farmer, W. R. & Liao, J. C. Precursor balancing for metabolic engineering of lycopene production in Escherichia coli. *Biotechnol Prog* **17**, 57-61, doi:10.1021/bp000137t (2001).

60 Pitera, D. J., Paddon, C. J., Newman, J. D. & Keasling, J. D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metab Eng* **9**, 193-207, doi:10.1016/j.ymben.2006.11.002 (2007).

61 Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotech* **21**, 796-802, doi:10.1038/nbt833 (2003).

62 Ruffner, D. E., Schmidt, E. W. & Heemstra, J. R. Assessing the Combinatorial Potential of the RiPP Cyanobactin tru Pathway. *ACS Synth Biol* **4**, 482-492, doi:10.1021/sb500267d (2015).

63 Zabriskie, T. M., Foster, M. P., Stout, T. J., Clardy, J. & Ireland, C. M. Studies on the solution- and solid-state structure of patellin 2. *J Am Chem Soc* **112**, 8080-8084, doi:10.1021/ja00178a035 (1990).

64 Walsh, C. T., O'Brien, R. V. & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem* **52**, 7098-7124, doi:10.1002/anie.201208344 (2013).

65 Williams, G. J. Engineering polyketide synthases and nonribosomal peptide synthetases. *Curr Opin Struc Biol* **23**, 603-612, doi:10.1016/j.sbi.2013.06.012 (2013).

66 Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **3**, 490-504, doi:10.1002/1439-7633(20020603)3:6<490::AID-CBIC490>3.0.CO;2-N (2002).

67 Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Ann Rev Biochem* **79**, 413-444, doi:10.1146/annurev.biochem.052308.105824 (2010).

68 Levengood, M. R., Knerr, P. J., Oman, T. J. & van der Donk, W. A. In vitro mutasynthesis of lantibiotic analogues containing nonproteinogenic amino acids. *J Am Chem Soc* **131**, 12024-12025, doi:10.1021/ja903239s (2009).

CHAPTER 2

SPECIES SPECIFICITY OF SYMBIOSIS AND SECONDARY METABOLISM IN ASCIDIANS

Manuscript reproduced with permission from:

Ma Diarey B Tianero, Jason C Kwan, Thomas P Wyche, Angela P Presson, Michael Koch, Louis R Barrows, Tim S Bugni and Eric W Schmidt. **Species specificity of symbiosis and secondary metabolism in ascidians.** *ISME Journal* 9, 615-628.

Note: My contribution to this paper was in design of the project, collection, processing, analysis of microbiome data, interpretation of results, and writing of the manuscript.

The ISME Journal (2014), 1–14

www.nature.com/ismej

ORIGINAL ARTICLE

Species specificity of symbiosis and secondary metabolism in ascidians

Ma Diarey B Tianero[1], Jason C Kwan[1,2], Thomas P Wyche[2], Angela P Presson[3], Michael Koch[4], Louis R Barrows[4], Tim S Bugni[2] and Eric W Schmidt[1]

[1]*Department of Medicinal Chemistry, L.S. Skaggs Pharmacy Institute, University of Utah, Salt Lake City, UT, USA;* [2]*School of Pharmacy, University of Wisconsin-Madison, Madison, WI, USA;* [3]*Study Design and Biostatistics Center, Division of Epidemiology, University of Utah, Salt Lake City, UT, USA and* [4]*Department of Pharmacology and Toxicology, L.S. Skaggs Pharmacy Institute, University of Utah, Salt Lake City, UT, USA*

Ascidians contain abundant, diverse secondary metabolites, which are thought to serve a defensive role and which have been applied to drug discovery. It is known that bacteria in symbiosis with ascidians produce several of these metabolites, but very little is known about factors governing these 'chemical symbioses'. To examine this phenomenon across a wide geographical and species scale, we performed bacterial and chemical analyses of 32 different ascidians, mostly from the didemnid family from Florida, Southern California and a broad expanse of the tropical Pacific Ocean. Bacterial diversity analysis showed that ascidian microbiomes are highly diverse, and this diversity does not correlate with geographical location or latitude. Within a subset of species, ascidian microbiomes are also stable over time ($R = -0.037$, P-value $= 0.499$). Ascidian microbiomes and metabolomes contain species-specific and location-specific components. Location-specific bacteria are found in low abundance in the ascidians and mostly represent strains that are widespread. Location-specific metabolites consist largely of lipids, which may reflect differences in water temperature. By contrast, species-specific bacteria are mostly abundant sequenced components of the microbiomes and include secondary metabolite producers as major components. Species-specific chemicals are dominated by secondary metabolites. Together with previous analyses that focused on single ascidian species or symbiont type, these results reveal fundamental properties of secondary metabolic symbiosis. Different ascidian species have established associations with many different bacterial symbionts, including those known to produce toxic chemicals. This implies a strong selection for this property and the independent origin of secondary metabolite-based associations in different ascidian species. The analysis here streamlines the connection of secondary metabolite to producing bacterium, enabling further biological and biotechnological studies.
The ISME Journal advance online publication, 29 August 2014; doi:10.1038/ismej.2014.152

Introduction

Filter-feeding chordates known as ascidians are found in all of the world's oceans (Shenkar and Swalla, 2011; Schmidt *et al.*, 2012). Ascidians are sessile, soft-bodied animals that are vulnerable to predation. Many ascidian species contain abundant, potently toxic secondary metabolites that are implicated in chemical defense. Indeed, there are several examples in which a defensive role has been experimentally assigned to individual secondary metabolites (Paul *et al.*, 1990; Joullie *et al.*, 2003). Strikingly, there are now several examples in which it has been shown that symbiotic bacteria produce potentially defensive metabolites, while the host animals do not (Supplementary Table S8) (Schmidt *et al.*, 2005; Donia *et al.*, 2006, 2008; Rath *et al.*, 2011; Donia *et al.*, 2011b, 2011c; Kwan *et al.*, 2012). Beyond defense, symbiotic metabolites in ascidians are also known to have other roles. For example, mycosporine amino acids are critical to the defense of tropical ascidians against ultraviolet irradiation, but they are produced by symbiotic bacteria (Hirose *et al.*, 2006; Donia *et al.*, 2011a). Thus, symbiotic bacteria contribute critical secondary metabolites that are apparently necessary for the survival of the animals.

In tropical ascidians of the Family *Didemnidae*, *Prochloron* spp. symbiotic bacteria produce a variety of toxic, cyclic peptides known as cyanobactins (Donia *et al.*, 2008), as well as ultraviolet-protectant mycosporine amino acids (Donia *et al.*, 2011a). *Prochloron* seems to be obligate for the host animals, yet horizontal transmission dominates for this group (Hirose and Fukuda, 2006; Hirose *et al.*, 2009; Kojima and Hirose, 2010). Nonetheless, isolated *Prochloron* chemistry is strongly ascidian species specific, implicating a chemical driver in the

Correspondence: EW Schmidt, Department of Medicinal Chemistry, L.S. Skaggs Pharmacy Institute, University of Utah, 307 SKH 30 South 2000 East, Salt Lake City, UT 84112, USA.
E-mail: ews1@utah.edu
Received 5 February 2014; revised 30 April 2014; accepted 2 June 2014

interaction (Donia *et al.*, 2011b). By contrast, certain groups of highly toxic polyketides are synthesized by genome-reduced, obligate intracellular alpha-proteobacteria, *Candidatus* Endolissoclinum faulkneri (Kwan *et al.*, 2012). This interaction has resulted in vertical transmission and co-speciation that has been ongoing for approximately 6–35 million years (Kwan *et al.*, 2012; Kwan and Schmidt, 2013). During that period, the production of toxic secondary metabolism has been maintained while the remainder of the symbiont's genome has degraded. Another proteobacterium, the gamma-proteobacterium *Candidatus* Endoecteinascidia frumentensis, produces potent ecteinascidins within the non-didemnid ascidian, *Ecteinascidia turbinata* (Rath *et al.*, 2011). Similarly, correlative evidence suggests that alpha-proteobacteria may produce polyketides in an Antarctic didemnid ascidian (Riesenfeld *et al.*, 2008). Finally, indirect but compelling evidence implicates symbiotic bacteria in the production of a variety of other ascidian metabolites (Schmidt and Donia, 2010; Xu *et al.*, 2012).

Although symbiotic bacteria have been shown to produce about 7% of known ascidian metabolites, the source of about 93% of described metabolites remains unknown. Additionally, it is probable that most ascidian metabolites remain undescribed. Here, we sought to examine the factors underlying production of secondary metabolites in ascidians globally, with the idea that this would be useful in rapidly identifying new bioactive natural products and in determining the true producers of ascidian compounds. We compared the microbiomes and metabolomes of ascidians from tropical, subtropical and temperate waters. We broadly studied the link between symbiotic bacteria and the presence of secondary metabolites. The long-term outcome of these studies will help establish an understanding of how symbiotic bacterial communities change and evolve across ascidian phylogeny as well as in time and space. Developing this big picture view provides a basis for collection strategies as well as increase our understanding of what factors might affect symbioses and natural products in ascidians.

Materials and methods

Sample collection and processing

Ascidian samples were collected by scuba diving and processed in the field, except for samples from Florida, which were shipped live by a professional collector. Each ascidian sample was divided into fractions that were preserved in RNAlater (Ambion, Austin, TX, USA) for the microbiome analyses and fractions that were frozen for the metabolomic analyses. In this study, we used samples collected from different locations (year of collection): Papua New Guinea (2011); Fiji (2006); Vanuatu (2008); Catalina Island, California (2010 and 2011); and Florida Keys (2011) (see Supplementary Table S1 for sampling details).

18S rRNA gene sequence analysis

Ascidian 18S rRNA genes were amplified from the same genomic DNA used for 16S rRNA gene sequencing. In cases where amplification was challenging, genomic DNA was extracted using an alternative method (Sokolov, 2000). 18S rRNA genes were amplified using combinations of the following: forward primers AscF3 and AscF1, reverse primers AscR1 and AscR3 (Yokobori *et al.*, 2006) and AscF2 and AscR5 (this study) (Supplementary Table S3). The amplicons were submitted for direct sequencing, leading to successful analysis of 17 out of the 32 samples. For the remaining 15, PCR products were cloned into Topo TA Zero blunt vector (Invitrogen, Carlsbad, CA, USA), and individual clones were sequenced. Primers AscF2 and AscR5 were used for samples that showed non-ascidian 18S products after the first round of sequencing (primers used for each sample are defined in Supplementary Table S1). The ascidian sequences obtained were aligned using MUSCLE (Edgar, 2004) and trimmed using Gblocks (Castresana, 2000). The parameters used for Gblocks were the following: minimum number of sequences for a conserved position = 25, minimum number of sequences for a flanking position = 40, minimum number for a contiguous non-conserved position = 8, minimum length of a block = 10, allowed gap position = none. Of the 2013 original positions, 739 (36%) were retained and used for construction of the phylogenetic tree. The 18S tree was constructed using FastTree MP (Price *et al.*, 2010) and visualized using the Interactive Tree of Life server (Letunic and Bork, 2011). Additional ascidian sequences included in the tree were retrieved from GenBank (see Supplementary Figure S1 for accession numbers). 18S sequences from this data set were deposited in Genbank under accession numbers KJ720704–KJ720729, KJ009376 (*L. patella* Fiji), KJ009378 (*L. patella* PNG_11040), KJ009377 (*L. patella* PNG_11033).

DNA extraction and preparation

Metagenomic DNA was extracted using previous methods (Schmidt and Donia, 2009) with modifications (see Supplementary Methods for details). The DNA obtained was further purified through either Genomic Tip (Qiagen, Valencia, CA, USA) or Genomic DNA Clean and Concentrator (Zymo Research, Irvine, CA, USA) kits and quantified using a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). The quality of DNA was checked by amplifying 16S rRNA gene PCR using 24F and 1492R primers (Supplementary Table S3). Samples that passed the test were used for 16S rRNA gene amplification and pyrosequencing at Research and Testing Laboratory (RTL, Lubbock, TX, USA). RTL amplified a portion of the 16S rRNA gene using primer forward primer 939F and reverse primer 1492R (Supplementary Table S3) spanning the V6–V9 hypervariable regions in the bacterial 16S rRNA gene and sequenced the resulting amplicons by 454 (Roche, Branford, CT, USA).

16S rRNA gene sequence analysis

After sequencing, the reads were denoised and checked for chimeras (UCHIME; Edgar *et al.*, 2011). Reads were then quality filtered, and the RTL-generated FASTA files with the corresponding quality scores and mapping files were demultiplexed in the QIIME analysis pipeline (Caporaso *et al.*, 2010b). All ascidian samples analyzed in this study were then joined together for operational taxonomic unit (OTU) picking, taxonomy assignment and further diversity analyses in QIIME. OTUs were assigned at 97% identity, corresponding approximately to a species-level assignment for bacteria. A representative set of UCLUST (Edgar, 2010) picked OTUs were aligned using PyNAST (Caporaso *et al.*, 2010a) and passed through the RDP classifier using the Greengenes reference sequences (version 2_10) (DeSantis *et al.*, 2006; Wang *et al.*, 2007). The BIOME or table format output from this analysis was then used as input for downstream analyses in QIIME and R (McMurdie and Holmes, 2013; Oksanen *et al.*, 2013; R Core Team, 2013). The analysis was done in parallel on Clovr 16S microbial diversity pipeline for comparison (Angiuoli *et al.*, 2011). Alpha and beta diversity analyses were performed within QIIME using *alpha_rarefaction.py* (parameters included the Phylogenetic Diversity (PD), Shannon, Simpson and Observed Species indices) and *beta_diversity_through_plots.py*, respectively. Principal coordinates analysis was then performed using the UniFrac distance matrix (Lozupone *et al.*, 2011) and visualized in R. Hierarchical clustering and heatmaps were done in R. For networking analysis, categories including location and different host categories (see below) were included in the mapping file. G tests for independence were done to test whether samples are more connected within these categories than expected by chance. Visualization of the networks was performed in Cytoscape (Kohl *et al.*, 2011) using the edge-weighted spring-embedded layout. Analysis of similarity (ANOSIM) calculations were performed to compare microbiome diversity across species and within species across time and location (Clarke, 1993; Chapman and Underwood, 1999). ANOSIM is a distribution-free analysis of variance method that relies on calculating a distance metric (we used Bray–Curtis dissimilarity values) and then comparing the difference between within- and between-sample distances using a permutation test approach (we required a minimum of two samples/group and used 100 000 permutations) (Clarke, 1993). A sensitivity analysis was performed using a similar method ('adonis' function) also available in the vegan R package. To compare the diversity measures between location and hosts, we used a permutation sampling test method that was developed specifically for comparing diversity measures (Pallmann *et al.*, 2012). The tests were performed comparing each host or location to the overall mean diversity (GrandMean option) using the mcpHill function in the SimBoot R package for locations and hosts with a minimum of two samples (Pallmann *et al.*, 2012). *P*-values were adjusted for multiple comparisons across groups (comparing each group with the overall mean) and across diversity measures (Shannon and Simpson). These sequence data have been submitted to SRA, under BioProjects accession SAMN02693397–SAMN02693428.

Ultra High Pressure Liquid Chromatography (LC)/high-resolution mass spectrometry (MS) analysis

LC/MS data (see Supplementary Methods for sample preparation) were acquired using a Bruker MaXis electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) mass spectrometer (Bruker, Billerica, MA, USA) coupled with a Waters Acquity UPLC system (Waters, Millford, MA, USA) operated by the Bruker Hystar software (Bruker). A gradient of methanol and H_2O (containing 0.1% formic acid) was used with a flow rate of $0.3\,ml\,min^{-1}$ on a RP C18 column (Phenomenex Kinetex 2.6 μm, 2.1×100 mm, Phenomenex, Torrance, CA, USA). The gradient went from 10% methanol/90% H_2O to 97% methanol/3% H_2O in 12 min, followed by 97% methanol/3% H_2O held for 3.5 min. Full-scan mass spectra (*m/z* 150–1550) were measured in the positive ESI mode. The mass spectrometer was operated using previously published parameters (Hou *et al.*, 2012). Tune mix (Agilent Technologies, Santa Clara, CA, USA; ESI-L low concentration) was introduced through a divert valve at the end of each chromatographic run for automatic internal calibration. Data-dependent MS/MS were acquired as previously published (Hou *et al.*, 2012).

LC/MS data processing and compound identification

LC/MS data was bucketed and Principal Component Analysis (PCA) was performed using Bruker ProfileAnalysis 2.0. The following parameters were applied to find molecular features: S/N threshold, 5; correlation coefficient threshold, 0.7; minimum compound length, 10 spectra; and smoothing width, 1. Bucket generation was performed in the retention time range 60–930 s and in a mass range from *m/z* 200 to 1500. All other bucket generation parameters were the same as previously published (Hou *et al.*, 2012). For further PCA analysis, the buckets obtained were filtered to intensities $>1.0e4$ as the threshold determined after inspection of individual chromatograms. PCA was then performed in R using the *prcomp()* function. The PCA biplot generated was used as a preliminary guide for identification of extract components that are correlated with certain clustering patterns of ascidian samples. For graphing and further clustering, principal components were chosen based on their eigenvalues or based on their contribution to the cumulative variance ($>80\%$). Hierarchical clustering was performed in R using the Euclidean

distance between the extracted values. For compound identification, mzxml files from LC/MS were processed in the MzMine software (Pluskal *et al.*, 2010).

Results

Phylogeny and diversity of ascidian samples

We examined the microbiomes and chemistry of 32 ascidian samples collected from four locations representing the temperate (Catalina Island, California), subtropical (Florida Keys), and tropical (Papua New Guinea, Fiji and Vanuatu) ocean regions at different times during the period 2006–2011 (Figure 1). We chose samples from the Family *Didemnidae* plus a few other representative sets from other ascidian families. Phylogenetic analysis using the 18S rRNA gene sequences confirmed that most of the samples are didemnid ascidians (21 samples) from the order Aplousobranchia (26 samples). Representative samples from the order Phlebobranchia (three individual samples) and Stolidobranchia (three samples) were also obtained (Supplementary Figure S1). Samples from the order Aplousobranchia include *Eudistoma* sp. (2 individuals), *Cystodytes* sp. (3 individuals), *Lissoclinum* sp. (9 individuals), *Trididemnum* sp. (2 individuals) and *Didemnum* sp. (10 individuals). Stolidobranchia samples include *Botryllus* (1), *Styela* (1) and *Pyura* (1) spp., while *Ecteinascidia* (2) and *Ascidia* (1) spp. represent the order Phlebobranchia.

In the course of sequencing 18S rRNA genes, we were also able to identify other eukaryotes that were associated with the ascidian tissues. These were obtained from many trials using different sets of 18S primers (Supplementary Table S3, see Methods for details). For example, we identified the crustacean species *Notodelphys prasina*, which were consistently associated with the *Cystodytes* sp. samples. Indeed, these crustaceans have been reported as seasonal parasites of *Cystodytes* sp. (Monniot *et al.*, 1991). Nematodes, apicomplexans, copepods and other uncultured eukaryotes are also among the organisms identified from the 18S rRNA gene sequencing (Supplementary Table S7).

Bacterial diversity

Using barcoded 16S rRNA gene sequencing (Roche 454), we obtained an average of 6784 sequences per sample, with a total of 217 703 reads from all ascidian samples. With an OTU defined as >97% 16S rRNA gene sequence identity, we observed a total of 3892 OTUs (10–804 OTUs per sample), indicating a rich and diverse bacterial population (Supplementary Table S2). Using the Shannon index and the other diversity measures, alpha diversity was compared according to variables such as location, ascidian host species and season (Figures 2a and b; Supplementary Table S4). We had hypothesized that there may be a difference in diversity by latitude and that this might in part explain the more abundant secondary metabolism ascribed to tropical ascidians. However, from the rarefaction curves there was no apparent correlation between the diversity of the microbiomes and geographical location (Figure 2a). We further confirmed this by showing that there was no statistical difference in the diversity between locations or latitudinal regions (CA, FL, PNG/FJ/VN; mcpHill analysis using the Shannon index: $P = 0.950–0.990$, Supplementary Figure S9). We also compared the diversity between host species in groups where we had at least two samples per species. There was also no statistical difference in diversity between the other

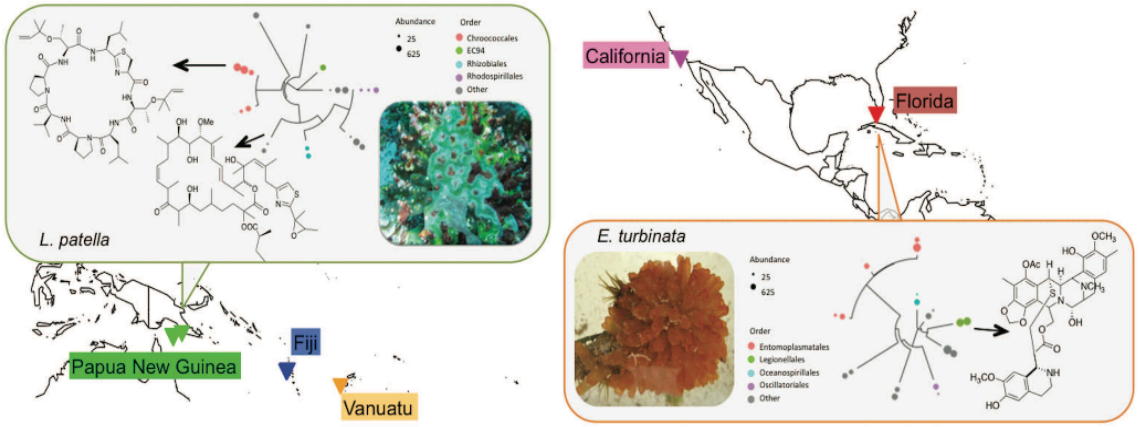


Figure 1 Collection sites and conception of this study. Samples were obtained from a wide variety of geographical locations (triangles). A subset of samples contained natural products where the producing symbiotic bacteria were already known (boxes). Also shown in boxes are phylogenetic trees of the top-10 bacteria found in each sample (see Supplementary Figure S2C for details). Within these top 10, the natural product-producing symbiont is a major representative. This information was compared with metabolomic and microbiome analysis from ascidians where the natural products and producing organisms are not known, with the purpose of understanding the global relationship between natural products and symbiotic bacteria in ascidians.

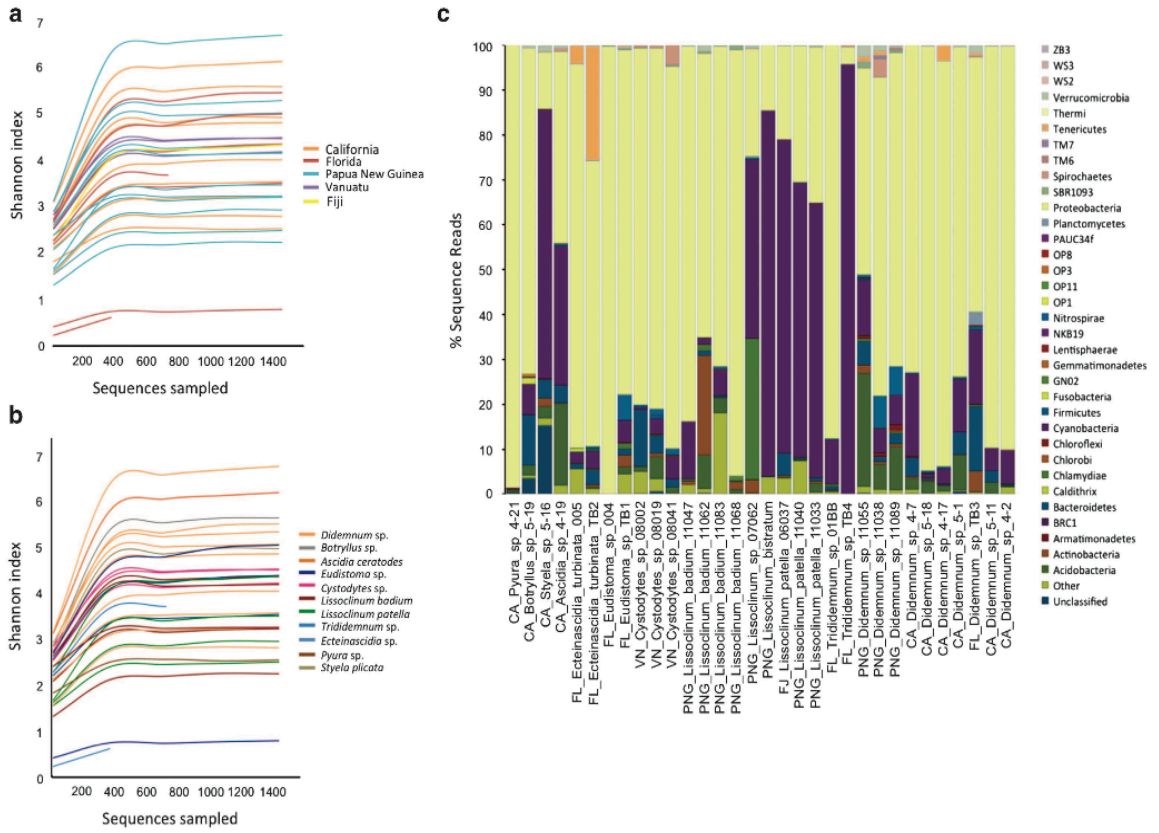


Figure 2 Alpha diversity of ascidian microbiomes assessed using Shannon diversity index showing the lack of correlation with location (**a**) and host ascidian species (**b**). (**c**) Bacterial composition of ascidian samples at the phylum level.

species ($P = 0.600$–1.000, Supplementary Figure S9), except for a group of *Didemnum* sp. from Papua New Guinea (*Didemnum* sp.2, Supplementary Figure S9), which had a higher diversity compared with the rest of the groups ($P = 0.008$). Results were similar for the Simpson diversity analysis. Two individual samples from this data set exhibited distinctly low diversity: *Eudistoma* sp. (41 OTUs of 13886 reads) and *Tridemnum* sp. (10 OTUs of 572 reads). More samples would be needed to determine whether these species are actually less diverse.

Host specificity of ascidian microbiomes

We aimed to determine which factors (host phylogeny, species identity, location, time/water temperature) were related to microbiome composition in ascidians. We first used two types of cluster analyses: networks (based upon OTU) and UniFrac (based upon phylogeny). A significant clustering by host species was observed from the network-based analysis (Figure 3a) ($G = 4.990$, $P = 0.010$). This was clearly supported by the dendrogram subclusters from the weighted Unifrac analysis in cases where we had multiple samples of the same hosts (Figure 3b). We further analyzed the differences in the community structure of the microbiomes between host species and found a statistically significant difference in the microbiomes across species collected in the same location at the same time (11 samples from 3 species were compared, ANOSIM: $R = 0.762$, $P = 0.002$, Supplementary Table S9). These results indicated that a major component of the ascidian microbiome was host specific and did not merely result from transient interactions with seawater or other environmental factors.

We further examined host specificity by testing microbiome stability over time and space. We compared a subset of ascidians from California that were collected in the spring (CA_*Didemnum*_sp_4.2 and CA_*Didemnum*_sp_4.17, CA_*Didemnum*_sp_4.7) and fall (CA_*Didemnum*_sp_5.1, CA_*Didemnum*_sp_5.11, and CA_*Didemnum*_sp_5.18) of 2011. There were no significant seasonal differences in the two sample sets (ANOSIM: $R = -0.037$, $P = 0.490$), suggesting the existence of a stable and host-specific microbiome over time. The lack of a change in the microbiome with time and space is further supported by the analysis of *L. patella* samples from the tropical Pacific. These samples were collected at widely different times (different years) and places, and yet they consistently clustered with a similar

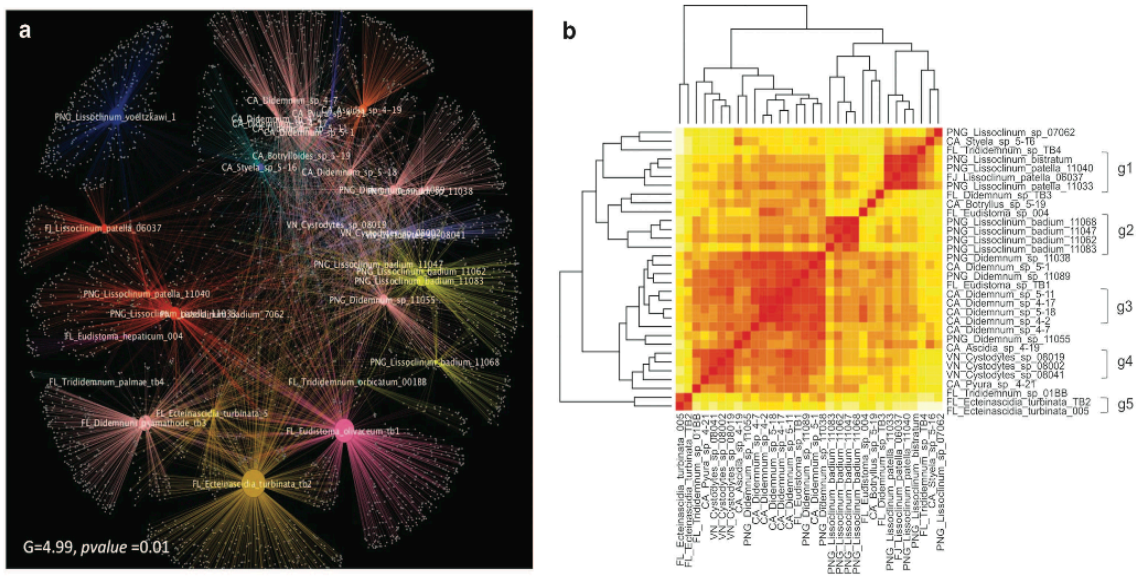


Figure 3 Host specificity of ascidian microbiomes assessed by UniFrac and network-based analyses. (**a**) Cytoscape network of ascidian microbiomes showing a significant clustering by host species. Categories are reflected in the color scheme, where identical species (defined in map 1 of Supplementary Table S5) have identical colors. (**b**) Unifrac-based hierarchical clustering of ascidians plotted in two symmetrical dimensions. Groups at the right most portion indicate the grouping used for the analysis of species-specific components of the microbiomes.

microbiome across time and space and grouped differently than other species collected at the same locations (Figure 3b). This finding is also consistent with recent reports of ascidian microbiomes, in which some of the same ascidian species we report here were previously examined in other locations (Erwin *et al*., 2013; Dishaw *et al*., 2014). Consistently similar bacteria are found across time and space in these studies, in an ascidian host-specific manner.

If the microbiome is intrinsic to the seawater rather than to ascidians, one would expect that ascidians sampled at a particular time and place would have similar microbiomes. Across samples from different ascidian species collected in California or Papua New Guinea, there was a statistically significant difference in microbiome types (CA samples $P=0.005$, PNG samples $P=0.04$). Across samples within a single ascidian species collected at these locations, there was no statistically significant difference in microbiome types ($P>0.05$), as expected (Supplementary Table S9). Taken together with the analysis described above, these results indicate that there is an ascidian species-specific microbiome, which appears to be stable across time and space, within the limits of samples studied so far. Below, we show that these species-specific bacterial sequences represent the most abundant (usually top 10) sequences present in the samples.

Geographical component of ascidian microbiomes

In addition to the host-specific signal, we also observed a significant grouping of the ascidian microbiomes by geographical location or latitudinal region (Figure 4a; $G=6.22$, $P=0.006$). This at first seemed contradictory to the findings of host specificity, prompting us to perform further analysis. In so doing, we found that the location-specific component of the microbiome differed from that found in the host-specific component. As further described below, the location-specific component includes the low-abundance sequences found in the microbiomes. It likely reflects the environment of the ascidians, including the seawater in which they live.

We first hypothesized that the location-specific component was due to the overlap between the ascidian species and location in our sample set, as some species were all collected from the same location (for example, *Cystodytes sp*., *L. badium*, *E. turbinata*). As an alternative hypothesis, because we used the 18S rRNA gene as the marker for phylogeny, we considered that the host speciation might not be optimally resolved. To address these possibilities, we repeated the analyses by defining host species at different levels (Supplementary Table S5) according to their position in the 18S phylogenetic tree and by excluding the ascidian groups that were only collected from a single location (Supplementary Table S5, Figure 4b). Despite some statistical differences (changes in G tests), the geographical signals for the microbiomes were maintained, disproving the hypotheses and indicating a likely geographical component of ascidian microbiomes. This was further examined by determining which microbial groups were generally shared by

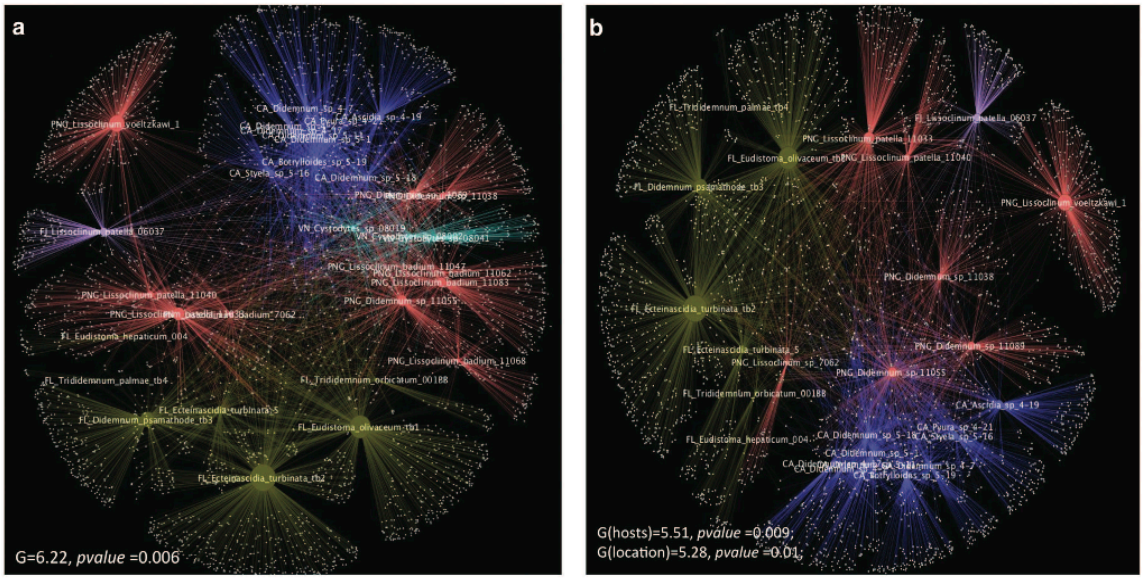


Figure 4 Cytoscape networks of ascidians evaluated using different categories of hosts and location. (**a**) Significant clustering of ascidians according to location, where each color indicates a location. (**b**) Ascidians are still clustered by location even after redefining the host species grouping at different levels and excluding ascidian groups that had only been collected at the same location. The grouping shown here is defined in map 3 of Supplementary Table S5.

most samples from those locations regardless of species. The geographical component was comprised of relatively rare sequences in the ascidian microbiome (described below). These sequences mostly represented relatively commonly sequenced bacterial strains found in seawater, and thus they may, at least in part, represent a component of the microbiome derived from local seawater. Many of these strains are the same as those found in other ascidian microbiomes, so that even if they are derived from the environment they are still largely associated with ascidians. This might be due to physical factors (filtration particle size; digestion efficiency in ascidians; and so on) or to biological factors, such as an actual symbiotic association between the organisms.

Bacterial composition of ascidians

We identified 18 known bacterial phyla, 16 candidate bacterial phyla and an unclassified bacterial group comprising 285 OTUs (Figure 2c). The most abundant phylum is *Proteobacteria* (2074 OTUs), comprising more than half of the total number of OTUs from all ascidian samples. *Proteobacteria* is the most abundant phylum in all samples with the exceptions of *L. patella*, *L. bistratum* and *Styela plicata*, in which *Cyanobacteria* is the most abundant group. We also detected other phyla, including *Acidobacteria*, *Nitrospira and Chloroflexi*, in relatively minor amounts in the samples, as well as other rare and minor (<1% of total sequences) candidate phyla, such as GN02, OP3, and PAUC34f (Figure 2c).

Below the phylum level, we observed little overlap between the kinds of bacteria that were specific to geographical location and those that were specific to host species. Bacteria associated with geographical locations were found in low abundance within animals, while those specific to a species were found in much greater abundance. This was extremely important for the natural products connection to symbiosis, as wherever we knew which bacteria produced ascidian natural products, they were among the most abundant (top-10 strains in terms of sequence abundance) in the animal (Figure 1; Supplementary Figure S2C). In other words, there was a strong connection between abundant symbionts and specific natural products.

Using the analyses employed, we were able to discern individual bacterial sequences responsible for location specificity (perhaps, environmentally derived sequences) from those that are host specific, which might be maintained in symbiosis with the hosts (Supplementary Table S6). Where the secondary metabolite producer was known in an ascidian species, this was among the major and species-specific strains found in that ascidian. For example, *E. turbinata* samples (Figure 3b, group 5) have two major, species-specific *Gammaproteobacteria*, one from the order *Legionellanes* (13–39% of sequenced reads, OTU 3550) and the other from an unclassified group (23–35% of sequenced reads, OTU 1589). The former is identical to *Candidatus* Endoecteinascidia frumentensis, which produces the potent anticancer metabolites, ecteinascidins within the ascidian *E. turbinata* (Rath *et al.*, 2011). In another secondary

metabolite-relevant example, *L. patella* from Fiji and Papua New Guinea (Figure 3b; group 1) differed from the other ascidians in the study by the high abundance of *Prochloron* spp. (58–80%). *Prochloron* spp. produce the family of cytotoxic cyclic peptides, cyanobactins, that are abundant, species-specific chemicals found in these ascidians (Schmidt *et al.*, 2005). In two *L. patella* samples, FJ_06037 and PNG_11033, we were also able to identify a species-specific population of *Alphaproteobacteria* (OTU 932), which was the patellazole producer *Ca.* E. faulkneri. The presence of these known secondary metabolites was individually confirmed in each of these samples.

Even in samples where the secondary metabolite producers are unknown, host-specific bacterial sequences could be clearly discerned. For example, sequences that were species specific to California *Didemnum* sp. (see Figure 3b, group 3) include *Gammaproteobacteria* from the family *Pseudomonadaceae* and a *Deltaproteobacteria* from the genus *Bdellovibrio*.

L. badium microbiomes (Figure 3b, group 2) shared a major *Deltaproteobacteria* component from the order Sva0853 (16–64% of sequenced reads), an uncultivated order that was previously reported from seawater and coldwater sediments (Shi *et al.*, 2011). The sequence itself was unique and represented an ascidian-specific OTU. Another group of *Deltaproteobacteria* in *L. badium* samples were the *Desulfobacteria* from the genus *Nitrospina* (Supplementary Figure S2; Supplementary Table S6). This is consistent with the recent report on the microbiome of ascidians from the Great Barrier Reef, where *Nitrospina* spp. were also found in high abundance in *L. badium* (Erwin *et al.*, 2013). *Nitrospina* spp. are nitrite oxidizers and are proposed to have key roles in nitrogen fixing and cycling within the ascidian tunics (Erwin *et al.*, 2013; Lucker *et al.*, 2013). Another group that was slightly enriched in *L. badium* samples specifically in PNG_11062 (22%) is the *Acidimicrobiales* class of actinobacteria, a group that was previously described as a sponge-specific bacterial group (Simister *et al.*, 2012). All the analyses for identifying the most abundant species-specific bacteria were further confirmed by Metastats analyses for the statistical identification of differentially abundance OTUs (data not shown).

By contrast, the location-specific component of ascidians represented sequences that are relatively common in the environment. Examples include OTU 3071, a *Gammaproteobacterium* from the order *Chromatiales* (average 0.16% in samples), and OTU 116, an *Alteromonas* sp. (average 0.10% in samples), both from California samples. In samples from Florida, unclassified *Gammaproteobacteria* (OTU 701, 0.17%), *Vibrio* spp. (OTU 766, 0.11%) and *Oceanospirillales* spp. (OTU 1931, 0.38%) are among the groups that are shared by the ascidians regardless of species.

Other notable bacteria found in sequenced samples

A subset of strains that might be of interest to the ascidian natural products community is described in this section. A nitrogen-fixing bacterium, *Mesorhizobium* sp., found in lower abundance in *L. badium*, was enriched in *Cystodytes* sp. samples (8–21%). This group has been reported from *Cystodytes* sp. before and was confirmed by BLAST analysis to the database sequences (Martinez-Garcia *et al.*, 2007). Marine *Mesorhizobium* spp. have been shown to produce quorum-sensing molecules (Krick *et al.*, 2007). The remainder of the *Cystodytes* sp. bacterial associates are distinct from those found in other ascidians. As an example, *Cystodytes* sp. samples contain *Microbulbifer* sp., a group known for their ability to degrade complex carbohydrates (Peng *et al.*, 2006), at a relatively higher abundance (3–7% of sequences) as compared with other samples (0–0.3% of sequences).

The lowest diversity of sequences was found in *Eudistoma* sp. 004, with 92% (out of 13 886 sequences) comprising a species from the *Endozoiconomonas* family of the order *Oceanospirillales* (Supplementary Figure S2B). This group of bacteria has been reported in association with other marine organisms, including gorgonians, sea anemones and corals (Bayer *et al.*, 2013; Correa *et al.*, 2013; Lema *et al.*, 2013). Interestingly, in most cases where they were present, they were also found to be dominant in the microbiome, leading to speculation about their possible symbiotic roles (Schuett *et al.*, 2007).

Sequences from Florida *Didemnum* sp. TB3 were 22% *Myxococcales* (Supplementary Figure S2A). Myxobacteria have rarely been found in marine habitats (Reichenbach, 1999). Myxobacteria are prolific producers of natural products, and the first marine obligate halophilic myxobacteria-derived natural products were only reported very recently (Felder *et al.*, 2013). *Didemnum* sp. TB3 also contained *Flavobacteria* from the genus *Tenacibaculum* (12%), a group that had been isolated from sponges and marine algae. Members of this group have also been reported as producers of hydroxamate siderophores (Fujita *et al.*, 2013).

The microbiome of the solitary ascidian *S. plicata* was recently described from a Mediterranean sample (Erwin *et al.*, 2013). We analyzed one sample of this species that we collected in California. Comparison of the composition at the phylum level of these samples showed that some of the major phyla such as *Proteobacteria* and *Bacteriodetes* are maintained, while some are not, perhaps owing to their ability to establish varied interactions upon introduction to new environments (Erwin *et al.*, 2013). Our sample contained high levels of sequences assigned to chloroplasts, which could indicate an algal origin, while the Mediterranean samples harbored high levels of *Planctomycetes*.

Bioactivity profile of ascidian chemical extracts
We assayed the chemical extracts of the ascidians against laboratory strains of microorganisms to test whether there are distinct differences in the potency and toxicity of metabolites from the tropical regions compared with the temperate regions. We tested for activities against different microbes, including *Staphylococcus aureus*, *Escherichia coli* and *Mycobacterium tuberculosis*. We also tested for cytotoxicity against a human CEM-TART cell line and for antifungal activity against *Candida albicans*. There is a clear geographical distribution of biological activity and potency in these samples, which reflects the results of the secondary metabolite analysis performed as described below (Supplementary Figure S8). In these assays, the most potent extracts were those of tropical ascidians while the least potent are those of temperate ascidians. Importantly, the toxicity of these samples was not correlated to diversity of the underlying microbiome but was independent of this variable. This indicates that strains with more diverse microbiomes do not produce more toxic compounds and serves to further reinforce the role of single, talented bacterial symbionts in the production of toxins (see Discussion).

Chemistry of ascidians
PCA and hierarchical clustering of the LC/MS profiles showed that ascidians generally grouped by geographical location rather than by species (Figure 5). Three major clusters are formed comprising California, Florida and a combination of Papua New Guinea and Vanuatu in a single cluster. We hypothesized that this pattern was due to lipids in the extracts, which likely vary, because different temperatures would require different components to maintain membrane fluidity (Parrish, 2013). Alternatively, it might be due to other effects, such as available food sources, although this seemed less likely due to the large geographical sampling range in the tropical Pacific. Using the PCA biplot of ascidians and compounds (Supplementary Figure S5A), we performed an initial identification of the extract components that led to the geographical clustering of the ascidians. The chromatographic peaks were assigned using Lipid Maps (Fahy *et al.*, 2007), and the resulting peaks assigned to lipids were then applied to PCA analysis. By using this putative lipid subset, we observed the same clustering by geographical location in the PCA (Supplementary Figure S5B). Conversely, we also subtracted the Lipid Maps-assigned peaks from the whole mzxml files and subjected the remaining data to PCA. Upon removing putative lipids from the analysis, the clustering according to geographical location was no longer observed. Instead, clustering was better correlated with animals species (see below).

A weakness of all MS-based approaches is that it is difficult to be sure that an ion has been accurately assigned. We therefore used further methods to determine whether ions preliminarily assigned as lipids did indeed truly represent lipids. We selected representative extracts from each PCA cluster and analyzed them using a validated metabolomics method, in which fragmentation of lipids (MS/MS profile) was compared with a large database of lipid fragment patterns. Through this analysis, we confirmed that the major components responsible for variance in the PCA were lipids. These included phospho-glycerolipids, glycerolipids and some straight chain fatty acids (Figure 5, Supplementary Figure S6).

We further analyzed the ascidian chemistry between samples within the same geographical location. We focused on Papua New Guinea samples where we have groups containing replicates of the same species and have characterized the natural products chemistry (Donia *et al.*, 2011b). By PCA, we found that the samples cluster by host species. For example, *L. patella* samples (one from Fiji and two from Papua New Guinea) formed a separate cluster that was strongly driven by secondary metabolites. These metabolites include members of the cyanobactin family: patellins 1, 2, and 5 (Carroll *et al.*, 1996); and the patellazoles (Corley *et al.*, 1988; Zabriskie *et al.*, 1988) (Figure 6, Supplementary Figure S7). Similarly, *L. badium* samples grouped together and in correlation with varamines (Molinski and Ireland, 1989) and some other unknown metabolites. When we included the *Cystodytes* sp. samples from Vanuatu in this analysis, these also strongly clustered together as expected, except for a single sample that was phylogenetically different than the others (Supplementary Figure S1). The grouping of *Cystodytes* samples was also caused by secondary metabolites, mostly pyridoacridine alkaloids (Marshall and Barrows, 2004), and the variations within the species were due to the abundance of different compound analogs. Specifically, VN_08002 and VN_08041 samples contained major amounts of cystodytins while VN_08019 mainly contained shermilamines in the extract (Cooray *et al.*, 1988; Kobayashi *et al.*, 1991).

In cases where they could be identified, secondary metabolites were also noticeably major components of extracts of other ascidians from Florida, such as the eudistomins (Kobayashi *et al.*, 1984) from *Eudistoma* sp. TB1. The extract of *L. bistratum* also contained major metabolites, including the bistramides and the bistratamides group of compounds (Degnan *et al.*, 1989; Biard *et al.*, 1994; Perez and Faulkner, 2003). However, in most samples, especially those from California, the majority of the components could not be assigned from databases, such as Antimarin and the Dictionary of Natural Products.

Discussion

We set out to determine whether there are any correlations between ascidian bacterial communities and the presence of abundant, bioactive

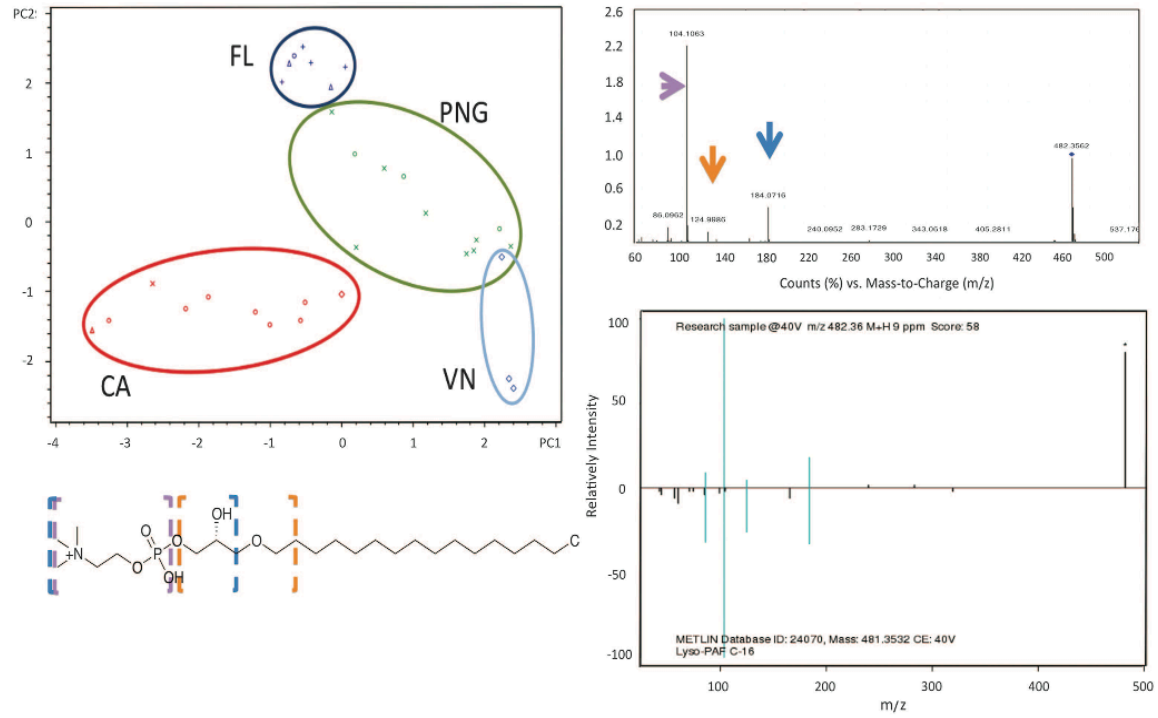


Figure 5 Geographical clustering of ascidian samples due to lipid components. (**a**) PCA plot of ascidians in the first two dimensions showing clustering by geographical location (FL, Florida; PNG, Papua New Guinea; VN, Vanuatu; CA, California). (**b**) Example of how lipids were identified in this study. LC-MS/MS analysis and comparison of fragmentation patterns and accurate mass to lipids in METLIN database allowed assignment of specific fragments (colored arrows) to their complementary molecular components within lipid structures (colored boxes surrounding lipid molecule). Although these fragments do not absolutely confirm lipid identity, they provide high confidence that the compound falls within the lipid family as shown. See Supplementary Figures S5 and S6 for more details of lipid analysis.

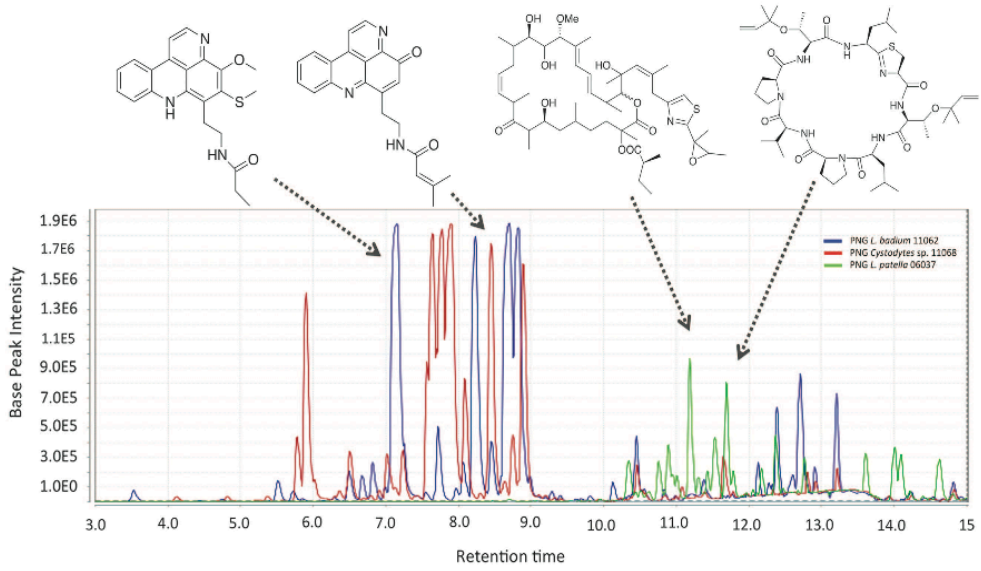


Figure 6 Total ion chromatograms of organic extracts of selected ascidians that are known for secondary metabolite production showing the abundance and host specificity of secondary metabolites in ascidian extracts. The secondary metabolites represent the major compounds in the extracts, and, as shown in Supplementary Figure S7, are highly species specific.

secondary metabolites. By examining phylogenetically diverse ascidians across diverse locations and a large latitudinal gradient, we hoped to differentiate factors related to location and habitat from those related to metabolism. At the level of common indices of overall bacterial diversity, ascidian microbiomes were virtually identically diverse across all sample types and habitats. With a few exceptions, ascidians generally maintain rich and diverse microbiomes in several widely different locations and environments and thus provide untapped sources of microbial diversity. Perhaps owing to their similar lifestyle, this diversity is comparable to that of sponges (Schmitt *et al*., 2012).

Our results show that chemical diversity in ascidians is not reflected at the microbiome diversity level but rather in specific interactions with talented bacterial producers. We propose that, in the tropical regions where there is more pressure for defensive compounds, selection leads to symbiosis with bacteria that produce defensive metabolites, leading to a higher fitness for the associated organisms. In this model, the host species-specific bacteria produce the species-specific secondary metabolites isolated from ascidians. The bacterial producers of ascidian compounds are only known in a few cases, including in several samples described in this study. The sources of by far most ascidian metabolites remain unknown. From this study, it is clear that ascidians do not have universal solutions to this problem and that ascidian species have widely different microbiomes, in which one or two talented producers may be responsible for producing the resulting metabolites.

Ascidian microbiomes were stable over time and space. The majority of these microbes compose the species-specific microbiome, which consisted of more abundant ('top-10') species of bacteria in each sample. These species-specific microbiomes often contained relatively rare types of bacteria. Low-abundance bacteria on the other hand tended to be relatively widespread species, which were strongly correlated with the geography of the samples. Such species may be associated with changes in seawater or environment, and their specific association with ascidians is uncertain.

Several studies have previously examined some of the species of ascidians included in this work, further extending that analysis to several different locations and water temperature regimes. Mediterranean *S. plicata* microbiomes were recently examined in detail (Erwin *et al*., 2013) and revealed key similarities with the California sample studied here. *L. badium* from the Great Barrier Reef (Erwin *et al*., 2013) had a very similar microbiome to samples examined here from Papua New Guinea. The *L. patella* microbiome has been examined by both whole genome sequencing methods in our laboatory using samples from Palau, Solomon Islands, Fiji and Papua New Guinea (Donia *et al*., 2011a, b) and by 16S-based methods using Great Barrier Reef samples (Behrendt *et al*., 2012). Here, we extended these analyses to further 16S-based methods across additional samples, further confirming the species specificity previously reported. Finally, *E. turbinata* has been examined by whole genome sequencing (Rath *et al*., 2011) and by 16S-based analysis (Moss *et al*., 2003; Perez-Matos *et al*., 2007), with deeper sequences obtained here. The comparison here of multiple species across time and space reveals common elements of diversity in ascidian microbiomes.

