

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Biosynthesis of natural products in marine bacteria : studies in molecular genetics, phylogeny and structural elucidation

### Permalink

<https://escholarship.org/uc/item/6x36s00v>

### Author

Sudek, Sebastian

### Publication Date

2006

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Biosynthesis of natural products in marine  
bacteria: studies in molecular genetics,  
phylogeny and structural elucidation**

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Marine Biology

by

Sebastian Sudek

Committee in charge:

Professor Margo G Haygood, Chair  
Professor Douglas Bartlett  
Professor Bianca Brahamsha  
Professor William Fenical  
Professor John Huelsenbeck



The dissertation of Sebastian Sudek is approved, and  
is accepted in quality and form for publication on  
microfilm:

---

---

---

---

---

Chair

University of California, San Diego

2006

## Table of contents

Signature page.....	iii
Table of Contents.....	iv
List of Abbreviations.....	v
List of Figures and Tables.....	vi
Acknowledgements.....	viii
Vita, Publications and Fields of Study.....	xiv
Abstract.....	xvi
1. Introduction: Biosynthesis of natural products in (symbiotic) marine bacteria.....	1
2. Identification of the <i>bry</i> polyketide synthase gene cluster from <i>E. sertula</i> : the putative biosynthetic pathway for bryostatin.....	26
3. Diversity of the symbiotic cyanobacterium <i>Prochloron sp.</i> in didemnid ascidians.....	57
4. Trichamide, a cyclic peptide from the bloom-forming cyanobacterium <i>Trichodesmium erythraeum</i> predicted from the genome sequence.....	86
Appendix: The tale of the <i>B. neritina</i> .....	113

## List of Abbreviations

ACP	acyl carrier protein
AT	acyl transferase
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
CA	California
CID	collision-induced dissociation
cDNA	complementary DNA
CoA	coenzyme A
DGGE	denaturing gradient gel electrophoresis
DH	dehydratase
DNA	deoxyribonucleic acid
ER	enoyl reductase
ESI	electron spray ionization
FT-MS	Fourier transform mass spectrometry
GTR	general time reversible
HMG	hydroxymethyl glutaryl
HMG-CS	hydroxymethyl glutaryl CoA synthase
HPLC	high-performance liquid chromatography
IRMPD	infrared multiphoton dissociation
ITS	internal transcribed spacer
KR	ketoreductase
KS	keto synthase
MS	mass spectrometry
m/z	mass/charge
MT	methyl transferase
NC	North Carolina
NRPS	non-ribosomal peptide synthase
ORF	open reading frame
PCR	polymerase chain reaction
PKS	polyketide synthase
PNG	Papua New Guinea
PS	pyran synthase
RNA	ribonucleic acid
RPA	ribonuclease protection assays
rRNA	ribosomal RNA
RT-PCR	reverse transcription polymerase chain reaction
SCUBA	self-contained underwater breathing apparatus
TE	thioesterase
tRNA	transfer RNA

## List of figures and tables

### Chapter 1

#### Figure

1.1	Structure of selected natural products.....	16
1.2	Modular organization of non-ribosomal peptide synthases and polyketide synthases.....	17
1.3	The putative first steps of bryostatin biosynthesis.....	18
1.4	Classic example of congruent phylogenetic trees.....	19
1.5	Overview of dissertation chapters.....	20

### Chapter 2

#### Table

2.1	Genes and PKS domains in the <i>bry</i> cluster.....	48
-----	--	----

#### Figure

2.1	Structure of bryostatins.....	49
2.2	The <i>bry</i> cluster in “deep” and “shallow” <i>E. sertula</i> .....	50
2.3	Proposed pathway for bryostatin biosynthesis.....	51
2.4	Maturation steps after PKS assembly.....	52

## Chapter 3

### Table

3.1	<i>Prochloron</i> samples analyzed in this study.....	73
3.2	Oligonucleotide primers used in this study.....	73