This study is the first to compare ascidian metabolomes and microbiomes across different species and geographical locations. As with the microbiomes, the metabolomes had location-specific and species-specific components. The metabolites responsible for most of the observed variance in PCA analysis were shown to be lipids. We propose that this difference may be due to the different temperatures at the varied sampling sites ($\sim$10 °C to $\sim$30 °C across the latitudinal gradient) (Parrish, 2013). Once the influence of the lipids were removed from the analysis, or focusing on samples from single regions, a strong species-specific component was revealed that was due largely to secondary metabolism (Figure 6). Thus both the microbiome and the major secondary metabolites were strongly species specific. In the relatively limited sample set studied, the tropical ascidians were highly bioactive, exhibiting potent toxicity to diverse cell types, whereas subtropical (Florida) samples were less toxic, and temperate samples nearly lacked toxicity. The compounds responsible for activity in tropical samples were mostly known compounds, exceptionally potent marine natural products such as patellazoles and pyridoacridines that may serve to defend their hosts. Within the subset of *L. patella*, secondary metabolism is highly correlated to phylogeny of the host even within this single species (Kwan *et al*., 2014). This analysis extends this observation across multiple species, showing that secondary metabolism is strongly related to ascidian phylogeny, and the apparently random distribution of secondary metabolites within a single species previously observed is likely due to lack of appropriate measures of phylogenetic divergence.

In this study, we provide an overview of the key players in the ecology of chemical defense in ascidians: bacteria and secondary metabolites. Bacterial diversity or abundance is not a marker for chemical diversity or the potency of secondary metabolites in ascidians. Instead, ascidians harbor species-specific microbiomes and species-specific secondary metabolites (often, defensive compounds). Where the producer of compounds is known, these producers represent one of the top-10 strains found in ascidian microbiomes. This implies that, in the search for natural products and their bacterial producers, efforts should be primarily focused on the most abundant bacteria in the

microbiota, such that only perhaps 10 strains may be examined to find a producer. The known bacterial producers of ascidian secondary metabolites are phylogenetically diverse, and they are ascidian species-specific and ancient in their distribution (Kwan and Schmidt, 2013). It is remarkable that many different associations with different bacteria have led to defensive chemicals in ascidians, implying a strong selection for this property and the independent origin of defensive associations in different ascidian species.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank James Cox and John Alan Maschek of the University of Utah Metabolomics core facility for assistance and help on the LC-MS/MS analysis. Appropriate permits were obtained for all collections reported in this manuscript. We are grateful to the staff of USC Wrigley Institute and, in particular, USC undergraduate Mariah Gill for aid with obtaining California samples. We thank Chris Ireland and Mary Kay Harper for the *L. voeltzkawi* sample that was obtained through the PNG-ICBG program and the Vanuatu samples that were collected through the NCDDG program. This investigation was supported by the University of Utah Study Design and Biostatistics Center, with funding in part from the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant 8UL1TR000105 (formerly UL1RR025764). This research was funded by NIH GM092009 and GM107557.

References

Angiuoli SV, Matalka M, Gussman A, Galens K, Vangala M, Riley DR *et al.* (2011). CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics* **12**: 356.

Bayer T, Neave MJ, Alsheikh-Hussain A, Aranda M, Yum LK, Mincer T *et al.* (2013). The microbiome of the Red Sea coral *Stylophora pistillata* is dominated by tissue-associated Endozoicomonas bacteria. *Appl Environ Microbiol* **79**: 4759–4762.

Behrendt L, Larkum AW, Trampe E, Norman A, Sorensen SJ, Kuhl M. (2012). Microbial diversity of biofilm communities in microniches associated with the didemnid ascidian *Lissoclinum patella*. *ISME J* **6**: 1222–1237.

Biard JF, Roussakis C, Kornprobst JM, Gouiffes-Barbin D, Verbist JF, Cotelle P *et al.* (1994). Bistramides A, B, C, D, and K: a new class of bioactive cyclic polyethers from *Lissoclinum bistratum*. *J Nat Prod* **57**: 1336–1345.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–267.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Carroll AR, Coll JC, Bourne DJ, MacLeod JK, Zabriskie TM. (1996). Patellins 1-6 and trunkamide A: novel cyclic hexa-, hepta- and octa-peptides from colonial ascidians, *Lissoclinum* sp. *Aust J Chem* **49**: 659–667.

Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.

Chapman MG, Underwood AJ. (1999). Ecological patterns in multivariate assemblages: Information and interpretation of negative values in ANOSIM tests. *Mar Ecol Prog Ser* **180**: 257–265.

Clarke KR. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18**: 117–143.

Cooray NM, Scheuer PJ, Parkanyi L, Clardy J. (1988). Shermilamine A: a pentacyclic alkaloid from a tunicate. *J Org Chem* **53**: 4619–4620.

Corley DG, Moore RE, Paul VJ. (1988). Patellazole B: a novel cytotoxic thiazole-containing macrolide from the marine tunicate *Lissoclinum patella*. *J Am Chem Soc* **110**: 7920–7922.

Correa H, Haltli B, Duque C, Kerr R. (2013). Bacterial communities of the gorgonian octocoral *Pseudopterogorgia elisabethae*. *Microb Ecol* **66**: 972–985.

Degnan BM, Hawkins CJ, Lavin MF, McCaffrey EJ, Parry DL, Watters DJ. (1989). Novel cytotoxic compounds from the ascidian *Lissoclinum bistratum*. *J Med Chem* **32**: 1354–1359.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Dishaw LJ, Flores-Torres J, Lax S, Gemayel K, Leigh B, Melillo D *et al.* (2014). The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* **9**: e93386.

Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J *et al.* (2006). Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* **2**: 729–735.

Donia MS, Ravel J, Schmidt EW. (2008). A global assembly line for cyanobactins. *Nat Chem Biol* **4**: 341–343.

Donia MS, Fricke WF, Partensky F, Cox J, Elshahawi SI, White JR *et al.* (2011a). Complex microbiome underlying secondary and primary metabolism in the tunicate-*Prochloron* symbiosis. *Proc Natl Acad Sci USA* **108**: E1423–E1432.

Donia MS, Fricke WF, Ravel J, Schmidt EW. (2011b). Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS One* **6**: e17897.

Donia MS, Ruffner DE, Cao S, Schmidt EW. (2011c). Accessing the hidden majority of marine natural products through metagenomics. *Chembiochem* **12**: 1230–1236.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Erwin PM, Pineda MC, Webster N, Turon X, Lopez-Legentil S. (2013). Down under the tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. *ISME J* **8**: 575–588.

Erwin PM, Pineda MC, Webster N, Turon X, Lopez-Legentil S. (2013). Small core communities and high variability in bacteria associated with the introduced ascidian *Styela plicata*. *Symbiosis* **59**: 35–46.

Fahy E, Sud M, Cotter D, Subramaniam S. (2007). LIPID MAPS online tools for lipid research. *Nucleic Acids Res* **35**: W606–W612.

Felder S, Dreisigacker S, Kehraus S, Neu E, Bierbaum G, Wright PR *et al.* (2013). Salimabromide: unexpected chemistry from the obligate marine myxobacterium *Enhygromxya salina*. *Chemistry* **19**: 9319–9324.

Fujita MJ, Nakano K, Sakai R. (2013). Bisucaberin B, a linear hydroxamate class siderophore from the marine bacterium *Tenacibaculum mesophilum*. *Molecules* **18**: 3917–3926.

Hirose E, Fukuda T. (2006). Vertical transmission of photosymbionts in the colonial ascidian *Didemnum molle*: the larval tunic prevents symbionts from attaching to the anterior part of larvae. *Zoolog Sci* **23**: 669–674.

Hirose E, Hirabayashi S, Hori K, Kasai F, Watanabe MM. (2006). UV protection in the photosymbiotic ascidian *Didemnum molle* inhabiting different depths. *Zoolog Sci* **23**: 57–63.

Hirose E, Neilan BA, Schmidt EW, Murakami A. (2009). Enigmatic life and evolution of *Prochloron* and related cyanobacteria inhabiting colonial ascidians. *Handbook on Cyanobacteria*. Nova Science Publishers Inc.: Hauppauge, NY, USA, pp 161–189.

Hou Y, Braun DR, Michel CR, Klassen JL, Adnani N, Wyche TP *et al.* (2012). Microbial strain prioritization using metabolomics tools for the discovery of natural products. *Anal Chem* **84**: 4277–4283.

Joullie MM, Leonard MS, Portonovo P, Liang B, Ding X, La Clair JJ. (2003). Chemical defense in ascidians of the didemnidae family. *Bioconjug Chem* **14**: 30–37.

Kobayashi J, Harbour GC, Gilmore J, Rinehart KL. (1984). Eudistomins A, D, G, H, I, J, M, N, O, P, and Q, bromo, hydroxy, pyrrolyl and iminoazepino.beta.-carbolines from the antiviral Caribbean tunicate *Eudistoma olivaceum*. *J Am Chem Soc* **106**: 1526–1528.

Kobayashi J, Tsuda M, Tanabe A, Ishibashi M, Cheng JF, Yamamura S *et al.* (1991). Cystodytins D-I, new cytotoxic tetracyclic aromatic alkaloids from the Okinawan marine tunicate *Cystodytes dellechiajei*. *J Nat Prod* **54**: 1634–1638.

Kohl M, Wiese S, Warscheid B. (2011). Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* **696**: 291–303.

Kojima A, Hirose E. (2010). Transfer of prokaryotic algal symbionts from a tropical ascidian (*Lissoclinum punctatum*) colony to its larvae. *Zoolog Sci* **27**: 124–127.

Krick A, Kehraus S, Eberl L, Riedel K, Anke H, Kaesler I *et al.* (2007). A marine *Mesorhizobium* sp. produces structurally novel long-chain N-acyl-L-homoserine lactones. *Appl Environ Microbiol* **73**: 3587–3594.

Kwan JC, Donia MS, Han AW, Hirose E, Haygood MG, Schmidt EW. (2012). Genome streamlining and chemical defense in a coral reef symbiosis. *Proc Natl Acad Sci USA* **109**: 20655–20660.

Kwan JC, Schmidt EW. (2013). Bacterial endosymbiosis in a chordate host: long-term co-evolution and conservation of secondary metabolism. *PLoS One* **8**: e80822.

Kwan JC, Tianero MDB, Donia MS, Wyche TP, Bugni TS, Schmidt EW. (2014). Host control of symbiont natural product chemistry in cryptic populations of the tunicate Lissoclinum patella. *PLoS One* **9**: e95850.

Lema KA, Willis BL, Bourne DG. (2013). Amplicon pyrosequencing reveals spatial and temporal consistency in diazotroph assemblages of the *Acropora millepora* microbiome. *Environ Microbiol;* e-pub ahead of print 24 December 2013; doi:10.1111/1462-2920.12366.

Letunic I, Bork P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475–W478.

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**: 169–172.

Lucker S, Nowka B, Rattei T, Spieck E, Daims H. (2013). The genome of nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front Microbiol* **4**: 27.

Marshall KM, Barrows LR. (2004). Biological activities of pyridoacridines. *Nat Prod Rep* **21**: 731–751.

Martinez-Garcia M, Diaz-Valdes M, Wanner G, Ramos-Espla A, Anton J. (2007). Microbial community associated with the colonial ascidian *Cystodytes dellechiajei*. *Environ Microbiol* **9**: 521–534.

McMurdie PJ, Holmes S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

Molinski TF, Ireland CM. (1989). Varamines A and B, new cytotoxic thioalkaloids from *Lissoclinum vareau*. *J Org Chem* **54**: 4256–4259.

Monniot C, Monniot F, Laboute P. (1991). *Coral Reef Ascidians of New Caledonia*. Editions de 'ORSTOM: Paris, France.

Moss C, Green DH, Perez B, Velasco A, Henriquez R, McKenzie JD. (2003). Intracellular bacteria associated with the ascidian *Ecteinascidia turbinata*: Phylogenetic and in situ hybridisation analysis. *Mar Biol* **143**: 99–110.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O' Hara RB *et al.* (2013). vegan: Community Ecology Package, R Package version 2.0-8 edn.

Pallmann P, Schaarschmidt F, Hothorn LA, Fischer C, Nacke H, Priesnitz KU *et al.* (2012). Assessing group differences in biodiversity by simultaneously testing a user-defined selection of diversity indices. *Mol Ecol Resour* **12**: 1068–1078.

Parrish C. (2013). Lipids in marine ecosystems. *ISRN Oceanogr* **2013**: article ID 604045.

Paul VJ, Lindquist N, Fenical W. (1990). Chemical defenses of the tropical ascidian *Atapozoa* sp. and its nudibranch predators *Nembrotha* spp. *Mar Ecol Prog Ser* **59**: 109–118.

Peng X, Adachi K, Chen C, Kasai H, Kanoh K, Shizuri Y *et al.* (2006). Discovery of a marine bacterium producing 4-hydroxybenzoate and its alkyl esters, parabens. *Appl Environ Microbiol* **72**: 5556–5561.

Perez LJ, Faulkner DJ. (2003). Bistratamides E-J, modified cyclic hexapeptides from the Philippines ascidian *Lissoclinum bistratum*. *J Nat Prod* **66**: 247–250.

Perez-Matos AE, Rosado W, Govind NS. (2007). Bacterial diversity associated with the Caribbean tunicate *Ecteinascidia turbinata*. *Antonie Van Leeuwenhoek* **92**: 155–164.

Pluskal T, Castillo S, Villar-Briones A, Oresic M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**: 395.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Rath CM, Janto B, Earl J, Ahmed A, Hu FZ, Hiller L *et al.* (2011). Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chem Biol* **6**: 1244–1256.

Reichenbach H. (1999). The ecology of the myxobacteria. *Environ Microbiol* **1**: 15–21.

Riesenfeld CS, Murray AE, Baker BJ. (2008). Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate *Synoicum adareanum*. *J Nat Prod* **71**: 1812–1818.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG *et al.* (2005). Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci USA* **102**: 7315–7320.

Schmidt EW, Donia MS. (2009). Chapter 23. Cyanobactin ribosomally synthesized peptides–a case of deep metagenome mining. *Methods Enzymol* **458**: 575–596.

Schmidt EW, Donia MS. (2010). Life in cellulose houses: symbiotic bacterial biosynthesis of ascidian drugs and drug leads. *Curr Opin Biotechnol* **21**: 827–833.

Schmidt EW, Donia MS, McIntosh JA, Fricke WF, Ravel J. (2012). Origin and variation of tunicate secondary metabolites. *J Nat Prod* **75**: 295–304.

Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, Lindquist N *et al.* (2012). Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.

Schuett C, Doepke H, Grathoff A, Gedde M. (2007). Bacterial aggregates in the tentacles of the sea anemone Metridium senile. *Helgol Mar Res* **61**: 211–216.

Shenkar N, Swalla BJ. (2011). Global diversity of Ascidiacea. *PLoS One* **6**: e20657.

Shi Y, Tyson GW, Eppley JM, DeLong EF. (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* **5**: 999–1013.

Simister RL, Deines P, Botte ES, Webster NS, Taylor MW. (2012). Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ Microbiol* **14**: 517–524.

Sokolov EP. (2000). An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. *J Mollusc Stud* **66**: 573–575.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Xu Y, Kersten RD, Nam SJ, Lu L, Al-Suwailem AM, Zheng H *et al.* (2012). Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *J Am Chem Soc* **134**: 8625–8632.

Yokobori S, Kurabayashi A, Neilan BA, Maruyama T, Hirose E. (2006). Multiple origins of the ascidian-*Prochloron* symbiosis: molecular phylogeny of photosymbiotic and non-symbiotic colonial ascidians inferred from 18S rDNA sequences. *Mol Phylogenet Evol* **40**: 8–19.

Zabriskie TM, Mayne CL, Ireland CM. (1988). Patellazole C: A novel cytotoxic macrolide from *Lissoclinum patella*. *J Am Chem Soc* **110**: 7919–7920.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)

Species specificity of symbiosis and secondary metabolism in ascidians

Ma. Diarey B. Tianero, Jason C. Kwan, Thomas P. Wyche, Angela P. Presson, Michael Koch, Louis R. Barrows, Tim S. Bugni, Eric W. Schmidt

Supplementary information index

Methods S1

DNA extraction. Ascidian tissue (~0.5 cm^3) was ground in liquid nitrogen in sterile mortar and pestle. The powder obtained was resuspended in TE (5 mL) with lysozyme (2 mg ml^{-1}) and incubated at 30 ºC for 1 hour. EDTA (0.5 M, 1.2 mL) was added to the solution and gently mixed, after which proteinase K (0.2 mg mL^{-1}, Qiagen) was added and the solution was incubated at 30 ºC for 5 minutes. For cell lysis, SDS (10%, 0.65 mL) was added and the solution was incubated at 37 ºC for 3-4 hours. After visibly complete lysis (clear solution), NaCl (5M, 1.2 mL) was added, mixed, and CTAB/NaCl (1 mL) solution was added before incubation at 65 ºC for 1 hour. Standard phenol/chloroform extraction was then performed (Sambrook and Russell 2006).

Methods S2

Processing and analysis of ascidian chemical extracts. Each ascidian was extracted for 1 hour with methanol (5 mL), and the methanol was removed and dried under vacuum. Each extract was dissolved in methanol:water (1:9; 1.1 mL) and loaded into a Gilson GX-271 liquid handling system. Samples (0.9 mL) were subjected to automated SPE (ThermoFisher Scientific HyperSep C18, 50 mg absorbent mass, 1 mL reservoir volume), washed with H_2O (1 mL), and eluted with 95% methanol (1 mL) directly into an LC/MS-certified vial. Following SPE, each extract was quantified using an evaporative light-scattering detector (ELSD) according to previously published methods.(Adnani et al 2012) Each extract was diluted with 95% methanol to a concentration of 30 ng/mL.

Methods S3

Representative samples from different locations were analyzed for putative lipid components on an Agilent MS 6520 Q-TOF mass spectrometer fitted to an Agilent 1290 UPLC. The extract components were separated on Acquity UPLC CSH C18 column (1.7 µM, 2.1 X100 mm) using the

following gradient at 0.3 mL min^{-1} (mobile phase A:B): t = 0, 85:15; t = 0-4 min, gradient to 70:30 = 4-5, gradient to 48:52; t = 5-22, gradient to 18:82; t = 22-23, gradient to 1:99; t = 23-30, 1:99. Mobile phase A consisted of acetonitrile:water (60:40 v/v) in 10 mM ammonium formate and 0.1% formic acid, and mobile phase B consisted of isopropyl alcohol:water (90:10 v/v) in 10 mM ammonium formate and 0.1% formic acid. The column temperature was set to 65 ^{o}C. Acquisition was performed in positive ESI mode acquiring from *m/z* 100-1700. The mass spectrometer was operated using the following parameters: dry gas temperature, 350 ^{o}C; dry gas flow, 11.1 L/min; nebulizer pressure, 24 psig; Vcap, 5000V; fragmentor, 250 V. Additional parameters (precursors p cycle, 2; absolute threshold, 200) were used for data-dependent MS/MS.

Methods S4

Antimicrobial broth dilution assays. The assay was performed as previously described(Koch et al 2010) with the exception that four wells per compound were used for each dilution for MIC determination. DMSO served as a negative control. Bacterial strains used, along with the positive control antibiotic for each strain, were *M. tuberculosis* H37Ra ATCC 25177 (rifampicin); *E. coli* ATCC 25922 (gentamicin); *B. subtilis* ATCC 6633 (gentamicin); and *S. aureus* ATCC 25923 (gentamicin) and *C. albicans* ATTC 90028 (itraconazole).

Cytotoxicity assay. Cytotoxicity testing was performed using CEM-TART lymphoblastoid cells(Chen et al 1992) cultured in RPMI-1640 with 20% fetal bovine serum and antibiotic/antimycotic supplement at 37 °C with 5% CO_2 in moisture saturated atmosphere in 96-well culture clusters as previously described (Lin et al 2013).

Table S1 : Sample collection details and 18S rRNA gene BLAST hits. * Indicates codes used in preliminary taxonomic identification of samples.

Code	ollection Date	Collection Site, coordinates	BLAST hits, GenBank accession number	% identity	18S primers
CA_Ascidia_sp_4-19	20-Feb-11	Catalina Island, 33° 26.95 ' N, 118° 29.94' W	*Ascidia ceratodes*, L12378.2	99	F1/R1
CA_Botryllus_sp_5-19	12-Oct-11	Catalina Harbor	*Bottrylus planus*, DQ346653.1	99	F3/R3
CA_Didemnum_sp_4-17	20-Feb-11	Catalina Island, 33° 26.95' N, 118° 29.94' W	*Didemnum molle*, AB211071.1	94	AscF2/ AscR5
CA_Didemnum_sp_4-2	19-Feb-11	Catalina Island, 33° 26.92' N, 118° 28.64' W	*Didemnum molle*, AB211071.1	92	F3/R3
CA_Didemnum_sp_4-7	19-Feb-11	Catalina Island, 33° 26.92' N, 118° 28.64' W	*Didemnum* sp. DidSA/57, AB211072.1	92	AscF2/ AscR5
CA_Didemnum_sp_5-1	10-Oct-11	Catalina Island, 33° 26.69' N, 118° 28.92' W	*Didemnum* sp. DidSA/57, AB211072.1	92	AscF2/ AscR5
CA_Didemnum_sp_5-11	11-Oct-11	Catalina Island, 33° 26.87' N, 118° 28.61' W	*Didemnum molle*, AB211071.1	94	AscF2/ AscR5
CA_Didemnum_sp_5-18	12-Oct-11	Catalina Island, 33° 26.21' N, 118° 27.48' W	*Didemnum molle*, AB211071.1	90	AscF2/ AscR5
CA_Pyura_sp_4-21	19-Feb-11	Catalina Island, 33° 26.67' N, 118° 29.417' W	*Pyura haustor*, AY903926.1	99	F3/R3
CA_Styela_sp_5-16	11-Oct-11	Catalina Harbor	*Styela plicata*, L12444.2	99	F3/R3
FJ_Lissoclinum_patella 06037		Fiji, 17°55' S, 177°16' E	*Lissoclinum patella,* AB211085.1	99	F1/R1
FL_Didemnum_sp_TB3 *(*D.psammathode*)	15-Aug-11	Florida Keys, 24 ° 41.672' N, 81° 26,797' W	*Didemnum* sp. Whangamata-EM-2003, AJ579861.1	89	AscF2/ AscR5
FL_Ecteinascidia_turbinata_5	10-Aug-11	Florida Keys, 24 ° 39.591' N, 81 ° 25.217' W	*Ecteinascidia turbinata*, FM244848.1	99	F3/R3
FL_Ecteinascidia_turbinata_TB2	15-Aug-11	Florida Keys, 24 ° 39.453' N, 81 ° 25.247' W	*Ecteinascidia turbinata,* FM244848.1	99	F3/R3
FL_Eudistoma_sp. _004 *(*E. hepaticum*)	28-Oct-10	Florida Keys, 24 ° 41.078 N, 81 ° 26.957 W	*Eudistoma gilboviride,* AB211069.1	89	AscF2/ AscR5
FL_Eudistoma_sp_TB1 *(*E. olivaceum*)	1-Sep-11	Florida Keys, 24 ° 41.685' N, 81 ° 26.840' W	*Eudistoma gilboviride,* AB211069.1	95	AscF2/ AscR5
FL_Trididemnum_sp_001BB *(*T. orbicatum)*	11-Oct-10	Florida Keys, 24 ° 37.487' N, 81 ° 27.443' W	*Trididemnum paracyclops,* AB211077.1	92	F1/R1
FL_Trididemnum_sp_TB4 *(*T. palmae*)	15-Aug-11	Florida Keys, 24 ° 37.487' N, 81 ° 27.443' W	*Trididemnum paracyclops,* AB211077.1	92	AscF2/ AscR5
PNG_Didemnum_sp_11038	9-Nov-11	Papua New Guinea, 10° 16' S, 145° 38' E	*Didemnum molle*, AB211071.1	94	AscF2/ AscR5
PNG_Didemnum_sp_11055	10-Nov-11	Papua New Guinea, 10° 2' S, 145° 33' E	*Didemnum* sp. DidSA/57, AB211072.1	95	AscF2/ AscR5
PNG_Didemnum_sp_11089	12-Nov-11	Papua New Guinea, 9° 36' S, 147° 18' E	*Didemnum molle,* AB211071.1	94	AscF2/ AscR5
PNG_Lissoclinum_patella_11033	9-Nov-11	Papua New Guinea, 10° 16' S, 145° 38' E	*Lissoclinum patella*, AB211085.1	99	F1/R1
PNG_Lissoclinum_patella_11040	9-Nov-11	Papua New Guinea, 10° 16' S, 145° 38' E	*Lissoclinum patella*, AB211085.1	99	F1/R1
PNG_Lissoclinum_vareau_11047	10-Nov-11	Papua New Guinea, 10° 8' S, 145° 35' E	*Lissoclinum badium*, AB211078.1	86	F3/R3
PNG_Lissoclinum_vareau_11062	11-Nov-11	Papua New Guinea, 10° 2' S, 145° 32' E	*Lissoclinum badium*, AB211078.1	86	F3/R3
PNG_Lissoclinum_vareau_11068	11-Nov-11	Papua New Guinea, 10° 2' S, 145° 32' E	*Lissoclinum badium*, AB211078.1	86	F3/R3

PNG_Lissoclinum_vareau_11083	12-Nov-11	Papua New Guinea, 9° 35' S, 147° 17' E	*Lissoclinum badium*, AB211078.1	87	F3/R3
PNG_Lissoclinum_bistratum *(*L. voeltzkowi*)	05-Oct-07	Papua New Guinea, 4° 8' S, 151° 34' E	*Lissoclinum bistratum,* AB211084.1	99	F1/R1
VN_Cystodytes_sp_08002	28-Oct-08	Vanuatu, 15° 32.496' S, 167°12.895' E	*Cystodytes* sp., FM244842.1	94	AscF2/ AscR5
VN_Cystodytes_sp_08019	1-Nov-08	Vanuatu, 15° 36.675' S, 167° 01.258' E	*Cystodytes* sp., FM244842.1	91	AscF2/ AscR5
VN_Cystodytes_sp_08041	4-Nov-08	Vanuatu, 15° 25.416' S, 167° 12.476' E	*Cystodytes* sp., FM244842.1	94	AscF2/ AscR5

Table S2. Microbiome statistics

Number of tunicate samples	32
Number of filtered reads	217073
Number of 97 % OTUs	3982
Counts statistics (min/max/average)	572/31660/6784
OTUs statistics (min/max/average)	10/804/207

Table S3. List of primers used in this study

Primer	Sequence (5'-3')	Source
AscF1	CTGGTTGATCCTGCCAG	Yokobori, 2006 (Yokobori et al 2006)
AscR1	CACCTACGGRWACCTTG	Yokobori, 2006 (Yokobori et al 2006)
AscF3	GATCCTGCCAGTAGTBATAT	Yokobori, 2006 (Yokobori et al 2006)
AscR3	TGATCCTTCTGCAGGTTCA	Yokobori, 2006 (Yokobori et al 2006)
AscF2new	CAAGGAAGGCAGCAGGCGCGCAAAT	This study
AscR5new	GCGGTGTGTACAAAGGGCAGGGA	This study
AscR3new	AAGGAATTGACGGAAGGGCACCACCAGGA	This study
18S1	CCTGGTTGATCCTGCCAG	Tsagkogeorga, 2009 (Tsagkogeorga 2009)
18S2	TAATGATCCATCTGCAGG	Tsagkogeorga, 2009 (Tsagkogeorga 2009)
27F	AGAGTTTGATCMTGGCTCAG	Lane, 1991 (Lane 1991)
1492R	CGGTTACCTTGTTACGACTT	Turner, 1999 (Turner et al 1999)
939F	TTGACGGGGGCCCGCACAAG	RTL in house primer

Table S4. Sequence counts and alpha diversity values of ascidian microbiomes measured using different diversity indices.

Sample	Counts	OTU	Shannon	Simpson	Chao1
CA_Ascidia_sp_4-19	2213	248	6.19	0.96	255.40
CA_Botryllus_sp_5-19	2406	220	5.68	0.94	229.67
CA_Didemnum_sp_4-17	3992	136	3.27	0.77	156.53
CA_Didemnum_sp_4-2	2959	106	2.87	0.68	117.12
CA_Didemnum_sp_4-7	2681	160	4.09	0.82	168.22
CA_Didemnum_sp_5-1	2478	170	4.86	0.92	182.00
CA_Didemnum_sp_5-11	3312	140	3.59	0.83	167.04
CA_Didemnum_sp_5-18	5068	149	3.15	0.67	152.89
CA_Pyura_sp_4-21	2841	58	2.54	0.73	60.50
CA_Styela_sp_5-16	2037	163	4.97	0.92	169.88
FJ_Lissoclinum_patella_06037	1594	209	4.35	0.84	393.53
FL_Didemnum_sp._tb3	7454	429	5.62	0.92	446.73
FL_Ecteinascidia_turbinata_5	10015	234	3.60	0.80	268.55
FL_Ecteinascidia_turbinata_tb2	31660	804	4.56	0.85	836.25
FL_Eudistoma_sp_004	13886	41	0.78	0.17	50.00
FL_Eudistoma_sp_tb1	12822	667	5.27	0.86	699.72
FL_Tridemnum_sp_001BB	1007	74	3.72	0.84	78.57
FL_Tridemnum_sp_tb4	572	10	0.61	0.15	10.25
PNG_Didemnum_sp_11038	14155	166	5.09	0.93	178.55
PNG_Didemnum_sp_11055	3372	290	6.79	0.98	300.44
PNG_Didemnum_sp_11089	5585	236	5.38	0.95	241.70
PNG_Lissoclinum_badium_11062	3916	92	4.20	0.90	98.18
PNG_Lissoclinum_badium_11068	8993	104	3.24	0.80	108.62
PNG_Lissoclinum_badium_11083	20324	278	2.40	0.56	284.36
PNG_Lissoclinum_badium_11047	1899	44	3.25	0.81	48.50
PNG_Lissoclinum_patella_11033	12851	347	3.08	0.65	368.72
PNG_Lissoclinum_patella_11040	13778	186	2.55	0.64	195.60
PNG_Lissoclinum_sp_7062	5528	114	4.39	0.88	121.00
PNG_Lissoclinum_bistratum	11015	475	3.68	0.64	779.59
VN_Cystodytes_sp_08002	2578	106	4.51	0.90	109.00
VN_Cystodytes_sp_08019	1610	97	4.49	0.92	100.62
VN_Cystodytes_sp_08041	2472	110	4.19	0.89	114.50

Table S5. Host species maps for networks analysis of ascidian microbiomes

SampleName	map1	map2	map3	map4
CA_Botrylloides_sp_5-19	Botryllus	Botryllus	Botryllus	Botryllus
CA_Pyura_sp_4-21	Pyura	Pyura	Pyura	Pyura
CA_Styela_sp_5-16	Styela	Styela	Styela	Styela
FL_Ecteinascidia_turbinata_5	Ecteinascidia_turbinata	Ecteinascidia_turbinata	Ecteinascidia_turbinata	Ecteinascidia_turbinata
FL_Ecteinascidia_turbinata_tb2	Ecteinascidia_turbinata	Ecteinascidia_turbinata	Ecteinascidia_turbinata	Ecteinascidia_turbinata
CA_Ascidia_sp_4-19	Ascidia	Ascidia	Ascidia	Ascidia
FL_Eudistoma_sp_004	Eudistoma_sp_1	Eudistoma_sp_1	Eudistoma_sp_1	Eudistoma_sp_1
FL_Eudistoma_sp_tb1	Eudistoma_sp_2	Eudistoma_sp_2	Eudistoma_sp_2	Eudistoma_sp_2
VN_Cystodytes_sp_08019	Cystodytes_sp	Cystodytes_sp_g2		
VN_Cystodytes_sp_08002	Cystodytes_sp	Cystodytes_sp_g1		
VN_Cystodytes_sp_08041	Cystodytes_sp	Cystodytes_sp_g1		
PNG_Lissoclinum_badium_11083	Lissoclinum_badium	Lissoclinum_badium_g1		
PNG_Lissoclinum_badium_11047	Lissoclinum_badium	Lissoclinum_badium_g2		
PNG_Lissoclinum_badium_11062	Lissoclinum_badium	Lissoclinum_badium_g2		
PNG_Lissoclinum_badium_11068	Lissoclinum_badium	Lissoclinum_badium_g2		
FJ_Lissoclinum_patella_06037	Lissoclinum_patella	Lissoclinum_patella	Lissoclinum_patella	Lissoclinum_patella
PNG_Lissoclinum_patella_11033	Lissoclinum_patella	Lissoclinum_patella	Lissoclinum_patella	Lissoclinum_patella
PNG_Lissoclinum_patella_11040	Lissoclinum_patella	Lissoclinum_patella	Lissoclinum_patella	Lissoclinum_patella
PNG_Lissoclinum_vareau_7062	Lissoclinum_sp	Lissoclinum_sp	Lissoclinum_sp	Lissoclinum_sp
PNG_Lissoclinum_bistratum	Lissoclinum_bistratum	Lissoclinum_bistratum	Lissoclinum_bistratum	Lissoclinum_bistratum
FL_Tridemnum_sp_001BB	Tridemnum_sp_1	Tridemnum_sp_1	Tridemnum_sp_1	Tridemnum_sp_1
FL_Tridemnum_sp_tb4	Tridemnum_sp_2	Tridemnum_sp_2	Tridemnum_sp_2	Tridemnum_sp_2
FL_Didemnum_sp_tb3	Didemnum_sp	Didemnum_sp_g1	Didemnum_sp	Didemnum_sp_g1
PNG_Didemnum_sp_11089	Didemnum_sp	Didemnum_sp_g1	Didemnum_sp	Didemnum_sp_g1
CA_Didemnum_sp_4-17	Didemnum_sp	Didemnum_sp_g1	Didemnum_sp	Didemnum_sp_g1
CA_Didemnum_sp_4-2	Didemnum_sp	Didemnum_sp_g1	Didemnum_sp	Didemnum_sp_g1
CA_Didemnum_sp_5-11	Didemnum_sp	Didemnum_sp_g1	Didemnum_sp	Didemnum_sp_g1
CA_Didemnum_sp_5-18	Didemnum_sp	Didemnum_sp_g2	Didemnum_sp	Didemnum_sp_g2
PNG_Didemnum_sp_11038	Didemnum_sp	Didemnum_sp_g2	Didemnum_sp	Didemnum_sp_g2
PNG_Didemnum_sp_11055	Didemnum_sp	Didemnum_sp_g2	Didemnum_sp	Didemnum_sp_g2
CA_Didemnum_sp_5-1	Didemnum_sp	Didemnum_sp_g2	Didemnum_sp	Didemnum_sp_g2
CA_Didemnum_sp_4-7	Didemnum_sp	Didemnum_sp_g2	Didemnum_sp	Didemnum_sp_g2
G(Host)	4.99	2.18	5.51	2.42
Pvalue	0.012	0.099	0.009	0.05
G(Location)	6.22	6.22	5.28	5.28
Pvalue	0.006	0.006	0.01	0.01

Table S6. Table of shared microbes (top 10) in locations (A-C) and hosts species (D-F) and their percent abundance.

A. California ascidians

#OTU ID	Consensus lineage	Average % abundance in samples	Present in % samples
denovo3071	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Chromatiales ; ; ;	0.16%	70
denovo116	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Alteromonadales ;Alteromonadaceae ;Glaciecola ;	0.10%	60
denovo3294	Bacteria ;Bacteroidetes ;Flavobacteriia ;Flavobacteriales ;Flavobacteriaceae	0.15%	60
denovo3389	Bacteria ;Cyanobacteria ;Chloroplast ;Stramenopiles ; ; ;	0.09%	60
denovo27	Bacteria ;Bacteroidetes ;Flavobacteriia ;Flavobacteriales ;Flavobacteriaceae ; ;	0.06%	50
denovo1982	Bacteria ;Proteobacteria ;Alphaproteobacteria ;Rhizobiales	0.05%	50
denovo2251	Bacteria ;Acidobacteria ;Sva0725 ;Sva0725 ; ; ;	0.21%	50
denovo2984	Bacteria ;Proteobacteria	0.06%	50
denovo3340	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Alteromonadales ;OM60 ; ;	0.04%	50

B. Florida ascidians

#OTU ID	Consensus Lineage	Average % abundance in samples	Present in % samples
denovo701	Bacteria ;Proteobacteria ;Gammaproteobacteria	0.17%	67
denovo766	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Vibrionales ;Vibrionaceae ;Vibrio	0.11%	67
denovo1101	Bacteria ;Proteobacteria ;Gammaproteobacteria	0.08%	67
denovo1332	Bacteria ;Proteobacteria ;Deltaproteobacteria ;Desulfobacterales ;Desulfobulbaceae ;	0.08%	67
denovo1589	Bacteria ;Proteobacteria ;Gammaproteobacteria	7.31%	67
denovo1834	Bacteria	0.38%	67
denovo1931	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Oceanospirillales ;Litoricolaceae ;Litoricola	0.38%	67
denovo2290	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Xanthomonadales ; ;	0.02%	67
denovo2327	Bacteria ;Proteobacteria ;Gammaproteobacteria	0.43%	67

C. Papua New Guinea ascidians

#OTU ID	Consensus Lineage	Average % abundance in samples	Present in % samples
denovo866	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Alteromonadales ;Alteromonadaceae ;Alteromonas ;	0.30%	42
denovo1357	Bacteria ;Cyanobacteria ;Oscillatoriophycideae ;Chroococcales ;Prochloraceae ;Prochloron ;	2.85%	42
denovo2569	Bacteria ;Proteobacteria ;Alphaproteobacteria	1.72%	33
denovo2672	Bacteria ;Proteobacteria ;Deltaproteobacteria ;Desulfobacterales ;Nitrospinaceae ;Nitrospina ;	3.33%	33
denovo2769	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Oceanospirillales ;Oceanospirillaceae ;Marinomonas ;	0.04%	33
denovo3942	Bacteria ;Proteobacteria ;Deltaproteobacteria ;Sva0853 ; ; ;	9.92%	33
denovo217	Bacteria ;Bacteroidetes ;Sphingobacteriia ;Sphingobacteriales ;Ekhidnaceae ; ;	0.10%	25
denovo283	Bacteria ;Proteobacteria ;Alphaproteobacteria ;Kordiimonadales ;Kordiimonadaceae ; ;	0.02%	25
denovo352	Bacteria	0.09%	25

D. *Lissoclinum patella* (group 1, Figure 3B)

#OTU ID	Consensus Lineage	Average % abundance in samples	Present in % samples
denovo1567	Bacteria ;Cyanobacteria ;Oscillatoriophycideae ;Chroococcales ;Prochloraceae ;Prochloron ;	46.08%	100
denovo2569	Bacteria ;Proteobacteria ;Alphaproteobacteria	6.86%	100
denovo2404	Bacteria ;Proteobacteria ;Gammaproteobacteria	6.17%	100
denovo1357	Bacteria ;Cyanobacteria ;Oscillatoriophycideae ;Chroococcales ;Prochloraceae ;Prochloron ;	10.76%	100
denovo883	Bacteria ;Proteobacteria ;Alphaproteobacteria	0.14%	100
denovo832	Bacteria ;Proteobacteria ;Alphaproteobacteria ;Rhodospirillales	2.77%	100
denovo1977	Bacteria ;Cyanobacteria ;Chloroplast ;Stramenopiles ; ; ;	0.12%	100
denovo231	Bacteria ;Cyanobacteria ;Oscillatoriophycideae ;Chroococcales ;Prochloraceae ;Prochloron ;	0.10%	100
denovo3369	Bacteria ;Cyanobacteria ;Oscillatoriophycideae ;Chroococcales ;Prochloraceae ;Prochloron ;	0.34%	100

E. *Lissoclinum badium* (group 2, Figure 3B)

#OTU ID	Consensus lineage	Average % abundance in samples	Present in % samples
denovo3942	Bacteria ;Proteobacteria ;Deltaproteobacteria ;Sva0853 ; ; ;	39.67%	75
denovo3967	Bacteria ;Proteobacteria ;Alphaproteobacteria	4.08%	75
denovo1567	Bacteria ;Cyanobacteria ;Oscillatoriophycideae ;Chroococcales ;Prochloraceae ;Prochloron ;	2.52%	75
denovo2672	Bacteria ;Proteobacteria ;Deltaproteobacteria ;Desulfobacterales ;Nitrospinaceae ;Nitrospina ;	9.99%	75
denovo279	Bacteria ;Proteobacteria	1.81%	75
denovo1477	Bacteria ;Proteobacteria ;Alphaproteobacteria ;Rhizobiales ;Phyllobacteriaceae ;Mesorhizobium ;	2.98%	75
denovo1436	Bacteria ;Proteobacteria ;Gammaproteobacteria	0.78%	75
denovo3289	Bacteria	0.72%	75
denovo2829	Bacteria ;Proteobacteria ;Alphaproteobacteria ;Kordiimonadales ;Kordiimonadaceae ; ;	0.50%	75

F. *Didemnum* sp. (group 3, Figure 3B)

#OTU ID	Consensus lineage	Average % abundance in samples	Present in % samples
denovo2476	Bacteria ;Proteobacteria ;Alphaproteobacteria	17.42%	100
denovo2071	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Pseudomonadales ;Pseudomonadaceae ;Pseudomonas	4.20%	100
denovo1698	Bacteria ;Proteobacteria ;Deltaproteobacteria ;Bdellovibrionales ;Bdellovibrionaceae ;Bdellovibrio ;	1.11%	100
denovo42	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Pseudomonadales ;Pseudomonadaceae	0.18%	100
denovo2984	Bacteria ;Proteobacteria	0.13%	100
denovo1982	Bacteria ;Proteobacteria ;Alphaproteobacteria ;Rhizobiales	0.09%	100
denovo3206	Bacteria ;Proteobacteria ;Gammaproteobacteria ;Oceanospirillales ;Endozoicimonaceae ; ;	0.15%	75
denovo980	Bacteria ;Proteobacteria ;Gammaproteobacteria	0.04%	75
denovo3508	Bacteria ;Proteobacteria ;Gammaproteobacteria	0.12%	75

Table S7. BLAST hits of 18S rRNA sequences of eukaryotes associated with ascidians

Ascidian sample	Symbiotic eukaryote	Genbank accession number
Eudistoma sp. 004	*Aiptasia mutabilis*	FJ489438.1
	Prostheceraeus vittatus	AJ312272.1
	Aiptasia pulchella	AY297437.1
Didemnum sp. 4-7	*Grateloupia luxurians*	U33132.1
Didemnum sp. 4-17	*Clytia gracilis*	DQ068053.1
	Clytia sp. AGC-200	AF358074.1
Didemnum sp. 5-1	*Baeria nivea*	AF182191.1
Didemnum sp. 5-11	*Selaginopsis cornigera*	Z92899.1
	Uncultured eukaryote clone SS1_E_02_11	EU050966.1
Didemnum sp. 5-18	*Igernella notabilis*	EU702420.1
	Rhacostoma atlantica	EU305501.1
	Melicertum octocostatum	AY920757.1
Cystodytes sp. 08002	*Lankesteria ascidiae*	JX187607.1
	Uncultured eukaryote clone SS1_E_02_1	EU050966.1
Cystodytes sp. 08019	*Corallochytrium limacisporum*	L42528.1
	Notodelphys prasina	JF781536.1
	Cyclopina gracilis	JF781537.1
Cystodytes sp. 08041	*Symbiodinium* sp.	AY165766.1
	Notodelphys prasina	JF781536.1
	Lankesteria ascidiae	JX187607.1
Didemnum sp. 11038	Uncultured eukaryote clone CC02A105.033	KF031747.1
	Cyclopina gracilis	JF781537.1
	Doridicola agilis	JF781541.1
Didemnum sp. 11055	*Cyclopina gracilis*	F781537.1
	Anticomidae sp. JCC29	HM564572.1
Didemnum sp.11089	*Melicertum octocostatum*	AY920757.1
	Dictyota sp. YMGTB002	AB087109.1
	Uncultured alveolate	FN263032.1
Eudistoma sp. TB1	*Calomicrolaimus parahonestus*	AY854218.1
	Sycettusa aff. hastifera OV-2012	JQ272322.1
	Spongia tubulifera	KC902150.1
	Grantessa sp. OV-2012 voucher GW979	JQ272312.1
Didemnum sp. TB3	*Posidonia australis*	GQ497582.1
	Spyridia filamentosa	EU718707.1
	Attheyella crassa	EU380307.1
Tridemnum sp. TB4	*Biemna fistulosa*	FR819688.1
	Thelepus crispus	JN936473.1
	Euplotes charon voucher JJM08102201	JF694043.1

Table S8. Summary of secondary metabolites and associated microbes in ascidians from this data set and related species.

Sample Code	Host Taxonomy	Location	Major associated secondary metabolites	Major symbiont characterized	Previous microbiome analysis of related species
CA_Ascidia_sp_4-19	*Ascidia ceratodes*	California			
CA_Botryllus_sp_5-19	*Bottrylus planus*	California			
CA_Didemnum_sp_4-17	*Didemnum* sp.	California			
CA_Didemnum_sp_4-2	*Didemnum* sp.	California			
CA_Didemnum_sp_4-7	*Didemnum* sp.	California			
CA_Didemnum_sp_5-1	*Didemnum* sp.	California			
CA_Didemnum_sp_5-11	*Didemnum* sp.	California			
CA_Didemnum_sp_5-18	*Didemnum molle*	California			
CA_Pyura_sp_4-21	*Pyura haustor*	California			
CA_Styela_sp_5-16	*Styela plicata*	California	plicatamide		Erwin et al, 2013
FL_Didemnum_sp_tb3	*Didemnum* sp.	Florida			
FL_Ecteinascidia_turbinata_5	*Ecteinascidia turbinata*	Florida	ecteinascidins	*Ca.* Endolisoclinum frumentensis	Rath 2011, Perez 2007, Moss 2003
FL_Ecteinascidia_turbinata_tb2	*Ecteinascidia turbinata*	Florida	ecteinascidins	*Ca.* Endolisoclinum frumentensis	Rath 2011, Perez 2007, Moss 2003
FL_Eudistoma_sp_004	*Eudistoma* sp.	Florida	eudistomins		
FL_Eudistoma_sp_tb1	*Eudistoma* sp.	Florida	eudistomins		
FL_Trididemnum_sp_001BB	*Trididemnum* sp.	Florida			
FL_Trididemnum_sp_tb4	*Trididemnum* sp.	Florida			
PNG_Didemnum_sp_11038	*Didemnum* sp.	Papua New Guinea			
PNG_Didemnum_sp_11055	*Didemnum* sp.	Papua New Guinea			
PNG_Didemnum_sp_11089	*Didemnum* sp.	Papua New Guinea			
PNG_Lissoclinum_patella_11033	*Lissoclinum patella*	Papua New Guinea	cyanobactins, patellazoles	*Prochloron* spp., *Ca*. Endolissoclinum faulkneri	Schmidt 2005, Donia 2008, Kwan 2012, Behrendt 2012
PNG_Lissoclinum_patella_11040	*Lissoclinum patella*	Papua New Guinea	cyanobactins, patellazoles	*Prochloron* spp.,	Schmidt 2005, Donia 2008, Kwan 2012, Behrendt 2012
FJ_Lissoclinum_patella_06037	*Lissoclinum patella*	Fiji	cyanobactins, patellazoles	*Prochloron* spp., *Ca*. Endolissoclinum faulkneri	Schmidt 2005, Donia 2008, Kwan 2012, Behrendt 2012
PNG_Lissoclinum_sp_7062	*Lissoclinum* sp.	Papua New Guinea			
PNG_Lissoclinum_badium_11047	*Lissoclinum badium*	Papua New Guinea	pyridoacridine alkaloids	*Nitrospina* sp.	Erwin 2013
PNG_Lissoclinum_badium_1	*Lissoclinum*	Papua New	pyridoacridine	*Nitrospina* sp.	Erwin 2013

1062	*badium*	Guinea	alkaloids		
PNG_Lissoclinum_badium_11068	*Lissoclinum badium*	Papua New Guinea	pyridoacridine alkaloids	*Nitrospina* sp.	Erwin 2013
PNG_Lissoclinum_badium_11083	*Lissoclinum badium*	Papua New Guinea	pyridoacridine alkaloids	*Nitrospina* sp.	Erwin 2013
PNG_Lissoclinum_bistratum	*Lissoclinum bistratum*	Papua New Guinea	cyanobactins, bistramides	*Prochloron* spp.,	
VN_Cystodytes_sp_08002	*Cystodytes* sp.	Vanuatu	pyridoacridine alkaloids		
VN_Cystodytes_sp_08019	*Cystodytes* sp.	Vanuatu	pyridoacridine alkaloids		
VN_Cystodytes_sp_08041	*Cystodytes* sp.	Vanuatu	pyridoacridine alkaloids		

Table S9. Statistical comparisons of ascidian microbiomes across species, and within the same species across locations and over time (at the same location). There was a statistically significant difference in ascidian microbiomes across species, and within the same species across locations, but no difference in microbiomes within a species collected at two different time points. Analyses were performed only on groups with a minimum of two samples using the ANOSIM method in the vegan R package(ANOSIM R, P-value1) with 100,000 permutations. A sensitivity analysis was performed using a second permutation ANOVA method "adonis" also available in vegan (R^2, P-value2).

Samples/Groups	# Samples	ANOSIM R	P-value1	R^2	P-value2
All Papua New Guinea	**11**	**0.762**	**0.002**	**0.426**	**0.002**
Didemnum	3				
Lissoclinum_badium	4				
Lissoclinum_patella	2				
All *Didemnum*	**9**	**0.846**	**0.011**	**0.252**	**0.012**
Papual New Guinea	3				
California	6				
California *Didemnum*	**6**	**-0.037**	**0.499**	**0.171**	**0.701**
Fall collection	3				
Spring collection	3				

References:

Adnani N, Michel CR, Bugni TS (2012). Universal quantification of structurally diverse natural products using an evaporative light scattering detector. *J Nat Prod* **75:** 802-806.

Chen H, Boyle TJ, Malim MH, Cullen BR, Lyerly HK (1992). Derivation of a biologically contained replication system for human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **89:** 7678-7682.

Koch M, Bugni TS, Sondossi M, Ireland CM, Barrows LR (2010). Exocarpic acid inhibits mycolic acid biosynthesis in Mycobacterium tuberculosis. *Planta Med* **76:** 1678-1682.

Lane DJ (1991). 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow,M. (ed). *Nucleic acid techniques in bacterial systematics*. John Wiley and Sons: New York. pp 115-175.

Lin Z, Koch M, Pond CD, Mabeza G, Seronay RA, Concepcion GP *et al* (2013). Structure and activity of lobophorins from a turrid mollusk-associated Streptomyces sp. *J Antibiot (Tokyo)*.

Sambrook J, Russell DW (2006). Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc* **2006**.

Tsagkogeorga G, Turon X, Hopcroft RH, Tilak MK, Feldstein T, Shenkar N, Loya Y, Huchon D, Douzery EJP, Delsuc F (2009). An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol Biol* **9**.

Turner S, Pryer KM, Miao VPW, Palmer JD (1999). Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis1. *J Eukaryot Microbiol* **46:** 327-338.

Yokobori S, Kurabayashi A, Neilan BA, Maruyama T, Hirose E (2006). Multiple origins of the ascidian-Prochloron symbiosis: molecular phylogeny of photosymbiotic and non-symbiotic colonial ascidians inferred from 18S rDNA sequences. *Mol Phylogenet Evol* **40:** 8-19.

Figure S1. The phylogenetic tree of ascidians using 18S rRNA gene sequences. Sequences obtained from this study are labeled with location (CA, FL, PNG, VN) and sample code. Other sequences were obtained from NCBI with the following accession numbers: *Megalocercus huxleyi* FM244868.1, *Herdmania litoralis* AB564300.1*, Pyura haustor* AY903926.1, *Pyura spinifera* JF961826.1, *Ecteinascidia turbinata* FM244848.1, *Phallusia fumigata* FM244844.1, *Ascidia ceratodes* L12378.2, *Aplidium pliciferum* AB211067.1, *Eudistoma gilboviride* AB211069.1, *Cystodytes* sp_TG-2008 FM244842.1, *Lissoclinum badium* AB211078.1, *Lissoclinum bistratum* AB211084.1*, Lissoclinum patella* AB211085.1, *Didemnum* sp DidSB AB211073.1, *Tridemnum paracyclops* AB211077.1, *Didemnum molle* AB211071.1, *Didemnum* sp DidSA/57 AB211072.1

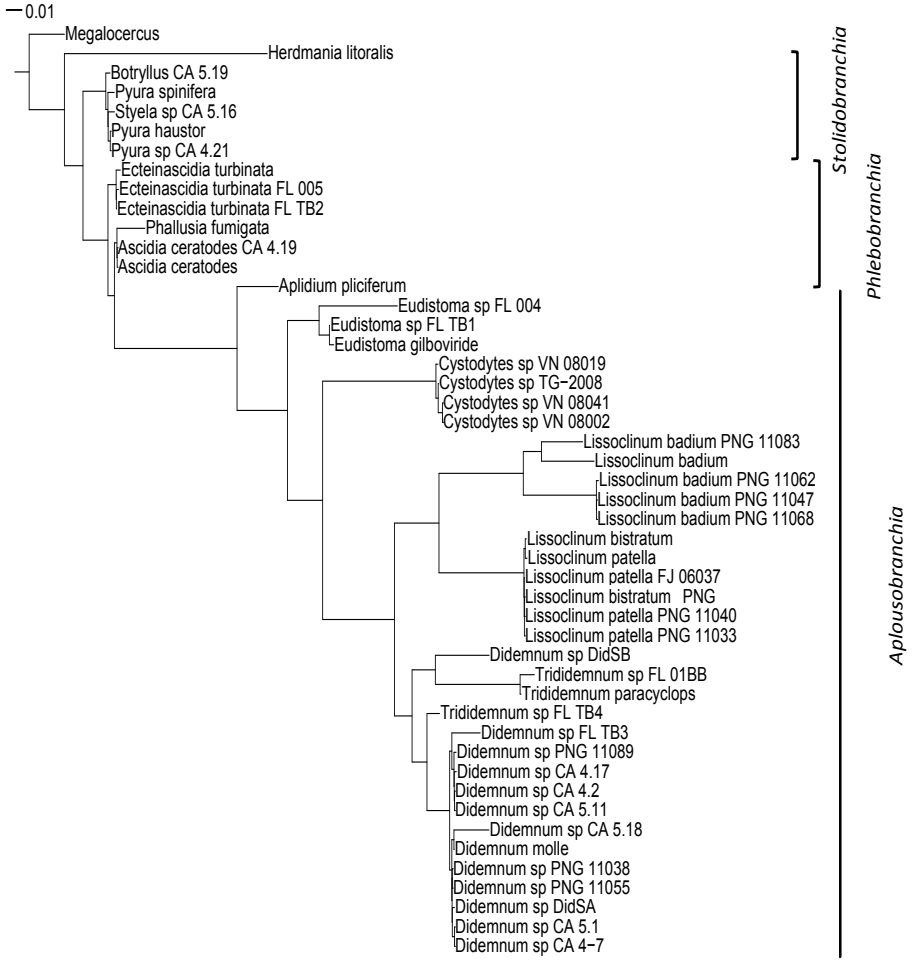


Figure S2. Abundance and order level composition of Deltaproteobacteria (A) and Gammaproteobacteria (B) showing ascidian species specificity of the most abundant groups. (C) Phylogeny of the top ten most abundant OTUs in *L. patella* and *E. turbinata* samples, arrows indicate known natural products producers. Bar plots and phylogenetic trees were generated using Phyloseq R package.

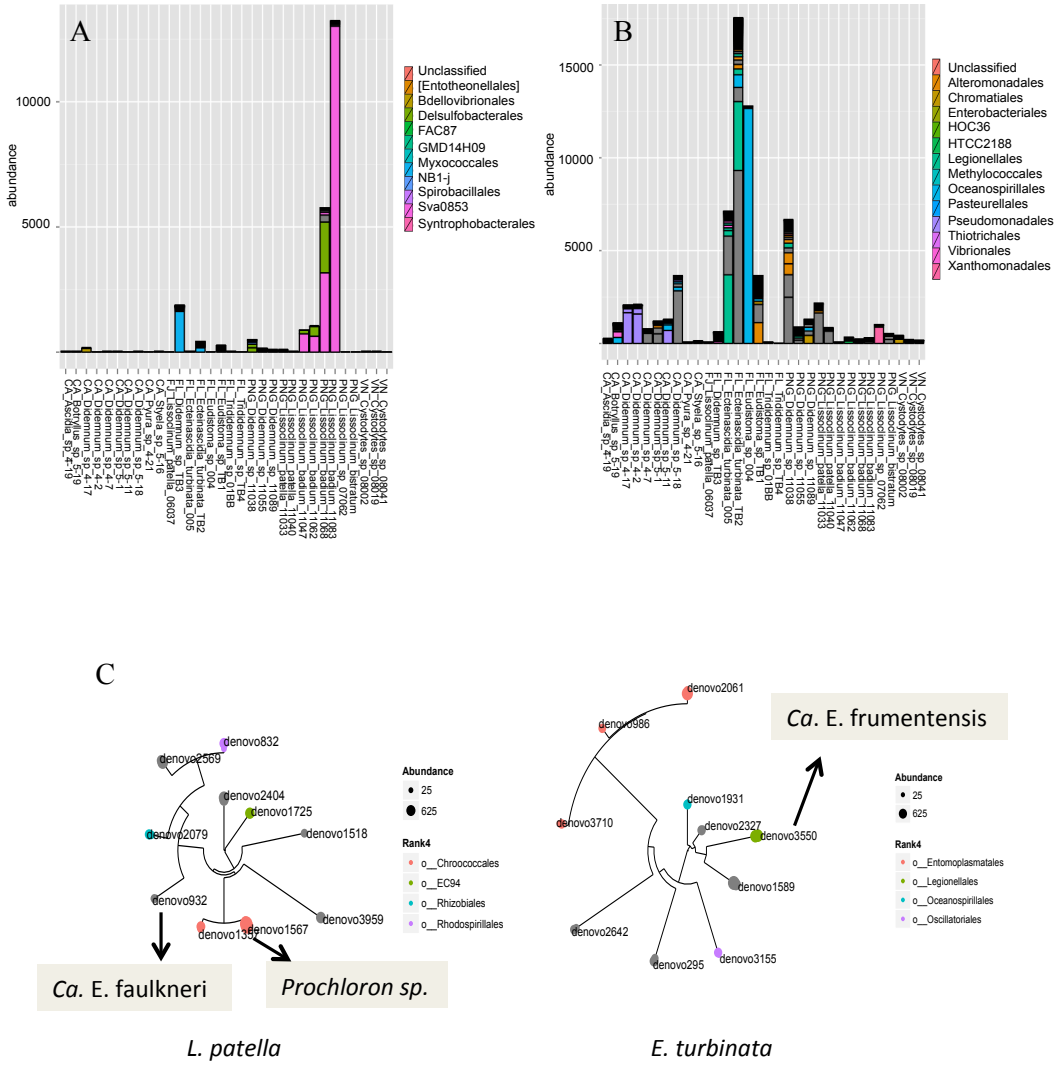


Figure S3. Effects of seasonal variation on the composition of ascidian microbiomes. The graph shows that the major phyla (filtered to top 20 percent) are maintained in ascidians collected from fall and spring of 2011 (x-axis), while the abundance of different species of bacteria varies.

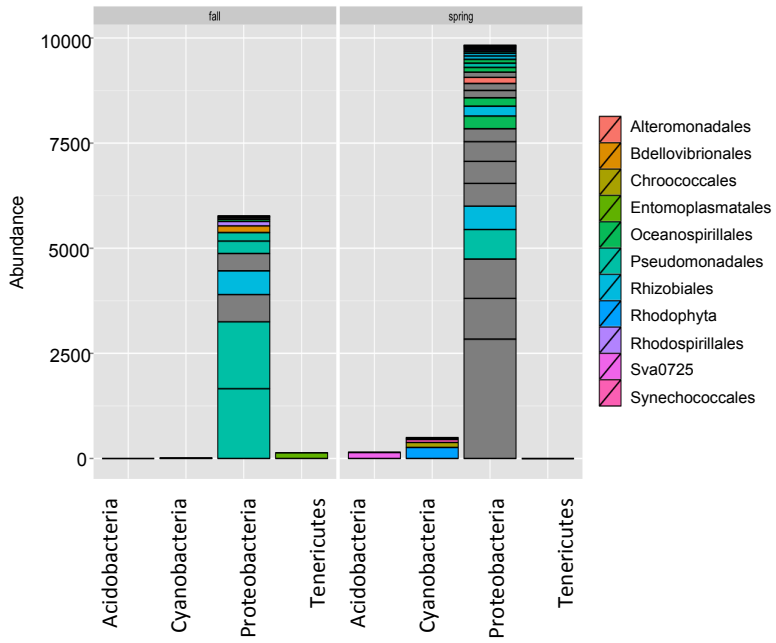


Figure S4. UniFrac-based 3D PCoA plots of all samples showing the clustering of samples according to (A) type of organisms (black: ascidians, red: mollusks) and (B) location (black: CA, green: FL, blue: PNG, light blue: VN, red: FJ).

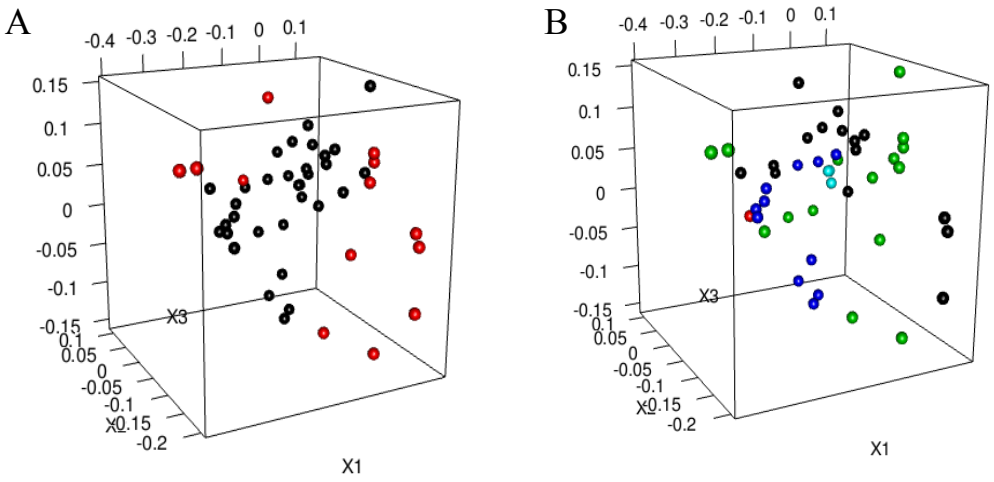


Figure S5. PCA biplots of (A) ascidian organic extracts showing correlation of samples from the same location (red arrows) with specific extract components (black print indicating retention time and m/z (B) resulting PCA biplot of assigned lipids after assignment of the LCMS data through Lipid Maps database showing the same pattern as A.

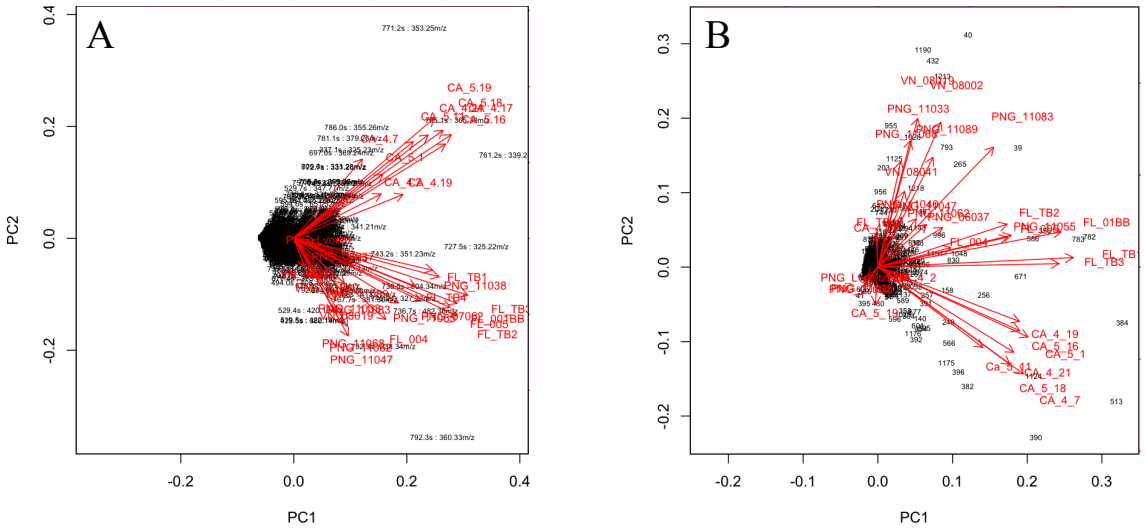
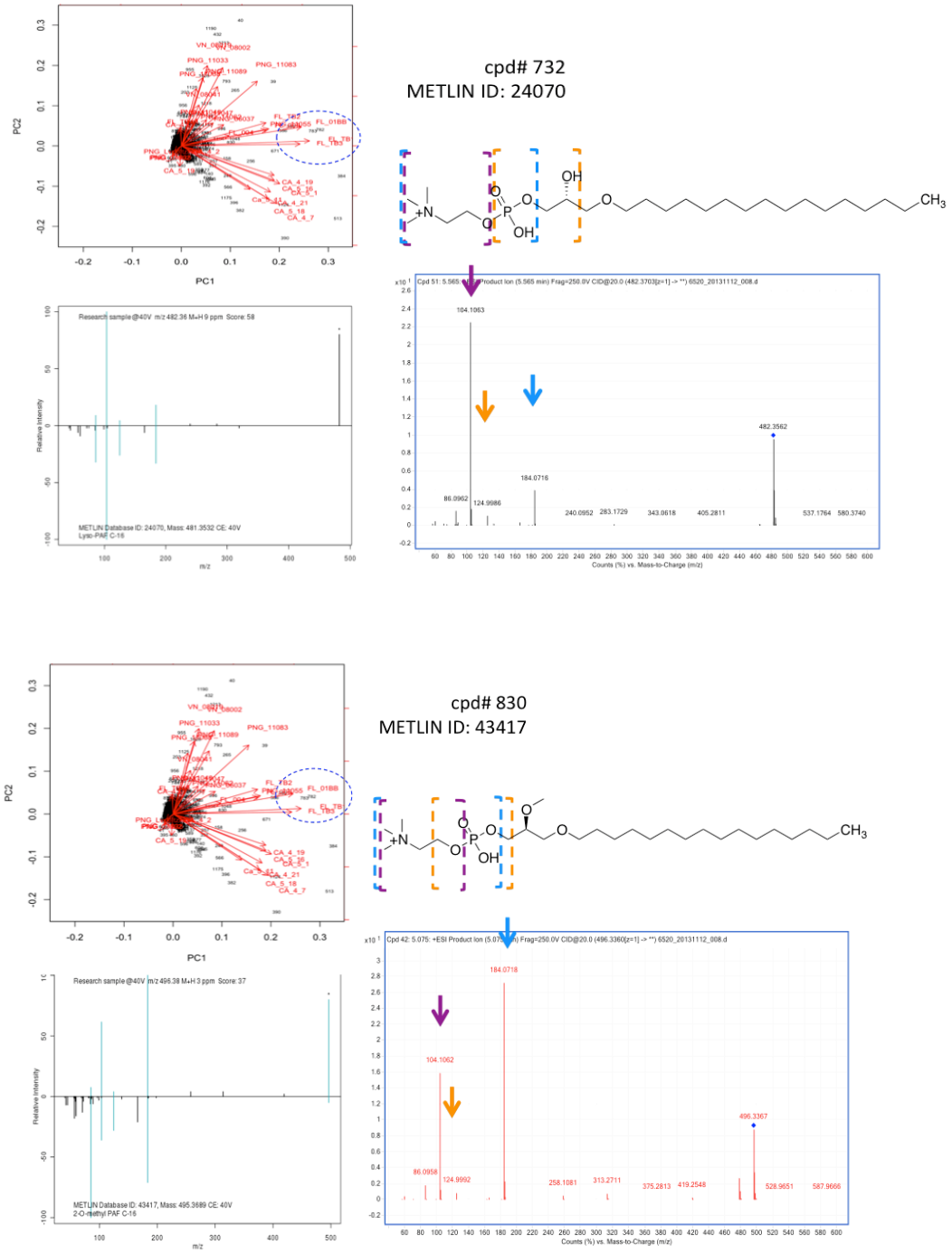
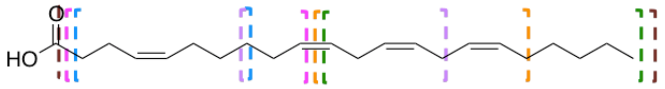
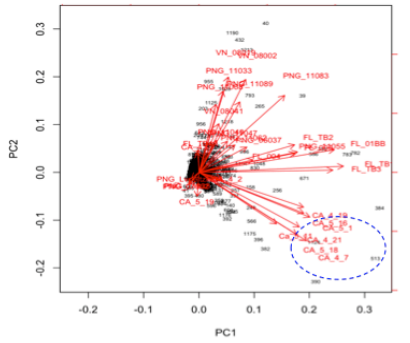


Figure S6. Example LC-MS/MS spectra of lipids that were correlated with extracts from different locations (PCA biplots are shown on the left panels). Compounds or compound families were identified by comparison of MS/MS fragmentation patterns and accurate masses to compounds in METLIN database (right panels).



METLIN ID: 45057





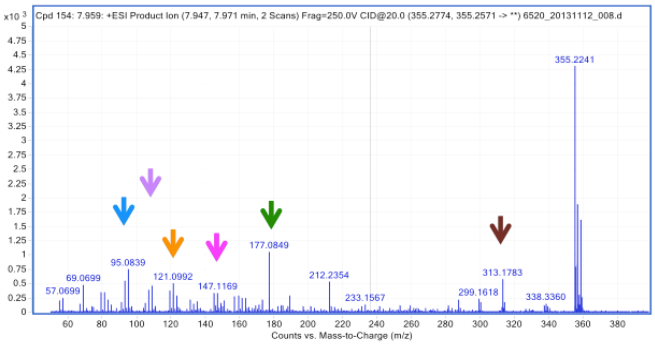


Figure S7A. PCA of samples from Papua New Guinea showing a clustering driven by secondary metabolites.

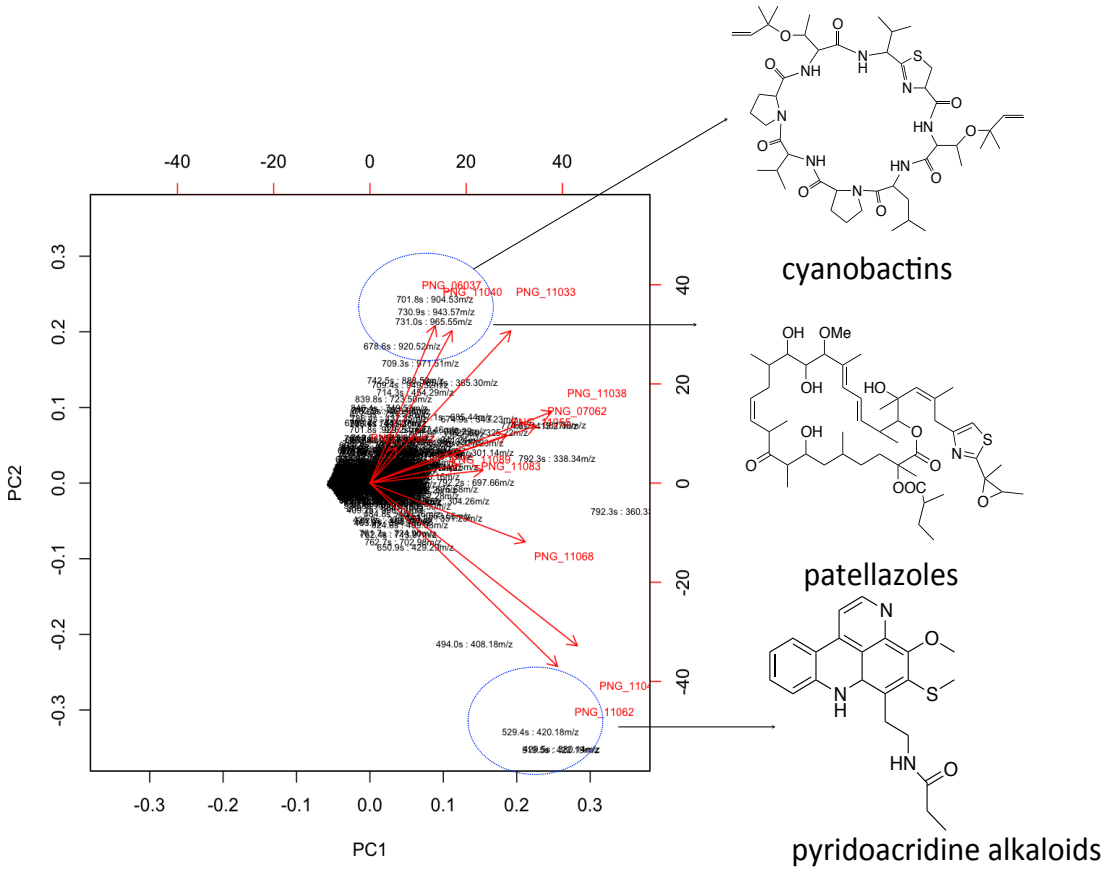


Figure S7B. Spectrum and selected ion chromatogram of varamine A from *Lissoclinum badium* samples.

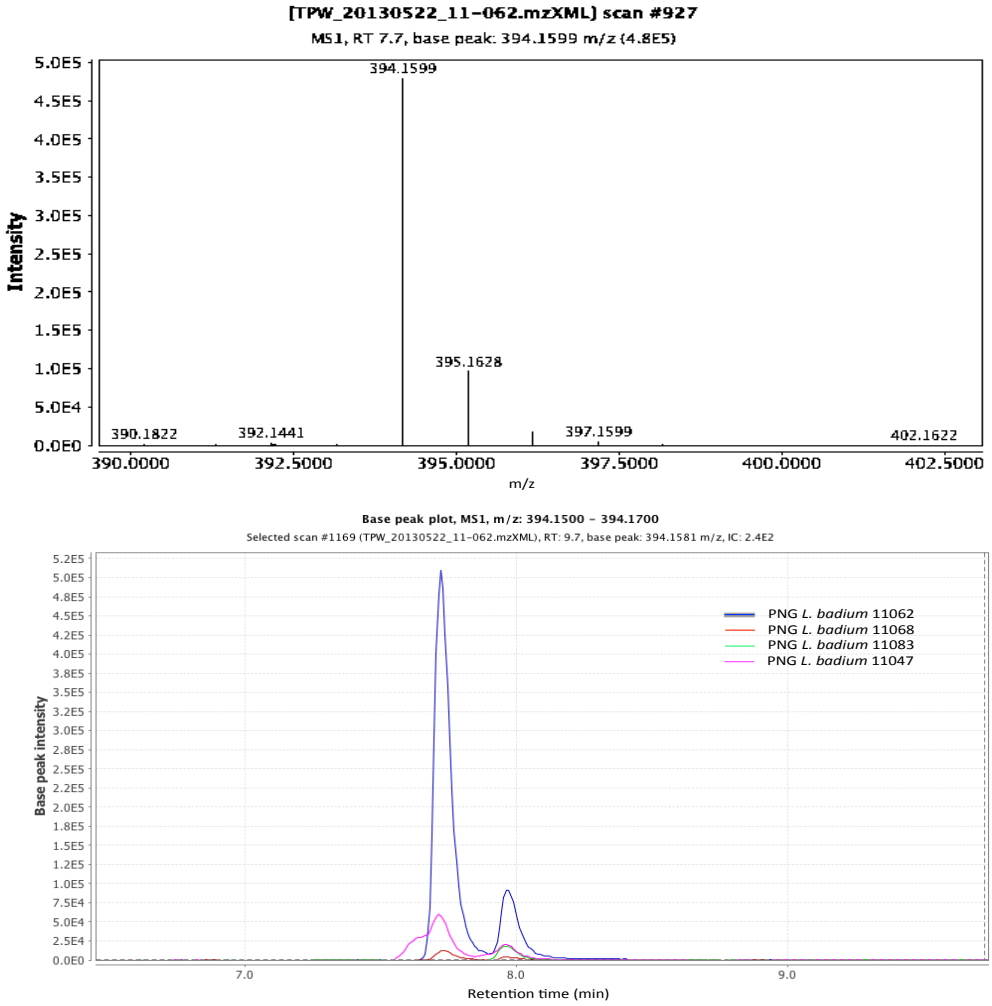


Figure S7C. Spectrum and selected ion chromatogram of varamine B from *Lissoclinum badium* samples.

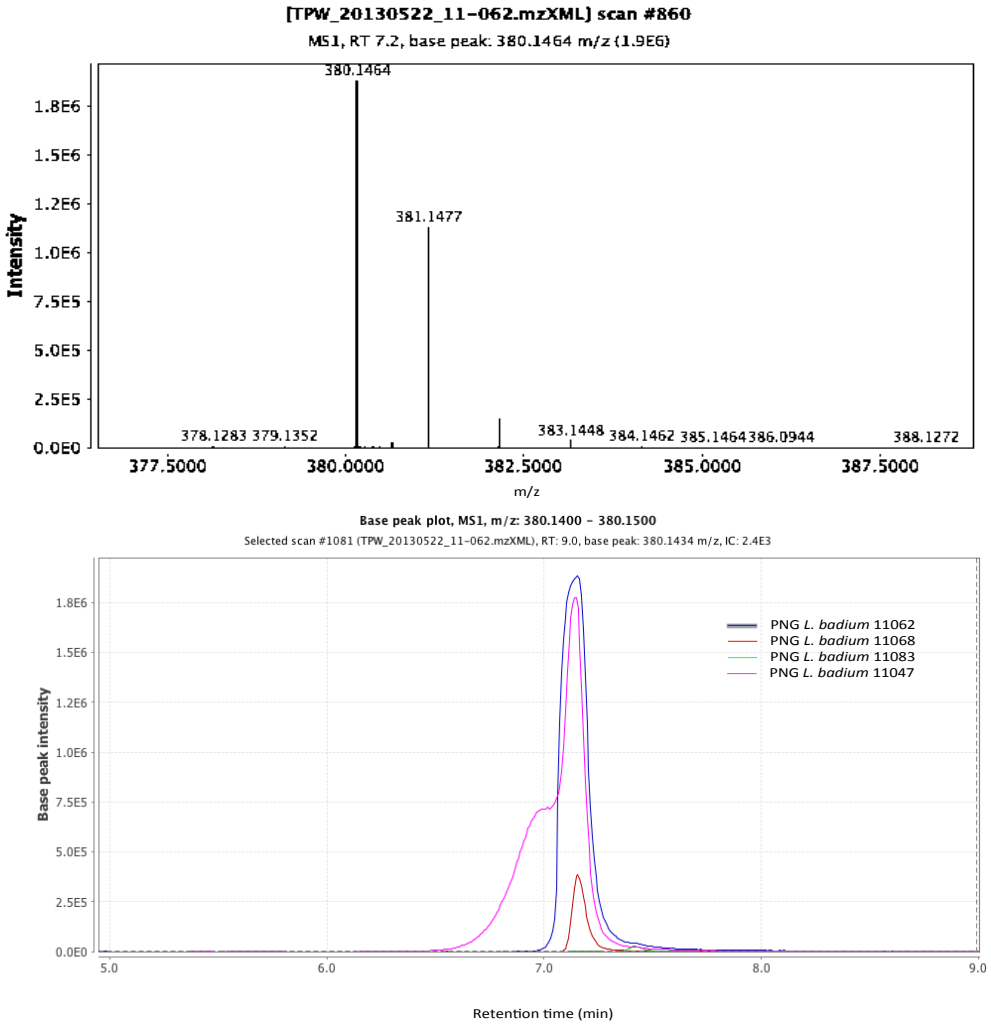


Figure S7D. Spectrum and selected ion chromatogram of diplamine from *Lissoclinum badium* samples.

\

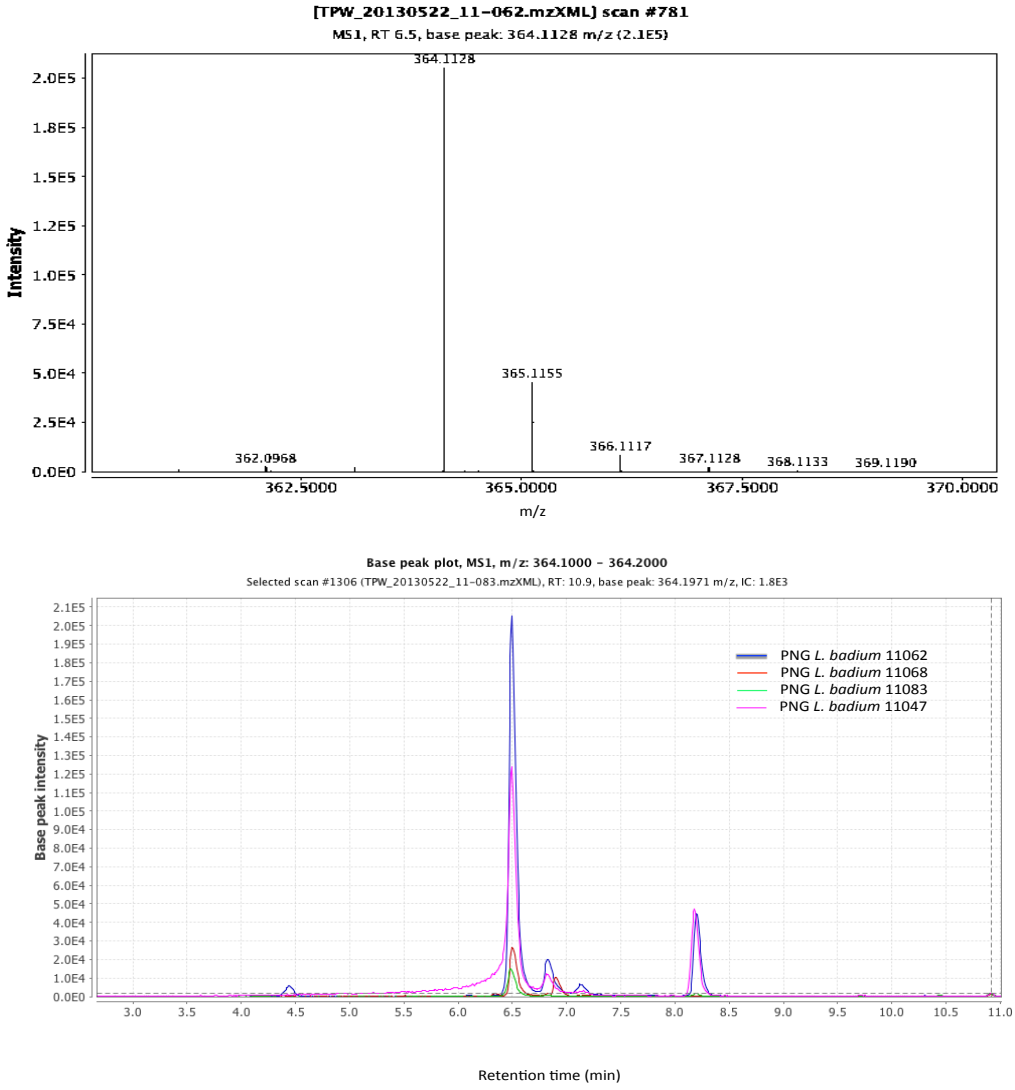


Figure S7E. Spectrum and selected ion chromatogram of lissoclin B from *Lissoclinum badium* samples.

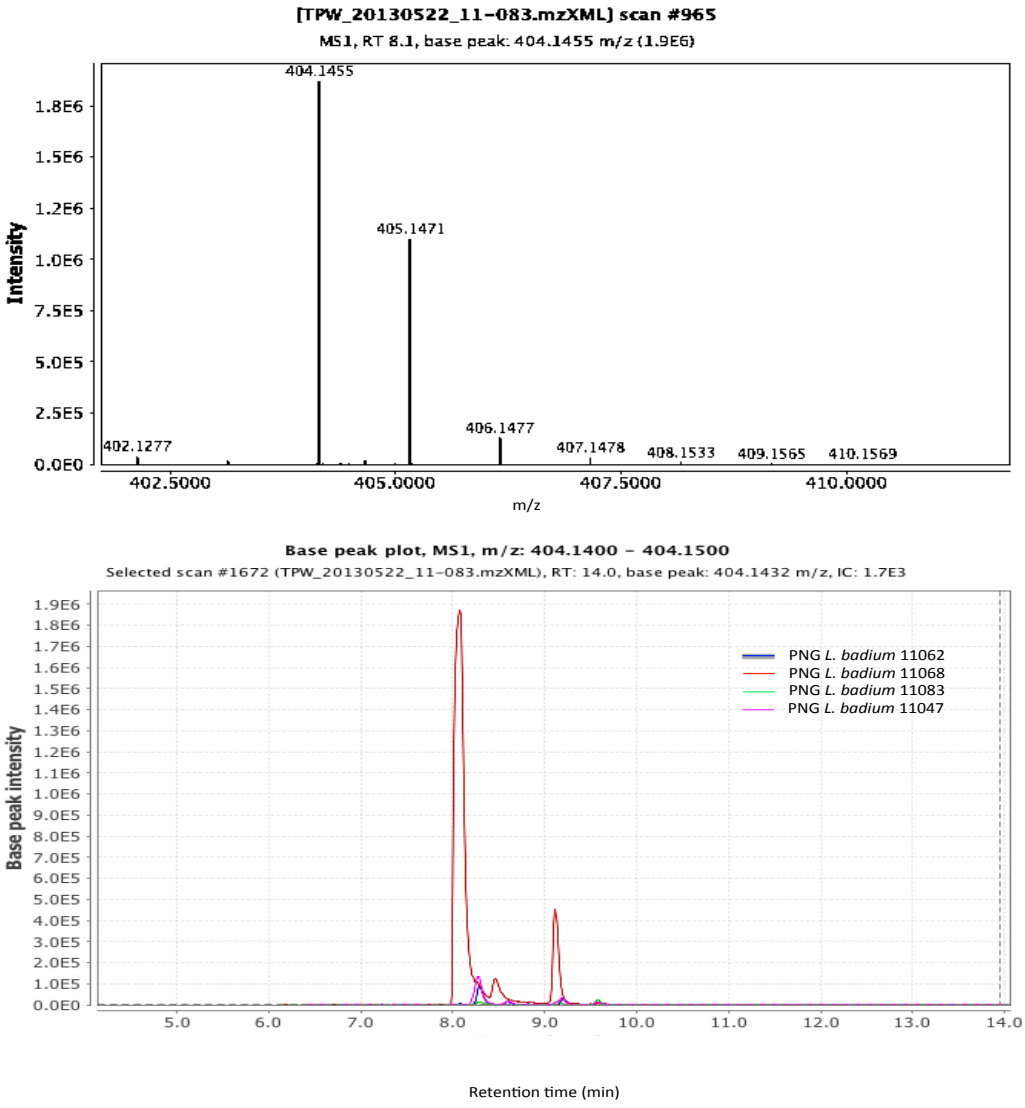


Figure S7F. Spectrum and selected ion chromatogram of patellazole A from *Lissoclinum patella* samples.

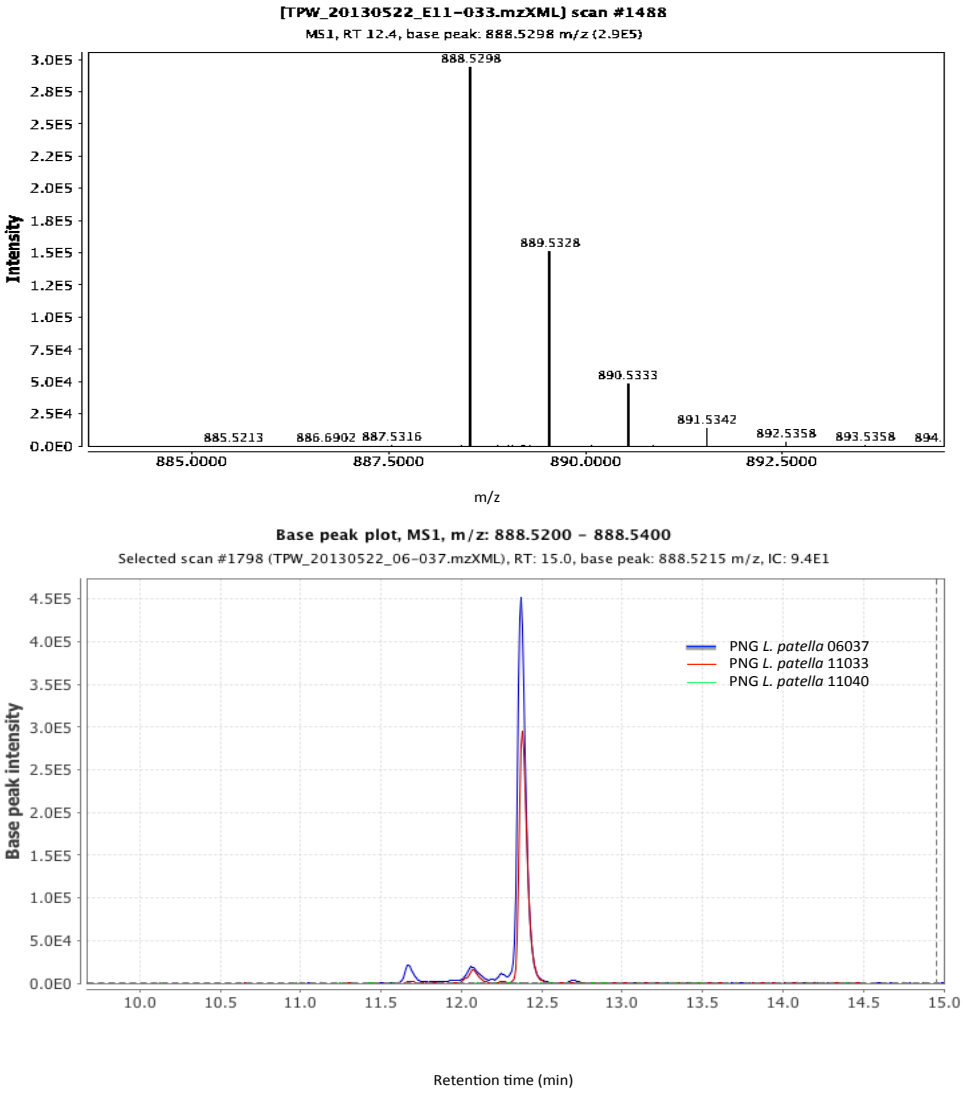


Figure S7G. Spectrum and selected ion chromatogram of patellin 3 from *Lissoclinum patella* samples.

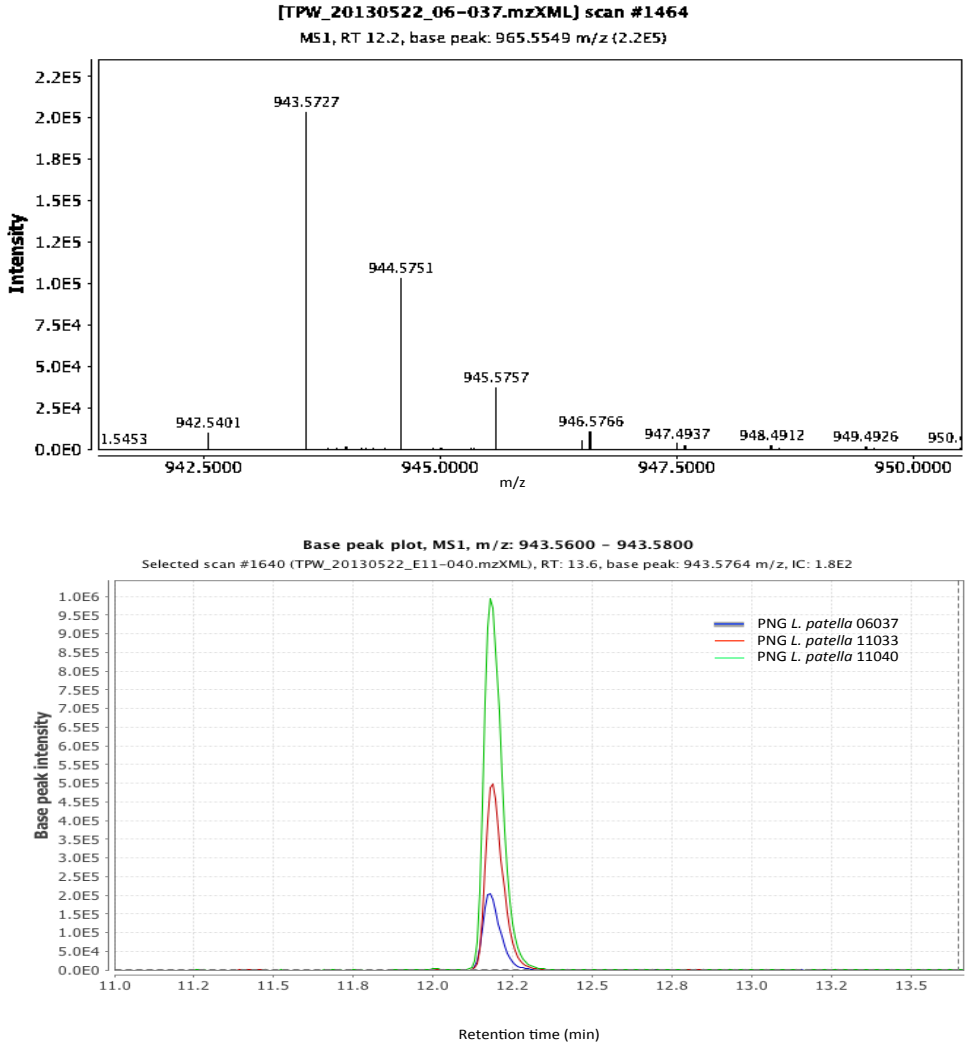


Figure S7H. Spectrum and selected ion chromatogram of cystodytin A from *Cystodytes* sp. samples.

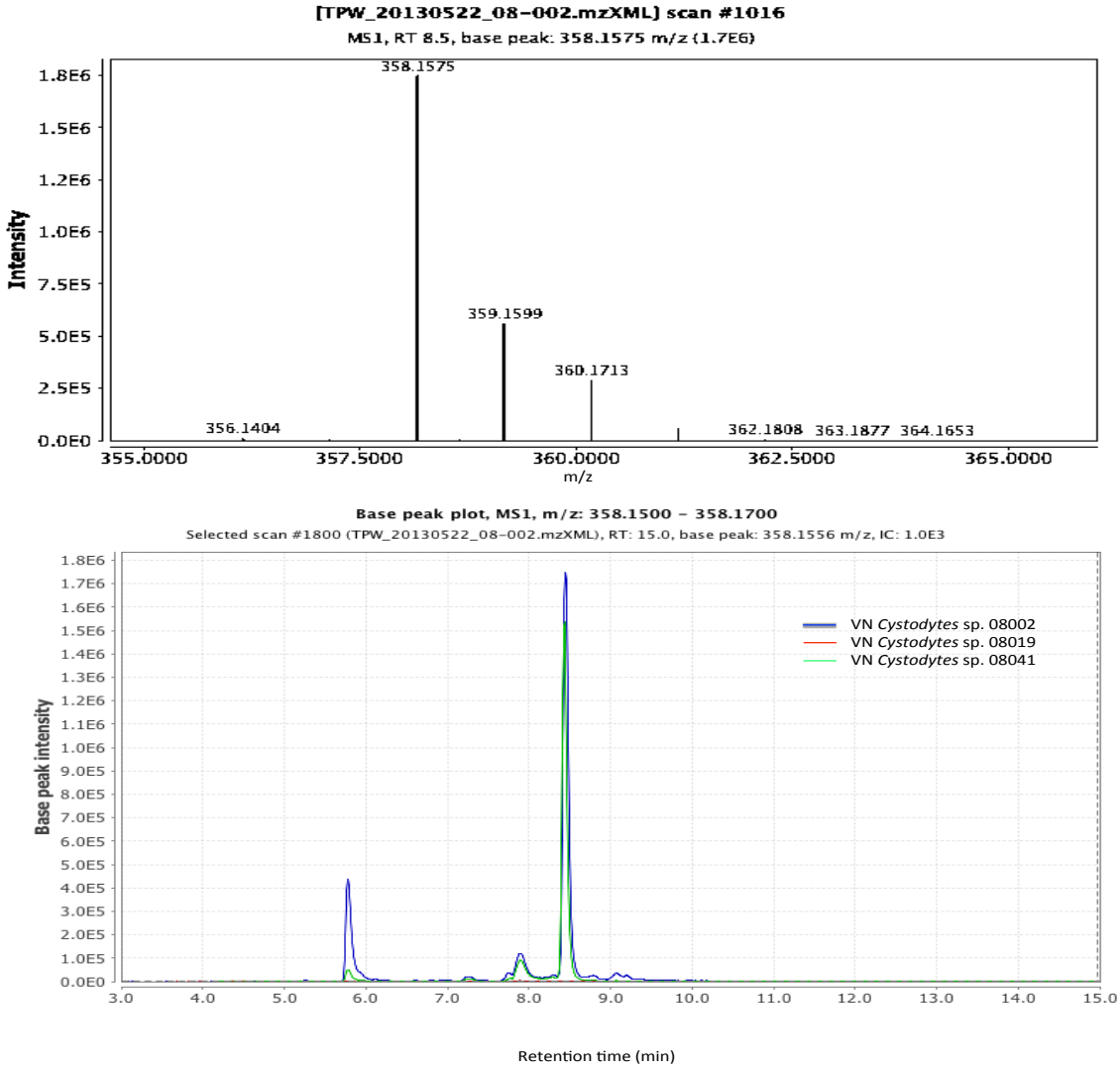


Figure S7I. Spectrum and selected ion chromatogram of cystodytin A from *Cystodytes sp* samples

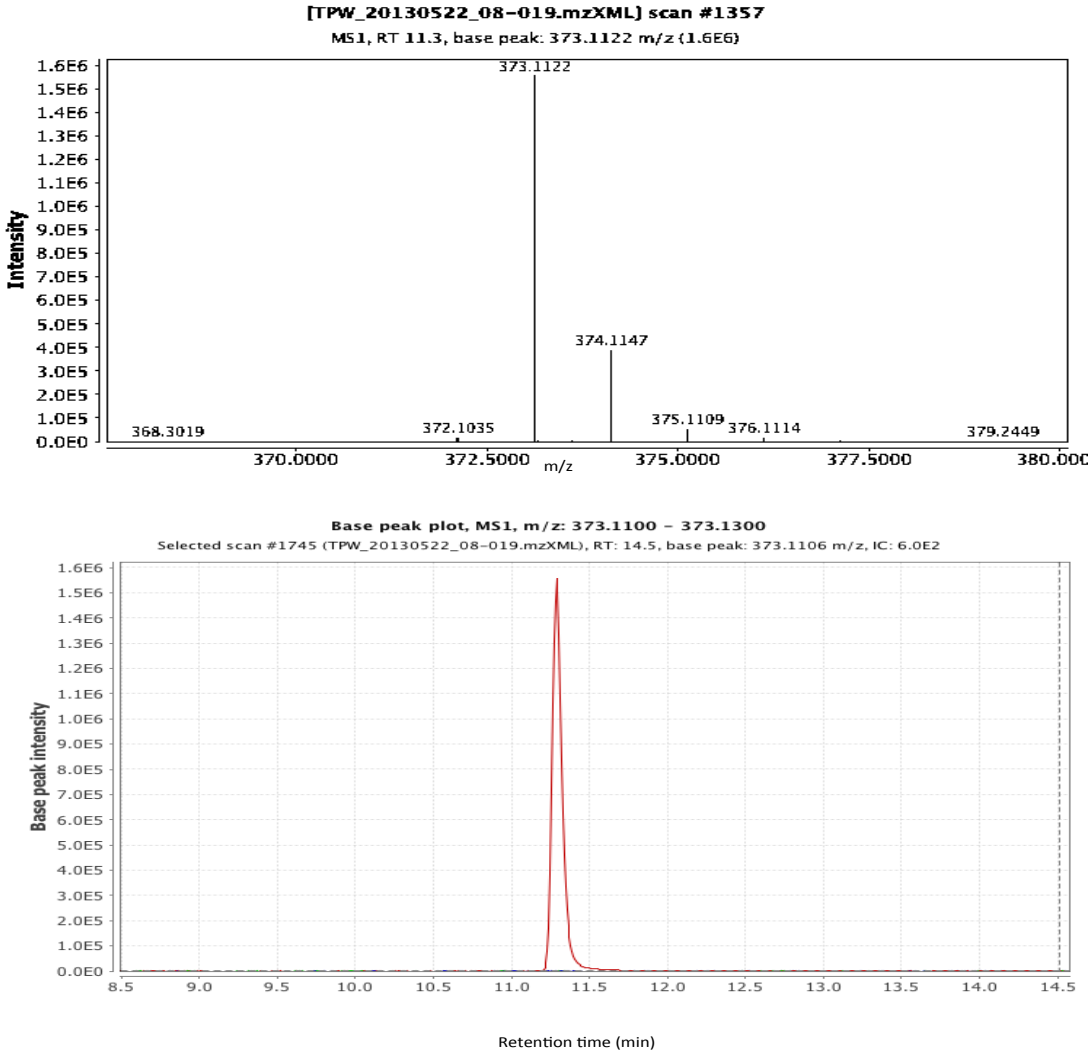


Figure S8. Bioactivity profile of ascidian extracts showing the distribution of bioactivity across different locations. Activities are shown as percent growth inhibition relative to standard compounds (see methods section).

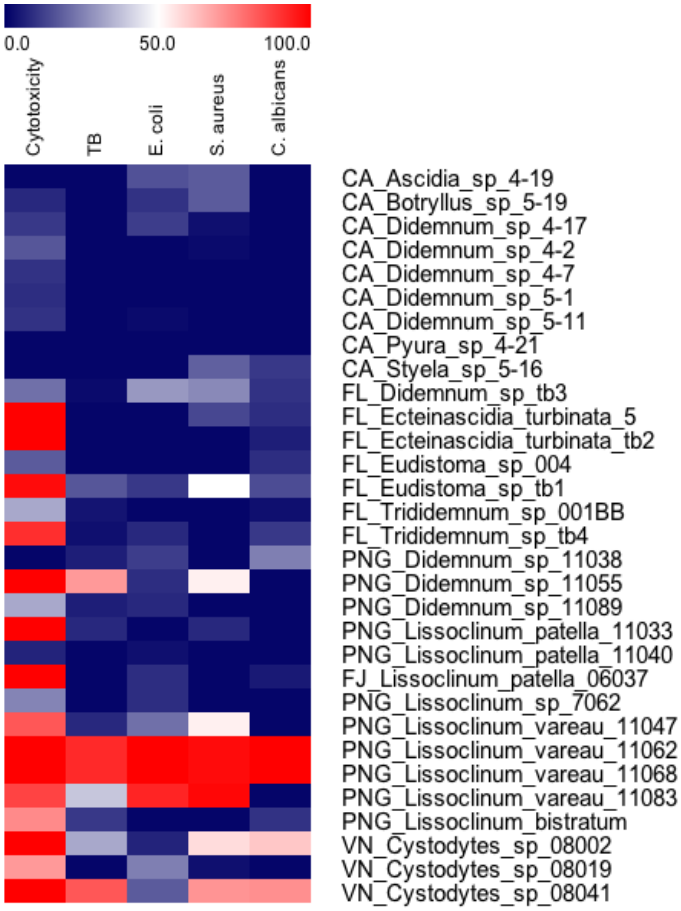
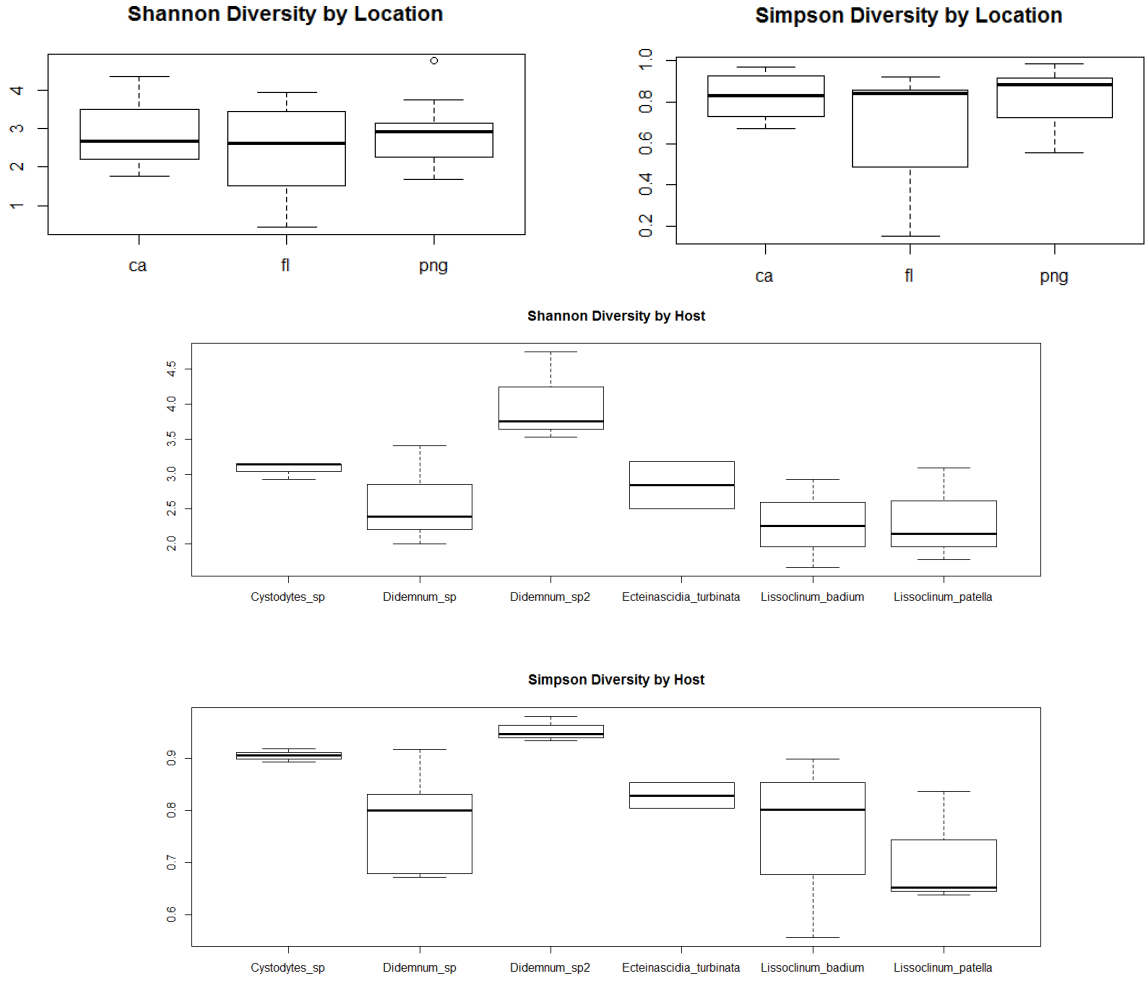


Figure S9. Group diversity analysis using the Shannon and Simpson indices. Calculations were made using the mcpHill function in the simboot R package. This function reports multiplicity adjusted p-values from distribution free tests (see methods for details).

Location (number of samples)	Pvalue(Shannon)	Pvalue(Simpson)
CA (10)	0.9958	0.9972
FL (7)	0.9564	0.6014
PNG (15)	0.9980	0.8104
Host (number of samples)		
Cystodytes sp. (3)	1.0000	0.9994
Didemnum sp. (6)	0.7678	0.6856
Didemnum sp._2 (3)	0.0082	0.0072
Ecteinascidia turbinata (2)	1.0000	0.9824
Lissoclinum badium (4)	0.6970	0.8482
Lissoclinum patella (3)	0.8678	0.7086



CHAPTER 3

METABOLIC ENGINEERING TO OPTIMIZE HETEROLOGOUS RIPP PRODUCTION IN *E. COLI*

Ma. Diarey B. Tianero,[1] Elizabeth Pierce,[1] John A. McIntosh,[1] Shrinivasan Raghuraman,[2] Baldomero M. Olivera,[2] Eric W. Schmidt[1]

[1]Department of Medicinal Chemistry, L.S. Skaggs Pharmacy Institute, University of Utah, Salt Lake City, UT, USA, [2] Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

Note: My contribution to this study was in design of the project, heterologous expression experiments and analysis, interpretation of results, and writing of the manuscript.

Abstract

The limited supply of natural products remains the single most important bottleneck in natural products drug discovery and development. Heterologous expression of biosynthetic pathways is often perceived as a ready solution to supply natural products from uncultivable microbes, but obtaining useful quantities of metabolites from heterologous expressions is far from straightforward, especially for complex, multistep pathways. Cyanobactins are among the few metabolites from 'unculturable' bacteria that have been successfully expressed in heterologous hosts. Cyanobactins belong to the growing class of ribosomally synthesized and posttranslationally modified peptides (RiPPs). One notable feature of RiPPs is that in some cases, combinatorial libraries of active compounds can be generated via the straightforward mutagenesis of precursor peptides. Cyanobactins are exemplary in this respect, as a recent study demonstrated that $>10^6$ compounds could be produced in this fashion. However, supply issues persist even in heterologous expression experiments, with typical titers ranging from 10-100 μg/L, with the highest reported yield 300 μg/L. Here we present a greatly improved platform for the production of cyanobactins in *E. coli*, resulting in a 100-fold increase in titer. To showcase this method, we apply heterologously produced cyanobactins to a broad phenotypic assay on mice dorsal root ganglia (DRG) neurons, and discover a new activity for the known natural product, patellin 2.

Introduction

During the past decade, natural products discovery has shifted from purely chemical isolation approaches to a hybrid of chemical and bioinformatic methods.[1-3] As a result, the number of biosynthetic pathways for both known and unknown compounds has vastly increased. However, the ability to rapidly discover novel compounds has thrown an old problem back into focus, namely that of how we can supply sufficient quantities of natural products for practical drug discovery and development. This problem is particularly severe for metabolites discovered in rare animals, or for those metabolites predicted from genome sequences. [4] We have provided an important and instructive example of how this problem can be addressed in our work on the discovery, characterization, and heterologous expression of cyanobactin pathways.[5-11] Cyanobactins are a broadly-distributed group of small (usually 6-10 amino acid) posttranslationally modified peptides with diverse bioactivities including anticancer,[12,13] antimalarial ,[14] and antiviral [15] activities.

In RiPP biosynthesis, the amino acid sequence of the resulting natural product is encoded on a precursor peptide that, after ribosomal synthesis, is modified by enzymes to add chemical diversity, or to render peptide structures more drug-like. [16] The modifying enzymes generally exhibit relaxed substrate specificity, such that mutations in the precursor peptides are easily tolerated, leading to new compound derivatives.[17,10,18] The biosynthetic steps in cyanobactin maturation have been characterized extensively. The promiscuity of the modifying enzymes in the cyanobactin pathways allows for the generation of libraries of derivatives, both natural[6 19] and engineered. [6,10,20] Assessment of the capacity of a cyanobactin pathway to generate compound libraries *in vivo* projects

that a simultaneous five-site saturation mutagenesis library of one precursor sequence could produce 2 x 10^5 cyanobactin derivatives.[20]

The most studied cyanobactin pathways to date are derived from unculturable cyanobacterial symbionts of marine tunicates of the genus *Prochloron*.[21,22] Given this obstacle to traditional fermentative approaches of natural product production and isolation, the alternative routes of chemical synthesis, heterologous expression, or *in vitro* synthesis are presently the only means of supplying cyanobactins. Considering that the medicinal appeal of cyanobactin pathways is tied to the possibility of creating cyanobactin libraries, heterologous expression is clearly preferable to chemical synthesis or *in vitro* reconstitution, as it allows the screening of huge libraries in easily transformable *E. coli* strains. [20] Accordingly, the cyanobactin pathways have been successfully manipulated in *E. coli*, and the potential of this heterologous system to produce novel derivatives has been demonstrated repeatedly. [23, 17,20] However, the yields are insufficient for practical purposes (drug screening, pharmacokinetics, animal testing etc.), as the best reported fermentation yield of cyanobactins is 300 μg/L.

In this work, we describe a greatly improved platform for the production of cyanobactins in *E. coli*. We achieve this through a combination of both exogenous and endogenous supply of additives and precursors. Specifically, we find that supplementing cultures with L-cysteine and co-expressing a previously reported heterologous mevalonate pathway[24] to increase isoprene availability lead to significant increases in compound production; these approaches combine synergistically to net a 100-fold increase in cyanobactin titer. We further demonstrate that mevalonate pathway coexpression improves the yield of both prenylated and unprenylated cyanobactins. We

showcase the usefulness of this expression system with the heterologous production and resulting discovery of the activity of the natural cyanobactin patellin 2 on dorsal root ganglia (DRG) neurons.

Results

Optimization of cyanobactin expression conditions

In the course of our optimization of the production of cyanobactins in *E. coli*, we have identified many different parameters that affect compound production. These include the volume of the liquid cultures, fermentation time, temperature, as well as different additives including amino acids, redox reagents, and metals. Previous publications have explored various strategies (i.e., different cell lines, promoters, media) to optimize cyanobactin expression. [23, 20] We compared expressions from 6 mL, 10 mL, 100 mL, and 250 mL expressions. We also performed 1L fermentations in flasks and in a 1 L fermentor. We determined that 6 mL cultures in 24-well plates gave us the maximum amount of compounds per volume of expression. We also compared the number days of expression from 2 to 7 days and determined that the maximum amount of compounds was obtained from 5 day expressions. This experimental set-up also allowed us to study the expressions under many different conditions in replicate experiments reducing the preparation and extraction time as well as materials cost.

We next examined how cyanobactin yield was affected by growth media composition as well as additives including metals, redox agents, amino acids and different cofactors (Table 3.1). We first found that a mixture of methionine and cysteine increased compound production. Further analysis with the individual amino acids showed

Table 3.1. Additives used in the initial optimization of production in *E. coli*

additive	concentration
glucose/glycerol	0.05%/0.5%
lactose	0.30%
glycerol	0.50%
amino acids	mix (17)
Fe(II)	1mM
Cu(II)	1mM
Zn(II)	1mM
MnCl2	1mM
Met/Cys	30mM(mix)
FMN	100nM

that cysteine was responsible for the observed effect, and leads to a ~10-fold increase in compounds by LC/MS (Figure 3.1). We therefore added cysteine (5 mM) in all succeeding experiments. The mechanism of increase in compound production due to cysteine will be described elsewhere.

Limited supply of isoprene precursors in *E. coli*

We previously observed that production of cyanobactins in *E. coli* yielded a mixture of cyanobactins with either zero, one, or two isoprene units added to Ser and Thr residues.[10 20] This is also apparent in Figure 3.1. We hypothesized that this was due to the limited availability of the necessary isoprene precursor, DMAPP in *E. coli*, and that increasing the intracellular DMAPP concentration would increase our yield of the desired, doubly prenylated cyanobactins. To increase supply of DMAPP, we co-expressed the second half of the mevalonate pathway (*mev*) from *Saccharomyces cerevisiae* that was previously engineered in *E. coli* for the production of amorphadiene, a precursor to the antimalarial drug artemisinin.[24] When using only the latter half of the *mevalonate* pathway, mevalonate must be supplied exogenously, as *E. coli* does not

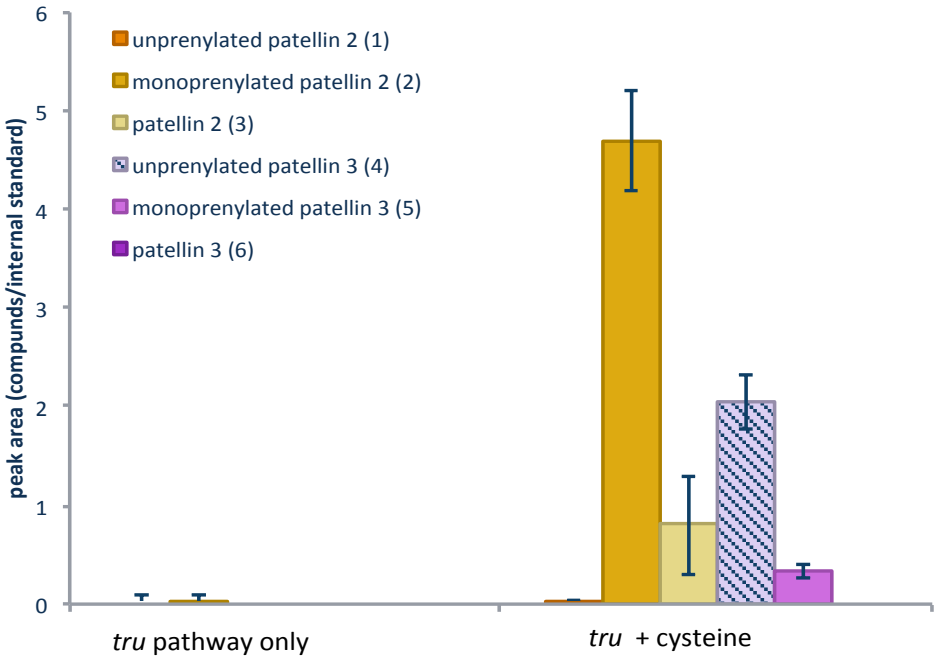


Figure 3.1. Increase in cyanobactin production upon addition of cysteine (5mM). Labels indicate the cyanobactin derivatives (compound number).

produce this metabolite. Our new expression system therefore involved the coexpression of *tru* with the mevalonate pathway (*mbi)* pathway. We supplemented these coexpressions with both mevalonate and cysteine (Figure 3.2). As anticipated, we observed an increased proportion of prenylated products in *tru/mbi* coexpressions (with added cysteine) when mevalonate was added (Figure 3.3). To further confirm that this was due to the addition of mevalonate, we fed ^{13}C-labeled mevalonate to *tru/mbi* coexpressions. Mass analysis showed that the labeled ^{13}C-labeled mevalonate was converted to DMAPP and was incorporated into the final compounds as evidenced by a +2 Da shift observable by LC-MS (Figure 3.4). This was also apparent in the masses of the fragment ions, showing the loss of one and two ^{13}C labeled isoprene units (inset). This demonstrated that the *mevalonate* pathway can be functionally coexpressed with the

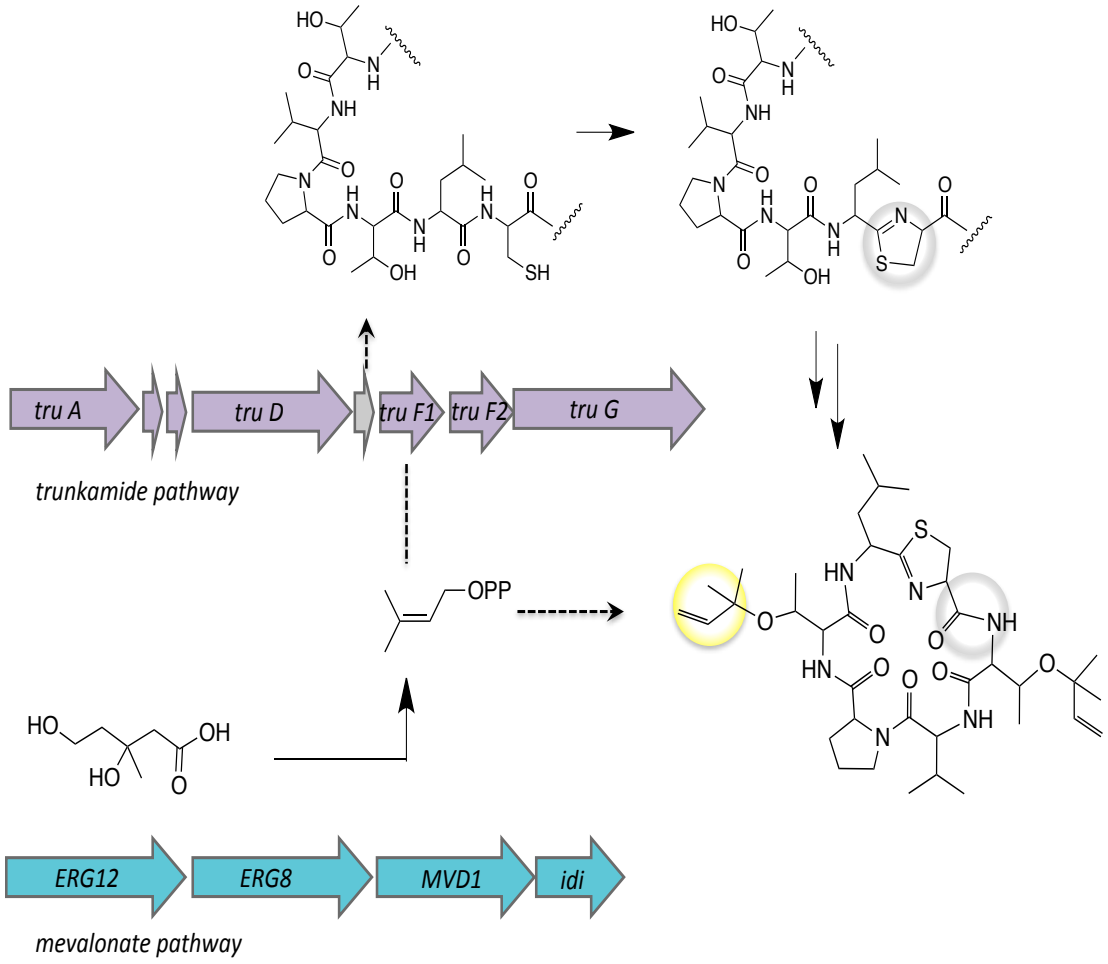


Figure 3.2. Expression system design using the combination of *tru* and *mevalonate* pathways for the production of fully prenylated cyanobactins

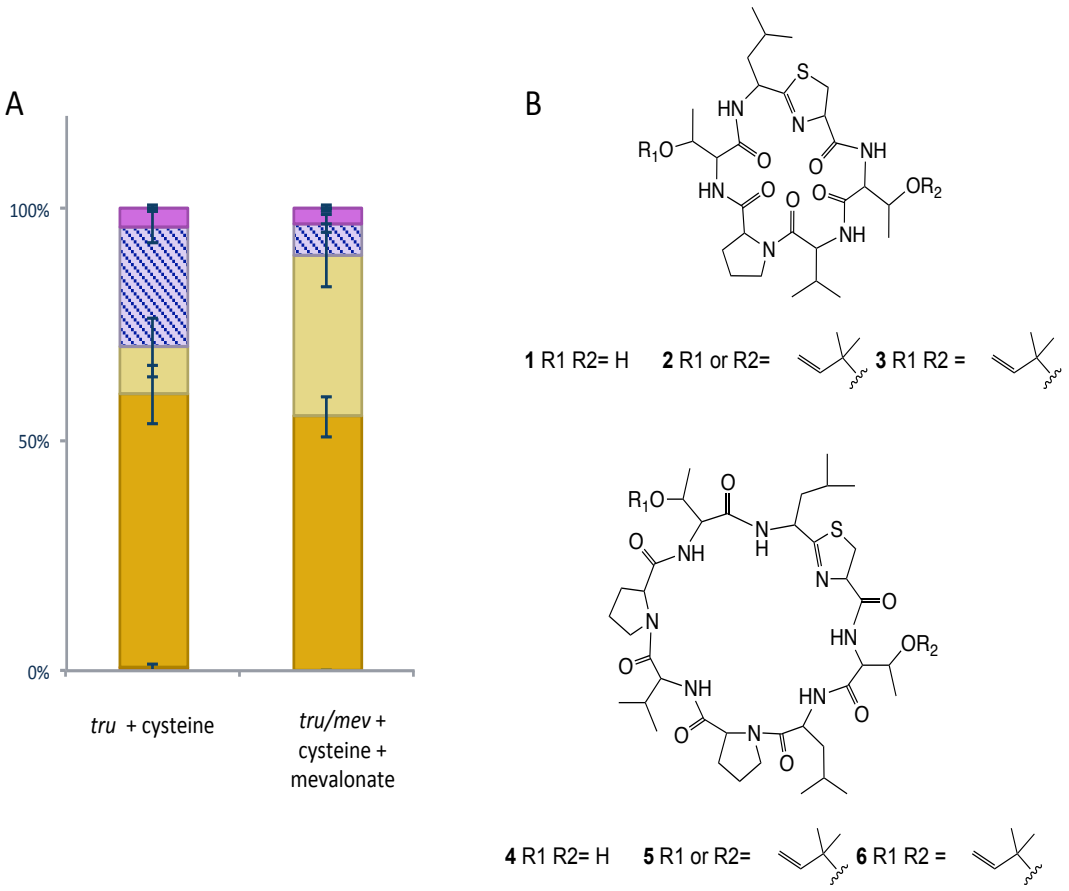


Figure 3.3 Limited amount of isoprene precursors in *E. coli*. A) Increase in prenylation levels in the presence of *mevalonate* pathway in *mbi* plasmid and mevalonate (10 mM). B) Structure of cyanobactins produced in *E. coli*.

tru pathway *in vivo* and can be used to increase the fraction of fully prenylated cyanobactins.

Effect of mevalonate pathway coexpression on cyanobactin titer

Although *mbi*-coexpression had the desired effect of increasing the fraction of fully prenylated compounds, we were somewhat concerned that metabolic stress due to coexpression of the *mevalonate* pathway might decrease overall cyanobactin titer. Thus, we were delighted to find that *mbi*-coexpression also resulted in a robust increase in the

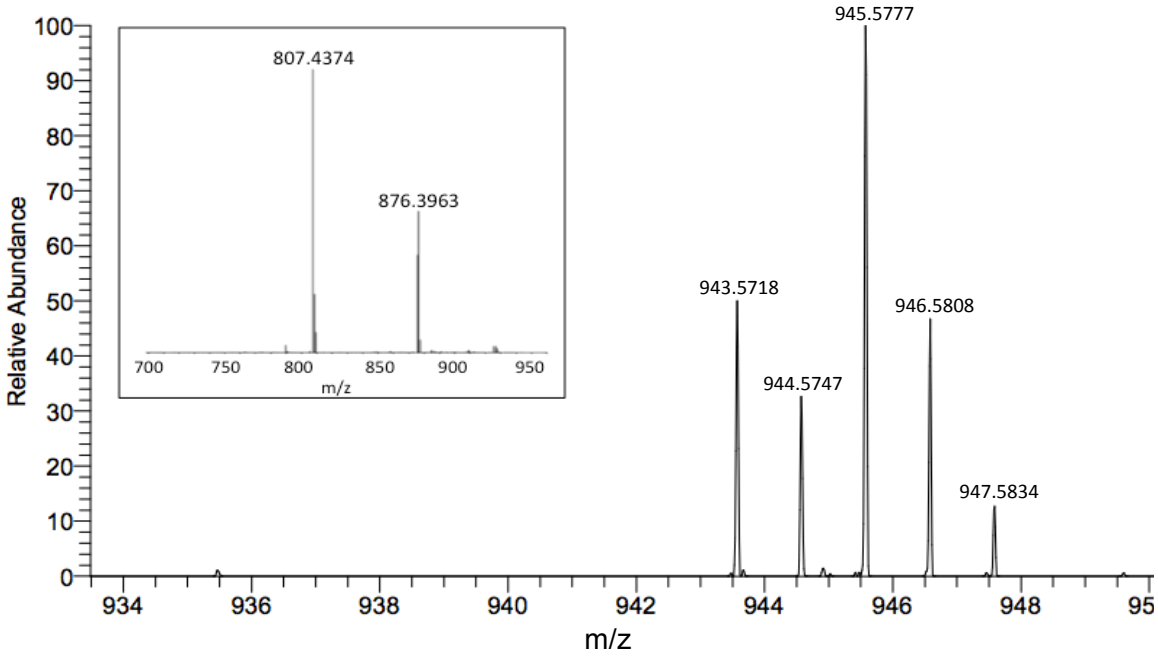


Figure 3.4. Incorporation of ^{13}C-DMAPP into cyanobactin prenyl groups; inset shows loss of a labeled prenyl group from MS fragmentation. The peak at 943.5718 m/z shows the unlabeled patellin 3 and the 945.5777 m/z shows the incorporation of two ^{13}C-DMAPP in patellin 3.

total amount of compounds. In the presence of mevalonate (20 mM) and cysteine (5 mM) (in *tru/mbi* coexpressions), the total amount of cyanobactins was increased consistently by about 100-fold compared to *tru* only expressions.

To test whether the increased titers were due to the products of the mevalonate pathway (i.e., increased DMAPP levels), and not from background effects due to coexpression with the *mbi* plasmid, we added increasing amounts of mevalonate into the *tru/mbi* expressions from 1 to 40 mM. A clear mevalonate-dependent increase in the total amount of cyanobactins was observed (Figure 3.5F-I). In expressions with 40 mM mevalonate, the total amount of compounds was estimated at 27 mg/L, nearly 100-fold higher than in previous reports. Notably, this strong increase in titer was only observed in

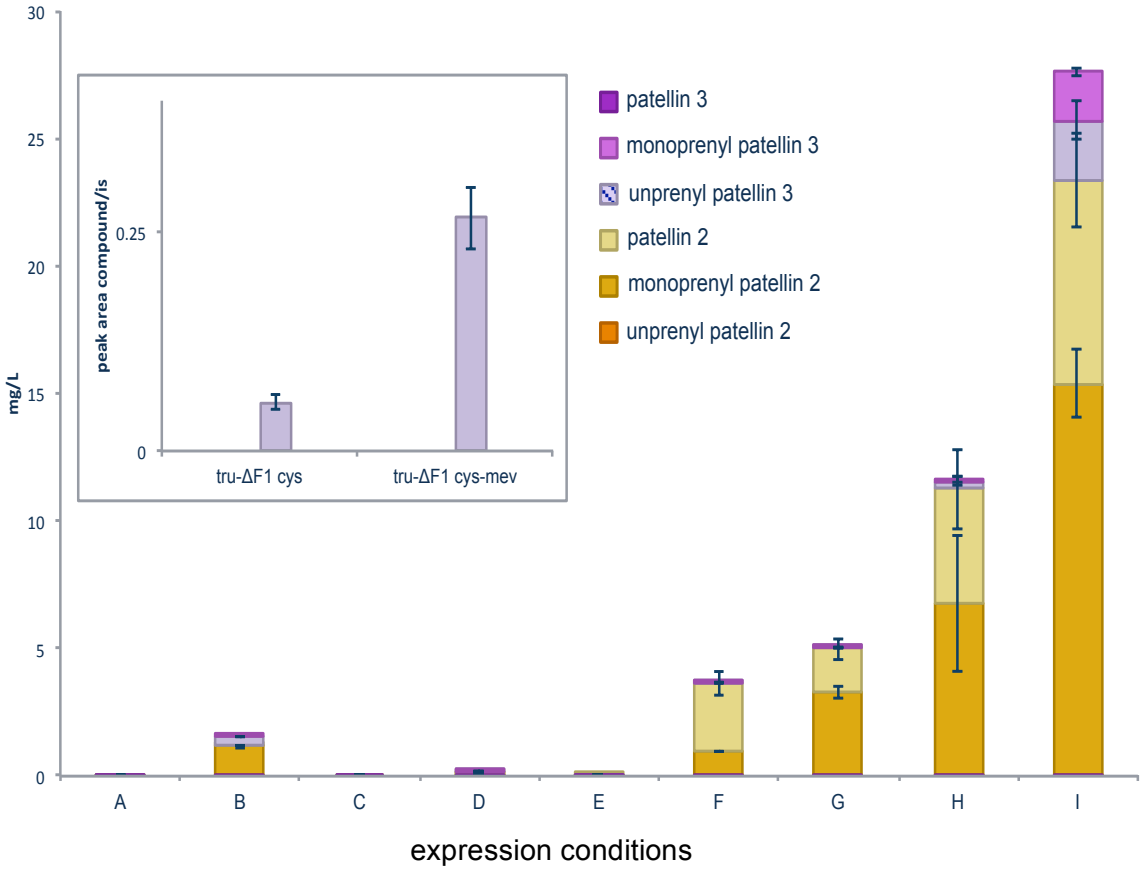


Figure 3.5. Increased cyanobactin production in the presence of mevalonate pathway and mevalonate. Y-axis shows that amount of total compounds produced in each expression condition (columns) in mg/L. Column A) *tru* B) *tru* + 5mM cys, C) *tru/mbi*, D) *tru/mbi* + 5mM cys, E) *tru/mbi* +10mM mevalonate, F) *tru/mbi* +5 mM cys + 5 mM mevalonate, G) *tru/mbi* + 5 mM cys + 10 mM mevalonate, H) *tru/mbi* + 5 mM cysteine + 20 mM mevalonate, I) *tru/mbi* + 5 mM cysteine + 40 mM mevalonate. Inset: comparison of *tru* and *truDF1* knockout expression showing the relative amounts of unprenylated patellin 3

the presence of both cysteine and mevalonate in *tru/mbi* expressions. Cysteine by itself consistently increased compound production, but to a much lesser extent (Figure 3.5B). Similarly, mevalonate pathway coexpression in the absence of cysteine gave only small increases in cyanobactin titer (Figure 3.5E).

The observation of an epistatic relationship between cys-supplementation and *mbi*-coexpression (i.e., the fact that *mbi*-coexpression only improved titer in the presence

of added cysteine) suggested to us that these two interventions might exert their effects on different stages of biosynthesis, with Cys-supplementation affecting earlier steps and *mbi*-coexpression affecting later steps of biosynthesis. Such a proposal would be consistent with the accepted order of biosynthesis, in which prenyltransfer is believed to be the final step in biosynthesis. [9]

Complicating this straightforward picture is the observation that all compounds, including unprenylated derivatives, are present at increased levels when the mevalonate pathway and mevalonate were added, indicating that the increases in cyanobactin titer cannot be explained solely by the increased availability of isoprene donors such as DMAPP. To further probe the role of the mevalonate pathway, we knocked out the functional pathway prenyltransferase, *truF1*. Consistent with the above results, *tru ΔF1*, showed mevalonate-dependent increases in cyanobactin expression, giving rise to only the unprenylated compounds (Figure 3.5, inset).

Certain families of cyanobactins are not modified by isoprene. For example, in the patellamide pathway, *pat*, threonine and serine residues are heterocylized to yield the corresponding oxazoline derivatives, and as a result are not subject to prenylation; moreover, the prenyltransferase homolog in the *pat* pathway, PatF, is apparently inactive as a prenyltransferase *in vitro*. Given the activating effect of the mevalonate pathway with the *tru ΔF1* pathway, we speculated that *mbi*-coexpression would similarly increase production of patellamides (as observed for unprenylated patellins). Surprisingly, we found that this was indeed the case: patellamide production was increased in proportion to the concentration of added mevalonate, despite the complete lack of prenylation in *pat*

pathway products (Figure 3.6). In the presence of cysteine, *pat* production is minimal and often undetectable and is increased to robustly detectable levels upon *mbi*-coexpression.

The above results suggest that the beneficial effect of mevalonate pathway expression on cyanobactin production may not be entirely exerted through the pathway itself. Rather, its effect may be at least partly indirect, acting through other metabolic pathways that may benefit from extra isoprenoid precursors. Channeling the precursors into these pathways would conceivably require longer precursors such as GPP and FPP, which would require both DMAPP and IPP for elongation. [25]

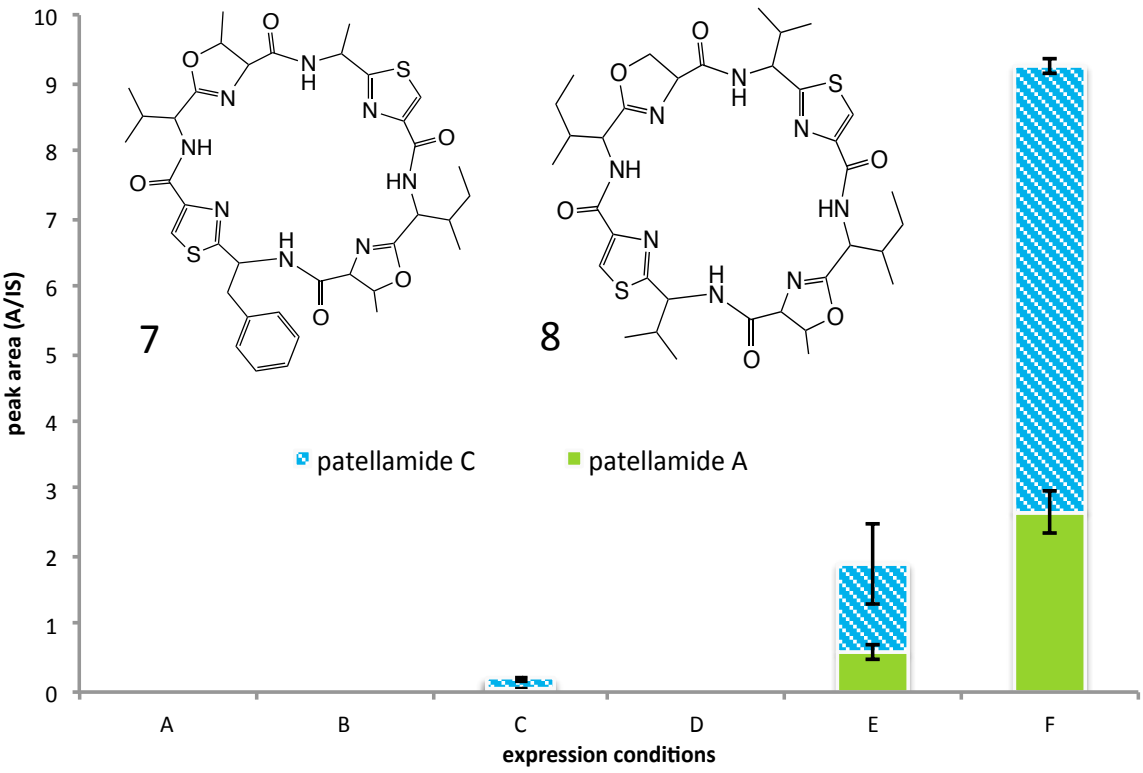


Figure 3.6. Increase in production of patellamides in the presence of mevalonate pathway and external mevalonate. Y-axis shows the ratio of compound peak area to the internal standard. Column A) *pat*, B) *pat* + 5mM cysteine, C) *pat/mbi* + 5mM cysteine, D) pat/mbi + 5mM mevalonate, E) *pat/mbi* + 5 mM cysteine + 5 mM mevalonate, F) *pat/mbi* + 5 mM cysteine + 10 mM mevalonate.

To test whether different isoprenoid species would exhibit defined effects in compound production, we co-expressed *tru* with three different versions of the mevalonate pathway: *mevb*, *mbi*, *mbis*, which convert mevalonate to IPP, DMAPP, and FPP, respectively. *mevb* lacks the isopentyl diphosphate isomerase, *idi*, for the conversion of IPP to DMAPP. We verified that the different expression strains contained these pathways by restriction digest of the overnight seed cultures. All three sets of mevalonate pathway enzymes, in the presence of cysteine and mevalonate (20 mM), increased compound production relative to controls (i.e, *tru* pathway only and empty plasmid pBBR) (Figure 3.7). *Tru/mevb* expressions yielded higher proportions of monoprenylated patellin 2, conceivably a result of the lack of *idi* to supply DMAPP for prenylation. The increase in *tru/mbis* might indicate that longer isoprenoid precursors could in part be responsible for the unexpected effects of the mevalonate pathway on cyanobactin expressions.

Effects of the mevalonate pathway coexpression on growth

To determine how the additives cysteine and mevalonate affect growth, we measured the growth (OD_{600}) of *E. coli* throughout the 5-day expression. *Tru* only and *tru*/pBBR expressions in the presence of cysteine, mevalonate, or both displayed similar growth patterns with a sharp log phase starting at 10 to 15 hours (Figure 3.8A) post inoculation. Without any additives, *tru/mevb, tru/mbi, and tru/mbis* expressions entered log phase at a similar time as controls (A and B, Figure 3.8) but reached a lower OD_{600} in stationary phase (Figure 3.8C-E), perhaps reflecting the burden of expressing additional enzymes. Cysteine in general did not affect growth measurably, nor did mevalonate. However, in the presence of mevalonate pathway enzymes and both additives (Figure

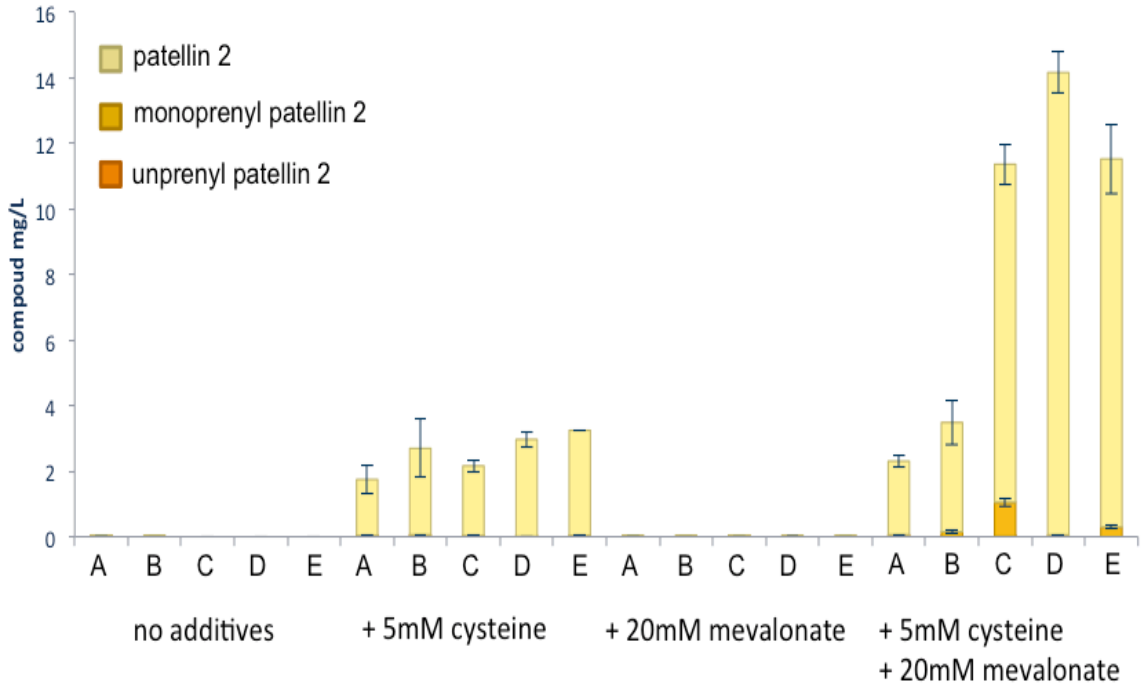


Figure 3.7. Increase in production of patellin 2 in the presence of mevalonate pathway variants. Y-axis shows the amount of compounds in mg/L . Expression conditions : A) *tru* B) *tru/pBBR* C) *tru/ mevb* D) *tru/mbi* E) *tru/mbis*. Columns 1-5, no additives; columns 6-10, with 5 mM cysteine; columns 11-15, with 20 mM mevalonate; columns 16-20, with 5 mM cysteine and 20 mM mevalonate.

3.8C-E), growth was clearly altered, characterized by a long lag phase after inoculation.

Activity of patellin 2 on DRG neurons

With a method to supply meaningful quantities of cyanobactins in hand, we sought next to study their bioactivity. While many cyanobactins have been reported to be bioactive, the range of reported activities is limited to relatively standard cytotoxicity and

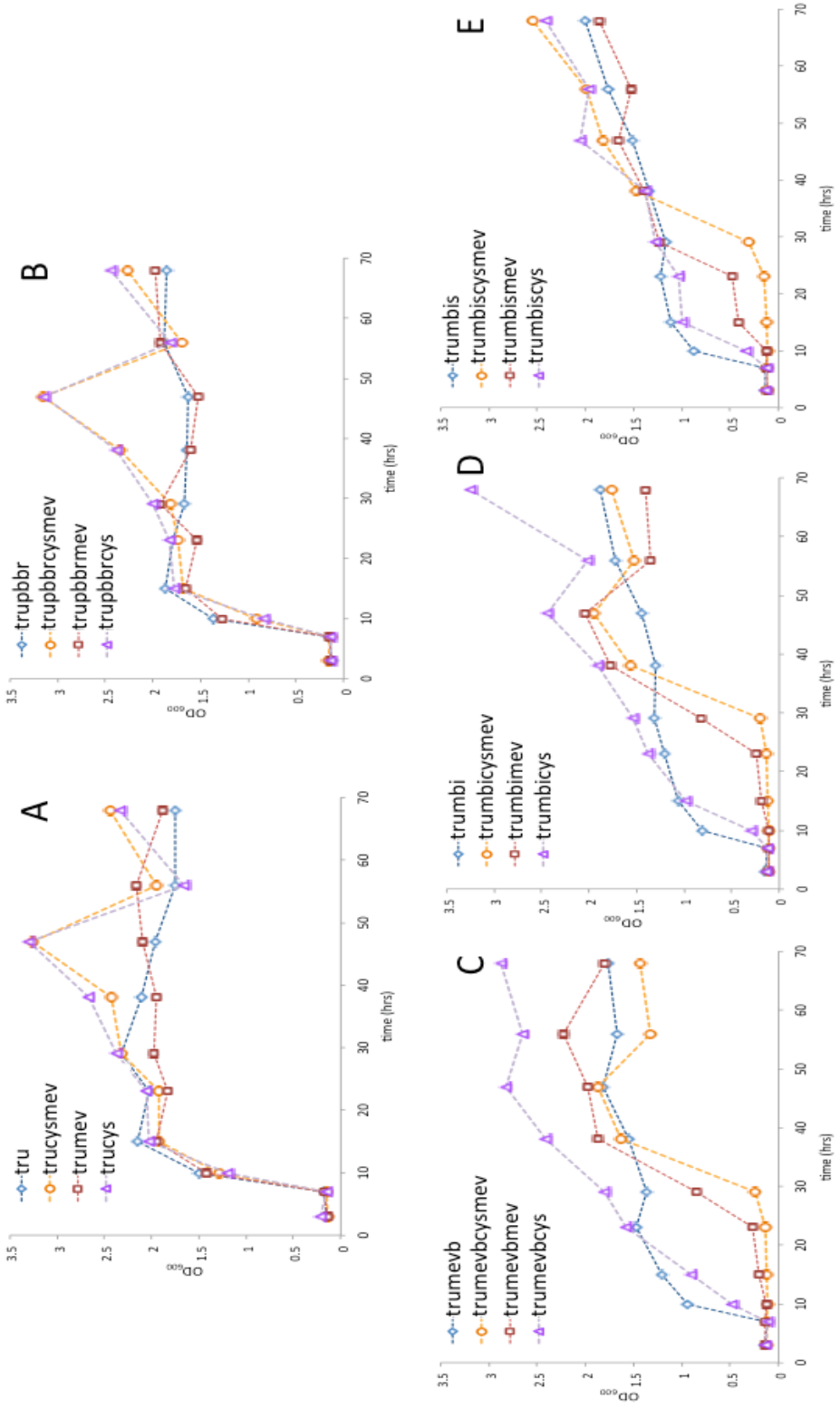
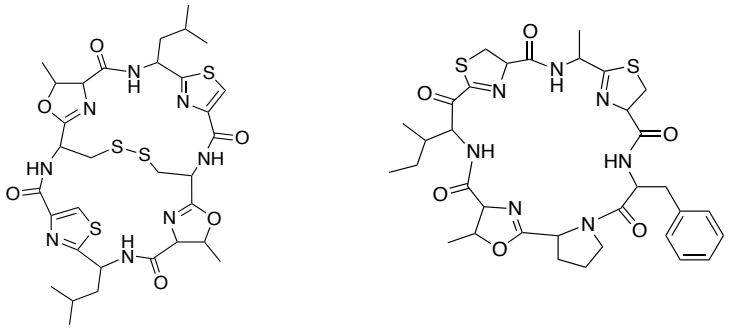


Figure 3.8. Growth of *E. coli* Dh10b in the presence of cysteine and mevalonate and A) *tru* pathway B) *tru/pbbr* C)*tru/mevb* D) *tru/mbi* E) *tru/mbis* pathways.

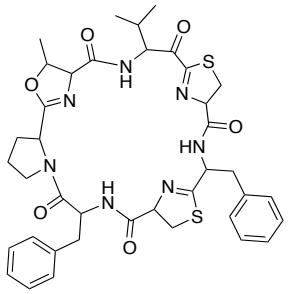
antibiotic assays.[26] Their ecological function within ascidians or free-living cyanobacteria remains unclear, [26] though a variety of hypotheses have been put forward, ranging from settlement inhibitors of competing organisms to antipredation defense.[27] To expand the range of potential pharmacological applications of cyanobactins, we chose to screen natural cyanobactins in a high-content phenotypic screen on mice dorsal root ganglion (DRG) neurons, which, given the breadth of the assay, has been described as a 'constellation pharmacology' approach.[28,29] In particular, the assay measures calcium influx into individual DRG neurons in a heterogenous population using a fluorescent probe, and can distinguish a broad variety of responses to compounds, including ion-channel agonists, antagonists, and a diverse array of other receptor ligands.[28] The goal was first to identify natural cyanobactin leads that exhibited neuroactivity on a subset of the heterogenous DRG population, and from there to use our heterologous system to supply compounds for follow-up studies into molecular mechanisms of action, and structure-activity relationships.

In our initial screen, we tested compounds **2**, **3**, **6**, patellamide C and a mixture (**7** and **8**), ulithiacyclamide[12] (**9**), lissoclinamide 3[30] (**10**), lissoclinamide 4[31] (**11**), and trunkamide [32] (**12**) at 10 μM concentrations (Figure 3.9). We observed clear activity for compounds **3** and **9**. Specifically, application of **3** caused changes in the $[Ca^{2+}]_i$ shown through enhanced response to KCl (25 mM) in some cells as well as direct perturbation of the baseline in others (Figure 3.10A). Compound 9 similarly enhanced the KCl-induced $[Ca^{2+}]_i$ upon application. Interestingly, compound **2** was inactive in this assay despite differing from **3** by only a single prenyl group. We chose to pursue **3**, as its production was already optimized in our heterologous expression system.

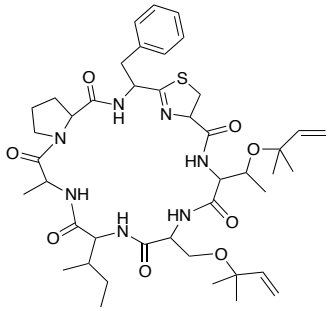


9 Ulithiacyclamide

10 Lissoclinamide 3



11 Lissoclinamide 4



12 Trunkamide A

Figure 3.9. Structures of cyanobactins screened in the DRG assay. Compounds were assayed at 10 μM concentrations. 2,3,and 6 are shown in Figure 3.3B.

To rule out the possibility of contamination by another extract component, we tested **3** purified from both the coral reef animal *Lissoclinum patella* and from recombinant *E. coli* expression. We observed that both were active and showed the same $[Ca^{2+}]_i$ phenotypic response in the same cells (Figure 3.10C). Further dose-response experiments separated the cells into subpopulations according to their sensitivity to **3**: in group **A** were cells that were dose-responsive showing an indirect and increasingly enhanced $[Ca^{2+}]_i$ signal from 3 μM **3,** in **B** were cells that were only affected at a high dose (30 μM) with indirect amplification of $[Ca^{2+}]_i$ signal, and in **C** were cells that were directly affected by a firing response at 30 μM (Figure 3.10D). The cells were also challenged with different pharmacological agents: menthol (cold, TRPM8), ATP

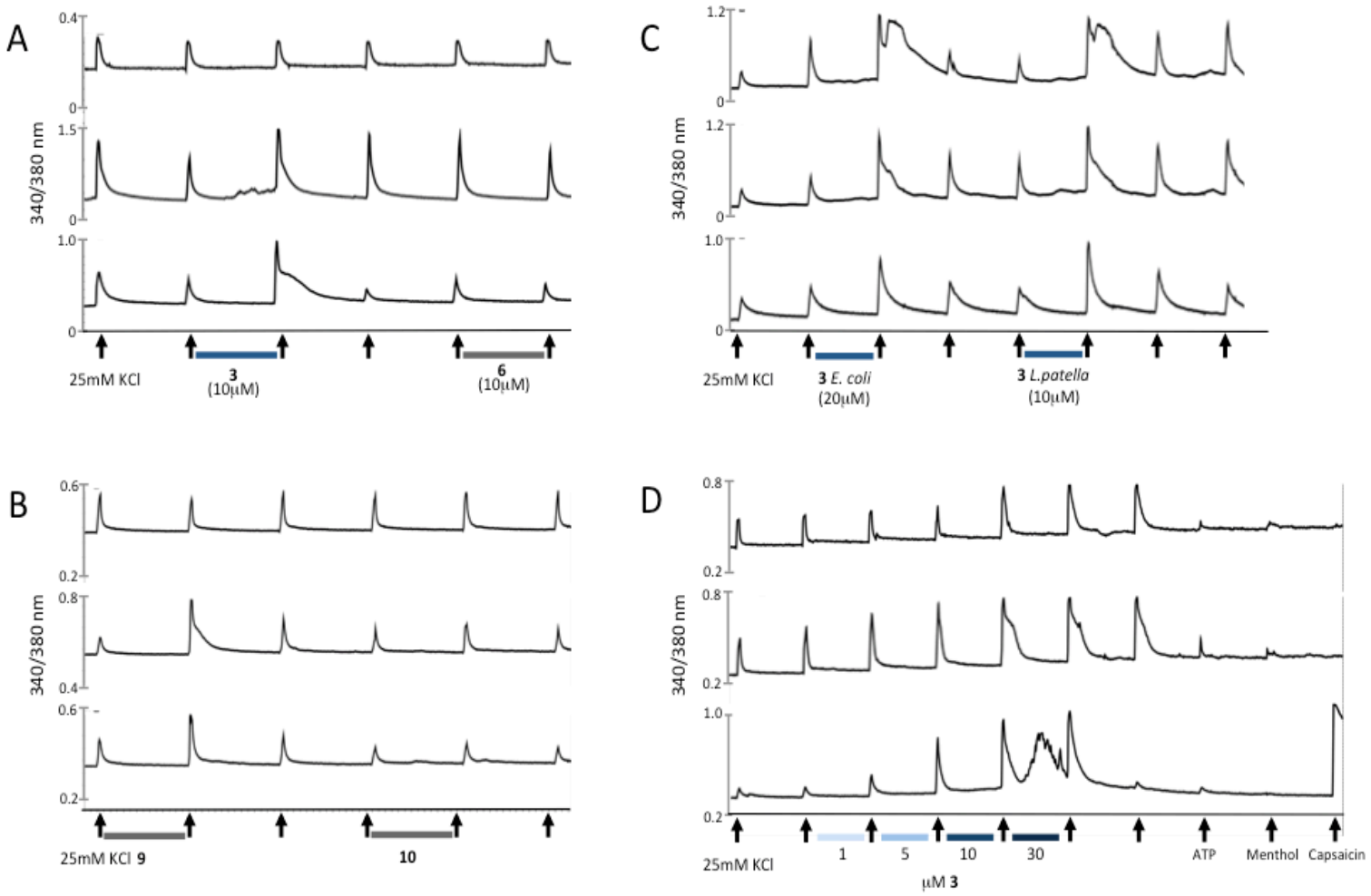


Figure 3.10. Calcium imaging experiments on mouse DRG cells. Enhanced depolarizing KCl response from 3 (A) and 9 (B) are shown in representative cells. C) Similar response from *L. patella* and *E. coli* derived **3**. D) Three types of responses to application of **3.**

(nociception, P2X), and capsaicin (heat, TRPV1). Table 3.2 shows the percent of cells, as classified by their response to **3** above, that responded to the application of the pharmacological agents. Cells from subpopulations B and C were highly correlated with TRPV1, indicating that **3** could affect nociceptors. The different responses in the sub-populations further suggests that **3** may act on multiple cell types that differ in the expression levels of its molecular target.

Discussion

Natural products are indispensible sources of drugs given their complex scaffolds and often potent activities against novel biological targets.[33] Most natural products, however, are left undeveloped because we lack a means to supply them. The engineering of natural product biosynthetic pathways into heterologous hosts has the potential to address these challenges by improving the supply of natural products, and also allowing the production of derivatives through simple genetic engineering.[34] However, many

Table 3.2. Summary of calcium imaging experiments with **3**. Values are presented as percentage of cells that exhibited a type of response (A-C) that are sensitive to pharmacological agents (columns 2-4). Response types to 3 are classified as: **A**, dose responsive from 3 mM to 30 mM; **B**, indirect through amplification of KCl response; **C**, direct effects on the baseline $[Ca^{2+}]_i$.

type of response to **3**	Capsaicin	ATP	Menthol
A	14%	29%	3%
B	69%	63%	0%
C	97%	78%	8%

heterologous expression experiments are conducted primarily for the purposes of verifying bioinformatics predictions of compounds, and thus are unlikely to result in consistent and robust supply of compounds. Thus, the continued development of methods to improve yields in heterologous expression is essential to ease this bottleneck in the search for medicinal natural products.

Here we have described a method for the improvement of the heterologous production of cyanobactins, resulting in a stable, 100-fold boost in titer. This increase was brought about through a combination of media optimization and improving the supply of endogenous precursors through the coexpression of an independent biosynthetic pathway. In particular, we have demonstrated that compound production is synergistically increased by the supplementation of the amino acid cysteine and by isoprene precursors supplied by the mevalonate pathway.

The effect of mevalonate pathway expression on cyanobactin yield is particularly interesting as we initially intended to increase DMAPP levels to support complete prenylation of cyanobactins. Although we observed a modest shift in the ratio of prenylated/unprenylated compounds upon mevalonate pathway coexpression, a greater effect was observed in the increase of the total amounts of compounds, including those that were unprenylated. We found that this effect depends on the concentration of the added mevalonate.

The increase in the amount of all the derivatives indicates that the increase in production is not directly related to compound prenylation. Indeed, knocking out the *tru* pathway prenyltransferase, *truF1*, showed that the levels of unprenylated compounds were still increased by mevalonate in a dose-dependent manner. Further, we have shown

a mevalonate-dependent increase in the production of the patellamides, which do not contain prenyl groups. Thus, it is likely that the effect of *mev*-coexpression is exerted through multiple channels, some of which act indirectly on the pathway enzymes. For example, excess isoprenoids could be channeled into pathways that in turn affect growth and fitness of the cells under conditions of compound production that are otherwise stressful. Indeed, through coexpressions with *tru/mbis*, we showed that channeling the isoprene intermediates into longer units also maintained compound production.

Given the growing importance of tools such as the *mev* pathway in metabolic engineering efforts, our results highlight the often unpredictable and nonadditive effects that can result. To this end, we are currently investigating the metabolome of *E. coli* under these different conditions with the goal of identifying the pathways and metabolites that change in response to *mev* coexpression. One other intriguing feature of our results is that the beneficial effects of *mev*-coexpression are dependent on the presence of cysteine. From our quantifications of many different experimental sets, we determined these effects are not merely additive. Cysteine by itself consistently increases compound production, while mevalonate exerts a significant effect only in the presence of added cysteine.

Altogether, the unanticipated effects of the mevalonate pathway on cyanobactin production opens new possibilities for heterologous production of RiPPs in *E. coli*, as even unprenylated cyanobactins were increased upon *mev* coexpression. To our knowledge, this is the first demonstration of such an 'off-target' effect of *mev* pathway coexpression, in which the titer of non-terpenoid natural products is observed to increase. Serendipity aside, this study also emphasizes the combination of both empirical (the

discovery of cysteine as an additive) and rational (supply of isoprene for full prenylation) approaches in the optimization of cyanobactin production.

Having established a high-yielding expression system of cyanobactins, we sought to readdress one of the fundamental goals of natural product discovery and engineering, that of discovering novel bioactivities. Thus, we applied several cyanobactins to a broad phenotypic assay, and uncovered the activity of patellin 2 in provoking action potentials in DRG neurons. Patellin 2 was first isolated and characterized from *L. patella* ascidian in 1990[35]; however, no biological activity has been reported on this compound until now. Although the neuronal subtype specificity and molecular target of patellin 2 is the subject of ongoing work, based on comparison of both expressed and natural compounds, we were able to establish that patellin 2 is active affecting different types of nociceptors. This represents the first DRG activity reported for a cyanobactin. These results emphasize the importance of broad pharmacological screens since the native targets for the vast majority of natural products are unknown, and further underline the potential of small, highly modified cyclic peptides as neuroactive agents.

In summary, we have established a method for the highest titer of cyanobactins so far. We have discovered, serendipitously, the positive effects of mevalonate pathway expressions to cyanobactin production. This method is the first step in removing the bottleneck for the development of cyanobactins as pharmacological and therapeutic agents, and for further exploration of their ecological and functional roles.

Materials and methods

Plasmids and strains

Plasmid pTru-SD contained the *tru* operon in Topo vector. *tru* contained the precursor peptide gene *TruE2*, which encodes for patellin 2 (**3**) and patellin 3 (**6**) in the core peptides. pMEVB, pMBI, and pMBIS, containing the second half of the mevalonate pathway were obtained from Addgene (plasmids 17819, 17816, and 17817, respectively). pMEVB contains *ERG12, ERG8, MVD1*; pMBI has the pMEVB genes but with *idi* appended after MVD1; pMBIS is pMBI with additional *ispa*. [24] Empty plasmid pBBR with a modified multiple cloning site was constructed from pMBI. pMBI was digested with KpnI and SacI. The vector backbone piece was ligated with a small piece of DNA constructed from overlapping primers to create an expanded multiple cloning site (forward: 5' AGTGTACAGGGCCCCCCCTCGAGGGTATCGATAAGC TTGATATCGAATTCCTGCAGTAGGAGGAATTAACCCATATGTC, reverse: GATGAGCT CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGG GTACCATGGACATATGGGTTAATTCCTC). pPat (*pat*) contains the patellamide pathway genes that were codon-optimized for *E. coli*. *pat* is in Topo vector. For the *truF1* knockout experiments, modified *tru* expression plasmid was made from fragments of the previously described vector,[17] containing *truA-D* and *truG* and gBlocks containing *truE, F1,* and *F2* (IDT, Coralville, IO). The gene sequences in the gBlocks were codon-optimized for expression in *E. coli*. This plasmid was assembled by recombination in *Saccharomyces cerevisiae* BJ4741 using previously described methods. [17] *truDF1* was made by Gibson assembly [36] of PCR products containing most of the modified *tru* plasmid in two pieces. The first piece consisted of half of the plasmid backbone and the

first part of the *tru* pathway (*truA-E*). The second piece contained the end of the tru pathway (*truF2* and *truG*) and the other half of the plasmid backbone. All vectors were partially sequenced to confirm that the pieces had been ligated correctly. All pathways were under the control of lac promoter and expressed constitutively in DH10B cells.

Chemicals and other materials

L-cysteine hydrochloride was obtained from Amresco (Solon, OH). Ampicillin and tetracycline were obtained from Sigma-Aldrich (St. Louis, MO). All solvents used for silica open column chromatography, HPLC, and HPLC/MS analyses were obtained from Fisher Scientific (Pittsburg, PA). *Lissoclinum patella* samples that were used for isolation of standards **3** and **6** were collected in Papua New Guinea with proper collection permits.

Fermentation conditions and compound expressions

Plasmid pTru-SD1 (*tru*) was transformed alone or co-transformed with pBBR, pMEVB (*mevb*), pMBI (*mbi*), or pMBIS (*mbis*) (Martin 2003). The resulting clones were then inoculated into liquid 2xYT broth (6 mL) in 24-well plates with the addition of appropriate amount of antibiotics (ampicillin: 50 mg/mL, tetracycline: 5 mg/mL), and grown overnight. Growing seed cultures (0.2% v/v) were then inoculated into fresh media containing antibiotics described above. Cysteine (5 mM) was added upon inoculation of expression cultures. Mevalonolactone (Sigma-Aldrich, St. Louis, MO) was hydrolyzed to mevalonate according published methods (Martin 2003) and was also added to the expression cultures at inoculation (5, 10, 20, 40 mM final concentrations). The cultures

were allowed to shake at 150 rpm, 30 °C for 5 days after which the cells were harvested by centrifugation at 4000 rpm, washed with saline solution (100 mM NaCl), and extracted with acetone (4 mL). The acetone extracts were then dried to yield the organic extracts for each 6 mL culture. In labeling experiments, ^{13}C mevalonate (1 mM) (Sigma-Aldrich, St. Louis, MO) was added at the beginning of the expressions as described above. To express patellamides, *pat* was co-transformed pMBI and processed as described above.

Isolation of patellin 2 from *E. coli* expressions

Overnight seed cultures of *E. coli* harboring *tru/mbi* were grown under appropriate antibiotic selection and inoculated into 1L 2xYT broth with antibiotics, cysteine (5 mM), and mevalonate (10 mM). The broth was distributed into 24 well plates containing 6 mL per well and grown for 5 days at 30 °C while shaking at 150 rpm. The cultures were harvested as described above, and the cells were extracted repeatedly with acetone. The combined acetone fractions were dried and fractionated by silica open column chromatography using previously described gradient and methods .[10] Fractions were analyzed by HPLC-MS for the presence of cyanobactins, and the compound containing fractions were subjected to reverse phase HPLC fractionation on an Eclipse XDB, 9.4 x 250 mm, 5m C18 column (Agilent Technologies, Santa Clara, CA) using the following solvent gradient: 5% B to 100% B (0-30 minutes), 100% B (30 to 40 minutes), and 100% to 5% B (40 to 45 mins); Solvent B consisted of acetonitrile and solvent A consisted of water. A flow rate of 2.5 mL min^{-1} was used. Compound **3** was purified on a Luna, 4.6 x 250 mm, 5 µm C18 Column (Phenomenex, Torrance, CA) using the solvent

gradient: 80% B to 94% B (0-15 minutes), 94% B (15-17 minutes), and 94% to 80% B (17-20 minutes), at a flow rate of 1ml min^{-1}; Solvent B consisted of acetonitrile and solvent A consisted of water.

Quantification by proton NMR

^{1}H NMR was used to quantify purified patellin 2. This information was used as the basis for the HPLC/MS calibration curve as well as for bioactivity assays. An external standard curve was made using the different amounts of 1,4-dinitrobenzene (Sigma-Aldrich, St. Louis, MO) using described methods. Briefly, 2-fold dilutions (1.2mM to 79.3 mM) of 1,4-DNB in DMSO-*d6* (Cambridge Isotope Laboratories Inc., Tewksbury, MA) were prepared in uniform 3 mM Kontes precision NMR tubes (Sigma- Aldrich, St. Louis, MO). ^{1}H NMR for each DNB sample was acquired on a Varian Inova 500 instrument (Agilent Technologies, Santa Clara, CA) with 32 scans using an 18 s relaxation delay (d1) as calculated from T1 using signal inversion. The integrals of the four DNB protons were obtained and used as the Y-values for the calibration curve (Figure 3.11). Patellin 2 was prepared using the same DMSO-*d6*. The 1H spectrum of patellin 2 was obtained using 32 scans and 4 s d1, which was determined as described above. Signals from a methyl protons of the valine residue were integrated and the values were compared with the external standards.

Quantification of cyanobactins from different expressions using HPLC/MS

Extracts were dissolved in methanol (500 mL) and analyzed by HPLC-ESI-MS using a Waters Micromass ZQ mass spectrometer. Samples were injected at uniform

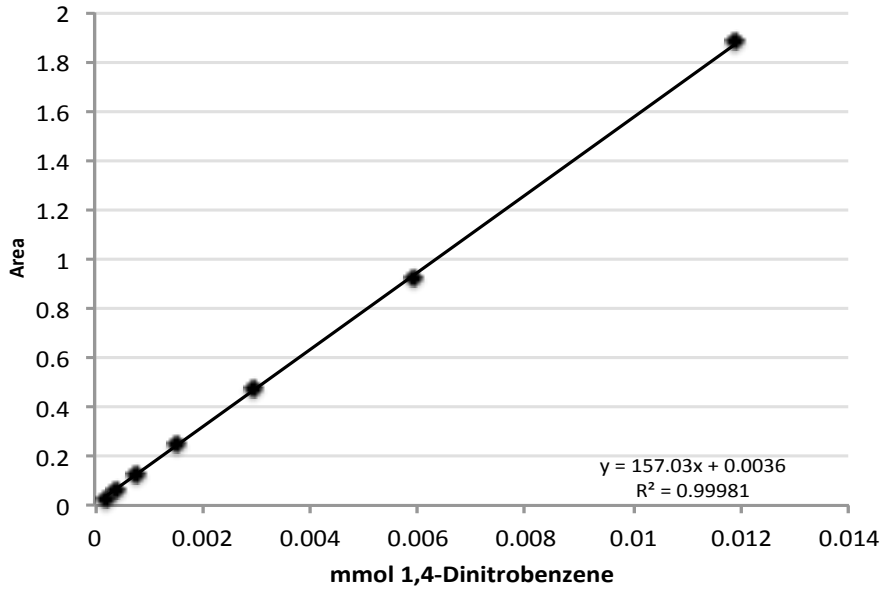


Figure 3.11. 1H NMR calibration curve using 1,4-Dinitrobenzene.

volumes (40 mL). A synthetic derivative of a heptamer cyclic peptide (7, Figure 3.12B) was used as an internal standard in all HPLC/MS analyses. The extracts were analyzed on a Zorbax Eclipse Plus 4.6 x 150 mm, 5 μm, C18 column (Agilent Technologies, Inc., Santa Clara, CA) using the following solvent gradient: 10% B (0-2 minutes), 10% B to 100% B (2-20 minutes), 100% B (21 to 30 minutes), and 100% to 10% B (30 to 35 mins). Solvent B consisted of acetonitrile with 0.05% (v/v) formic acid; solvent A consisted of water with 0.05% (v/v) formic acid. The peaks of interest were selected from the chromatograms and the resulting peak areas for each compound in the extract were obtained. The ratio of the peak area of the compound of interest to the peak area of the internal standard was then taken in each sample and the amount of each compound was calculated using the calibration curve that was generated using different concentrations of

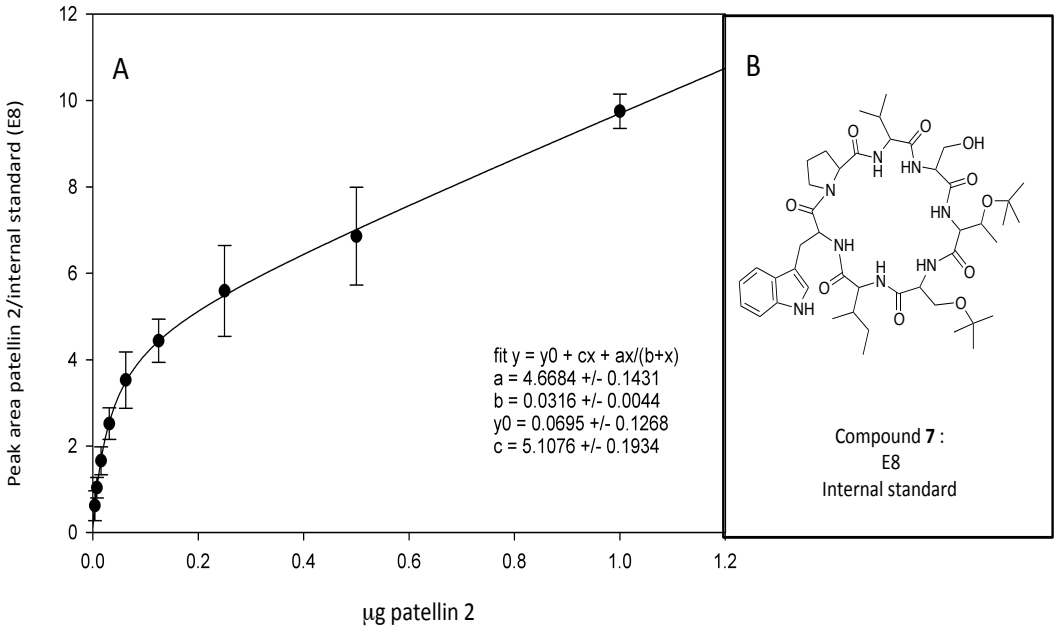


Figure 3.12. A) LC/MS calibration curve of patellin 2. Y-axis represents the ratio of the peak area of different amounts (X-axis) of patellin 2 to the internal standard (panel B). B) structure of E8 internal standard.

purified patellin 2 (Figure 3.12a). All expressions were done in triplicate. Around 70 independent expression experiments (average of 8 different conditions in each set) were done over the course of optimization, confirming the reproducibility of the results.

Growth experiments

Samples were grown in 6 mL volumes, in quadruplicates sets, in 24-well plates as described above. Aliqouts (100 mL) were drawn and transferred to a 96-well plate containing 2xYT (100 mL) and the OD was measured on a microplate reader. OD_{600} was measured every 3-4 hours for 120 hours. The remaining cultures were extracted for compound production and analyzed for relative abundance of cyanobactins between each condition.

Calcium imaging on mouse dorsal root ganglion (DRG) cells

The DRG assay has been described previously.[37] [28] Briefly, the cells in their growth media were loaded with Fura-2-AM for 1 h at 37° C and incubated further for 30 minutes at room temperature. The media was then replaced with a CSF for calcium imaging at room temperature. Changes in cytosolic calcium concentration were monitored over time through the ratio of fluorescence intensities at 510 nm obtained from intermittent (typically once every 2 s) excitation by 340-nm and 380-nm light (labeled as " 340/380 nm " in the y-axis of calcium-imaging figures. The increase or decrease in calcium levels is represented by an upward or downward deflection of the trace. High concentrations of KCl (25 mM) were used to determine indirect responses elicited by **3**.
[28]

References

1 Luo, Y., Cobb, R. E. & Zhao, H. Recent advances in natural product discovery. *Curr Opin Biotechnol* **30**, 230-237, doi:10.1016/j.copbio.2014.09.002 (2014).

2 Deane, C. D. & Mitchell, D. A. Lessons learned from the transformation of natural product discovery to a genome-driven endeavor. *J Ind Microbiol Biotechnol* **41**, 315-331, doi:10.1007/s10295-013-1361-8 (2014).

3 Helfrich, E. J., Reiter, S. & Piel, J. Recent advances in genome-based polyketide discovery. *Curr Opin Biotechnol* **29**, 107-115, doi:10.1016/j.copbio.2014.03.004 (2014).

4 Wilson, M. C. & Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* **20**, 636-647, doi:10.1016/j.chembiol.2013.04.011 (2013).

5 Schmidt, E. W. *et al.* Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proc Natl Acad Sci U S A* **102**, 7315-7320, doi:10.1073/pnas.0501424102 (2005).

6 Donia, M. S. *et al.* Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* **2**, 729-735, doi:10.1038/nchembio829 (2006).

7 McIntosh, J. A. *et al.* Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. *J Am Chem Soc* **132**, 15499-15501, doi:10.1021/ja1067806 (2010).

8 McIntosh, J. A. & Schmidt, E. W. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *Chembiochem* **11**, 1413-1421, doi:10.1002/cbic.201000196 (2010).

9 McIntosh, J. A., Donia, M. S., Nair, S. K. & Schmidt, E. W. Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *J Am Chem Soc* **133**, 13698-13705, doi:10.1021/ja205458h (2011).

10 Tianero, M. D., Donia, M. S., Young, T. S., Schultz, P. G. & Schmidt, E. W. Ribosomal route to small-molecule diversity. *J Am Chem Soc* **134**, 418-425, doi:10.1021/ja208278k (2012).

11 Sardar, D., Pierce, E., McIntosh, J. A. & Schmidt, E. W. Recognition sequences and substrate evolution in cyanobactin biosynthesis. *ACS Synth Biol* **4**, 167-176, doi:10.1021/sb500019b (2015).

12 Ireland, C., Scheuer,P.J. Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate. *J. Am. Chem. Soc.* **102**, 5688-5691 (1980).

13 Ireland, C. M., Durso, A.R., Newman, R.A, Hacker, M.P. Antineoplastic cyclic peptides from the marine tunicate *Lissoclinum patella*. *J. Org. Chem* **47**, 1807-1811 (1982).

14 Linington, R. G. *et al.* Venturamides A and B: antimalarial constituents of the panamanian marine Cyanobacterium Oscillatoria sp. *J Nat Prod* **70**, 397-401, doi:10.1021/np0605790 (2007).

15 Mitchell, S. S., Faulkner, D. J., Rubins, K. & Bushman, F. D. Dolastatin 3 and two novel cyclic peptides from a palauan collection of Lyngbya majuscula. *J Nat Prod* **63**, 279-282 (2000).

16 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108-160, doi:10.1039/c2np20085f (2013).

17 Donia, M. S., Ravel, J. & Schmidt, E. W. A global assembly line for cyanobactins. *Nat Chem Biol* **4**, 341-343, doi:10.1038/nchembio.84 (2008).

18 Zhang, Q., Yang, X., Wang, H. & van der Donk, W. A. High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. *ACS Chem Biol* **9**, 2686-2694, doi:10.1021/cb500622c (2014).

19 McIntosh, J. A., Lin, Z., Tianero, M. D. B. & Schmidt, E. W. Aestuaramides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived claisen rearrangements. *ACS Chem Biol* **8**, 877-883, doi:10.1021/cb300614c (2013).

20 Ruffner, D. E., Schmidt, E. W. & Heemstra, J. R. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synth Biol* **4**, 482-492, doi:10.1021/sb500267d (2015).

21 Donia, M. S. *et al.* Complex microbiome underlying secondary and primary metabolism in the tunicate-Prochloron symbiosis. *Proc Natl Acad Sci U S A* **108**, E1423-1432, doi:10.1073/pnas.1111712108 (2011).

22 Donia, M. S., Fricke, W. F., Ravel, J. & Schmidt, E. W. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS One* **6**, e17897, doi:10.1371/journal.pone.0017897 (2011).

23 Donia, M. S., Ruffner, D. E., Cao, S. & Schmidt, E. W. Accessing the hidden majority of marine natural products through metagenomics. *Chembiochem* **12**, 1230-1236, doi:10.1002/cbic.201000780 (2011).

24 Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotech* **21**, 796-802, doi:10.1038/nbt833 (2003).

25 Rodriguez-Concepcion, M., Boronat, A. in *Isoprenoid synthesis in plants and microorganisms* Vol. 1 (ed T. Bach, Rohmer, M.) 1-16 (Springer-Verlag, 2013).

26 Donia, M. S. & Schmidt, E. W. in *Comprehensive Natural Products II* (eds Chief Editors in, xA, Mander Lew, & Liu Hung-Wen) 539-558 (Elsevier, 2010).

27 Sera, Y., Adachi, K., Fujii, K. & Shizuri, Y. Isolation of Haliclonamides: new peptides as antifouling substances from a marine sponge species, Haliclona. *Mar Biotechnol (NY)* **4**, 441-446, doi:10.1007/s10126-001-0082-6 (2002).

28 Teichert, R. W. *et al.* Functional profiling of neurons through cellular neuropharmacology. *Proc Natl Acad Sci U S A* **109**, 1388-1395, doi:10.1073/pnas.1118833109 (2012).

29 Teichert, R. W., Schmidt, E. W. & Olivera, B. M. Constellation pharmacology: a new paradigm for drug discovery. *Annu Rev Pharmacol Toxicol* **55**, 573-589, doi:10.1146/annurev-pharmtox-010814-124551 (2015).

30 Rashid, M. A., Gustafson, K. R., Cardellina, J. H., 2nd & Boyd, M. R. Patellamide F, A new cytotoxic cyclic peptide from the colonial ascidian Lissoclinum patella. *J Nat Prod* **58**, 594-597 (1995).

31 Degnan, B. M. *et al.* New cyclic peptides with cytotoxic activity from the ascidian Lissoclinum patella. *J Med Chem* **32**, 1349-1354 (1989).

32 Carroll, A. R. *et al.* Patellins 1-6 and trunkamide A: novel cyclic hexa-, hepta- and octa-peptides from colonial ascidians, Lissoclinum sp. *Aust J Chem* **49**, 659-667 (1996).

33 Cragg, G. M. & Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim Biophys Acta* **1830**, 3670-3695, doi:10.1016/j.bbagen.2013.02.008 (2013).

34 Ongley, S. E., Bian, X., Neilan, B. A. & Muller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep* **30**, 1121-1138, doi:10.1039/c3np70034h (2013).

35 Zabriskie, T. M., Foster, M. P., Stout, T. J., Clardy, J. & Ireland, C. M. Studies on the solution- and solid-state structure of patellin 2. *J Am Chem Soc* **112**, 8080-8084, doi:10.1021/ja00178a035 (1990).

36 Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods in enzymology* **498**, 349-361, doi:10.1016/B978-0-12-385120-8.00015-2 (2011).

37 Raghuraman, S. *et al.* Defining modulatory inputs into CNS neuronal subclasses by functional pharmacological profiling. *Proc Natl Acad Sci U S A* **111**, 6449-6454, doi:10.1073/pnas.1404421111 (2014).

CHAPTER 4

RIBOSOMAL ROUTE TO SMALL MOLECULE DIVERSITY

Manuscript reproduced with permission from:

Ma. Diarey B. Tianero, Mohamed S. Donia, Travis S. Young, Peter G. Schultz, Eric W. Schmidt. **Ribosomal Route to Small-Molecule Diversity**. *J. Am. Chem. Soc*. 2012, 134, 418−425

Note: My contribution to this paper was in the heterologous expression of compounds, analysis of data, interpretation of results, and writing the manuscript.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

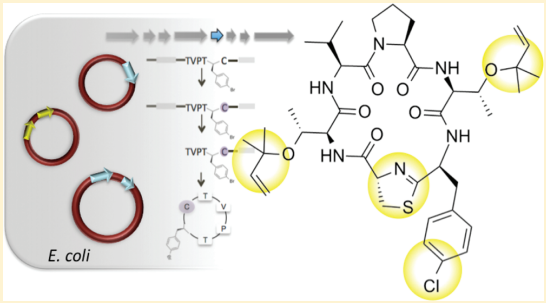
Ribosomal Route to Small-Molecule Diversity

Ma. Diarey B. Tianero,[†] Mohamed S. Donia,[†,§] Travis S. Young,[‡,⊥] Peter G. Schultz,[‡] and Eric W. Schmidt*[,†]

[†]Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, United States
[‡]The Scripps Research Institute, La Jolla, California 92037, United States

Supporting Information

ABSTRACT: The cyanobactin ribosomal peptide (RP) natural product pathway was manipulated to incorporate multiple tandem mutations and non-proteinogenic amino acids, using eight heterologous components simultaneously expressed in *Escherichia coli*. These studies reveal the potential of RPs for the rational synthesis of complex, new small molecules over multiple-step biosynthetic pathways using simple genetic engineering.



INTRODUCTION

A major goal of synthetic biology is to engineer the synthesis of organic compounds in vivo.[1] Ribosomal peptide (RP) natural products provide a relatively simple starting point for such studies. In RP pathways, short precursor peptides are ribosomally translated and subsequently modified by enzymes into complex natural products.[2] The products of RP pathways are often elaborately tailored, so that in extreme cases (such as pyrroloquinoline quinone),[3] the compounds are no longer easily recognizable as arising from amino acids. A further advantage is that RP biosynthetic enzymes generally exhibit relaxed substrate specificity.[2,4] Because of this, often sequence-identical enzymes are capable of accepting a vast array of peptide substrates.[5,6] Here, we seek to develop methods for making unnatural RPs and to better understand substrate selectivity of RP pathways. In so doing, we hope to enable synthetic biological approaches to the directed creation of diverse structures using ribosomally translated starting materials. In addition to these design and synthesis goals, RPs are often potently bioactive against diverse targets, and thus understanding and manipulating individual RP pathways is a major goal in its own right.

In Nature, RPs occupy many different chemical classes, with many different posttranslational modifications.[2] These classes have been examined by observing natural variation, as well as by manipulating biosynthetic genes and by performing biochemical and substrate analyses. Observation of Nature indicates that at least several pathways handle extremely diverse substitutions. For example, >60 cyanobactin genes have been isolated from symbiotic bacteria living in marine animals, where the enzymes are essentially identical yet the products are extremely sequence-diverse.[5] RP biosynthetic genes have already been modified in several ways, most of which have focused on point mutations of pathways that require one or two enzymatic steps for maturation.[7−10] Single or multiple mutations have been introduced into lantibiotics, and more recently, multiple tandem substitutions have been introduced into lasso peptides.[12] These mutations are usually aimed either to improve the known biological activity, or occasionally even to design a new biological activity.[13] A few mutational studies have focused on complex pathways with multiple biosynthetic steps.[5,14,15] For example, the integrin-binding RGD motif has been introduced into cyclic peptides using cyanobactin pathway enzymes[5] and lasso peptide enzymes.[13] However, overall the field is still distant from achieving the goal of wholesale introduction of new motifs at will, especially in the more complex, multistep pathways.

A further goal in this field has been to incorporate motifs other than the canonical 20 amino acids into the backbone of RPs. Since the posttranslational enzymes exhibit extremely broad substrate tolerance, such changes would greatly extend the capabilities of RP pathways in synthetic biology. In many cases, synthetic substrates have been fed to RP enzymes in vitro. A potentially more powerful approach from the synthetic biology point of view is to simply incorporate non-proteinogenic amino acids in vivo. Indeed, non-proteinogenic amino acids have been added to RPs in vivo, although only in single-step processes involving lanthionine synthase or inteins.[16,17] Incorporation of such "unnatural" amino acids is considered desirable because it provides unprecedented control over chemical structure in vivo.

Here, we sought to explore the mutation and addition of non-proteinogenic amino acids with a much more complex

Received: September 1, 2011
Published: November 22, 2011

dx.doi.org/10.1021/ja208278k | J. Am. Chem.Soc. 2012, 134, 418−425

pathway to cyanobactins, which are highly modified, macrocyclic RPs.[18,19] Cyanobactins usually have four types of posttranslational modifications, with up to eight individual modified amino acids, encompassing up to 18 separate biochemical transformations (Figure 1). Heterocylization and

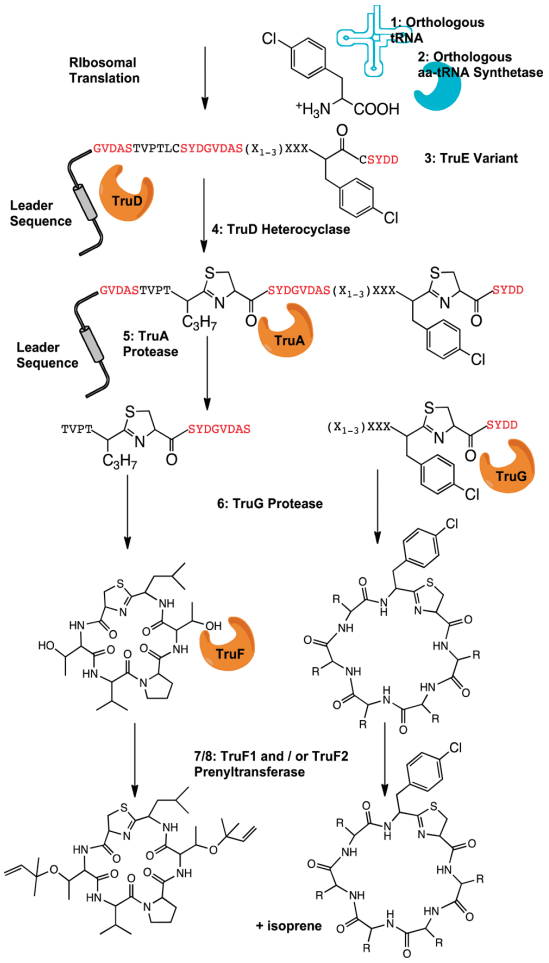


Figure 1. Biosynthesis of *tru* pathway derivatives in *Escherichia coli*. Numbered components indicate eight individual elements including enzymes, tRNA, or precursor peptide (TruE variants) that are required for synthesis of mature natural products. The orthologous tRNA and amino acyl tRNA synthetase components required to introduce nonproteinogenic amino acids are shown in blue. Enzymes from the *tru* pathway are indicated in orange. The precursor peptide TruE is shown with a helical, 35-amino acid leader sequence indicated schematically.[11] X indicates any amino acid.

oxidation, N- and C-terminal proteolysis, macrocyclization, and prenylation lead to final natural products.[5,6,20−24] In the *pat* and *tru* cyanobactin pathways, core sequences encoding six- to eight-amino acid cyclic peptides are contained in defined cassettes in a precursor peptide (Figures 2 and 3).[5,6] The 18−24 nucleotide hypervariable regions are flanked by invariant sequences that direct enzymatic modification.[5] Thus, the synthesis of new compounds requires only simple genetic engineering.[6]

In the *tru* pathway, originally cloned from uncultivated symbiotic bacteria in a marine animal, a ~70 amino acid precursor peptide (TruE) is synthesized that encodes two cyanobactins (Figure 1).[6] Heterocyclase TruD acts on Cys residues in specific positions, synthesizing thiazoline.[23] Subsequently, TruA protease removes the leader sequence and affords two free N-termini for macrocyclization.[22] TruG protease cleaves C-terminal recognition sequences in tandem with macrocyclization to provide small, cyclic peptides.[20,22,23] Finally, TruF1/TruF2 decorate the cyclic peptides with isoprene on specific Ser and Thr residues.[24] In this manner, we have previously successfully produced coral reef-derived marine natural products patellin 2 (**1**), trunkamide (**2**) and patellin 3 (**3**) by heterologous expression in *E. coli*. Optimal production of wild type compounds **1**−**3** requires five days, emphasizing the complexity of the process.[25]

Some ascidian cyanobactins, and especially **2**, are considered as potential anticancer leads.[26,27] Cyanobactins in the patellamide class are potent inhibitors of drug efflux.[28,29] Therefore, synthesis of derivatives in the manner described herein is expected to lead to bioactive molecules that are amenable to drug development. More importantly, as shown by the integration of RGD motifs into cyanobactins, we are interested in exploring the mutability of these pathways because they can be used to install desired or random features into natural product-like structures, using simple genetic engineering.

RESULTS

Cloning and Expression Strategy. The pUC-based vector ptru-SD1 (Symbion Discovery, Inc.) encodes *tru* biosynthetic enzymes and a copy of the precursor peptide TruE1, encoding patellins 2 (**1**) and 3 (**3**), under control of the *lac* promoter (Figure 2). To encode new cyanobactin derivatives, we constructed a second vector (ptruE), which is compatible with ptru-SD1 and can be used for coexpression experiments. ptruE contains only the *truE* gene, under control of the *lac* promoter. The first cassette of ptruE encodes **3**, while the second cassette can be varied to synthesize novel compounds (Figures 1 and 2, Tables 1 and 2). In this system, compounds **1** and **3** are always synthesized in *E. coli* and serve as internal controls to show that the *tru* pathway is functional. In addition, because both plasmids are under control of the constitutive *lac* promoter, no induction is necessary, and optimum production requires five days of fermentation at 30 °C. The general system was previously optimized, and we showed that addition of inducers or repressors serve to decrease yields.[25] Thus, after five days without induction, internal controls **1** and **3** are produced, and possibly new derivatives encoded on ptruE.

In this study, the sequence in the second cassette of ptruE was manipulated using cloning, while the first cassette was kept constant to produce positive control **1**. The fidelity of this process was determined by gene sequencing, where the ptruE derivatives were Sanger sequenced. Thus, each *E. coli* expression clone encoded all necessary *tru* enzymes, internal standards leading to production of **1** and **3**, and a gene for one of the new compounds **4**−**22**. In expression experiments, production of **1** and **3** would indicate that enzymes were functional and active and would also provide internal calibration for yield determination, while new compounds would be expressed only if the sequences were substrates for all *tru* enzymes.

dx.doi.org/10.1021/ja208278k | *J. Am. Chem.Soc.* 2012, 134, 418−425

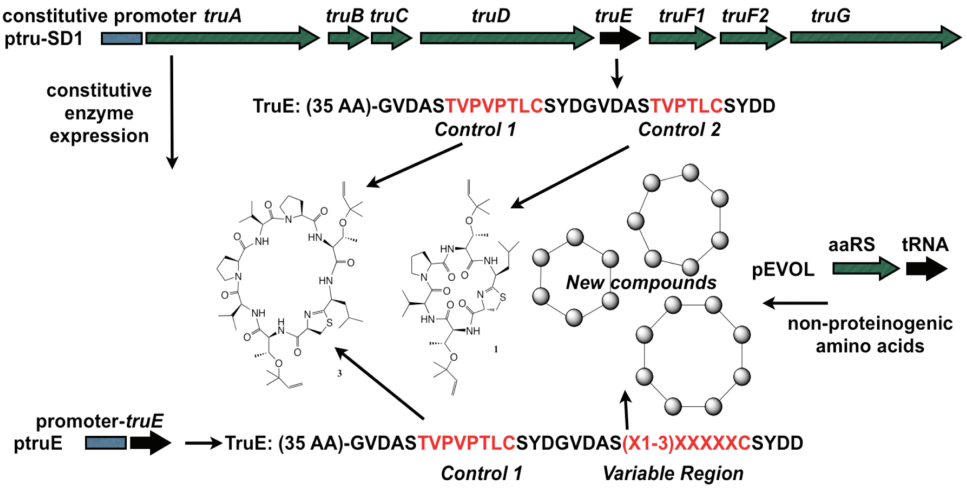


Figure 2. Expression strategy. Enzymes and two control molecules are synthesized constitutively from the vector ptru-SD1 (top). An internal control sequence and a variable region, that can lead to peptide libraries, is synthesized from ptruE (bottom). Optionally, non-proteinogenic amino acids can be included using a third vector, pEVOL (right).

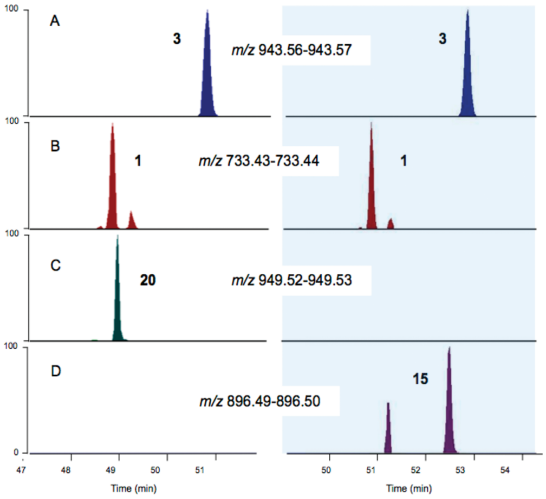


Figure 3. Expression of mutant cyanobactins. *E. coli* cultures expressing **20** (left) and **15** (right) were harvested and analyzed by FT-MS. Extracted ion chromatograms showed that each culture expressed control compounds **3** (A) and **1** (B). In cells containing plasmids encoding **20**, compound **20** could be detected (C), but not **15** (D), nor any other cyanobactin **4**–**22**. Similarly, in cells encoding **15**, only compound **15** (D) but not **20** (C), nor any other cyanobactin, could be detected. Thus, each experiment contained two internal positive controls and 19 external negative controls.

We constructed mutants based upon known natural products, compounds **1**-**3**, which we previously identified in uncultivated symbiotic bacteria living in marine animals and expressed successfully in *E. coli*.[6] In addition, **1**–**3** are hexa-, hepta-, and octameric, respectively, so that mutations explored a range of different substrate and product sizes. We sought primarily to make a series of mutants that would broadly explore sequence selectivity of the *tru* pathway. Out of the constellation of possible mutations that could answer this question, we picked representative derivatives that were also interesting to us because they helped to answer other pressing scientific questions about the cyanobactins pathways (see Discussion). Thus, the results of these experiments would have broad application in synthetic biology and specific application to understanding this interesting group of natural products, which are very broadly distributed.

Table 1. Design Strategy for Mutants

#	sequence[a]	design principle[b]	size[c]
1	TVPTLC	Patellin 2 natural product[41]	6
4	TLATIC	Genome mining product[25]	6
5	IVPPFC	Genome mining product[30]	6
6	TTVTAC	Genome mining product[30]	6
7	VTPFVC	Mollamide B natural product[42]	6
8	IPGSLC	Keenamide natural product[37]	6
9	TVPT-**pBrF**-C	Patellin 2 + non-proteinogenic	6
10	TVPT-**pClF**-C	Patellin 2 + non-proteinogenic	6
11	TVPT-**pOMeF**-C	Patellin 2 + non-proteinogenic	6
12	TVPT-**pN$_3$F**-C	Patellin 2 + non-proteinogenic	6
2	TSIAPFC	Trunkamide natural product	7
13	TSIAPLC	Trunkamide point mutant	7
14	TSIASFC	Trunkamide point mutant	7
15	TVIAPFC	Trunkamide point mutant	7
16	ISFPCIP	Mollamide natural product[43]	7
17	IPISFPC	Mollamide natural product[43]	7
18	TSIAP-**pBrF**-C	Trunkamide + non-proteinogenic	7
3	TVPVPTLC	Patellin 3 natural product	8
19	TLPVPTVC	Patellin 4 natural product	8
20	TVPVPSFC	Patellin 5 natural product	8
21	CTLCCTLC	*pat-tru* hybrid (ulithiacyclamide)[44,5]	8
22	VTACITFC	*pat-tru* hybrid (patellamide C)[45,5]	8

[a].Wild-type compounds **1**–**3** were mutated, with changes to wild type shown in bold. [b].Compounds were designed based on known natural products that had not been previously expressed, or based on sequences identified by genome mining (see Discussion for full description). [c].Macrocylic ring size, in number of amino acids.

Chemical Analysis and Isolation. After five days of fermentation, the pelleted *E. coli* cells were extracted with methanol. The organic extracts were partially purified and then analyzed using HPLC-ESI-MS. In all cases, heterologous expression of control compounds **1** and **3** was confirmed, with the compounds eluting with the same profile as standards of authentic **1** and **3**, which we obtained from a marine animal

dx.doi.org/10.1021/ja208278k | *J. Am. Chem.Soc.* 2012, 134, 418–425

Table 2. Expression Yields in *E. coli*

#	sequence	yield new (μg/L)[a]	yield (**1**) (μg/L)	yield (**3**) (μg/L)	total (μg/L)	rel yield (%)
1	TVPTLC	Std				
4	TLATIC	8.8/16[a]	26/13	15/5.9	85	64[b]
5	IVPPFC	90/NA	47/10	74/22	240	160
6	TTVTAC	12/19	87/24	140/50	330	28
7	VTPFVC	ND	190/74	320/110	700	0
8	IPGSLC	ND	150/37	170/41	400	0
9	TVPT-BrF-C	ND/6.2	120/64	52/33	280	3.3
10	TVPT-ClF-C	3.2/28	170/130	100/70	500	10
11	TVPT-OMeF-C	ND/5.8	120/32	120/70	350	4.0
12	TVPT-N$_3$F-C	ND/0.8	99/64	74/55	290	5.2
2	TSIAPFC	Std				
13	TSIAPLC	44/130	86/37	44/6.6	350	140
14	TSIASFC	54/13	74/24	48/29	240	68
15	TVIAPFC	ND	51/11	56/16	130	0
16	ISFPCIP	ND	3.9/ND	1.2/1.9	7.0	0
17	IPISFPC	ND	96/84	62/17	260	0
18	TSIAP-BrF-C	5.0[c]	97/79	66/33	280	2.8
3	TVPVPTLC	Std				
19	TLPVPTVC	7.7/10	40/68	38/10	170	17
20	TVPVPSFC	41/33	170/55	120/85	500	33
21	CTLCCTLC	ND	7.7/0.6	13/3.0	24	0
22	VTACITFC	4.8/2.7	30/29	10/1.9	78	13

[a]Yields shown are for new compounds, compound **1**, compound **3**, and the total of all compounds found in a single expression experiment. Yields are reported as fully prenylated/−1 prenyl; i.e., for compound **4**, 8.8 μg L^{-1} of diprenylated and 16 μg L^{-1} of monoprenylated compounds were produced. [b]Rel yield (%) indicates the total amount of new compounds, divided by the total amount of compound **1**, found in an individual sample (times 100 to give percentage). This gives the most accurate assessment of relative yield of new compounds between experiments. [c]Only the nonprenylated **18** was identified, indicating that **18** is not a substrate for TruF1/F2.

as previously described (Figure 3).[6] We previously showed that isoprene is readily lost from Ser and Thr cyanobactin derivatives under standard MS conditions.[25] Thus, loss of isoprene reliably indicates the formation of mature cyanobactins, and this loss is not observed in any natural *E. coli* compound. In addition, we observed ions representing a total of 16 out of 22 recombinant cyanobactins. A table was constructed in which the recombinant sequence determined by DNA sequencing was used to predict a unique mass for the new cyanobactin and for the loss of one or more isoprene groups from each predicted new compound. The prediction ions were only observed in expression experiments involving the sequence in question, and not in other experiments, so that we essentially had 19 negative control experiments for each compound produced (Figure 3). In addition, if prenylated, the ions readily fragmented to lose the predicted numbers of isoprene groups, in contrast to all other *E. coli* metabolites. Finally, in most cases we observed incomplete prenylation, so that mono-, di- and sometimes triprenylated derivatives were formed in *E. coli*. In total, we identified 21 new compounds, produced in *E. coli*, that matched the masses predicted from DNA sequencing.

To further confirm the expression of the predicted compounds, they were subjected to analysis by high-resolution LC-FT-ICR-MS/MS, using previously established methods.[24,25,30] The ions did indeed reflect the predicted compounds, to <2 ppm value, with loss of isoprene observed in MS/MS. When no isoprene was present on a compound, the fragmentation pattern reflected the sequence of the peptide. Thus, because of the numerous internal and external controls, the sequencing data, and the well-validated MS data, we had high confidence about the identity of expressed products.

Finally, to further demonstrate that the recombinant compounds were successfully produced, we selected a set of representative compounds, **1**, **10**, and **14**, for NMR analysis. These compounds were purified to homogeneity from the *E. coli* cell pellets, and their ^{1}H NMR spectra were obtained (Figures S4−S7 and Table S1 in Supporting Information [SI]). In all cases, the NMR spectra matched those predicted for the new compounds, showing that they were the predicted cyanobactins.

Expression of Mutant Cyanobactins. Hexameric patellin 2 (**1**) was manipulated to create compounds **4**−**8**, which were triple or quadruple mutants of the wild type sequence (Tables 1 and 2). The triple mutants (**4**−**6**) were successfully synthesized in *E. coli*, while the quadruple mutant **7** was not detected. We previously reported successful expression of **4**,[31] while all other compounds **5**−**22** are reported here for the first time.

In contrast to hexapeptide derivatives, for which diverse sequences are known, there are few known heptameric *tru* derivatives. Therefore, we created single-point mutants **13**−**15** of heptameric trunkamide (**2**). Only **13** and **14** were successfully processed in vivo. Pentuple mutants **16** and **17** were also not formed. Octapeptide selectivity was examined by synthesizing double mutants **19** and **20**, pentuple mutant **21**, and hextuple mutant **22**. Surprisingly in comparison with our experience with heptapeptides, **19**, **20**, and **22** were synthesized.

If more than one prenylation event was possible, we detected both singly and multiply prenylated derivatives. (For derivatives **9**, **11**, and **12** we detected only monoprenylated compounds, while for **18**, no prenylation was detected, indicating that the molecule was not a substrate for TruF1/F2.) This was true even for the wild type compounds in *E. coli*, showing that complete prenylation is a limiting step in this host. We also observed this pattern in systems that produce a lower cyanobactin yield, indicating that it may be an intrinsic pathway property and not linked to the amount of dimethylallylpyr-

ophosphate available in *E. coli*. In known natural products, TruD heterocyclase only modifies the C-terminal Cys residue. However, we previously showed that TruD modifies internal Cys residues in unnatural substrates in vitro.[16] The primary sequence of compound **22** was derived from the *pat* biosynthetic pathway,[23] in which all Cys and Thr/Ser residues are heterocyclic. However, in **22**, two Cys residues were heterocyclic, but the two Thr residues were prenylated. This reveals that the *pat* and *tru* pathways may be hybridized to create diverse new derivatives.

Incorporation of Non-Proteinogenic Amino Acids. We sought to incorporate non-proteinogenic amino acids using an orthogonal tRNA/aminoacyl-tRNA synthetase (tRNA/aaRS) pair that incorporates a specific unnatural amino acid in response to nonsense or frameshift mutation.[24] First, we showed that non-proteinogenic amino acids could be incorporated into the TruE precursor peptide. The plasmid pEVOL-pAcF which can specifically incorporate The non-proteinogenic amino acid *p*-acetyl-phenylalanine (*p*-AcF) was coexpressed with plasmids encoding *truE* derivatives, which were C-terminally His-tagged and in which each position in patellin 2 was individually mutated to the amber codon TAG (Figure S3, SI). *p*-AcF was added to the fermentations. SDS-PAGE of Ni purified lysates revealed that *p*-AcF could be incorporated in all six positions in the context of a precursor peptide (Figure S3, SI). These experiments showed that there was no problem in placing non-proteinogenic amino acids into the TruE precursor peptide backbone, in the absence of modifying enzymes. However, we still needed to determine whether the modifying enzymes could accept the highly unnatural amino acids and carry them through 4−5 biosynthetic steps to the mature unnatural natural products.

We therefore attempted to coexpress this same system with ptru-SD1 and other ptru vectors. However, in extensive initial experiments with non-proteinogenic amino acids, including *p*-AcF and others, could not be incorporated into cyanobactins using standard conditions. Several enzymes must be temporally coordinated in order to produce the mature cyanobactins over the course of our five-day fermentation. In extensive previous work with this pathway, changing the coordinated regulation of this pathway in any way severely impacted product yield.[25] For example, even providing each gene in the eight-gene *tru* pathway with its own heterologous promoter was unsuccessful. It is clear that the complex regulation of this multistep pathway remains poorly understood.

Previously, it was noted that incorporation of non-proteinogenic amino acids into single proteins over relatively long (14 h) induction times could be optimized with two copies of aaRS, one of which was constitutive and one of which was inducible with arabinose.[32] In fact, ∼1 day is the longest expression period to which this methodology has been previously applied, to the best of our knowledge. The five-day fermentation time for heterologous expression of *tru* is much longer than this, and in addition, the complex pathway regulation of *tru* is poorly understood. Therefore, we reasoned that adjustment of aaRS induction might improve cyanobactin production. Specifically, we thought that constitutive expression of aaRS might be better matched to the constitutive expression of *tru*. The plasmid pEVOL-pCNF (which encodes a polyspecific aaRS that incorporates a variety of Phe derivatives but not Phe or Tyr to any significant extent)[25] was used. We varied expression of the arabinose-inducible copy of aaRS in pEVOL-pCNF using different concentrations of inducer (L-arabinose). Optimal incorporation of *p*-chloro-phenylalanine was achieved in the absence of arabinose, while increasing arabinose concentrations and thus the flux of aaRS incrementally decreased product yield (Figure 4). With optimized conditions in hand, pEVOL-pCNF was used in the preparative synthesis of **9**−**12**, in which Leu of patellin 2 (**1**) was replaced with substituted phenylalanine derivatives, and **18**, in which Phe of trunkamide (**2**) was replaced with *p*-bromophenylalanine (Figure 2 and Table 1). Out of the many possible amino acids that might be incorporated into engineered natural products using pEVOL-pCNF, we chose a limited subset for this work not including *p*-AcF.

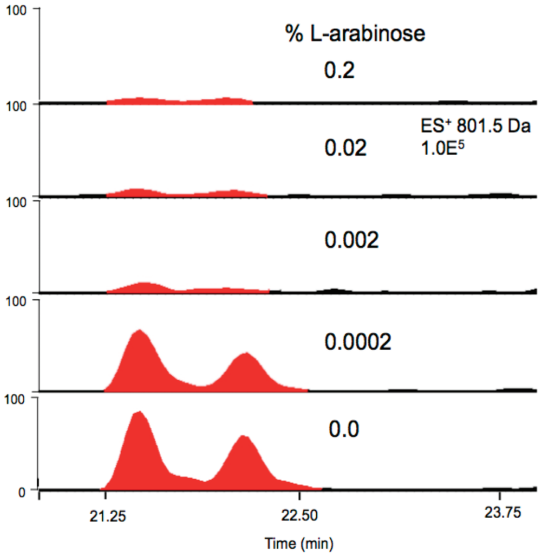


Figure 4. Optimization of non-proteinogenic amino acid incorporation. Compound **10** was synthesized in *E. coli* and analyzed by LC-ESI-MS. Optimum production was achieved in the absence of inducer (arabinose), whereas higher concentrations of inducer completely repressed synthesis. Two peaks are present because multiple stereoisomers are present (the α-proton adjacent to thiazoline is labile, and Pro undergoes *cis*−*trans* isomerization in this family).[26].

Yield Measurement. Yields were determined for all heterologous expression experiments. First, purified **1** was quantified by NMR integration using several different concentrations of the standard, 1,4-dinitrobenzene,[33] to give an isolated yield of 260 μg from a 10-L fermentation. This method is much more accurate than other methods for microgram quantities of material, since it measures the relative molar amounts of compound directly using a calibrant.

With a standard concentration of **1** available, we used known concentrations of **1** as internal and external standards for HPLC-ESI-MS experiments. These experiments used crude fractions containing recombinant peptides, so that they much more closely resembled the native production yield, and not merely the isolated yield. For example, the area under the curve was integrated to show that 1.7 mg of **1** was produced in the 10 L *E. coli* expression described above, but the isolated yield was only ∼10% after several purification steps. An additional 1.3 mg of the monoprenylated variant of **1** was also produced in this 10 L fermentation, leading to an overall yield of 3 mg of **1** variants.

dx.doi.org/10.1021/ja208278k | *J. Am. Chem.Soc.* 2012, 134, 418−425

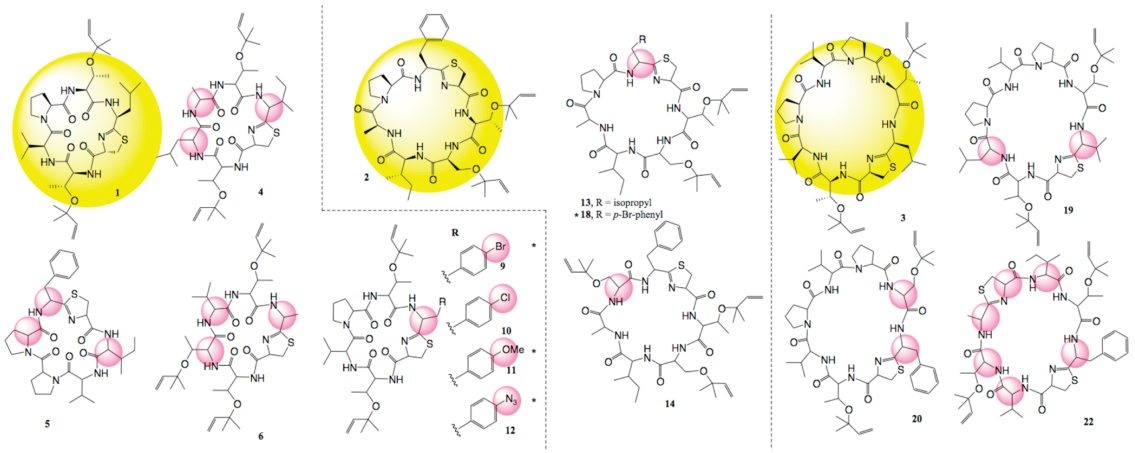


Figure 5. Compounds synthesized in this study. Yellow indicates wild-type compounds, while pink bubbles indicate mutations that deviate from wild type. Hexa- (left), hepta- (mid), and octa- (right) peptide derivatives were synthesized using these methods. (*) Indicates compounds for which only monoprenylated (**9**, **11**, and **12**) or nonprenylated (**18**) derivatives were identified. For all other compounds, both singly, doubly, and sometimes triply prenylated products were identified.

Because **1** was produced in all expression experiments reported herein, the compound could also be used to estimate the production level of other recombinant products. Yields of most compounds were in the range of 1−170 μg L^{-1}. (see Table 1). Compound **2** was also present as an internal control. A potential weakness of this method is that the relative ionization efficiencies of the compounds are likely to differ. However, the ratios were also compared to those achieved using FT-ICR-MS, showing a consistent relative ratio of ions in both methods. Moreover, in previous studies we have not seen a large range of ionization efficiencies for these similar compounds, indicating that the reported yields are good estimates. The relative yields determined by MS were also reflected in the isolated yields of **10** and **14**.

DISCUSSION

In this study, we exploited the relaxed substrate selectivity of the *tru* pathway to generate diverse mutants, despite the need to go through a complex biosynthetic route. Although not all derivatives can be made, this study shows that the *tru* posttranslational machinery exhibits broad substrate tolerance. We also report the first in vivo incorporation of non-proteinogenic amino acids into multistep RP products. Notably, control of six different enzymes, two precursor peptides, and a tRNA molecule was required for synthesis of the reported derivatives. Incorporated, non-proteinogenic amino acids were successfully carried through a multistep pathway with 12 individual enzymatic transformations. This was accomplished using eight enzyme *tru* domains encoded in six *tru* proteins, whereas previously at most one enzyme with two domains has been used.[11] We expect that the methods here will be widely useful in the synthesis of diverse RP derivatives encoding non-proteinogenic amino acids.

A simple but important advance was the observation that low-level, constitutive expression provides excellent incorporation of non-proteinogenic amino acids in natural product derivatives. Prior to this report, the described technology had been used routinely for incorporation of amino acids into proteins, and a few reports exist for incorporation into short peptides which undergo one further posttranslational modification. In these cases, the relative simplicity of the heterologous expression systems made them very tolerant of overexpression. In more complex systems, our data suggest that producing too much of the aaRS (or any other component in a complex pathway) is detrimental to efficient product synthesis.

There are now many different methods for implanting non-proteinogenic amino acids into ribosomal peptides, some of which involve changing the ribosomal code and some of which involve synthetic modifications.[34] Notably, recently the concepts of multiple orthogonal substitution and residue-specific substitution have been elaborated.[35,36] The former enables different non-proteinogenic amino acids to be introduced into single peptides, while the latter enables all residues of a specific type to be replaced by non-proteinogenic amino acids. These methods are compatible with the ideas developed here, but they will undoubtedly require some further study to be applied to highly modified RPs, since so far they have not been applied to such compounds.

In addition to exploring incorporation of diverse substitutions, these experiments were designed to provide proof-of-concept for key technological advances in marine natural products, in five major areas. First, we have recently applied a genomic method, in which we sequence genes from marine animals and their symbiotic bacteria to discover new TruE-like peptides.[25,30] Subsequently, we find that the predicted peptides are the major natural products found in whole animals. However, the compounds are exceptionally limited in supply, and at most micrograms are available from the natural sources. As mentioned above, previously we expressed **4** to supply the compound. Here, we wished to test the broad applicability of the method by expressing compounds **4**-**6**, which were first identified by genomics and which were successfully made by *E. coli* here.

Second, there were a series of previously described, rare compounds from Nature, including mollamide (encoded by **16** and **17**) and keenamide (encoded by **8**).[19,37] These compounds were reported to exhibit interesting bioactivity but are in very short supply. However, these compounds were not produced by our *tru* pathway in *E. coli*, indicating that they are probably synthesized by a different variation of the *tru* pathway found in

dx.doi.org/10.1021/ja208278k | *J. Am. Chem.Soc.* 2012, 134, 418−425

Nature. Third, we wished to express known marine animal compounds, patellins 4 and 5 (**19** and **20**),[19,38] for which we had not previously found genes; these were successfully made.

Fourth, trunkamide is a potential anticancer agent of interest because of its unusual profile in the National Cancer Institute's 60-cell line panel. We tested its potential for derivatization by making a series of point mutants, **13**−**15**. Interestingly, addition of a Ser led to a third prenylation event in **14**, and **13** was successfully produced. However, a simple Ser-Val substitution in **15** led to loss of product.

Fifth, for compounds **21** and **22**, we wished to cross the *tru* pathway with the *pat* pathway that leads to patellamides. Patellamides are not prenylated, and instead Ser and Thr residues are cyclized to yield oxazoline residues.[5,39,40] Moreover, often 2 thiaz*ole* residues exist in the *pat* products, instead of 1 thiaz*oline* as found in all known *tru* pathway relatives. The genetic and biochemical basis of these differences have been thoroughly established. We have been very interested in determining the "portability" of enzymes; that is, can individual RP enzymes be moved from one pathway to another? As a first step to establish this fact, we cloned patellamide precursor peptides that would normally yield the compounds ulithiacyclamide and patellamide C into the *tru* pathway background, to give derivatives **21** and **22**. **21** could not be produced, but **22** was successfully synthesized and contained 2 thiazoline residues, the first time this modification has been reported in *tru* derivatives. Both Thr residues were prenylated, rather than heterocyclic as they would be in the natural product.

This report represents a significant expansion of the toolkit available for directed, posttranslational modification of peptides in living cells. Additionally, it provides the first technical guidelines for performing complex manipulations with multistep RP pathways. Ultimately, we hope to improve control over RP processing for directed synthetic biology of fine chemicals and pharmaceuticals. This technology has numerous possible applications, including supply of marine natural products and improvement of anticancer properties of cyanobactins. More importantly, such technology allows simple genetic engineering tools to be applied to creation of wholly new drug motifs that combine the advantages of peptide technology and synthetic chemistry. For example, intein-circularized and phage display libraries encode an enormous sequence diversity which can be used to attack diverse biological targets,[40] but the resulting compounds are generally not drugs because of poor pharmacological properties. The addition of many posttranslational modifications promises to afford the same sequence diversity, but with products that are more "drug-like". Here, we use the multistep cyanobactin pathway to explore the chemistry that necessarily underlies these downstream applications.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Methods, supporting data, and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

ews1@utah.edu

Present Addresses

[§]Department of Bioengineering and Therapeutic Sciences and California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California 94158, United States.

[⊥]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, United States.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

E.W.S. is a co-owner of Symbion Discovery, Inc.

■ ACKNOWLEDGMENTS

This work was funded by NIH GM071425 and GM071425-S1 to E.W.S. and NIH GM097206 to P.G.S. We thank K. Parsawar and C. Nelson (University of Utah Mass Spectrometry Core Facility) for acquiring high-resolution mass spectra, J. Skalicky (University of Utah) for help with NMR, and D. E. Ruffner (Symbion Discovery, Inc.) for providing ptru-SD1.

■ REFERENCES

(1) Mitchell, W. *Curr. Opin. Chem. Biol.* **2011**, *15*, 505.
(2) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26*, 537.
(3) Velterop, J. S.; Sellink, E.; Meulenberg, J. J.; David, S.; Bulder, I.; Postma, P. W. *J. Bacteriol.* **1995**, *177*, 5088.
(4) Oman, T. J.; van der Donk, W. A. *Nat. Chem. Biol.* **2010**, *6*, 9.
(5) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729.
(6) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341.
(7) Widdick, D. A.; Dodd, H. M.; Barraille, P.; White, J.; Stein, T. H.; Chater, K. F.; Gasson, M. J.; Bibb, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4316.
(8) Kelleher, N. L.; Hendrickson, C. L.; Walsh, C. T. *Biochemistry* **1999**, *38*, 15623.
(9) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477.
(10) Melby, J. O.; Nard, N. J.; Mitchell, D. A. *Curr. Opin. Chem. Biol.* **2011**, *15*, 369.
(11) Houssen, W. E.; Wright, S. H.; Kalverda, A. P.; Thompson, G. S.; Kelly, S. M.; Jaspars, M. *ChemBioChem* **2010**, *11*, 1867.
(12) Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 5016.
(13) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. *Angew. Chem., Int. Ed.* **2011**, *50*, 8714.
(14) Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. *J. Am. Chem. Soc.* **2010**, *132*, 7519.
(15) Li, C.; Zhang, F.; Kelly, W. L. *Mol. Biosyst.* **2011**, *7*, 82.
(16) Shi, Y.; Yang, X.; Garg, N.; van der Donk, W. A. *J. Am. Chem. Soc.* **2011**, *133*, 2338.
(17) Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11052.
(18) Sivonen, K.; Leikoski, N.; Fewer, D. P.; Jokela, J. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 1213.
(19) Donia, M. S.; Schmidt, E. W. In *Comprehensive Natural Products II*; Mander, L, Liu, H.-W., Eds.; Elsevier: Oxford, 2010; Vol 2, p 539.
(20) Lee, J.; McIntosh, J.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122.
(21) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 4089.
(22) McIntosh, J. A.; Robertson, C. R.; Agarwal, V.; Nair, S. K.; Bulaj, G. W.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 15499.
(23) McIntosh, J. A.; Schmidt, E. W. *ChemBioChem* **2010**, *11*, 1413.
(24) McIntosh, J. A.; Donia, M. S.; Nair, S. K.; Schmidt, E. W. *J. Am. Chem. Soc.* **2011**, *133*, 13698.

(25) Donia, M. S.; Ruffner, D. E.; Cao, S.; Schmidt, E. W. *ChemBioChem* **2011**, *12*, 1230.
(26) Carroll, A. R.; Coll, J. C.; Bourne, D. J.; MacLeod, J. K.; Zabriskie, T. M. *Aust. J. Chem.* **1996**, *49*, 659.
(27) Wipf, P.; Uto, Y. *J. Org. Chem.* **2000**, *65*, 1037.
(28) Aller, S. G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P. M.; Trinh, Y. T.; Zhang, Q.; Urbatsch, I. L.; Chang, G. *Science* **2009**, *323*, 1718.
(29) Williams, A. B.; Jacobs, R. S. *Cancer Lett.* **1993**, *71*, 97.
(30) Donia, M. S.; Fricke, W. F.; Ravel, J.; Schmidt, E. W. *PLoS One* **2011**, *6*, e17897.
(31) Donia, M. S.; Schmidt, E. W. *Chem. Biol.* **2011**, *18*, 508.
(32) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2010**, *395*, 361.
(33) Rundlof, T.; Mathiasson, M.; Bekiroglu, S.; Hakkarainen, B.; Bowden, T.; Arvidsson, T. *J. Pharm. Biomed. Anal.* **2010**, *52*, 645.
(34) Budisa, N. *Angew. Chem., Int. Ed.* **2004**, *43*, 6426.
(35) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464*, 441.
(36) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Curr. Opin. Chem. Biol.* **2010**, *14*, 774.
(37) Wesson, K. J.; Hamann, M. T. *J. Nat. Prod.* **1996**, *59*, 629.
(38) Fu, X.; Do, T.; Schmitz, F. J.; Andrusevich, V.; Engel, M. H. *J. Nat. Prod.* **1998**, *61*, 1547.
(39) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7315.
(40) Tavassoli, A.; Benkovic, S. J. *Nat. Protoc.* **2007**, *2*, 1126.
(41) Zabriskie, T. M.; Foster, M. P.; Stout, T. J.; Clardy, J.; Ireland, C. M. *J. Am. Chem. Soc.* **1990**, *112*, 8080.
(42) Donia, M. S.; Wang, B.; Dunbar, D. C.; Desai, P. V.; Patny, A.; Avery, M.; Hamann, M. T. *J. Nat. Prod.* **2008**, *71*, 941.
(43) Carroll, A. R.; Bowden, B. F.; Coll, J. C.; Hockless, D. C. R.; Skelton, B. W.; White, A. H. *Aust. J. Chem.* **1994**, *47*, 61−69.
(44) Ireland, C.; Scheuer, P. J. *J. Am. Chem. Soc.* **1980**, *102*, 5688.
(45) Ireland, C. M.; Durso, A. R.; Newman, R. A.; Hacker, M. P. *J. Org. Chem.* **1982**, *47*, 1807.

Supporting Information for:

Ribosomal route to small molecule diversity

Ma. Diarey B. Tianero, Mohamed S. Donia, Travis S. Young, Peter G. Schultz, and Eric W. Schmidt

Table of Contents

Materials and Methods

Plasmids. ptru-SD1 was obtained from Symbion Discovery, Inc., in which the *tru* operon is controlled by *lac*. This vector performs identically to the previously reported TOPO-tru, except that yield is improved. ptruE is derived from the previously reported pRSF-true2 vector in which the TruE peptide is present in the pRSF vector backbone and TruE expression is controlled by *lac*.[1] Restriction sites were introduced so that the second cassette of TruE2 could be replaced with desired variants. Top strand and bottom strand oligonucleotides encoding the desired cyanobactin variant were synthesized at the University of Utah Peptide Core Facility. The two strands were annealed ligated to ptruE digested with the corresponding restriction enzymes. T4 ligase (NEB) was used for cloning and *E. coli* TOP10 (Invitrogen) was used for the propagation of the plasmids. All constructs were verified by DNA sequencing. Using this conventional cloning method, ptruE-X1-ptruEX18 and ptruE-amber1 and ptruE-amber2 were constructed and verified by sequencing. These vectors are identical to previously reported pRSF-true2,[1] except that the second cassette encodes the new cyanobactin variant.

Expression of cyanobactin variants. ptru-SD1 and ptruE wild type or mutant vectors were co-transformed into Top10 *E. coli* cells (Invitrogen) following the manufacturer's protocol and plated on LB agar plates supplemented with ampicillin (50 μg ml^{-1}) and kanamycin (25 μg ml^{-1}). Because expression from individual colonies is variable,[1] for each mutant five to ten colonies were picked and grown overnight in 2xYT broth (5 ml). These seed cultures were pooled, and pooled broth (1 ml) was inoculated into 2xYT broth (100 ml) in a 250 ml flask. All 2xYT media contained ampicillin (50 μg ml^{-1}), kanamycin

(25 μg ml^{-1}), and a proprietary Symbion Mix that slightly improves yield but does not affect *lac* expression. The cells were cultivated for five days at 30 ^{o}C and 200 rpm, after which they were collected by centrifugation at 3320 x *g* for 20 minutes. The cell pellet was washed with NaCl (100 mM) and extracted by sonication in methanol (75 ml) for 30 minutes. The suspension was filtered, Diaion HP20 (5% v/v) was added to the filtrate, and the mixture was dried under vacuum to remove most of the methanol. The resin was transferred into vac elut cartridges (Varian) and washed with water, followed by a step gradient of methanol (25%, 50%, 100%) in water. The 100% methanol fraction was dried under vacuum and re-suspended in methanol (500 μl). An aliquot (50 μl) of this extract was passed through a small plug of end-capped C_{18} resin and analyzed by HPLC-ESI-MS using a Waters Micromass ZQ mass spectrometer and a Phenomenex C_{18} column (10 cm x 4.6 mm, 5μm). A linear gradient was employed at 0.2 ml min^{-1}, from 10-100% acetonitrile in water (0.1% formic acid) over 20 minutes, followed by 15 minutes at 100% acetonitrile.

Co-expression of truE with amber suppressor tRNA/aaRS. Seven pRSF-TruE constructs containing all six possible amber stop codon mutants in the TVPTLC sequence and the wild type (WT) sequence were independently co-transformed with pEVOL-*p*AcF (incorporates *p*-acetylphenylalanine in response to the amber stop codon) in BL21(DE3) cells. Single colonies were picked for expression in LB media containing kanamycin (25 μg ml^{-1}), and chloramphenicol (25 μg ml^{-1}) and grown overnight. Cultures (250 ml) were grown to OD_{600}=0.8 and induced with IPTG (1 mM) and arabinose (0.02%) and supplemented with *p*AcF (1 mM). One culture of the pRSF-WT variant was not induced

with IPTG (indicated as uninduced in Figure S3). After 12 hours, cultures were pelleted, lysed in Bugbuster cell lysis reagent (5 ml, Novagen), and purified using Ni-NTA resin (0.25 ml, Qiagen) following protocols in the Qiagen protein purification handbook for native purification. Eluents were concentrated 5-fold before loading onto a 4-20% SDS-PAGE gel (Novagen) with DTT (20 mM). The gel was stained with a solution of Coomassie Blue R-250 (0.25%, Sigma) and washed before imaging. The presence of full-length TruE protein in constructs containing an amber mutation indicated successful incorporation of the *p*AcF unnatural amino acid.

In vivo synthesis of cyanobactins with non-protein amino acids. ptru-SD1 (Symbion Discovery, Inc.), ptruE-amber, and pEVOL-pCNF containing the amber suppresor tRNA/aaRS were co-transformed in Top10 *E. coli* cells following the manufacturer's protocol and plated on LB agar plates supplemented with ampicillin (50 μg ml^{-1}), kanamycin (25 μg ml^{-1}), and chloramphenicol (25 μg ml^{-1}). The resulting clones were treated as described in *Expression of cyanobactin variants*, except that chloramphenicol (12.5 μg ml^{-1}) and non-proteinogenic amino acids (Chemimpex; 2 mM) were also added to the culture broth.

To optimize induction, experiments were run as described above, except that growth was monitored until OD_{600}=0.8, at which time L-arabinose was added at varying concentrations (0%, $2x10^{-4}$%, $2x10^{-3}$%, $2x10^{-2}$%, $2x10^{-1}$%). Expression using pEVOL was optimized using *p*-chloro-L-phenylanine (*p*-ClF) (2mM), which was added in batches. At inoculation, *p*-ClF was added to 0.7 mM and growth was monitored until

OD_{600}=0.8 at which time L-arabinose and *p*-ClF (1.3 mM) were added. Later, this batch addition was found not to make a difference, so that exogenous amino acids (2 mM) were subsequently added upon inoculation.

Characterization of fermentation products. Fractions containing new and known cyanobactin derivatives were partially purified using HP20 and C_{18} resins, as described above, and then analyzed by HPLC FT-ICR-ESI-MS at the University of Utah Mass Spectrometry Core Facility using a LTQ-FT (ThermoElectron), with MS/MS using collision induced dissociation (Table S2 and Figure S4).

*Preparative fermentation of TVPT-pClF-C (**11**) and TSIASFC (**15**).* Seed cultures were used to inoculate 2xYT broth (10 l) containing additives as above in a 14 l fermentor (Bioflo 110, New Brunswick Scientific). Fermentation was performed over five days at 30 ^{o}C and 200 rpm, after which the cells were collected by centrifugation at 4424 x *g* and washed with NaCl (100 mM). The pellet was extracted with acetone (1.5 l). Diaion HP20 (5% v/v) was added, and the slurry was dried under vacuum. The resin was then washed extensively with water, followed by 40% acetone. Finally, an extract enriched with desired compounds was obtained by elution with acetone (500 ml). The vacuum-dried acetone extract was fractionated by flash chromatography on silica gel (50 g, 230-400 mesh) pre-equilibrated in 9:1 hexane: ethyl acetate. A step gradient was employed consisting of the following solvents: 1) 9:1 hexane: ethyl acetate (100 ml); 2) 1:1 hexane: ethyl acetate (100 ml); 3) ethyl acetate (150 ml); 4) 10:1 dichloromethane: methanol (200 ml); 5) 5:1 dichloromethane: methanol (200 ml). Fractions 3 and 4, containing cyanobactin derivatives, were further purified by HPLC on a Hitachi LaChrom Elite

system using two rounds of chromatography. The first employed a gradient of 35-100 % acetonitrile (aq) over 40 minutes (2 ml min^{-1}) on a Supelco HS Discovery (25 cm x 10 mm, 5 μm) C_{18} HPLC column. The second used 72 % acetonitrile (aq; 1 ml min^{-1}) on a Phenomenex Gemini (25 cm x 4.6 mm, 5μm) C_{18} HPLC column. NMR data were obtained using a Varian INOVA 600 (^{1}H 600 MHz). For spectral data, see Figures S5 and S6.

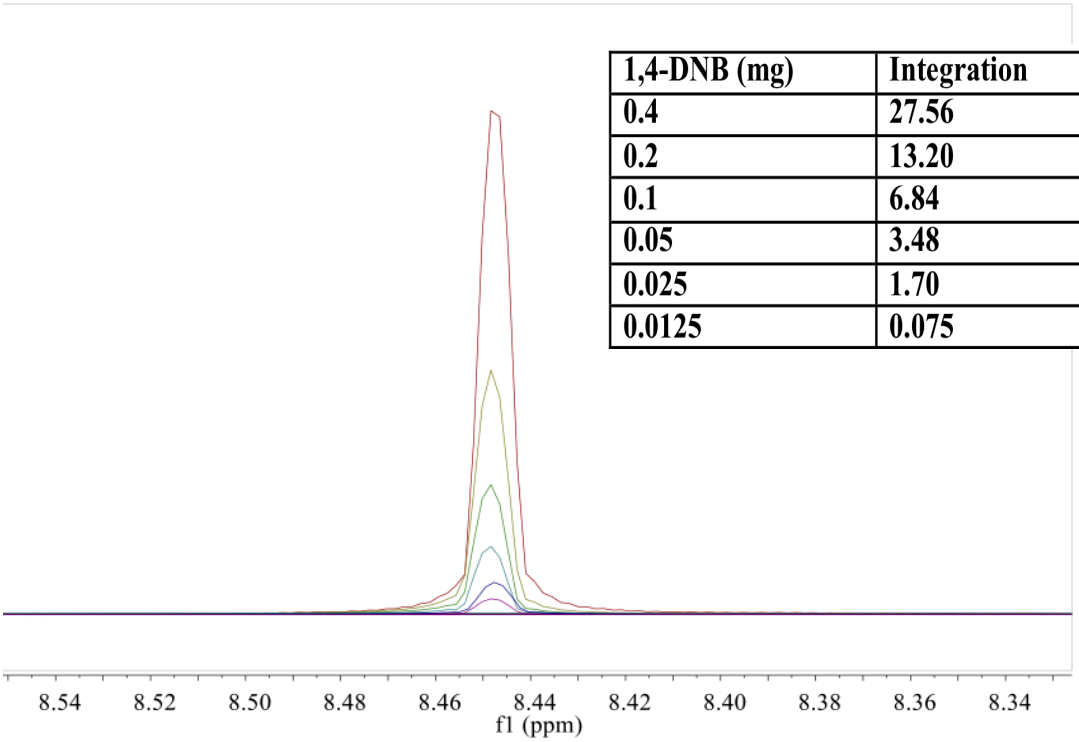
Yield determination. Yield of purified products could be determined by the usual methods, but we also sought more accurate determination using spectroscopic techniques, detailed below.

Quantification by NMR. ^{1}H NMR quantification was performed on a Varian INOVA digital 500 MHz NMR spectrometer at 25 ^{o}C. Quantification was performed using an external standard, 1,4-dinitrobenzene (DNB) (Sigma-Aldrich), following a previously validated method.[2] The standard **1** and different dilutions of DNB were prepared in DMSO-d_6 (Cambridge Isotopes) (120 mL) in identical, vacuum-dried NMR tubes (3mm). All samples were treated identically by placing into the probe and tuning and matching to identical values. All ^{1}H NMR experiments were run using the same conditions, with repeated single 90^o pulse sequence, D1(*5T1*) of 4s, and 512 scans as determined for appropriate signal averaging. Calibration of the p90^o pulse length for each samples was done by determination of the null (360^o) pulse length (pw=pw/4). T1 relaxations for both the **1** and DNB were obtained by inversion recovery method.

A two-fold dilution series of DNB (6 samples from 0.4 to 0.0125 mg) was prepared in DMSO-d_6 from a stock solution (34 mg/mL). The singlet signal at 8.4 ppm corresponding to four equivalent protons of DNB was integrated and used to generate a standard curve under various conditions (Figure S1). The ^{1}H NMR spectrum of **1** was measured under identical conditions, and well-separated NH (d 7.4X ppm) and methyl (d 0.73 ppm) signals were integrated and compared to the standard curve of DNB (Figure S1). Under several different conditions, reproducible values were obtained using this method, which met the previously described accuracy level.[2]

Quantification by HPLC-MS. The quantified NMR sample measured above was used directly from the NMR tube to generate highly accurate standards for MS. To generate a standard curve, a serial dilution of known concentrations of **1** (five different concentrations, from 0.26 to 0.032 mg in 40 mL MeOH) was used (Figure S2). This curve was generated both in pure MeOH and by adding the diluted standard to crude of the E. coli expression constructs reported above. Samples were then run under the LC-MS conditions described above. The area under the curve was determined by obtaining the spectrum corresponding to **1** over its ~2.5 min elution time, then multiplying this time by the average signal intensity at m/z = 733.5 over that time. Analysis was performed in duplicate and the integrations for each point were averaged. A plot of area vs. concentration was generated, indicating that the signal response to concentration was linear and that **1** provided a robust internal standard over the concentration range being measured. The resulting linear equation was used to calculate the concentrations of the recombinant and standard compounds (Figure S2).

Figure S1. A. Stacked ^{1}H-NMR spectra and integration values of 1,4-dinitrobenzene at different concentrations. **B**. Integration of sample **1** (CH_3).



1,4-DNB (mg)	Integration
0.4	27.56
0.2	13.20
0.1	6.84
0.05	3.48
0.025	1.70
0.0125	0.075

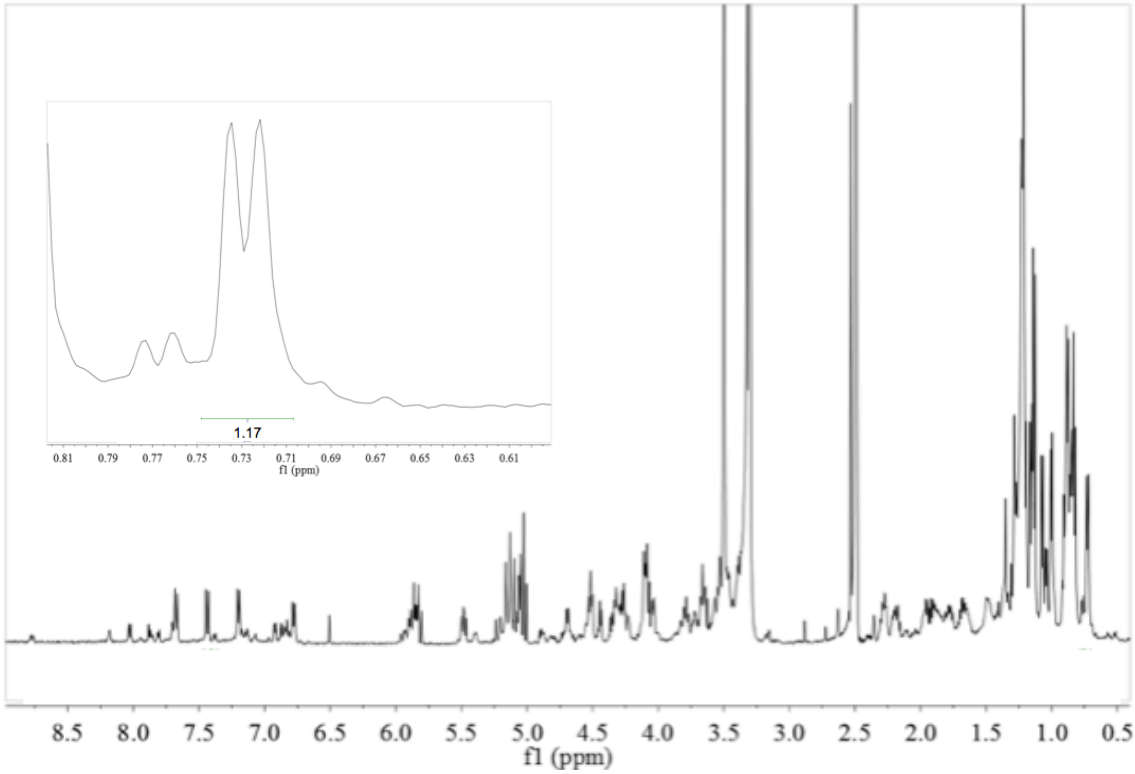


Figure S2. Standard curve with internal standard **1** (0.262 to 0.032 μg) added to IVVPFC crude extracts.

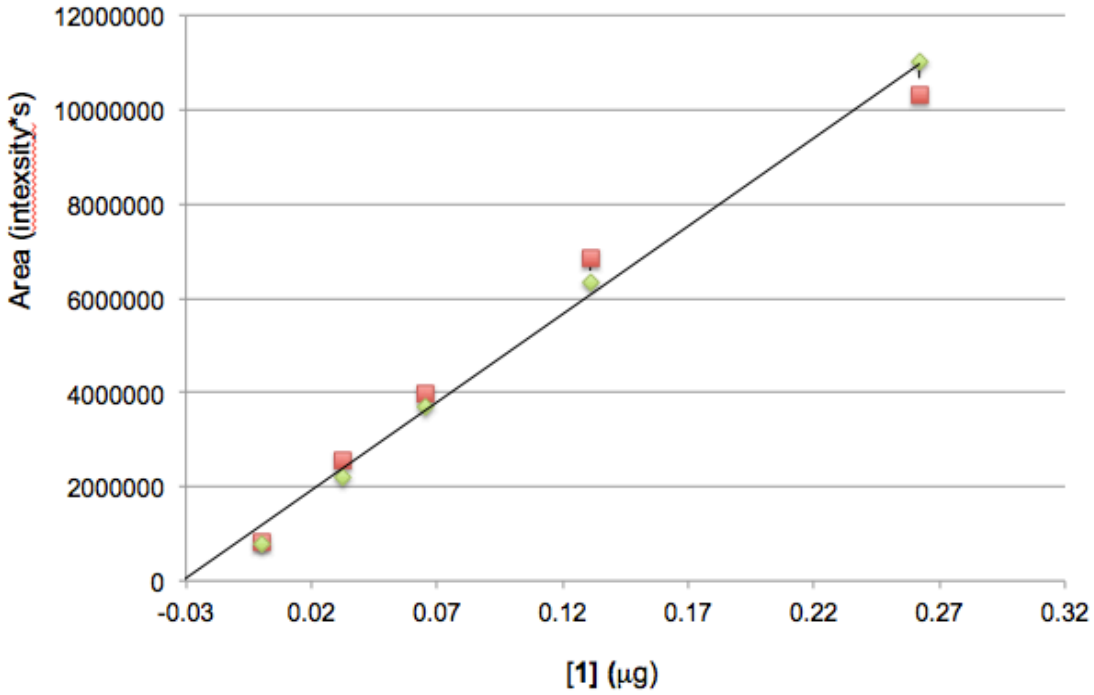


Table S1. Calculated and observed masses for compounds synthesized in *E. coli*. Prenylation state varied, and often derivatives were found that were not completely prenylated. Compounds with 1, 2 or 3 prenyl groups were identified as mixtures in single expression experiments.

Compound	calculated $[M+H]^+$	[a]observed $[M+H]^+$	[c]observed $[M+H-C_{10}H_{16}]^+$	[d]observed $[M+H-C_5H_8]^+$
TLATIC (**4**)	721.4317	721.4333	585.31	653.25
TLATIC (**4**) (i)[b]	653.3691	653.3701		585.31
IVPPFC (**5**)	639.3323	639.3325	-	-
TTVTAC (**6**)	763.4422	763.4431	627.21	695.20
TTVTAC (**6**) (i)	695.3796	695.3801	559.20	627.20
TVPT-Br-C (**9**) (i)	777.2640	777.2644	-	711.15
TVPT-pClF-C (**10**)	801.3770	801.3793	665.28	733.20
TVPT-pClF-C (**10**) (i)	733.3144	733.3158		665.27
TVPT-pOMeF-C (**11**) (i)	729.3640	729.3654	-	661.30
TVPT-pN3F-C (**12**) (i)	740.3548	740.3559	-	672.27
TSIAPLC (**13**)	804.4688	804.4699	668.30	736.34
TSIAPLC (**13**) (i)	736.4062	736.4065	-	668.30
TSIASFC (**14**)	896.4950	-	-	-
TSIASFC (**14**) (i)	828.4324	828.4341	692.26	760.21
TSIAP-pBrF-C(**18**)(ii)	780.2850	780.2388	-	-
TLPVPTVC (**19**)	929.5528	-	-	-
TLPVPTVC (**19**) (i)	861.4902	861.4914	-	793.38
TVPVPSFC (**20**)	949.5215	949.5228	813.41	881.45
TVPVPSFC (**20**) (i)	881.4589	881.4603	-	813.43
VTACITFC (**22**)	939.4830	939.4832	803.32	871.27
VTACITFC (**22**) (i)	871.4205	871.4225	-	803.33

a) all masses observed with +/- 2 ppm error
b) (i) indicates compounds produced with 1 less prenyl group
c) - $C_{10}H_{16}$ indicates loss of 2 prenyl groups (MS/MS)
d) - C_5H_8 indicates loss of 1 prenyl group (MS/MS)

Figure S3. Co-expression of precursor peptide constructs with pEVOL-*p*AcF where X indicates the position of the amber stop codon in the truE primary amino acid sequence. The wild type (WT) sequence contains no amber stop codon (sequence TVPTLC). The arrow indicates the position of full length TruE protein.

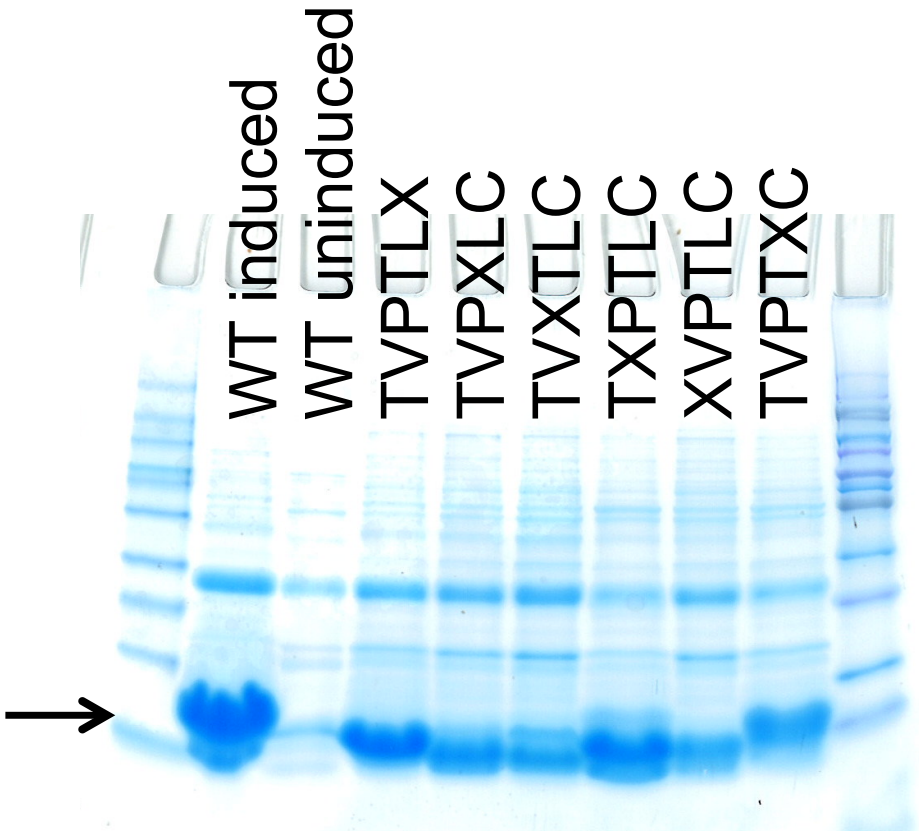
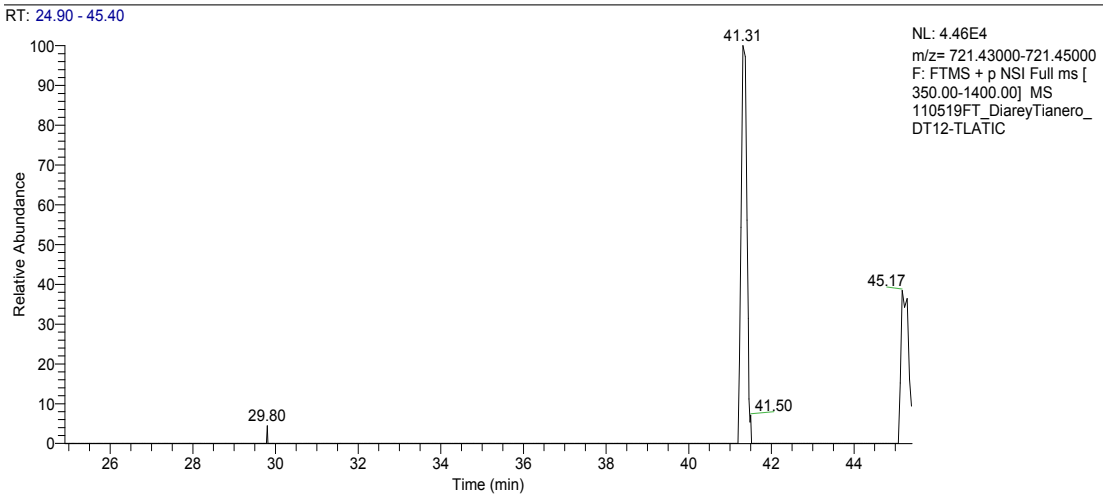
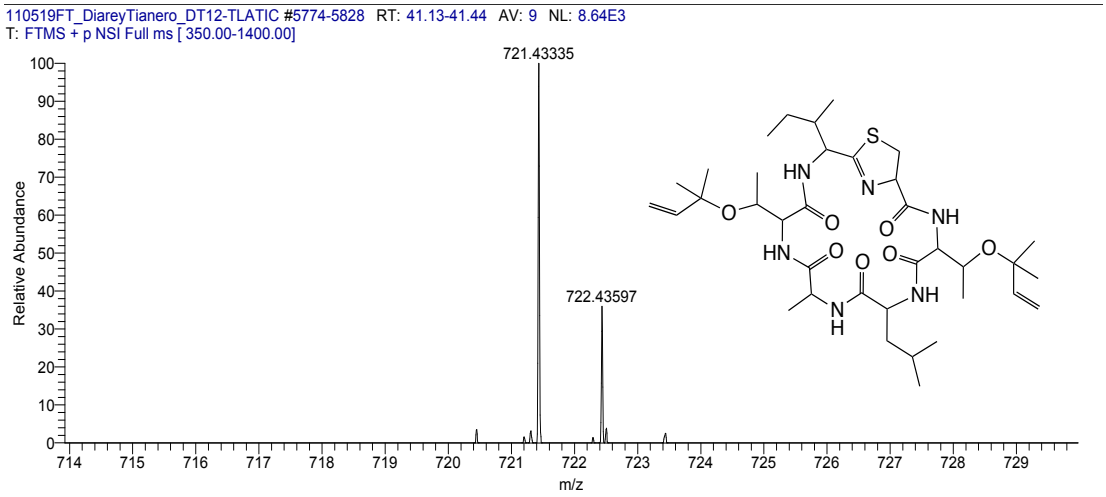


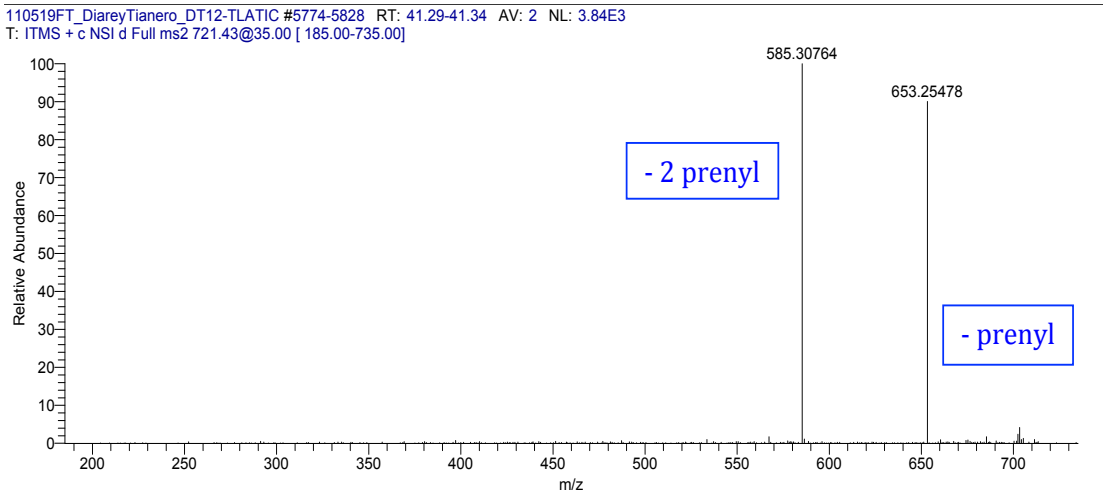
Figure S4.
A1 TLATIC (**4**) chromatogram
A2 TLATIC (**4**) spectrum
A3 TLATIC (**4**) MS-MS spectrum
B1 TLATIC (**4**) (i) chromatogram
B2 TLATIC (**4**) (i) spectrum
B3 TLATIC (**4**) (i) MS-MS spectrum
C1 IVPPFC (**5**) chromatogram
C2 IVPPFC (**5**) spectrum
C3 IVPPFC (**5**) MS-MS spectrum
D1 TTVTAC (**7**) chromatogram
D2 TTVTAC (**7**) spectrum
D3 TTVTAC (**7**) MS-MS spectrum
E1 TTVTAC (**7**) (i) chromatogram
E2 TTVTAC (**7**) (i) spectrum
E3 TTVTAC (**7**) (i) MS-MS spectrum
F1 TVPT-pBrF-C (**9**) (i) chromatogram
F2 TVPT-pBrF-C (**9**) (i) spectrum
F3 TVPT-pBrF-C (**9**) (i) MS-MS spectrum
G1 TVPT-pClF-C (**10**) chromatogram
G2 TVPT-pClF-C (**10**) spectrum
G3 TVPT-pClF-C (**10**) MS-MS spectrum
H1 TVPT-pClF-C (**10**) (i) chromatogram
H2 TVPT-pClF-C (**10**) (i) spectrum
H3 TVPT-pClF-C (**10**) (i) MS-MS spectrum
I1 TVPT-pOMeF-C (**11**) (i) chromatogram
I2 TVPT-pOMeF-C (**11**) (i) spectrum
I3 TVPT-pOMeF-C (**11**) (i) MS-MS spectrum
J1 TVPT-pN_3F-C (**12**) (i) chromatogram
J2 TVPT-pN_3F-C (**12**) (i) spectrum
J3 TVPT-pN_3F-C (**12**) (i) MS-MS spectrum
K1 TSIAPLC (**13**) chromatogram
K2 TSIAPLC (**13**) spectrum
K3 TSIAPLC (**13**) MS-MS spectrum
L1 TSIAPLC (**13**) (i) chromatogram
L2 TSIAPLC (**13**) (i) spectrum
L3 TSIAPLC (**13**) (i) MS-MS spectrum and assignments
M1 TSIASFC (**14**) chromatogram
M2 TSIASFC (**14**) spectrum
M3 TSIASFC (**14**) MS-MS spectrum
N1 TSIASFC (**14**) (i) chromatogram
N2 TSIASFC (**14**) (i) spectrum
N3 TSIASFC (**14**) (i) MS-MS spectrum

O1 TSIAS-pBrFC (**18**) (i) chromatogram
O2 TSIAS-pBrFC (**18**) (i) spectrum
O3 TSIAS-pBrFC (**18**) (i) MS-MS spectrum
P1 TLPVPTVC (**19**) chromatogram
P2 TLPVPTVC (**19**) spectrum
P3 TLPVPTVC (**19**) MS-MS spectrum
Q1 TLPVPTVC (**20**) (i) chromatogram
Q2 TLPVPTVC (**20**) (i) spectrum
Q3 TLPVPTVC (**20**) (i) MS-MS spectrum
R1 TVPVPSFC (**21**) (i) chromatogram
R2 TVPVPSFC (**21**) (i) spectrum
R3 TVPVPSFC (**21**) (i) MS-MS spectrum
S1 VTACITFC (**22**) chromatogram
S2 VTACITFC (**22**) spectrum
S3 VTACITFC (**22**) MS-MS spectrum
T1 VTACITFC (**22**) (i) chromatogram
T2 VTACITFC (**22**) (i) spectrum
T3 VTACITFC (**22**) (i) MS-MS spectrum

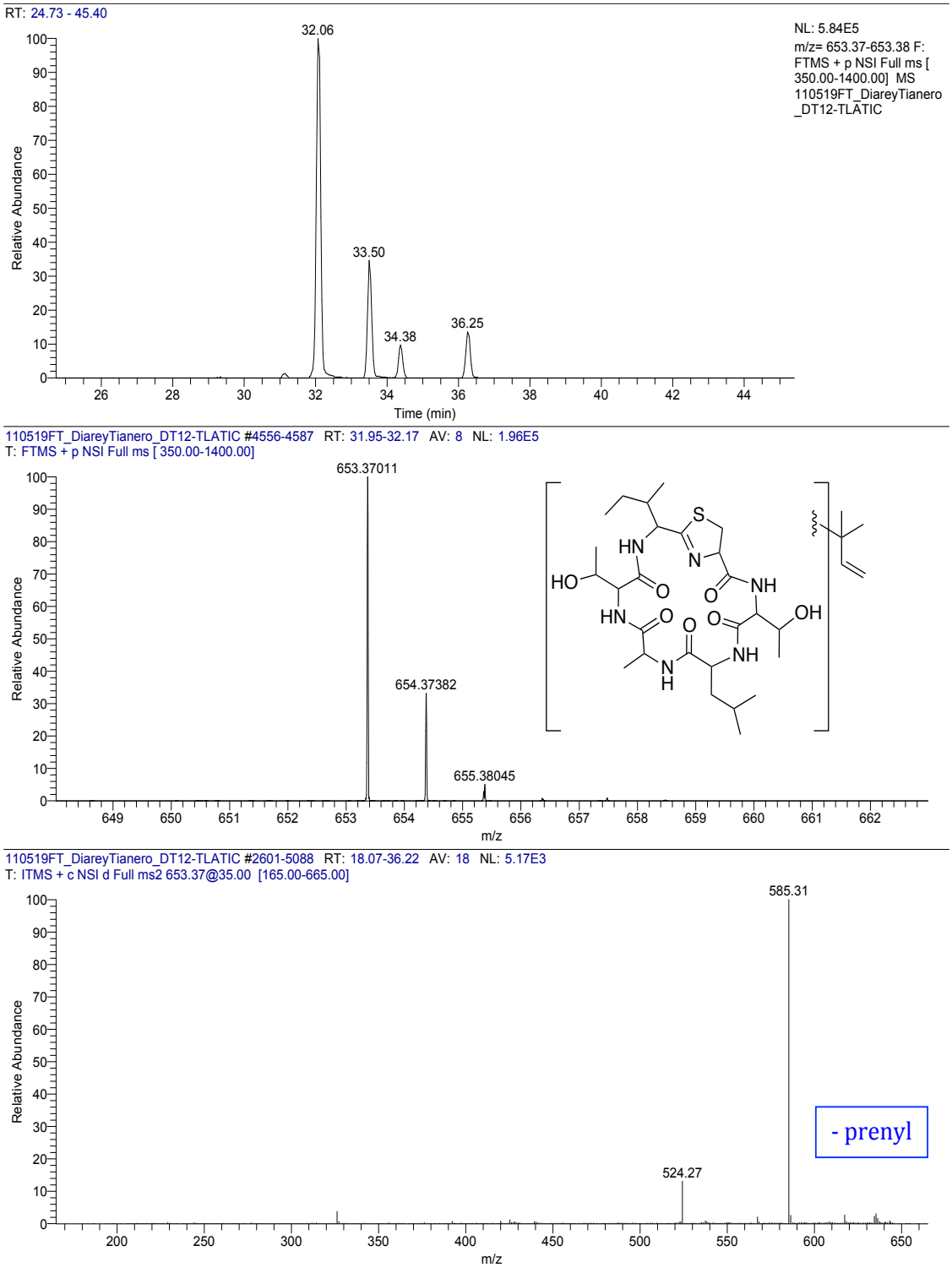
S4. A1. LC-FT-ICR selected ion chromatogram for TLATIC (**4**)
A2. LC-FT-ICR spectrum of TLATIC (**4**)
A3. MS-MS spectrum of TLATIC (**4**)



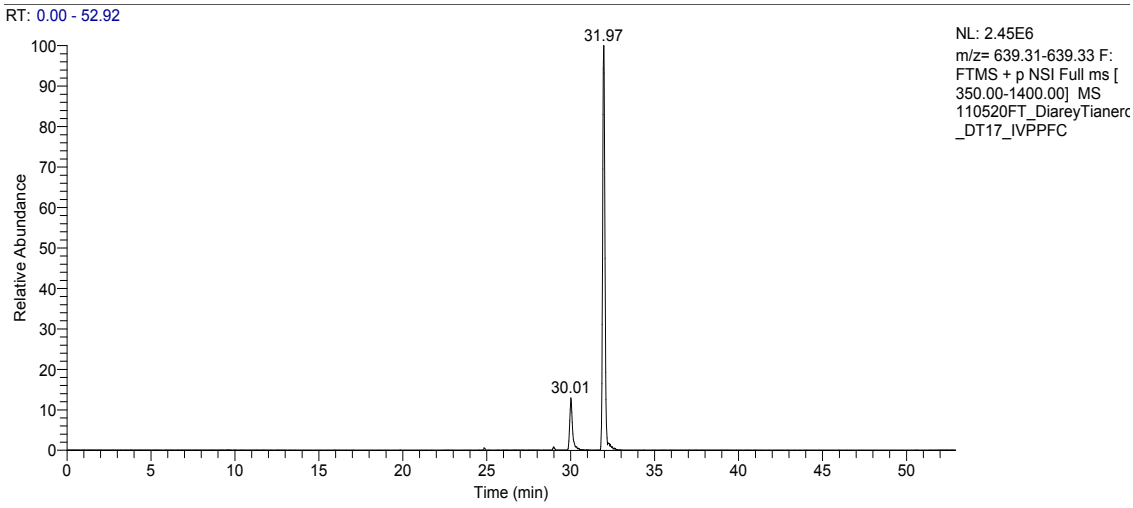


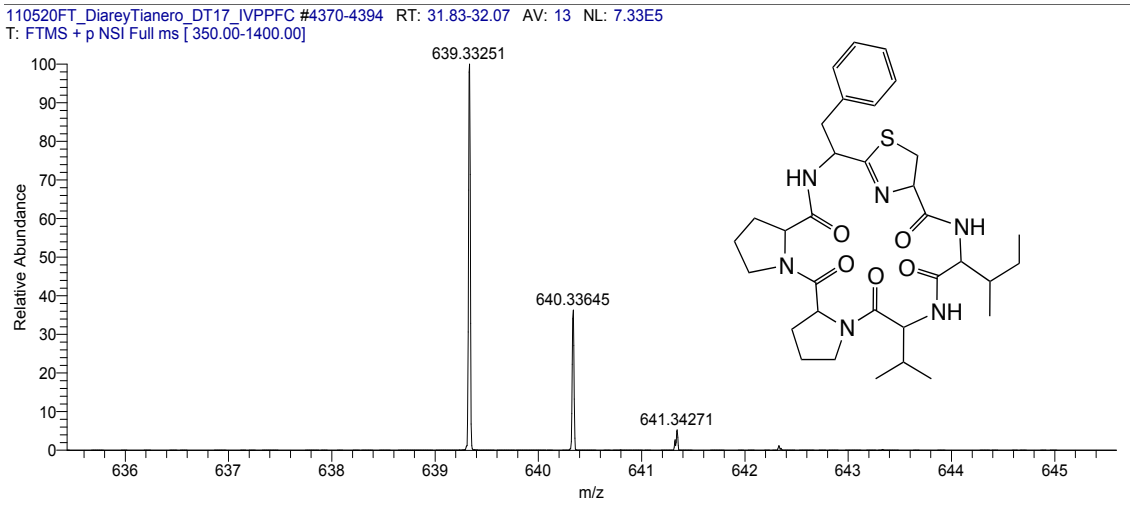


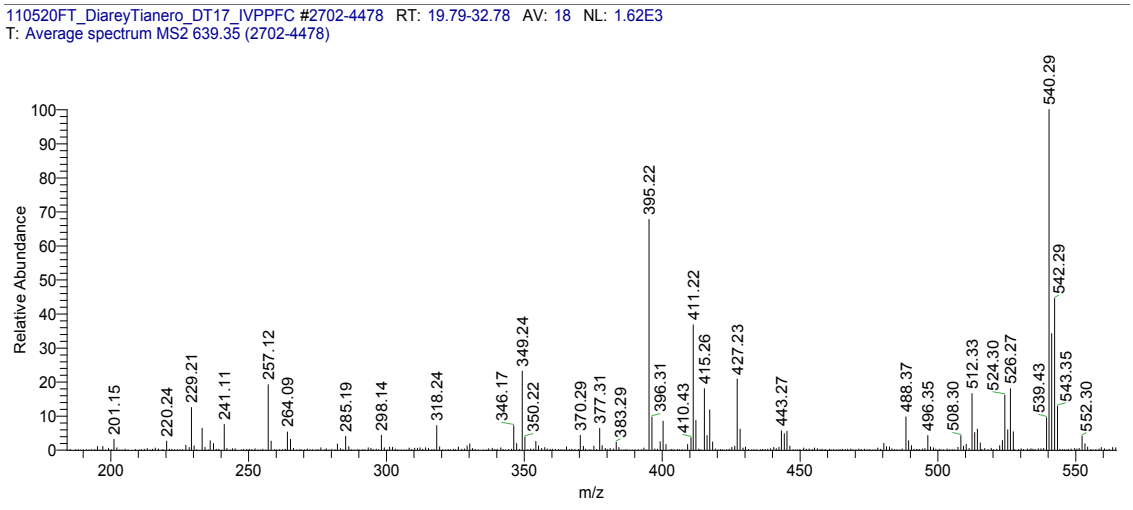
S4. B1. LC-FT-ICR selected ion chromatogram for TLATIC (**4**) **(i)**
B2. LC-FT-ICR spectrum of TLATIC (**4**) **(i)**
B3. MS-MS spectrum of TLATIC (**4**) **(i)**



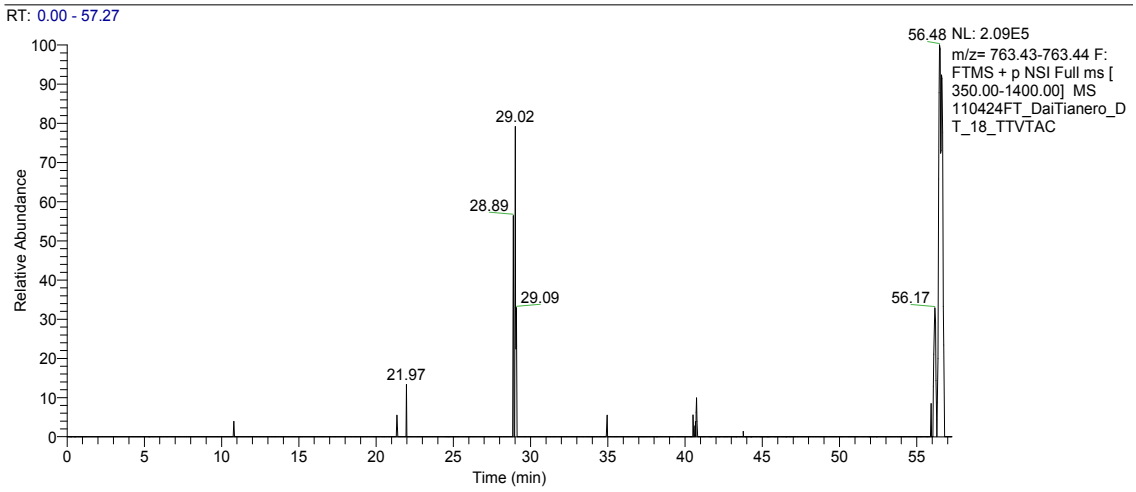
S4. C1. LC-FT-ICR selected ion chromatogram for IVPPFC (**5**)
C2. LC-FT-ICR spectrum of IVPPFC (**5**)
C3. MS-MS spectrum of IVPPFC (**5**)

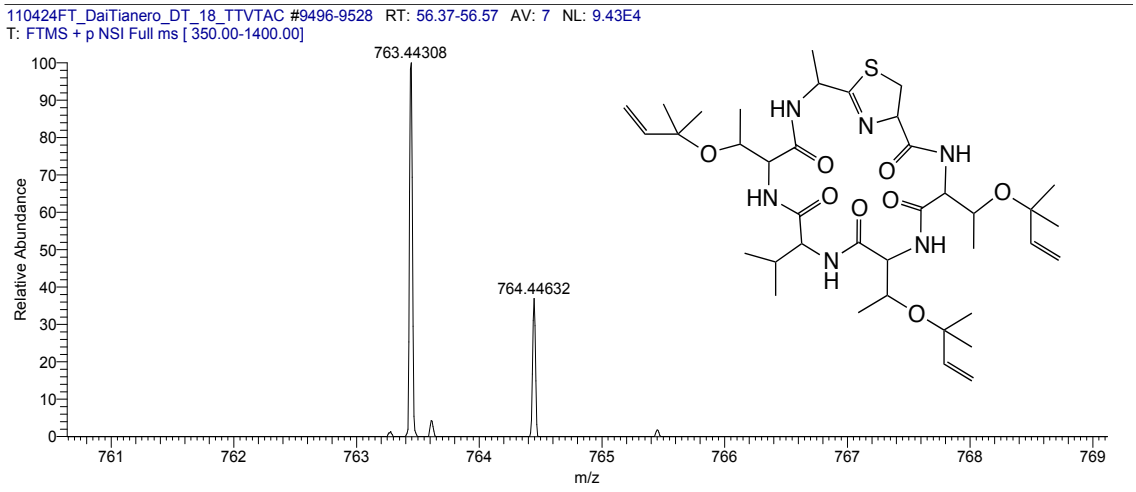


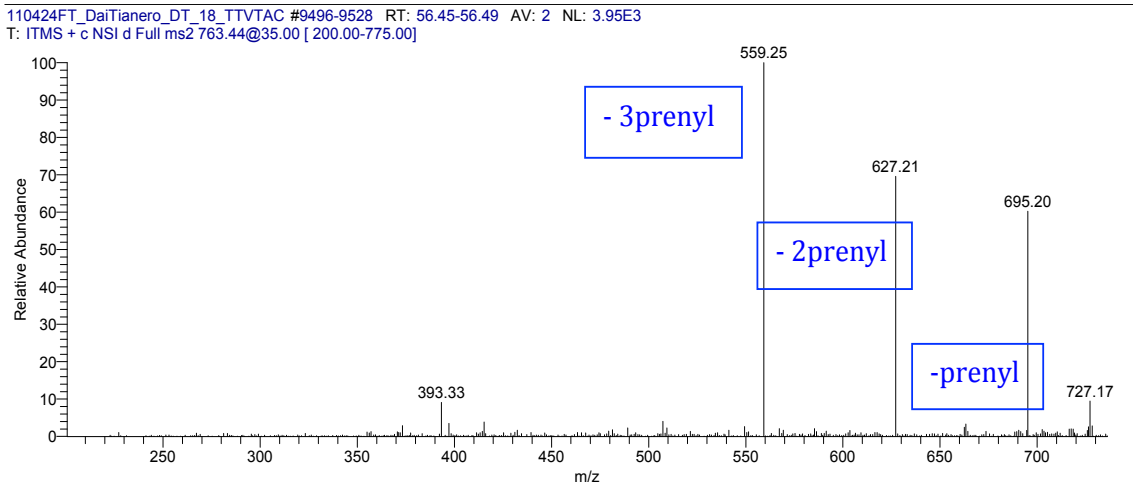




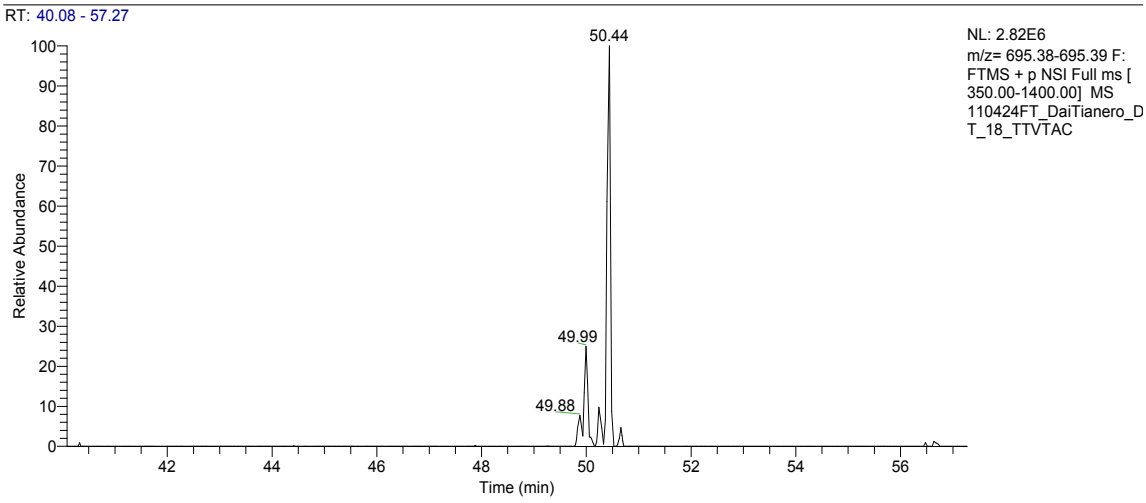
S4. D1. LC-FT-ICR selected ion chromatogram for TTVTAC (**7**)
D2. LC-FT-ICR spectrum of TTVTAC (**7**)
D3. MS-MS spectrum of TTVTAC (**7**)

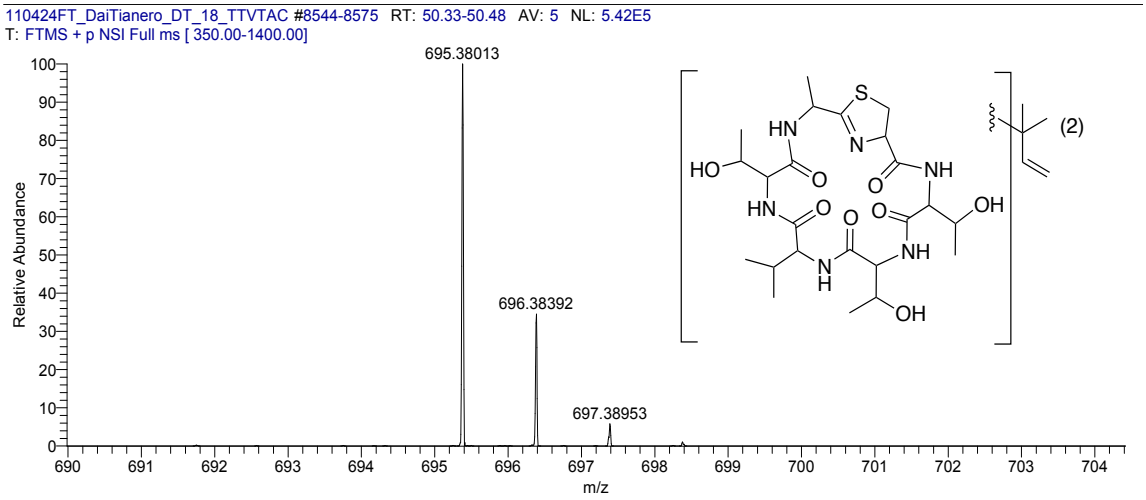


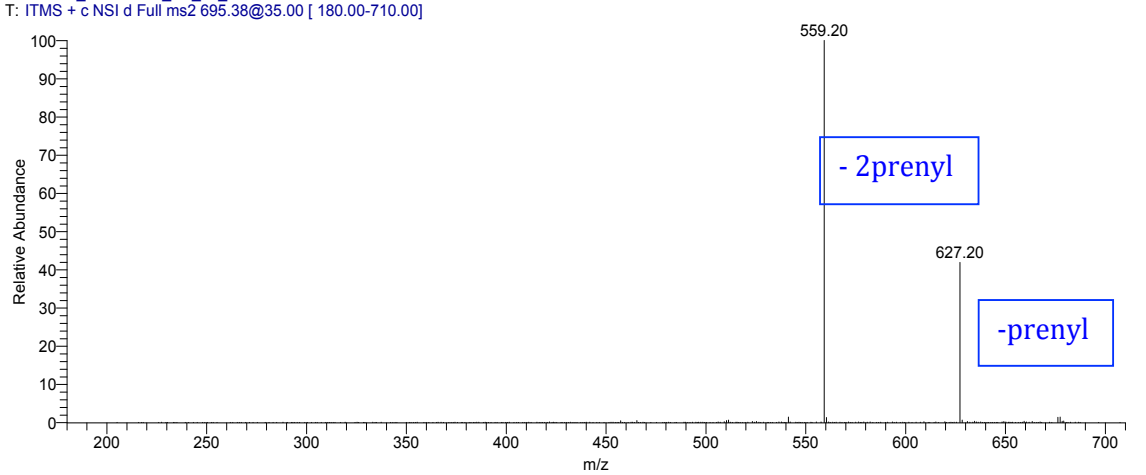




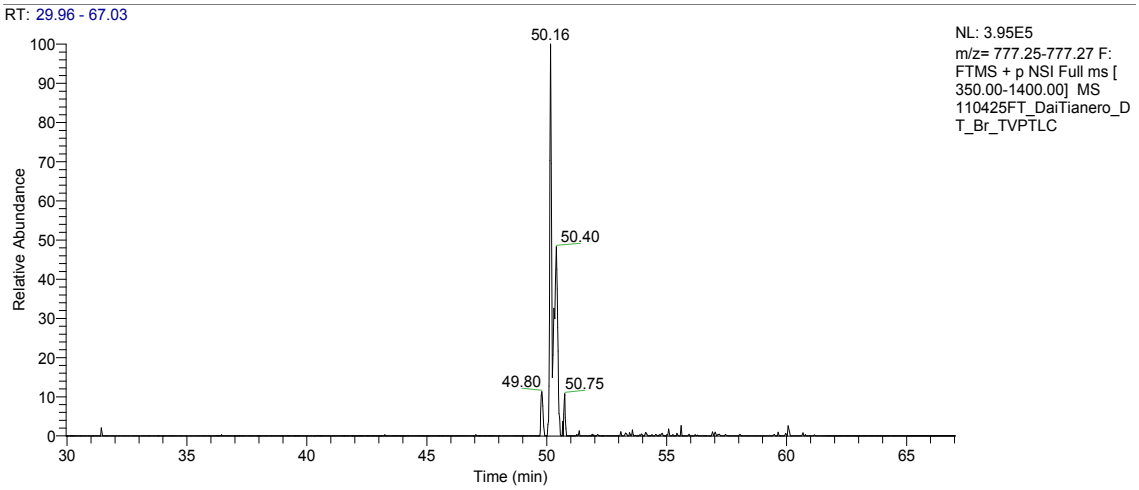
S4. E1. LC-FT-ICR selected ion chromatogram for TTVTAC (**7**) (i)
E2. LC-FT-ICR spectrum of TTVTAC (**7**) (i)
E3. MS-MS spectrum of TTVTAC (**7**) (i)

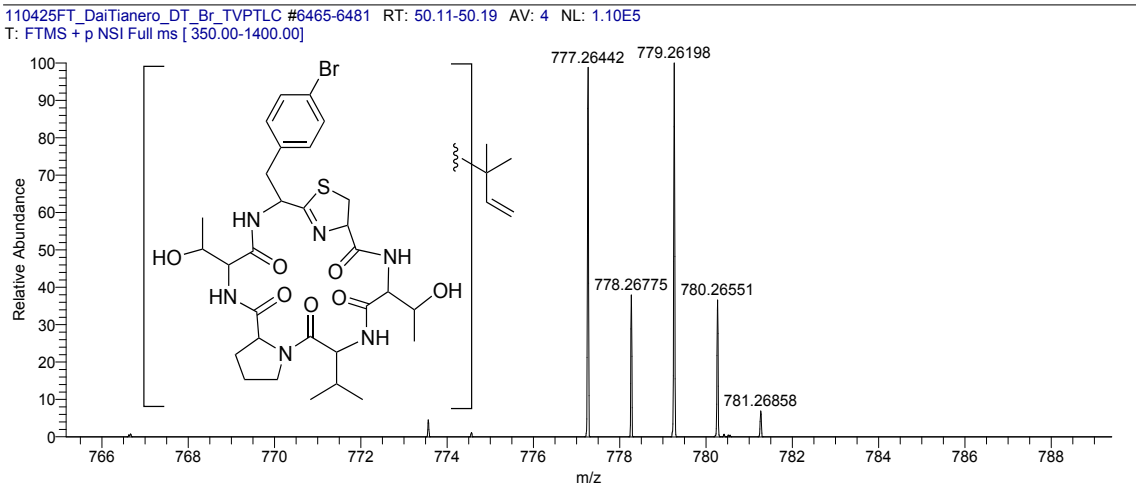


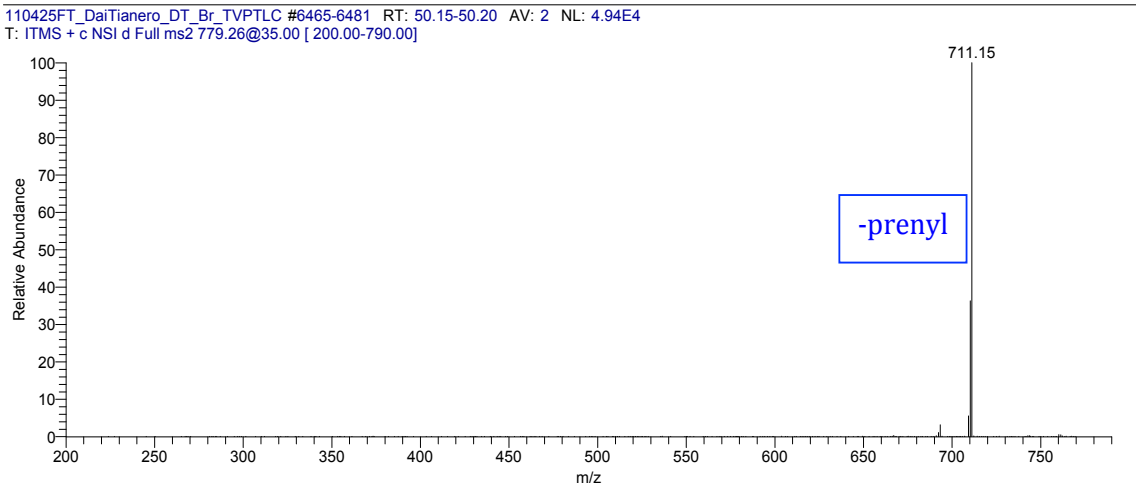




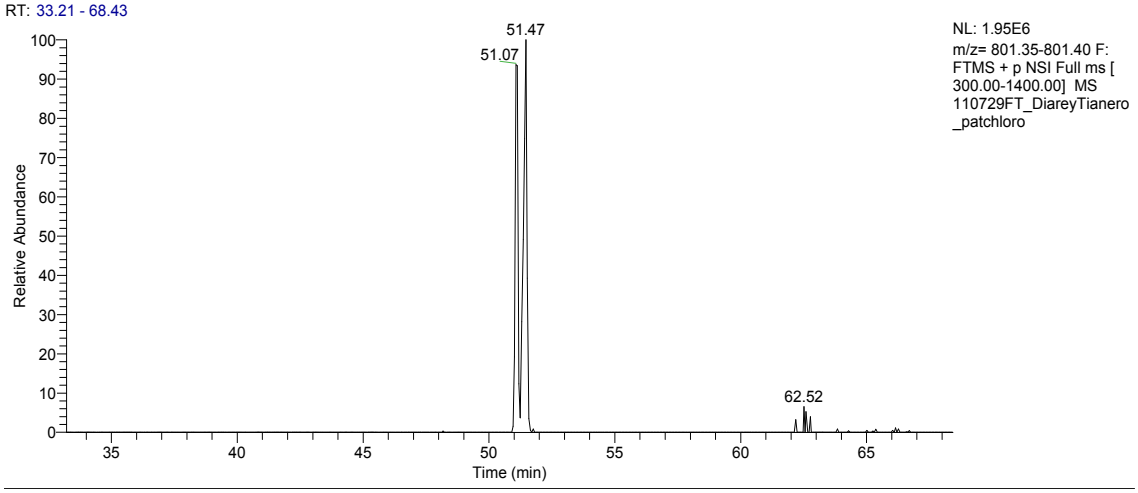
S4. F1. LC-FT-ICR selected ion chromatogram for TVPT-pBrF-C (**9**) (i)
F2. LC-FT-ICR spectrum of TVPT-pBrF-C (**9**) (i)
F3. MS-MS spectrum of TVPT-pBrF-C (**9**) (i)

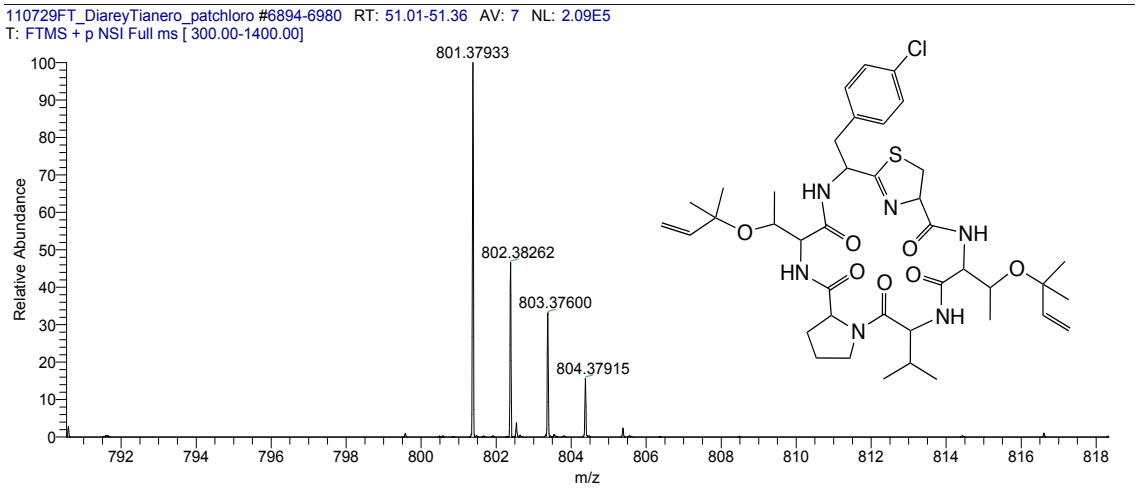


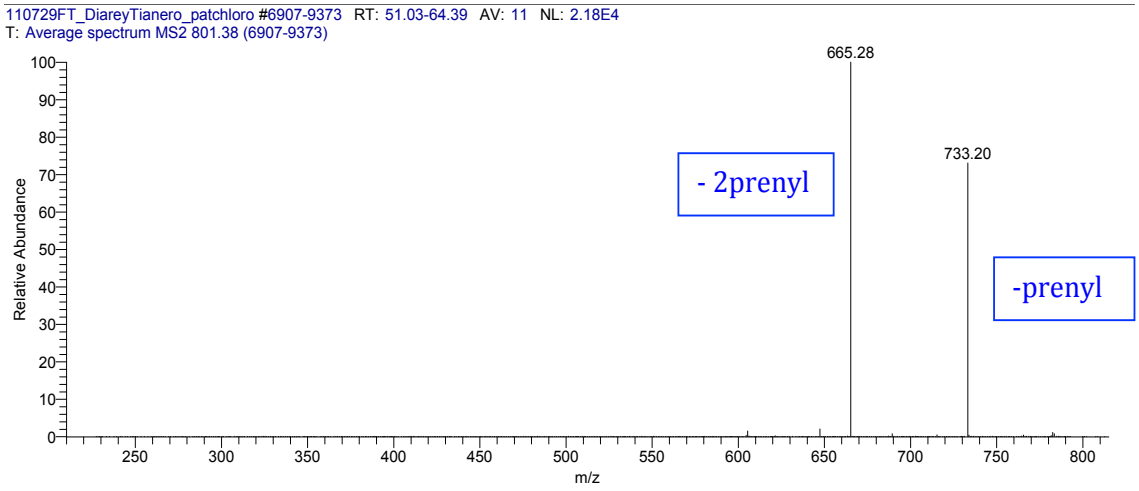




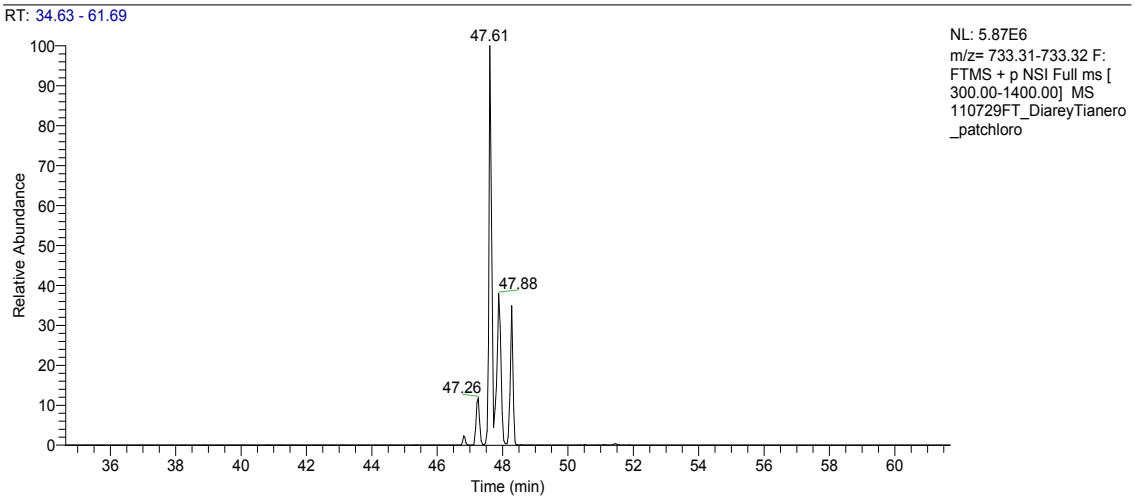
S4. G1. LC-FT-ICR selected ion chromatogram for TVPT-pClF-C (**10**)
G2. LC-FT-ICR spectrum of TVPT-pClF-C (**10**)
G3. MS-MS spectrum of TVPT-pClF-C (**10**)

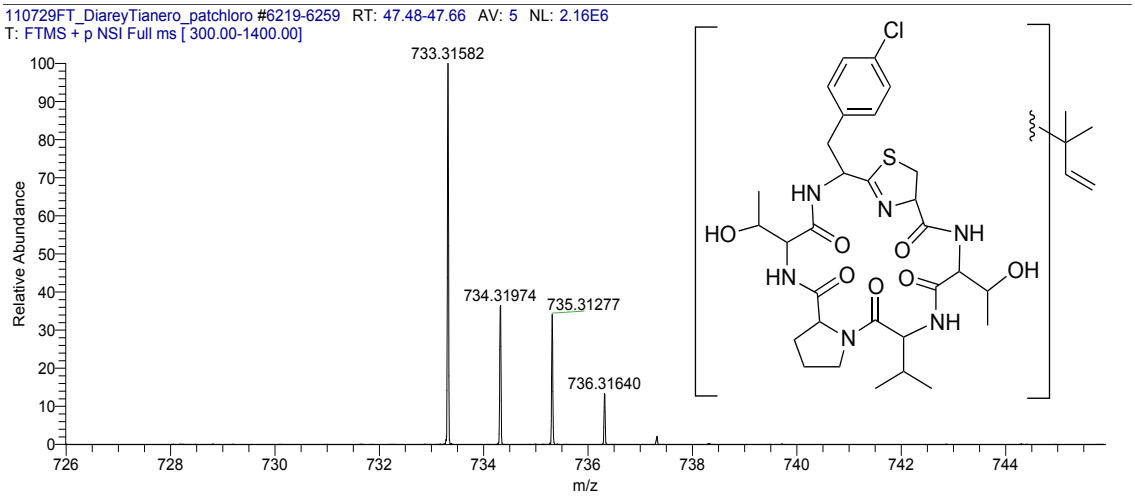


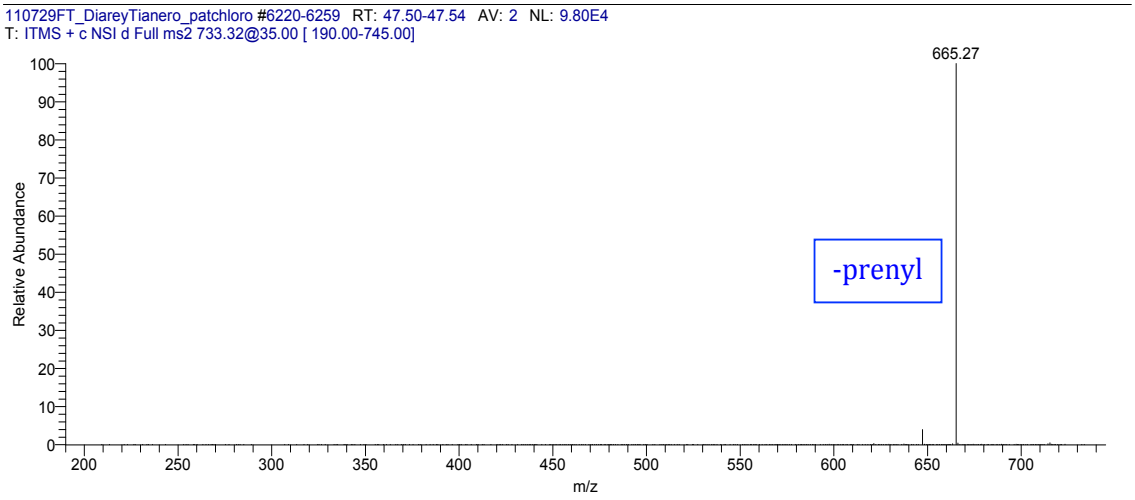




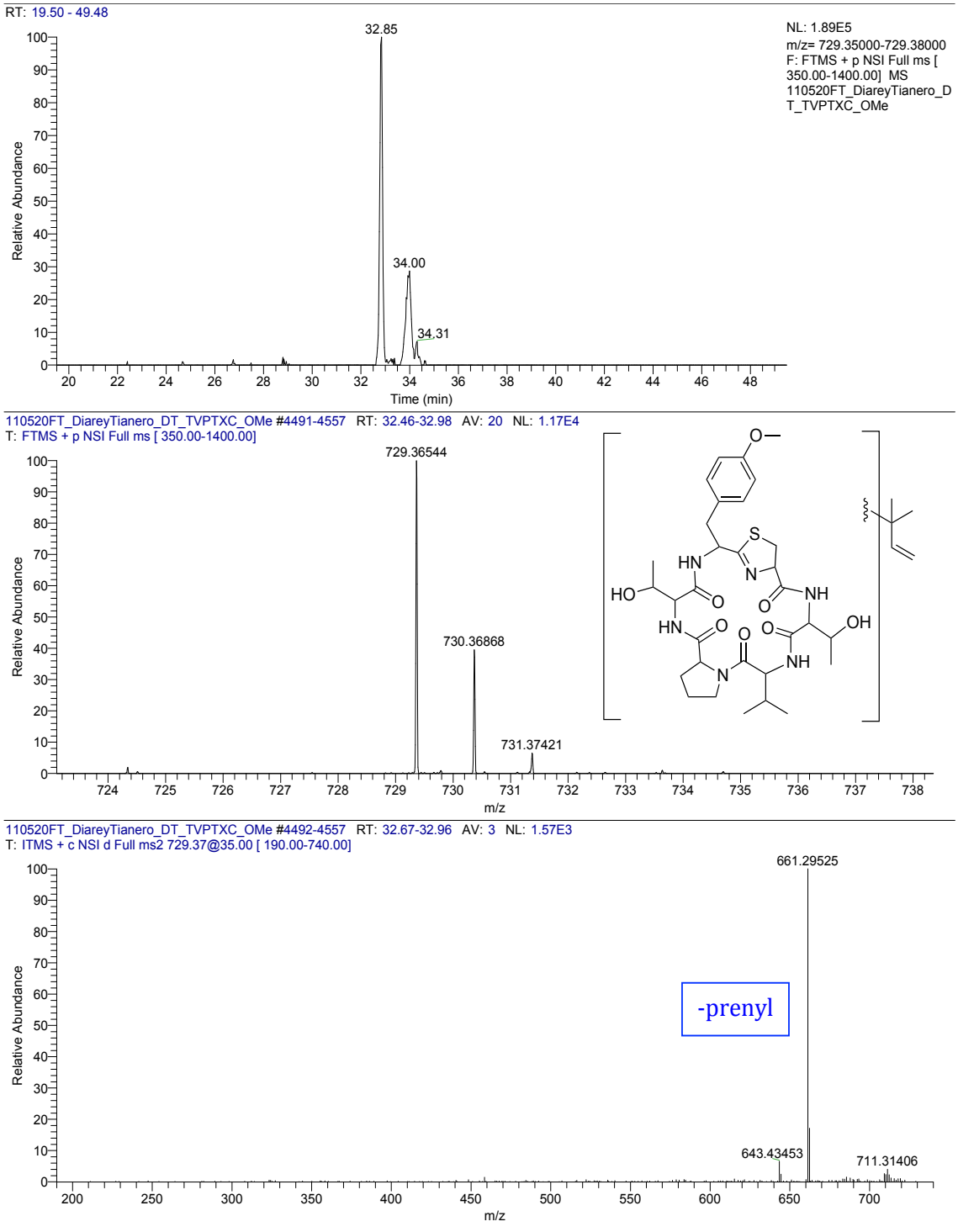
S4. H1. LC-FT-ICR selected ion chromatogram for TVPT-pClF-C (**10**) (i)
H2. LC-FT-ICR spectrum of TVPT-pClF-C (**10**) (i)
H3. MS-MS spectrum of TVPT-pClF-C (**10**) (i)



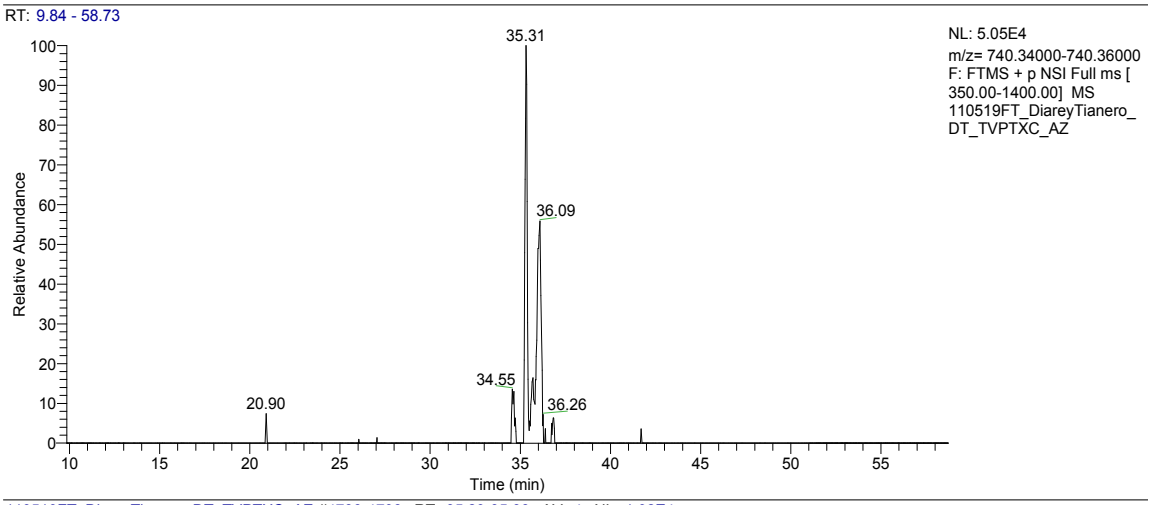


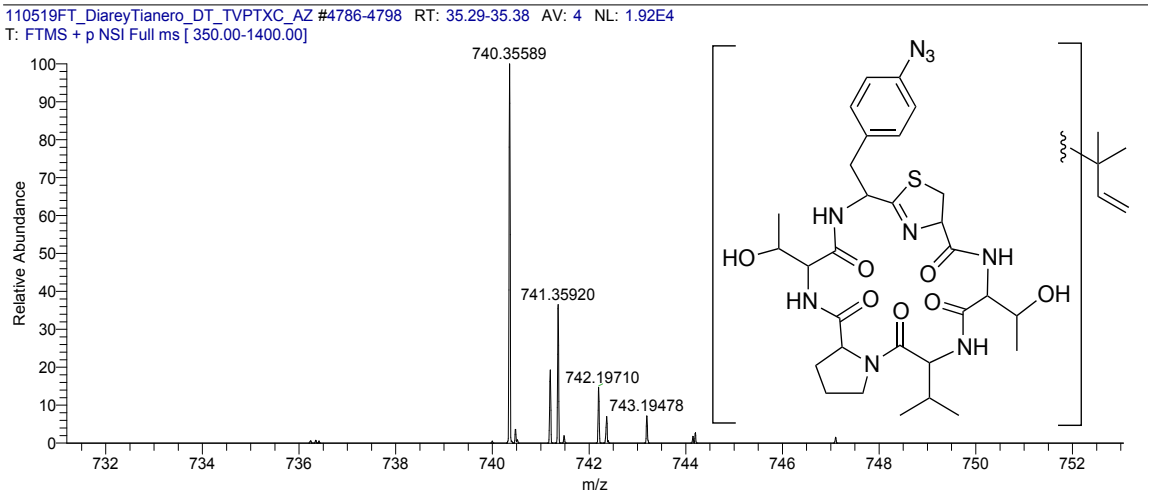


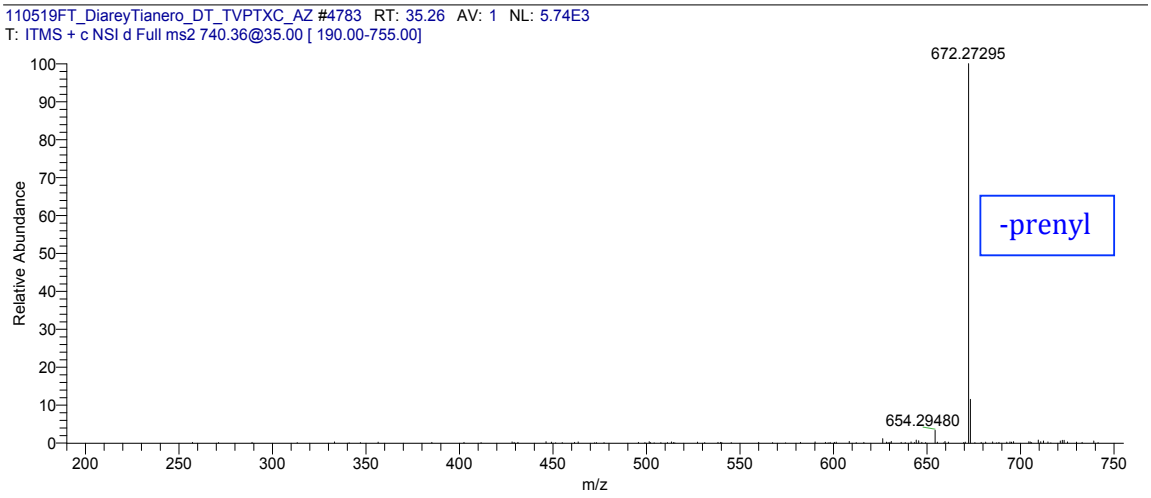
S4. I1. LC-FT-ICR selected ion chromatogram for TVPT-pOMeF-C (**11**) (i)
I2. LC-FT-ICR spectrum of TVPT-pOMeF-C (**11**) (i)
I3. MS-MS spectrum of TVPT-pOMeF-C (**11**) (i)



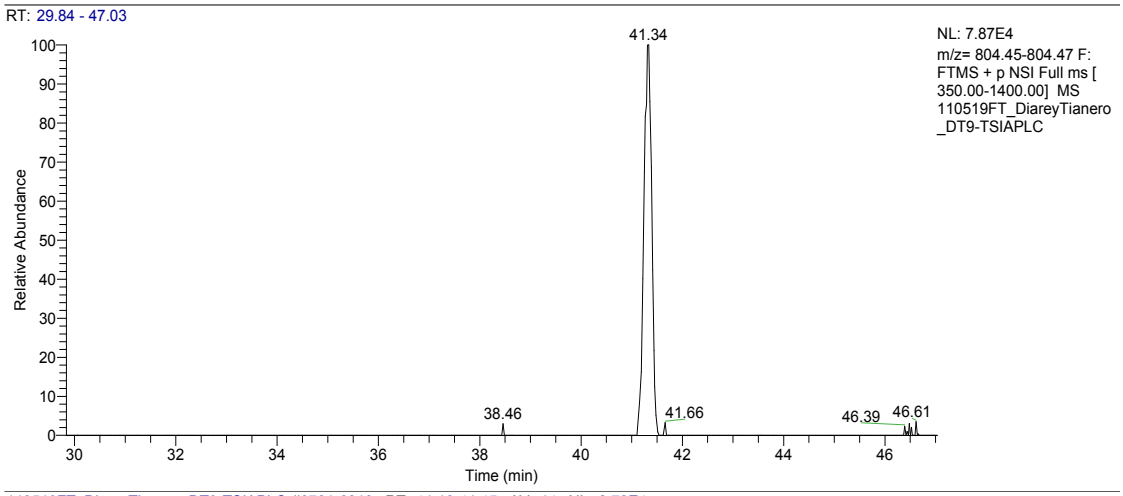
S4. J1. LC-FT-ICR selected ion chromatogram for TVPT-pN_3F-C (**12**) (i)
J2. LC-FT-ICR spectrum of TVPT-pN_3F-C (**12**) (i)
J3. MS-MS spectrum of TVPT-pN_3F-C (**12**) (i)

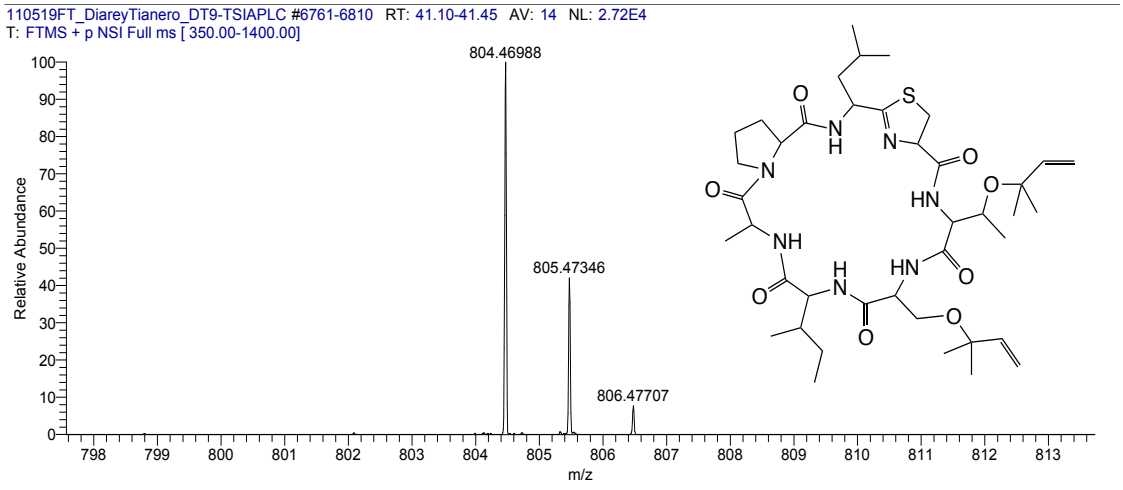


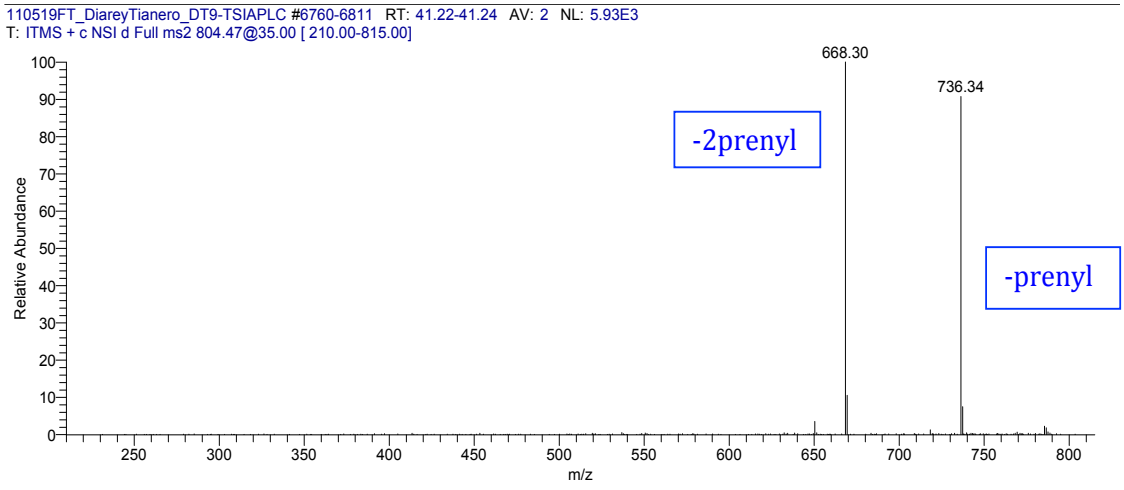




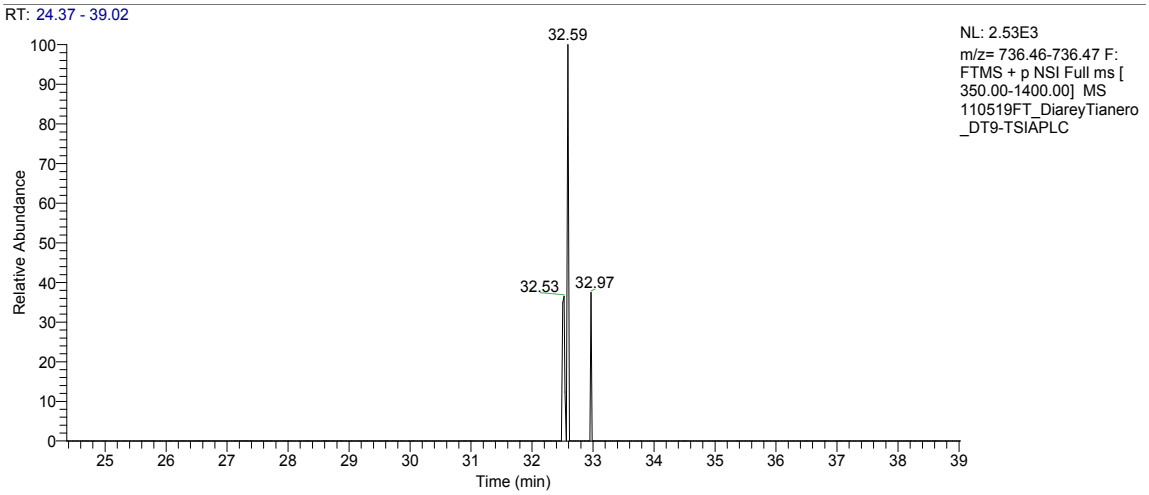
S4. K1. LC-FT-ICR selected ion chromatogram for TSIAPLC (**13**)
K2. LC-FT-ICR spectrum of TSIAPLC (**13**)
K3. MS-MS spectrum of TSIAPLC (**13**)

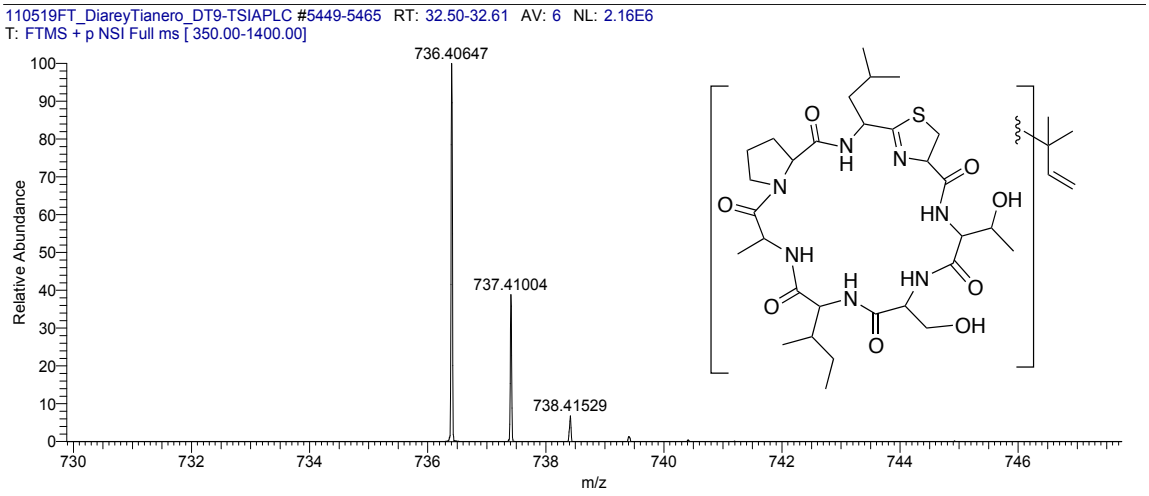


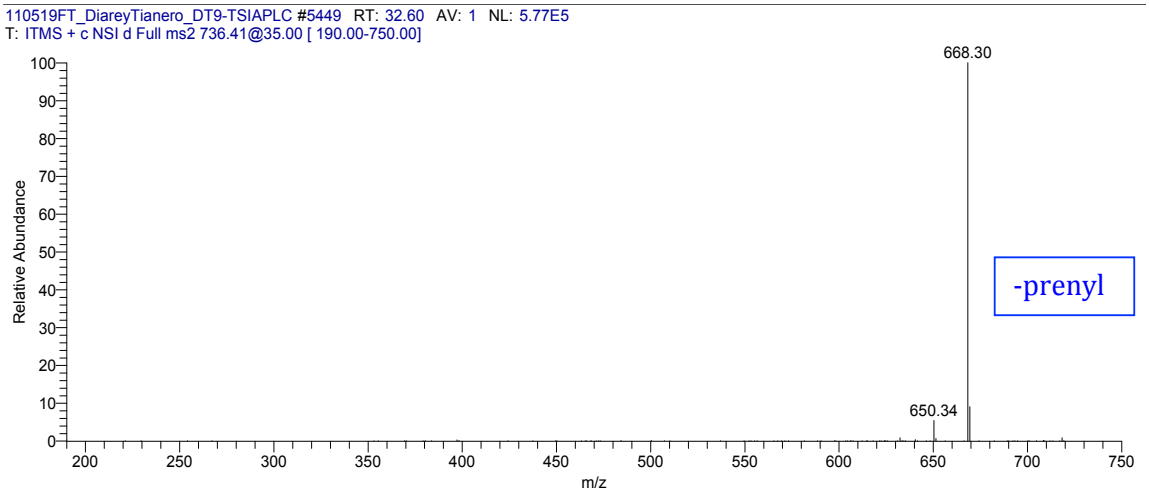




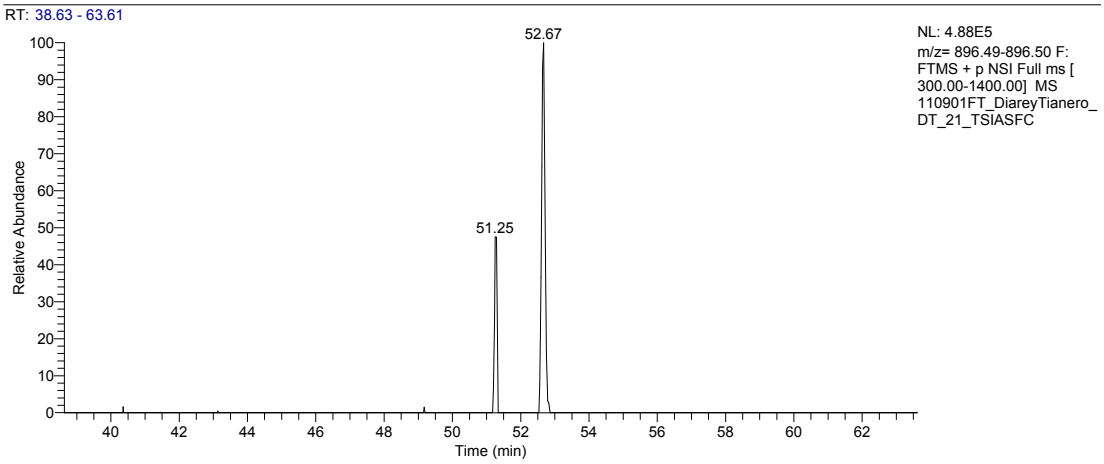
S4. L1. LC-FT-ICR selected ion chromatogram for TSIAPLC (**13**) (i)
L2. LC-FT-ICR spectrum of TSIAPLC (**13**) (i)
L3. MS-MS spectrum of TSIAPLC (**13**) (i)

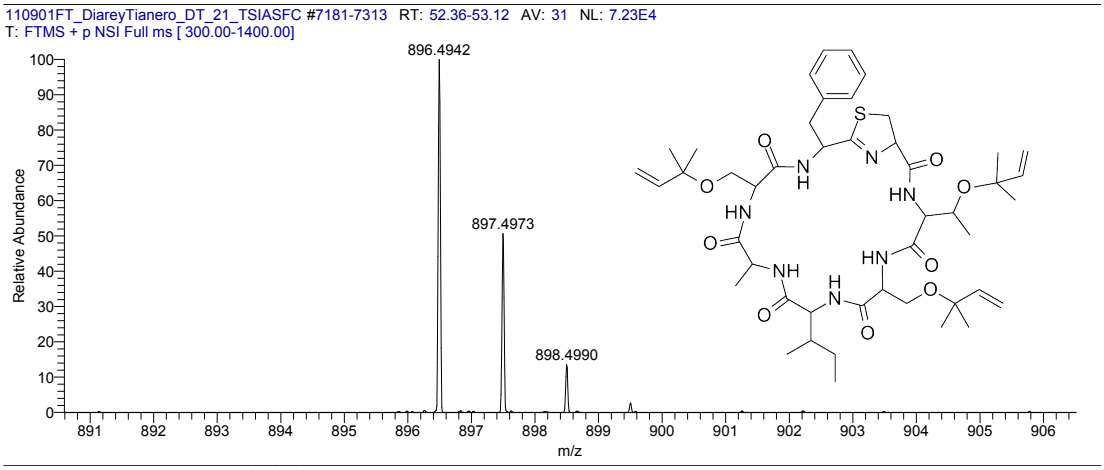


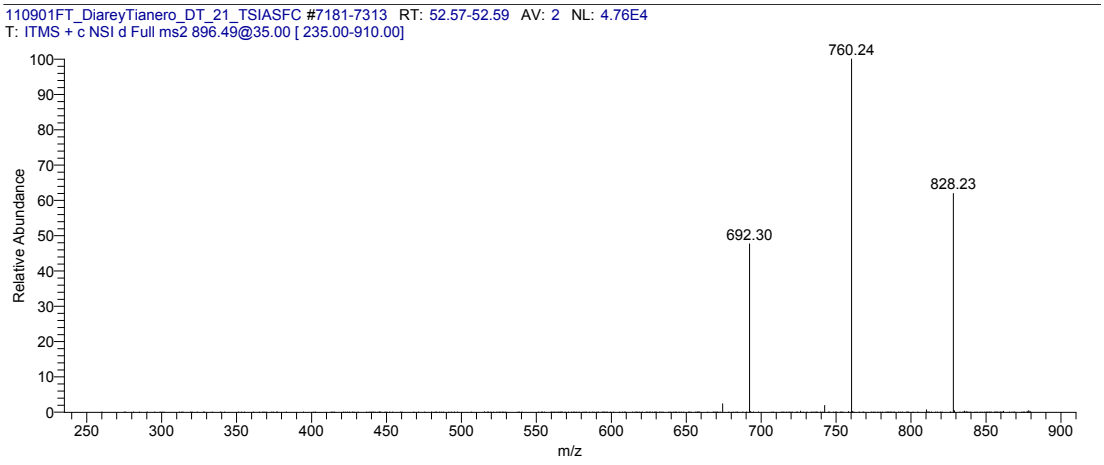




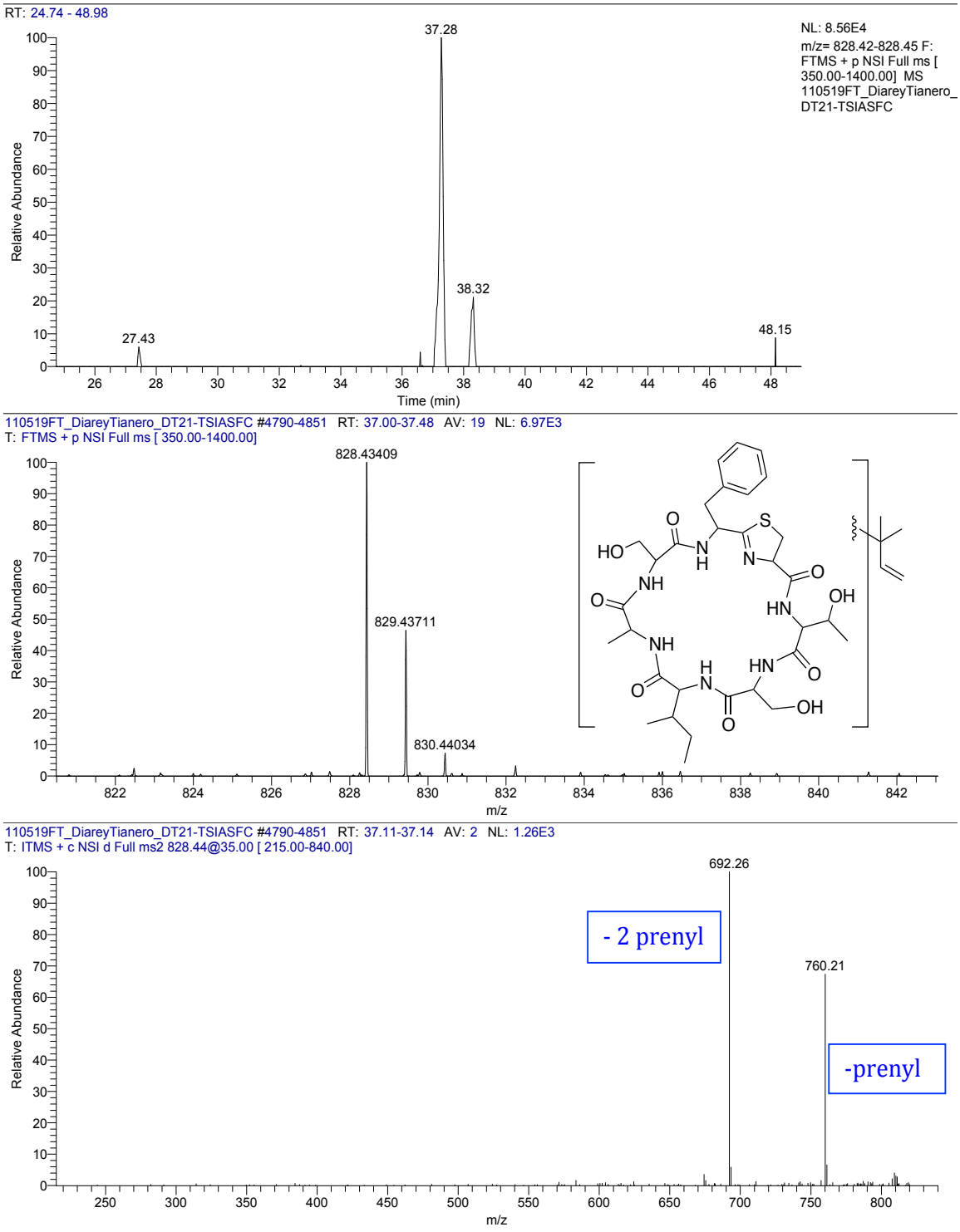
S4. M1. LC-FT-ICR selected ion chromatogram for TSIASFC (**14**)
M2. LC-FT-ICR spectrum of TSIASFC (**14**)
M3. MS-MS spectrum of TSIASFC (**14**)



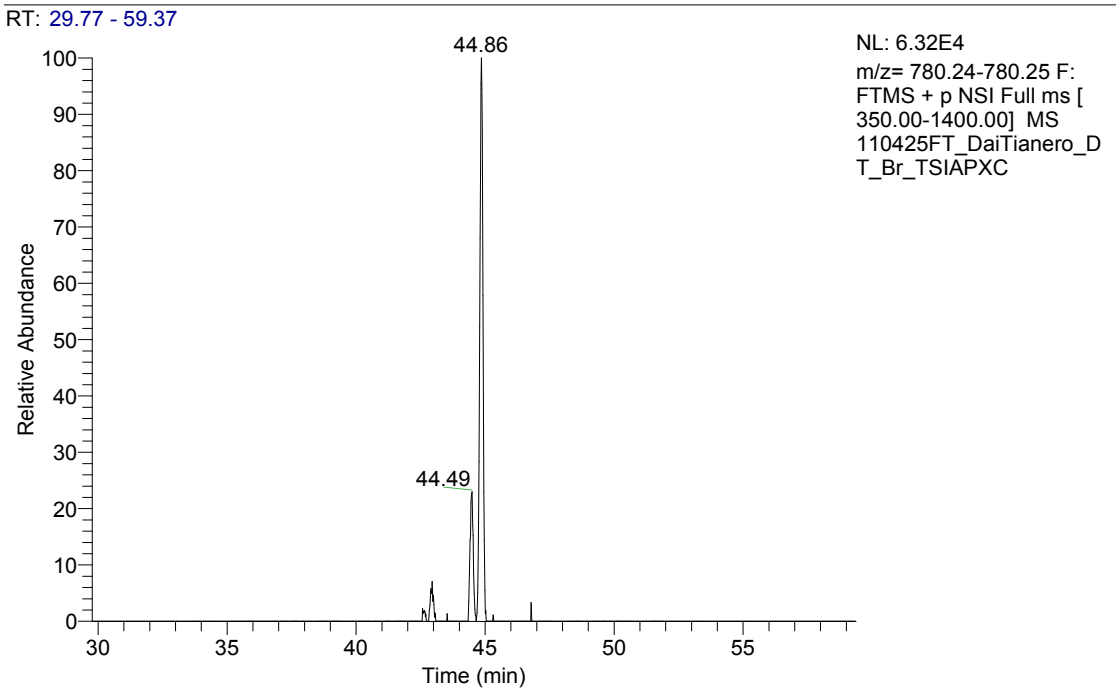


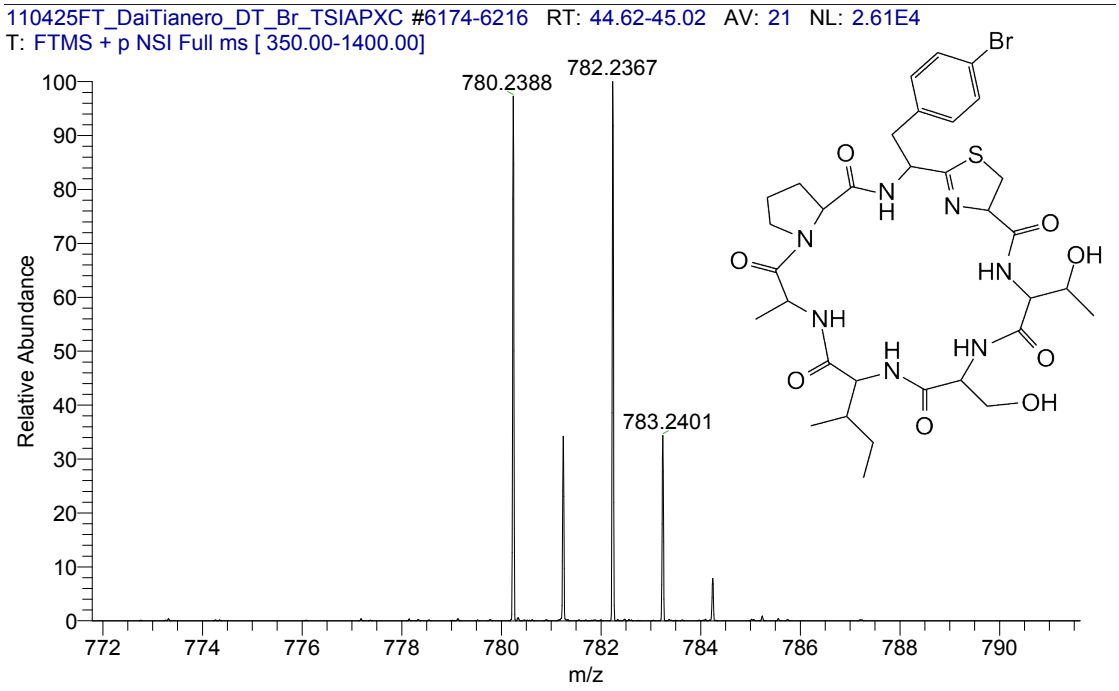


S4. N1. LC-FT-ICR selected ion chromatogram for TSIASFC (**14**) (i)
N2. LC-FT-ICR spectrum of TSIASFC (**14**) (i)
N3. MS-MS spectrum of TSIASFC (**14**) (i)

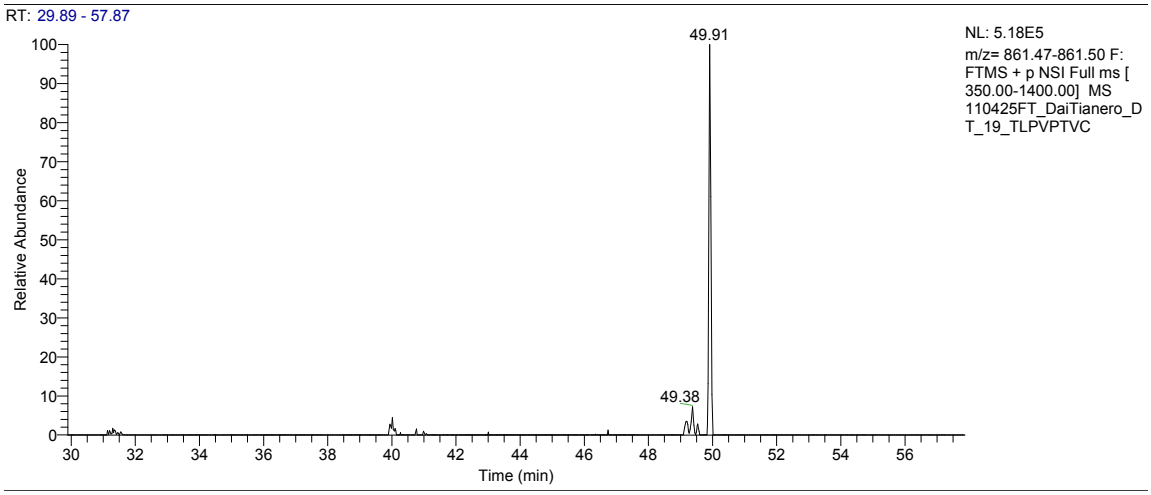


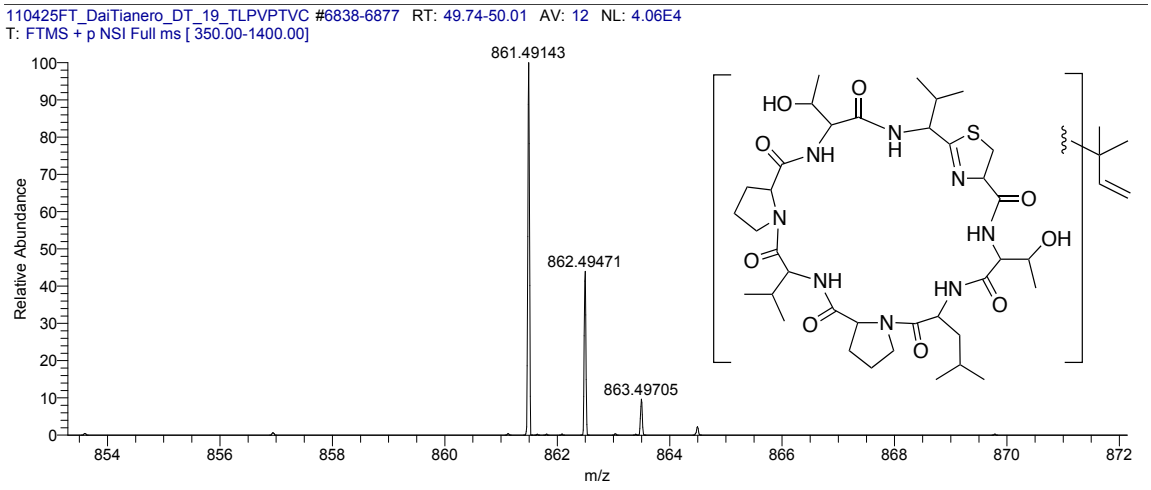
S4. O1. LC-FT-ICR selected ion chromatogram for TSIAP-pBrF-C (**18**) (ii)
O2. LC-FT-ICR spectrum of TSIA-pBrF-C (**18**) (ii)
O3. MS-MS spectrum of TSIA-pBrFC (**18**) (ii)

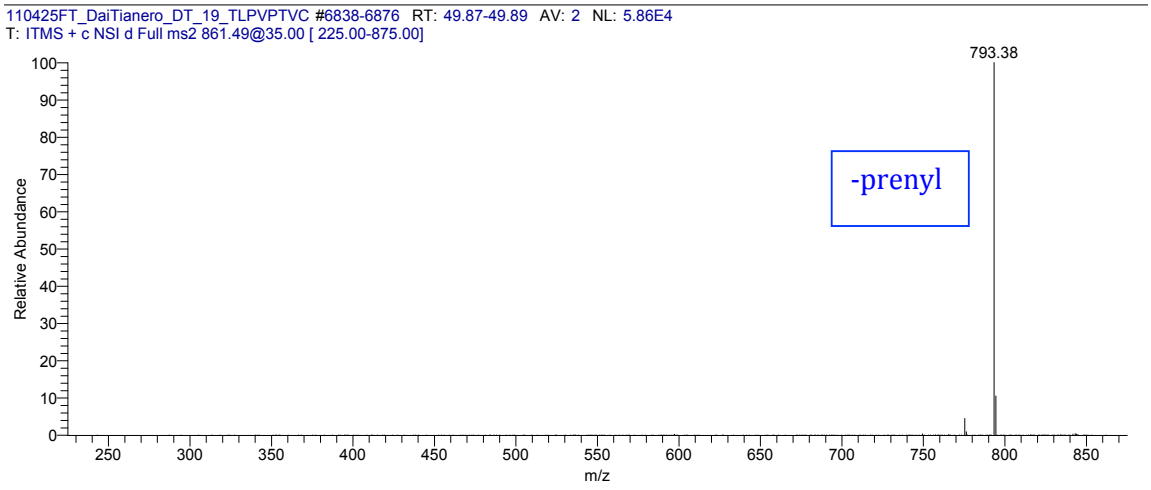




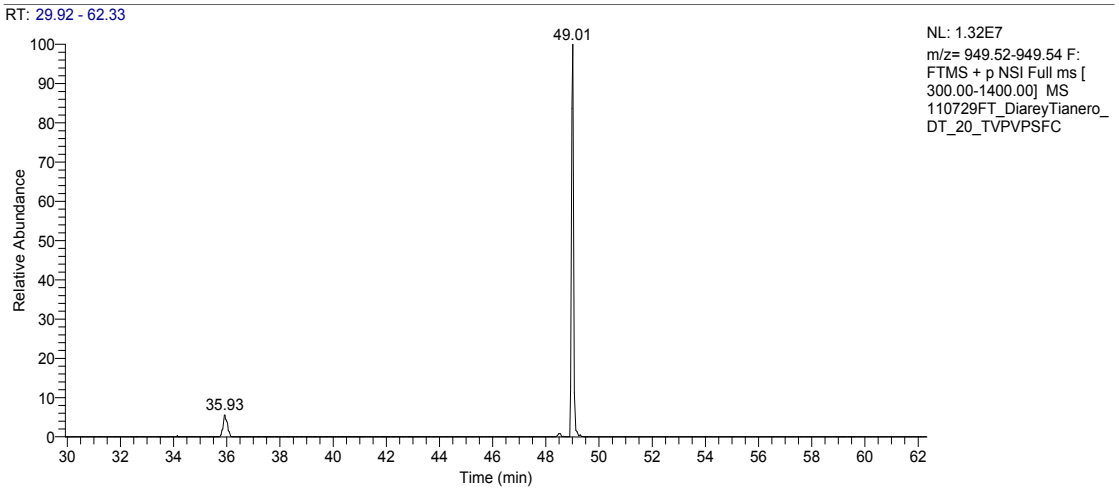
S4. P1. LC-FT-ICR selected ion chromatogram for TLPVPTVC (**19**) (i)
P2. LC-FT-ICR spectrum of TLPVPTVC (**19**) (i)
P3. MS-MS spectrum of TLPVPTVC (**19**) (i)

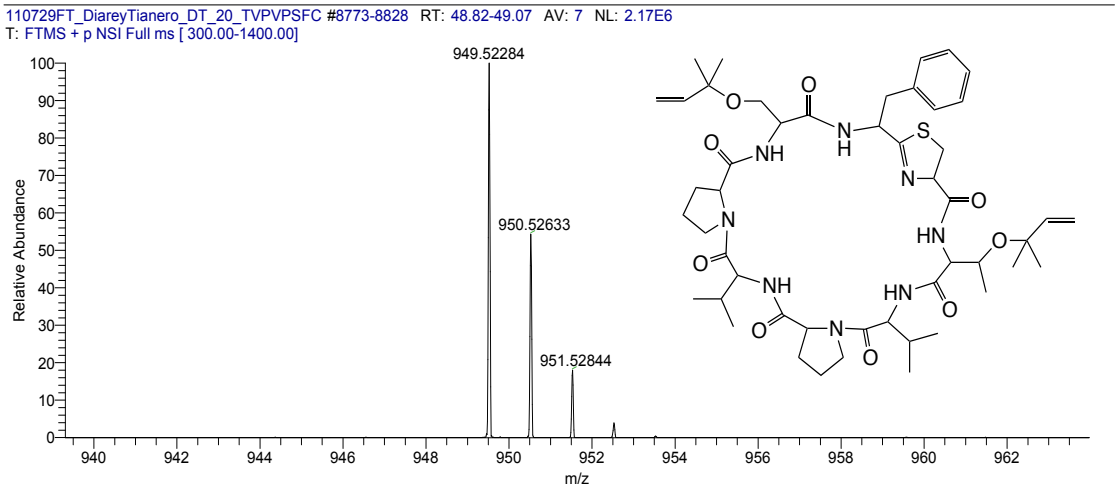


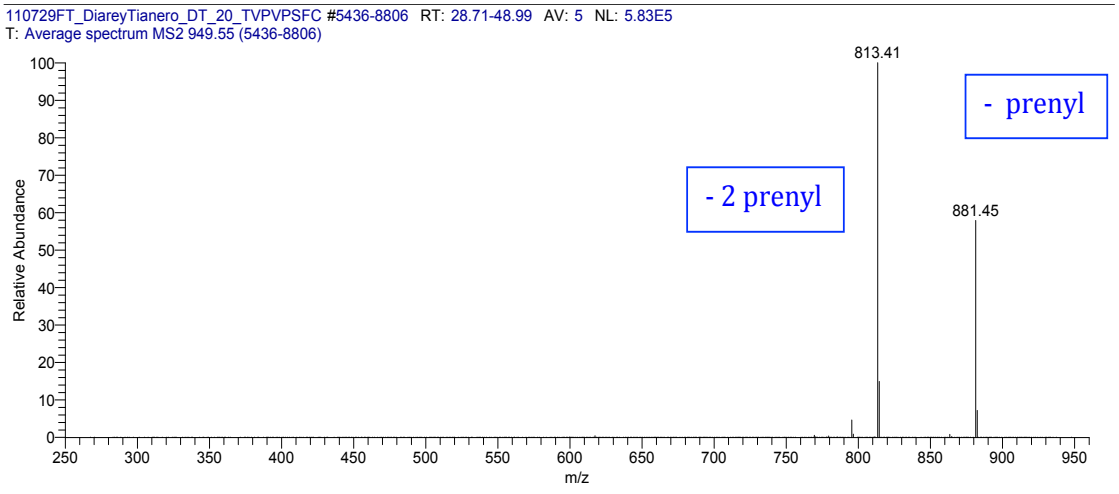




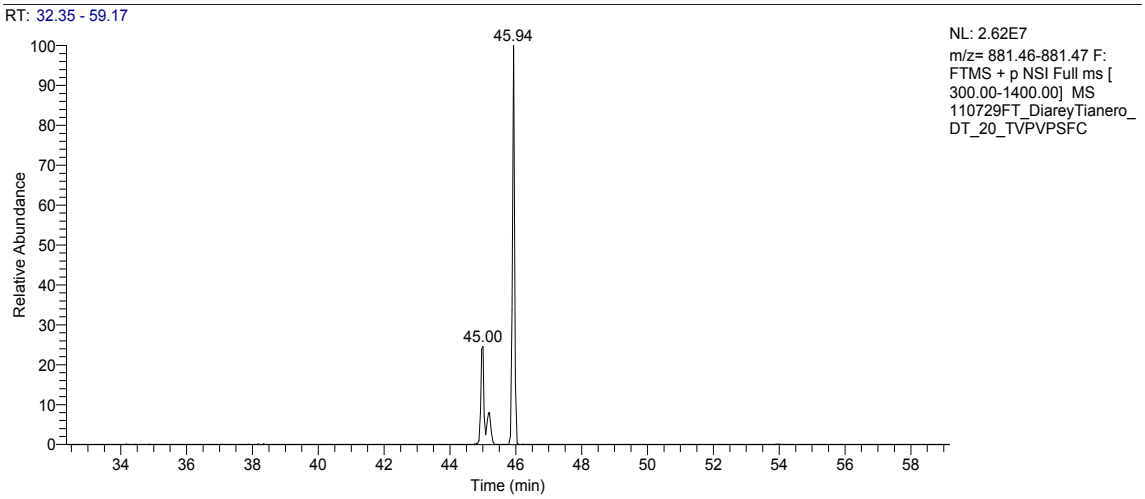
S4. Q1. LC-FT-ICR selected ion chromatogram for TVPVPSFC (**20**)
Q2. LC-FT-ICR spectrum of TVPVPSFC (**20**)
Q3. MS-MS spectrum of TVPVPSFC (**20**)

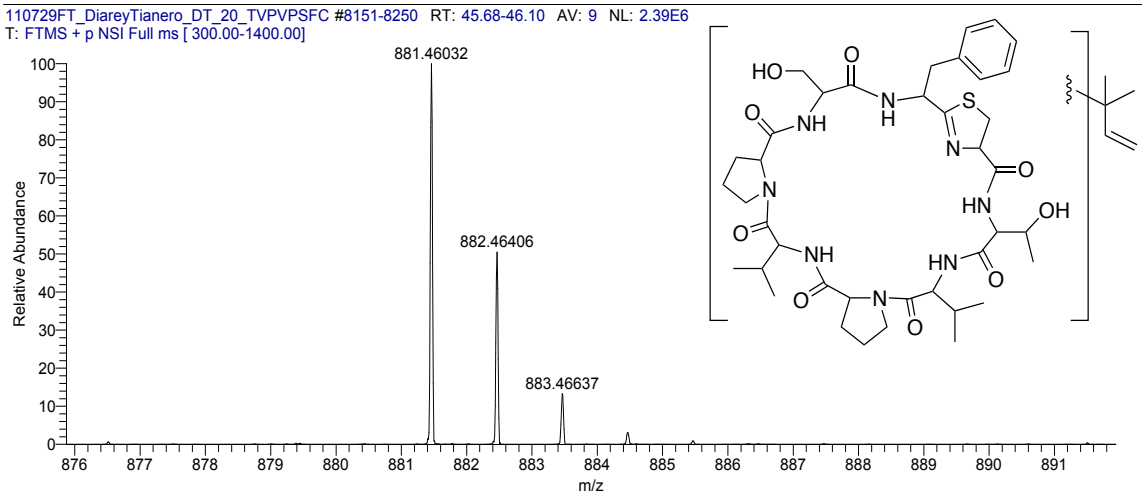


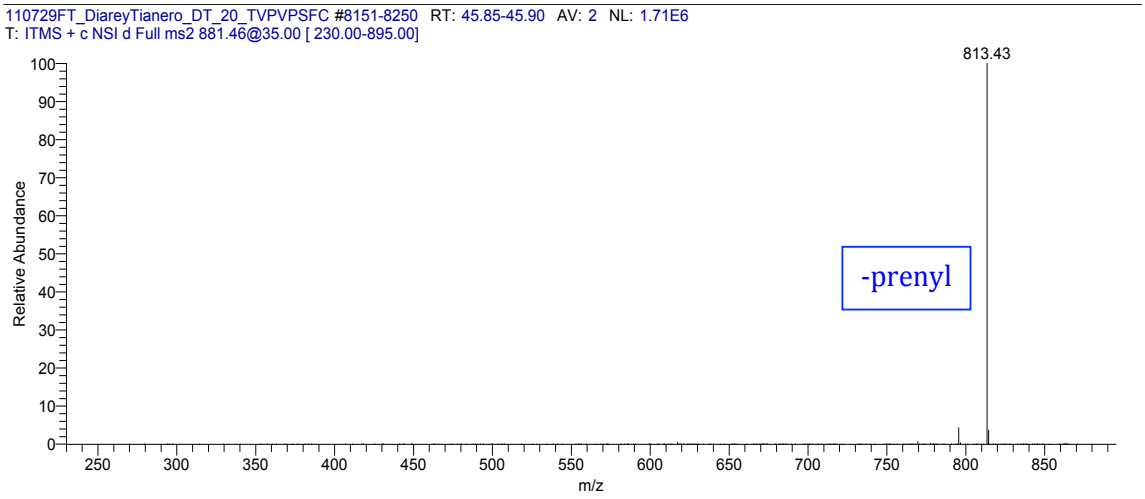




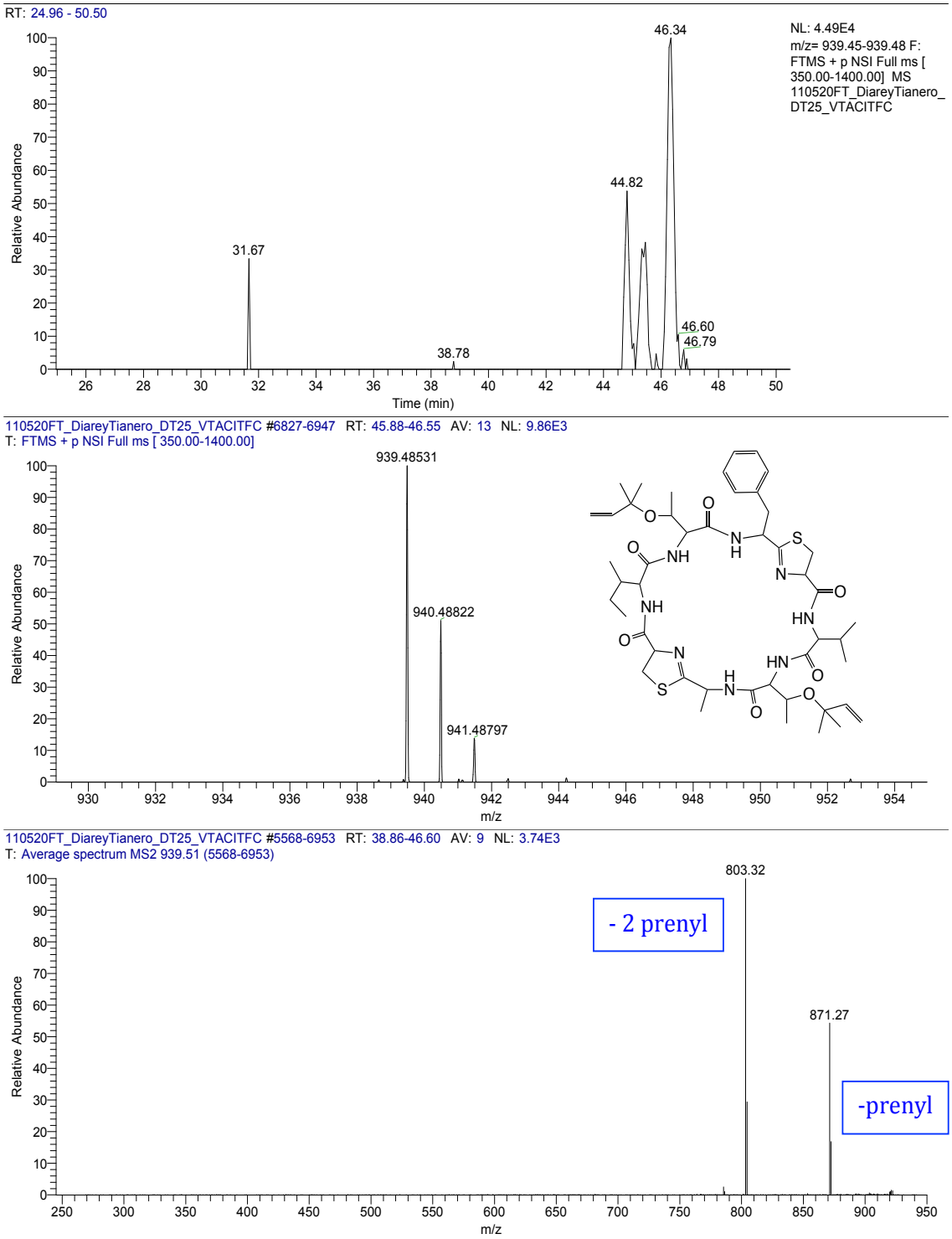
S4. R1. LC-FT-ICR selected ion chromatogram for TVPVPSFC (**20**) (i)
R2. LC-FT-ICR spectrum of TVPVPSFC (**20**) (i)
R3. MS-MS spectrum of TVPVPSFC (**20**) (i)



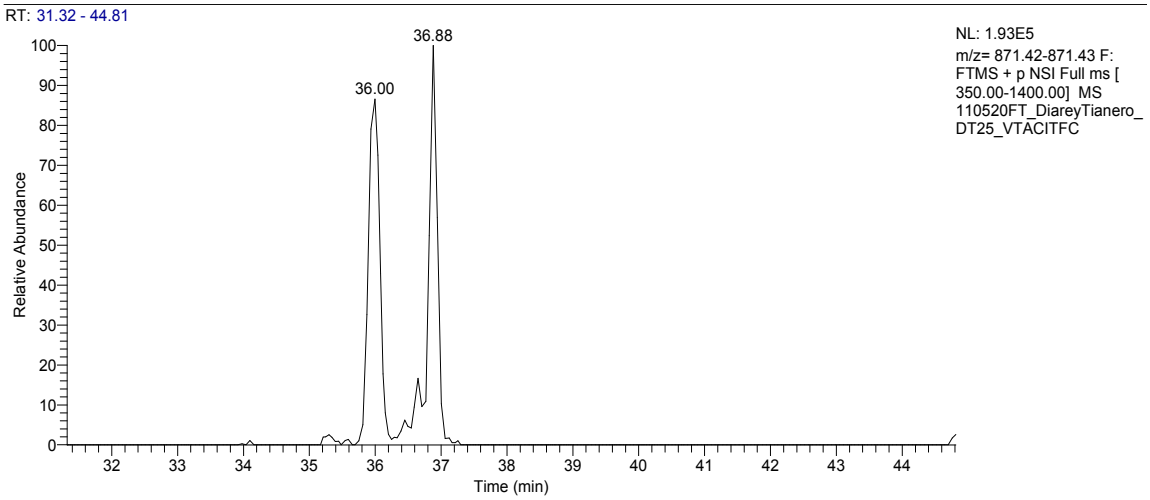


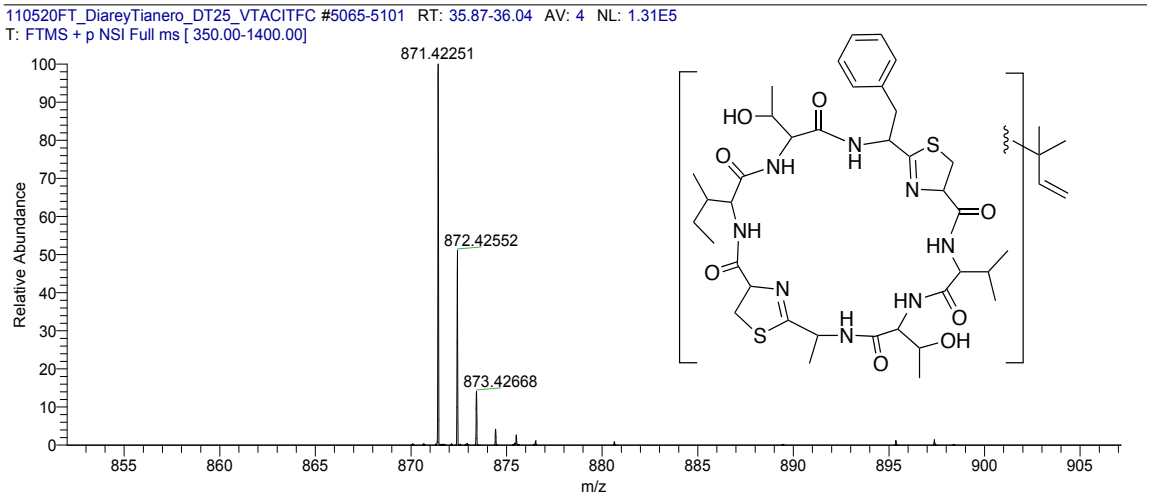


S4. S1. LC-FT-ICR selected ion chromatogram for VTACITFC (**22**)
S2. LC-FT-ICR spectrum of VTACITFC (**22**)
S3. MS-MS spectrum of VTACITFC (**22**)



S4. T1. LC-FT-ICR selected ion chromatogram for VTACITFC (**22**) (i)
T2. LC-FT-ICR spectrum of VTACITFC (**22**) (i)
T3. MS-MS spectrum of VTACITFC (**22**) (i)





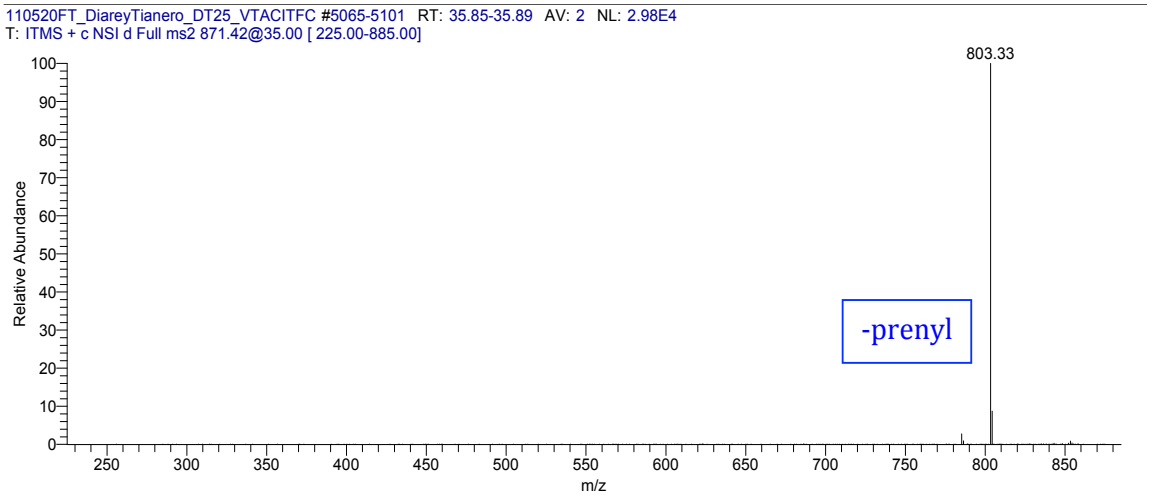


Figure S5.
A1. ^{1}H NMR spectrum of TSIASFC (**15**) in DMSO-d_6 at 600 MHz

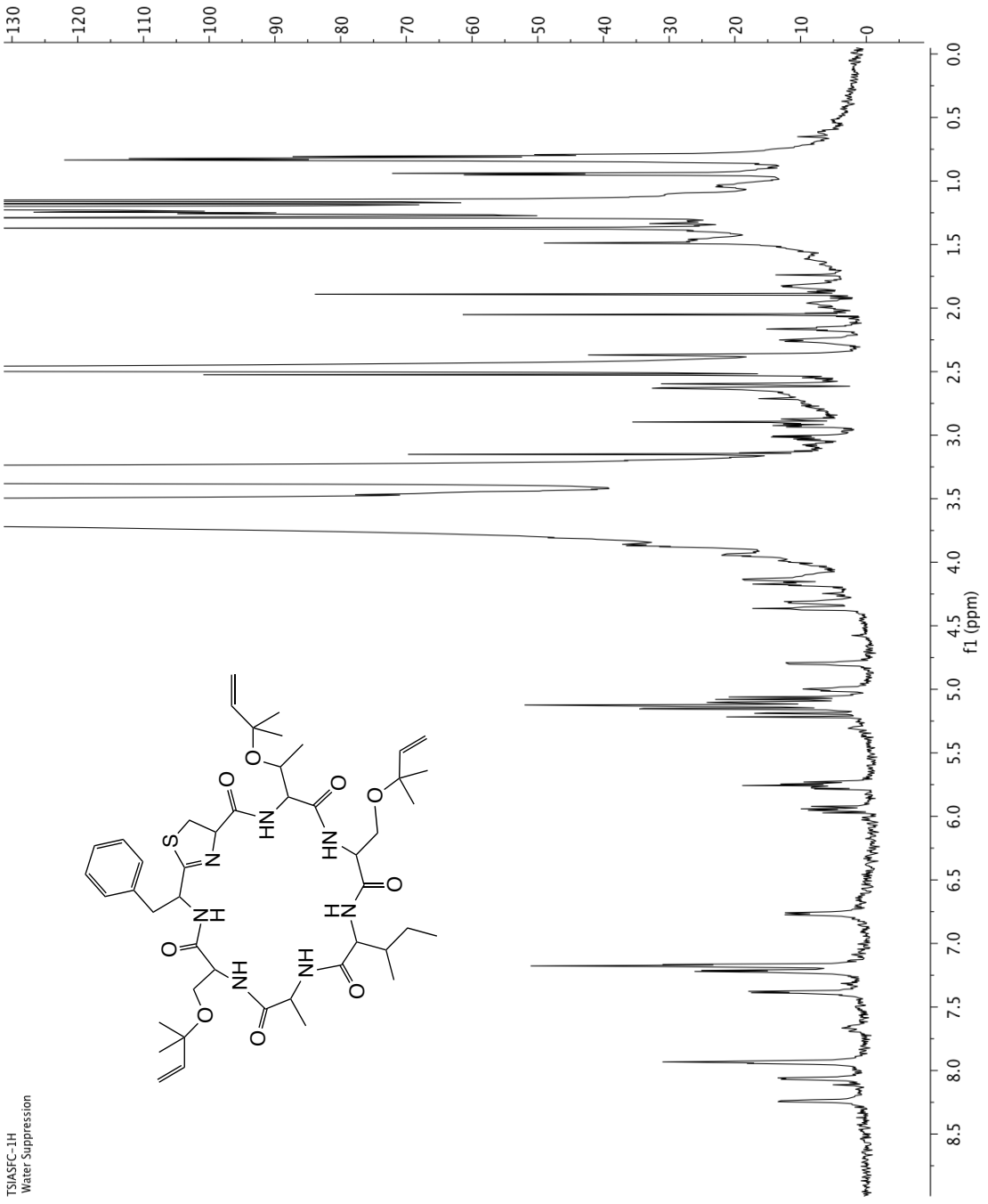


Figure S5.
B1. COSY spectrum of TSIASFC (**15**) in DMSO-d_6 at 600 MHz [3]

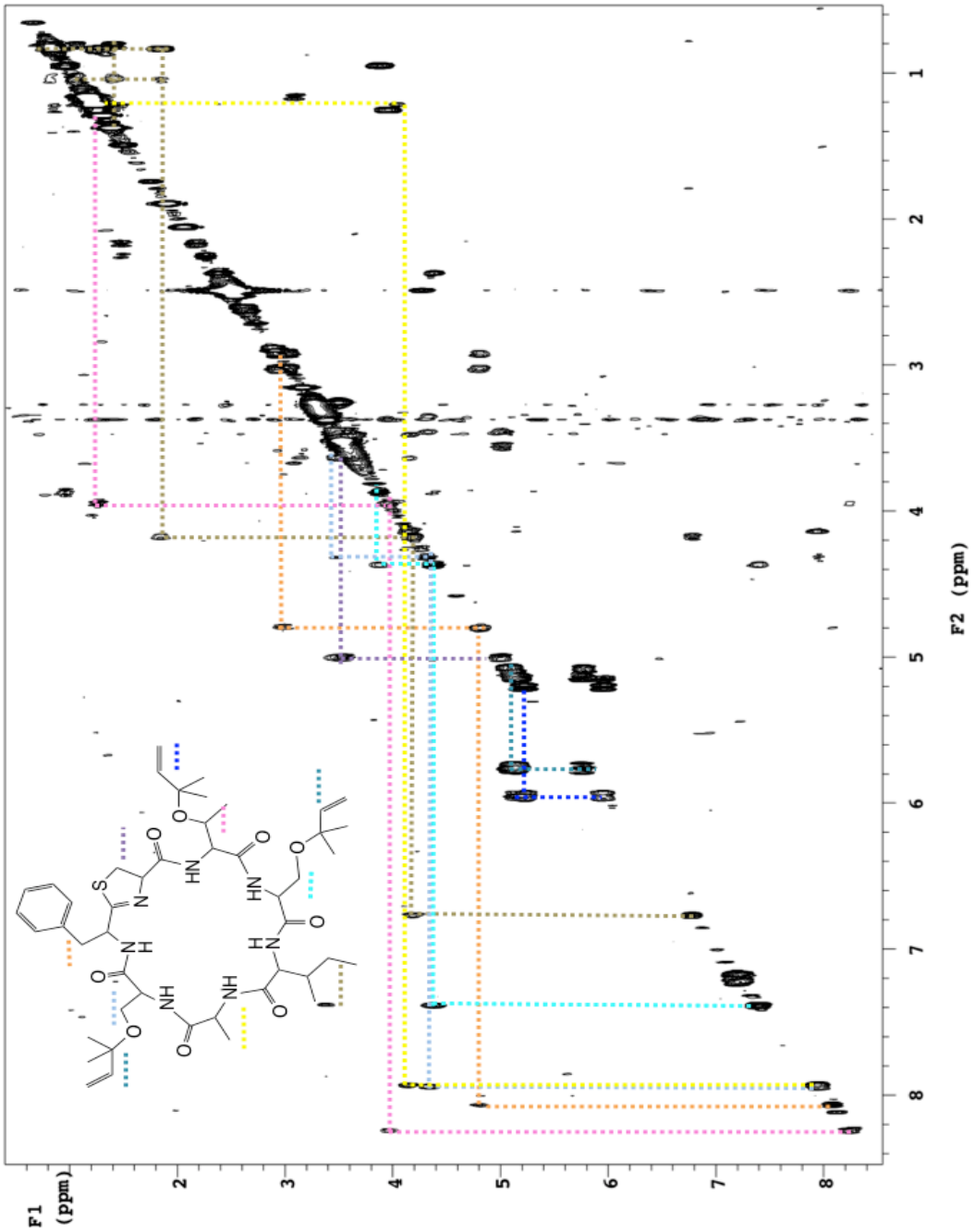


Figure S6.
A1. ^{1}H NMR spectrum of TVPT-pClF-C(**11**) in DMSO-d_6 at 600 MHz

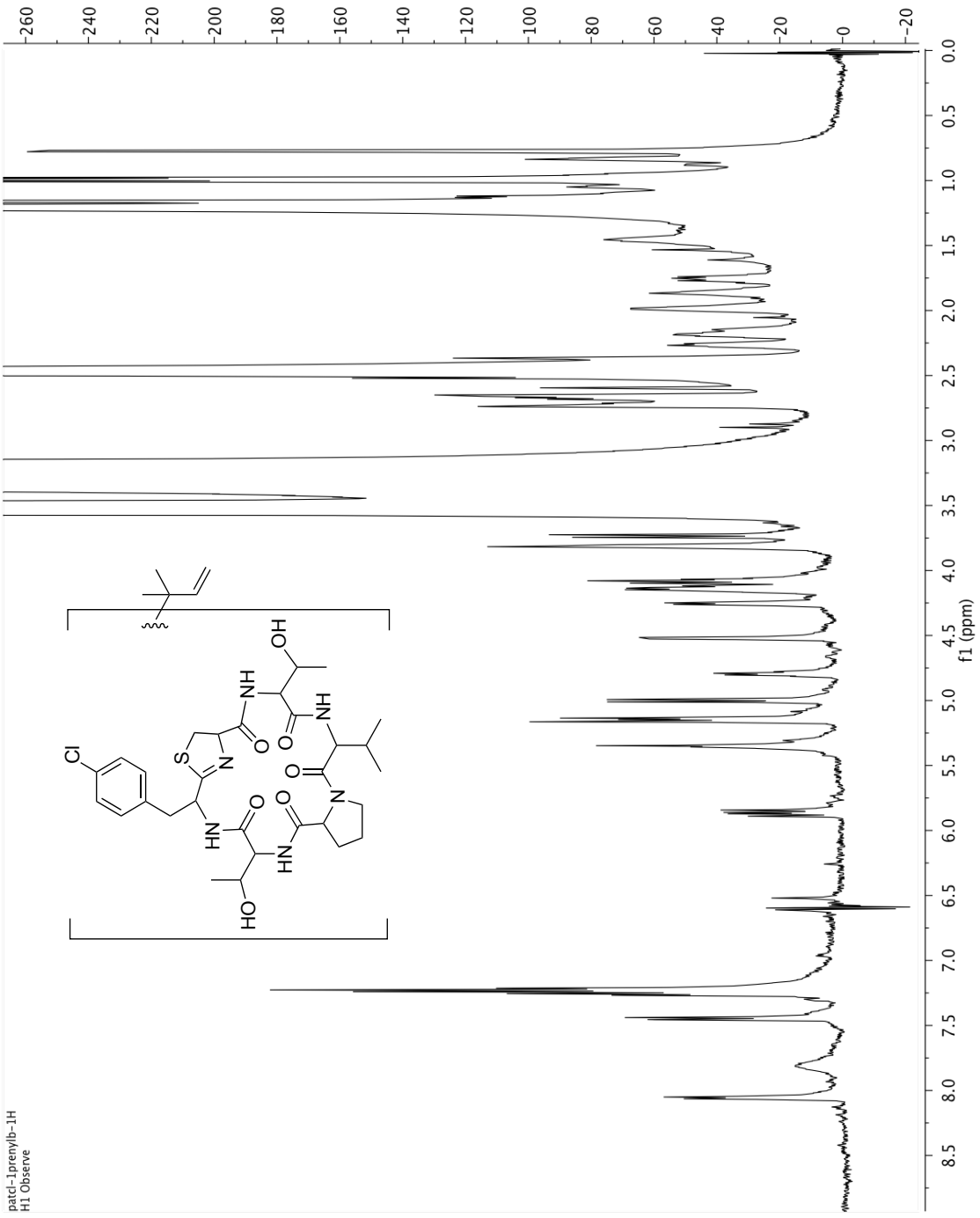


Figure S6.
B1. COSY spectrum of TVPT-pClF-C(**11**) in DMSO-d_6 at 600 MHz [4]

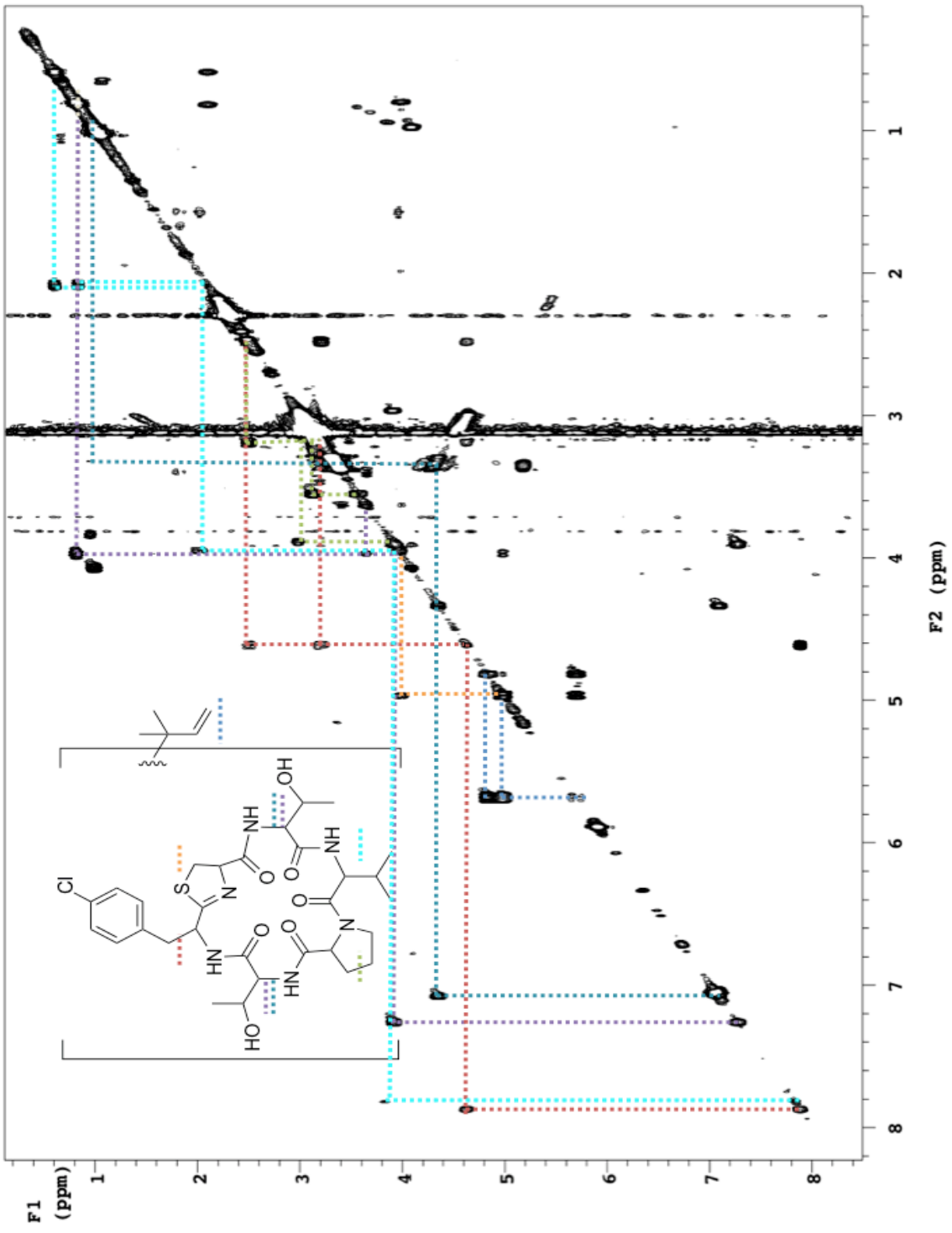
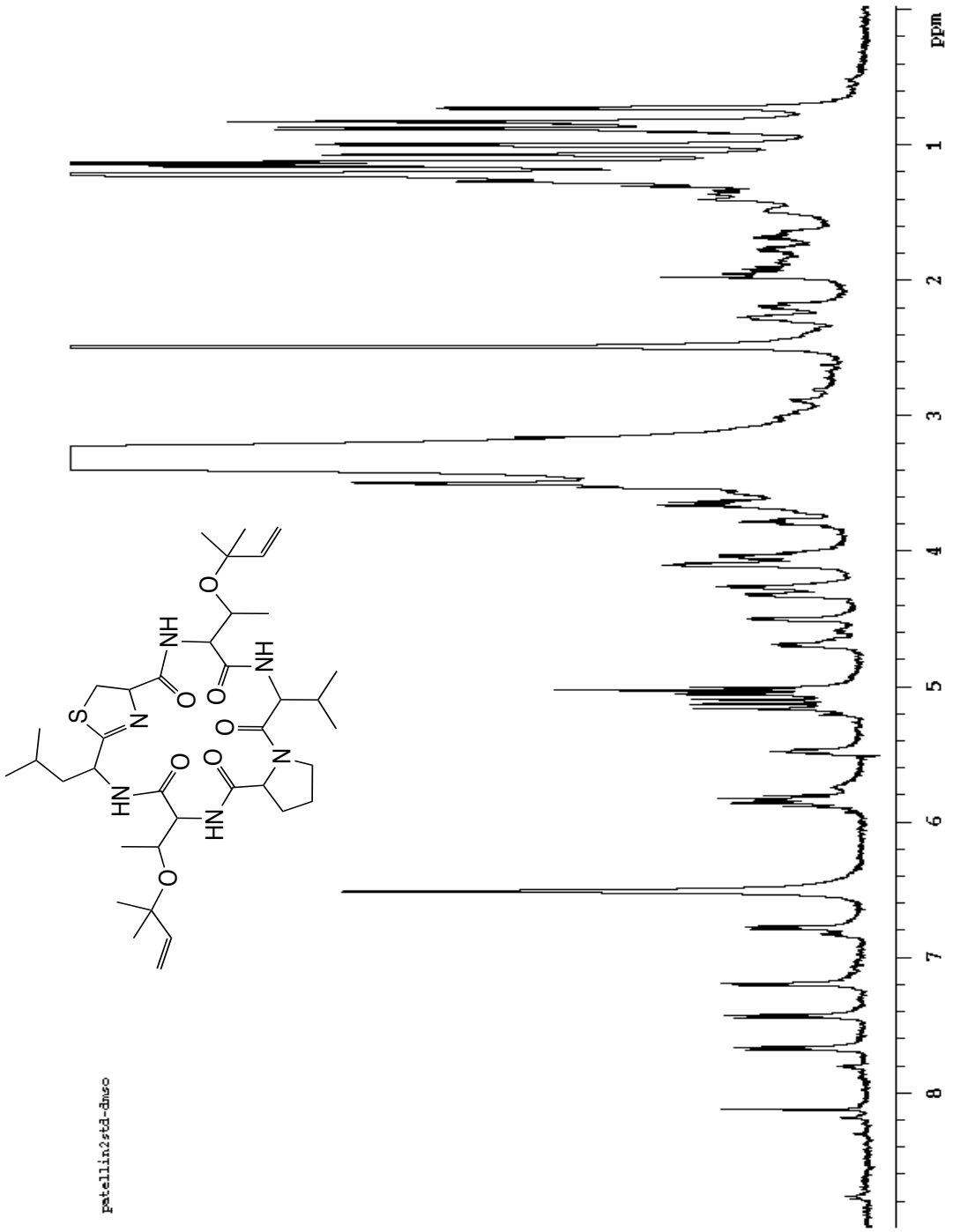


Figure S7.
B1. COSY spectrum of Patellin 2 (**1**) in DMSO-d_6 at 600 MHz



(1)Donia, M. S.; Ruffner, D. E.; Cao, S.; Schmidt, E. W. *ChemBioChem* **2011**, *12*, 1230.
(2)Burton, I. W., Quilliam, M.A., Walter,J.A. *Anal. Chem.* **2005**, *77*, 3123.
(3)Wipf, P.; Uto, Y. *J. Org. Chem.* **2000**, *65*, 1037.
(4)Zabriskie, T. M.; Foster, M. P.; Stout, T. J.; Clardy, J.; Ireland, C. M. *J. Am. Chem. Soc.* **1990**, *112*, 8080.

CHAPTER 5

CONCLUSIONS

Conclusions

We set out to address the two outstanding problems in natural products drug discovery: the variability in occurrence and supply of natural products. Both of these challenges create a significant hindrance to the development of natural products into useful drugs. Although geographical and sample-to-sample variability in occurrence can occur in terrestrial natural products, [1] it can be particularly widespread for those natural products derived from marine organisms. Several reasons may explain variability of occurrence for marine natural products, such the complexity of ocean habitats [2,3] (local predation, competition, flux of microbes from seawater) as well as difficulties assigning species among heteromorphic organisms such as sponges. [4,5] Answering the question of 'who makes what, when, and why?' is therefore highly pertinent to understanding the distribution of natural products and their potential.

There is increasing evidence that symbiotic bacteria are the actual producers of most marine invertebrate-derived NPs. [6-9] We thus proposed that symbiotic bacteria are central to distribution of marine natural products. Further, we proposed that the factors that control symbiosis in turn control the distribution of natural products. Piece by piece, this hypothesis has been addressed in individual studies of chemical symbiosis. We asked if there are global patterns and what they would mean for drug discovery – global in this sense means examining all symbionts and chemicals within different animals and the changes in their distribution across time, location, and hosts. It was therefore critical that we use sequencing platforms capable of identifying thousands of bacteria in each sample. This approach allowed us to compare their general characteristics as a community to the other variables that were being tested. Statistical methods that are independent of

metabolite identity were also critical for the overall analysis of chemistry between many samples.

We found that in ascidians, the hosts generally select their symbionts and secondary metabolites. In known cases, the symbionts that are natural product producers are the most abundant components of the microbiome. This is an important finding in the era of genomics-based natural product discovery, not just in ascidians but also in other marine invertebrates. Marine invertebrates are microbial hotspots, [10,11] containing hundreds or thousands of different species of bacteria. Here we show that in mining for microbial producers of natural products, efforts should be focused on the most abundant microbes in the composition. Studying the latitudinal influence of geography to the symbionts and chemistry was very important as it revealed that the selection pressure between these regions could influence chemistry and symbiosis in imbalanced ways. While the symbionts generally maintained host specificity, the potency of secondary metabolites increased in tropical samples. Therefore, within the limits of our sampling and analysis, symbiosis is generally host-specific while secondary metabolism can be location specific when there are large biogeographical divides such as water temperature and species diversity.

While it is tempting to generalize our results to a broader section of marine invertebrates, their symbionts, and associated metabolites, it is important to reiterate the limits of our data. In particular, our tests on the potency of extracts were limited to a select few strains, which may not capture ecologically relevant activities of the extracts from temperate ascidians. Additionally, since almost none of the chemistry can be identified from the temperate ascidians in the database collections, it is difficult to say

precisely what differences in chemistry exist between closely related species in temperate and tropical regions. In this respect, methods that can reveal structural relatedness without dependence on complete identification and isolation of compounds, such as molecular networking, will be useful in future studies. [12]

To address the supply of symbiotic natural products, we used the heterologous expression of cyanobactins as a model case. Cyanobactin pathways are not only amenable to library generation; they are also amenable to metabolic engineering, specifically through precursor supply and simple fermentation optimization. Serendipitously, we discovered that increasing the precursor DMAPP improved cyanobactin production, in a manner that is independent of compound prenylation. The scope of this effect is of great interest in that it might be applicable to other heterologous systems such as to other RiPPs, or other natural product pathways in general. The ability to supply cyanobactins will now allow further practical investigations on their chemistry and biology. Indeed this method allowed us to uncover the first activity for patellin 2 using a high content phenotypic assay. This opens the possibility of finding more activities for 'orphan' cyanobactins and may lead to a clearer picture of their ecological roles. We further demonstrated the capacity of cyanobactin pathways to generate unusual derivatives containing nonproteinogenic amino acids. By incorporating unnatural amino acids to the cyanobactin backbone, we successfully combined the structural complexity of non-ribosomal peptides with a multistep, substrate tolerant ribosomal pathway.

References

1 Marienhagen, J. & Bott, M. Metabolic engineering of microorganisms for the synthesis of plant natural products. *J Biotech* **163**, 166-178, doi:10.1016/j.jbiotec.2012.06.001 (2013).

2 Hay, M. E. Marine chemical ecology: what's known and what's next? *J Exp Mar Biol Ecol* **200**, 103-134, doi:http://dx.doi.org/10.1016/S0022-0981(96)02659-7 (1996).

3 Li, J. W. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-165, doi:10.1126/science.1168243 (2009).

4 Cardenas, P., Perez, T. & Boury-Esnault, N. Sponge systematics facing new challenges. *Adv Mar Biol* **61**, 79-209, doi:10.1016/B978-0-12-387787-1.00010-6 (2012).

5 Morrow, C. & Cardenas, P. Proposal for a revised classification of the Demospongiae (Porifera). *Front Zool* **12**, 7, doi:10.1186/s12983-015-0099-8 (2015).

6 Florez, L. V., Biedermann, P. H., Engl, T. & Kaltenpoth, M. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat Prod Rep*, doi:10.1039/c5np00010f (2015).

7 Piel, J. Metabolites from symbiotic bacteria. *Nat Prod Rep* **26**, 338-362, doi:10.1039/b703499g (2009).

8 Schmidt, E. W. Trading molecules and tracking targets in symbiotic interactions. *Nat Chem Biol* **4**, 466-473, doi:10.1038/nchembio.101 (2008).

9 Schmidt, E. W., Donia, M. S., McIntosh, J. A., Fricke, W. F. & Ravel, J. Origin and variation of tunicate secondary metabolites. *J Nat Prod* **75**, 295-304, doi:10.1021/np200665k (2012).

10 Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* **10**, 641-654, doi:10.1038/nrmicro2839 (2012).

11 Erwin, P. M., Pineda, M. C., Webster, N., Turon, X. & Lopez-Legentil, S. Down under the tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. *Isme J*, doi:10.1038/ismej.2013.188 (2013).

12 Nguyen, D. D. *et al.* MS/MS networking guided analysis of molecule and gene cluster families. *Proc Natl Acad Sci U S A* **110**, E2611-2620, doi:10.1073/pnas.1303471110 (2013).