### Figure

3.1	<i>Didemnum molle</i> , <i>Lissoclinum patella</i> and symbiotic <i>Prochloron sp.</i> ....	74
3.2	Schematic of a didemnid ascidian.....	75
3.3	Patellamides A and C.....	76
3.4	rRNA operon of <i>Prochloron sp.</i> .....	77
3.5	Denaturing gradient gel electrophoresis of <i>Prochloron sp.</i> .....	77
3.6	Pairwise distance matrix of <i>Prochloron spp.</i> 16S rRNA sequences.....	78
3.7	Partial 16S alignment.....	79
3.8	Partial 16S-23S ITS alignment.....	79
3.9	Phylogenetic tree of <i>Prochloron spp.</i> based on the 16S rRNA gene.....	80
3.10	Phylogenetic tree of <i>Prochloron spp.</i> based on 16S-23S ITS.....	81
3.11	Overview of the didemnid – <i>Prochloron sp.</i> symbiosis.....	82

## Chapter 4

### Table

4.1	The <i>tri</i> cluster proteins and their homologs.....	102
4.2	Mass spectrometry data of trichamide and fragments.....	103

### Figure

4.1	The <i>tri</i> gene cluster.....	104
4.2	Alignment of the precursor peptides PatE and TriG.....	104
4.3	Structure of trichamide.....	105
4.4	Proposed biosynthetic pathway to trichamide.....	106
4.5	Fourier transform mass spec of crude <i>T. erythraeum</i> extracts.....	107
4.6	Mass spec fragmentation pattern under CID.....	108
4.7	Mass spec fragmentation pattern under IRMPD.....	109



## Acknowledgements

Over the course of the last five years, many people have influenced and shaped my dissertation research as well as my personality in their different ways and I am deeply thankful to them. Margo Haygood, my Ph.D advisor, accepted me as her graduate student without a personal meeting or a talk on the phone. Quite a gamble and I hope she thinks it paid off, I certainly consider joining her lab the right decision. In countless one-on-one meetings, she has continuously provided feedback, helped develop research ideas and challenged me to start thinking as an independent scientist. She also provided encouragement when things looked dire to me (I will remember not to be so pessimistic in the future) and showed infectious enthusiasm about cool new results. Her grants covered most of my salary as well as allowed me to attend a number of scientific conferences.

All members of my committee have been very supportive. Bill Fenical was a great resource for natural product chemistry both in terms of advice as well as generous offers to use his machinery. He certainly is the fastest replier when it comes to email and always finds a way to squeeze a meeting into his busy schedule. Doug Bartlett has a tremendous knowledge in microbiology that he is willing to share and put to use by suggesting experimental approaches. This is coupled with great kindness and genuine caring for graduate students. Bianca Brahamsha also had great suggestions for experiments (unfortunately there wasn't enough time to follow up on some) and provided long-term incubator space for the *Trichodesmium* cultures. John

Huelsenbeck is an intimidatingly sharp yet friendly computational phylogeneticist.

Both in his class as well as in personal meetings I learnt to appreciate how much more there is to phylogeny than just hitting the “make tree” button.

The Haygood and Tebo labs had many talented and fun people over the years and provided the right atmosphere both for making serious progress as well as deal with the daily craziness of scientific research. Grace Lim taught me the do's and don'ts of DGGE as well as being a discussion partner for research ideas. Koty Sharp was always a patient listener to toil over PCR difficulties and she was very willing to lend out her Buzz Lightyear figurine that could magically make experiments work. Christine Anderson shared the joys and frustrations of the bryostatin project that dominated a long part of our dissertation research. Her calmness and keen sense of humor helped through many rough spots. Laura Murphy was another bryostatin struggler and had the right attitude to keep things in perspective. She and her husband Bob also introduced me to the beer-on-the-beach style of Independence Day celebration. Cheers! The Manganese-and-other-metal from the Tebo lab made wandering over there to use the common equipment an entertaining trip. Overall they also ensured a more reasonable male to female ratio. Hope Johnson is an accomplished surfer and made me go out north of the pier many times. Greg Dick kept reminding me that alles in butter ist...and in the end he was right. Brian Clement introduced me to my future wife; more importantly, he helped start our relationship by hosting a veritable spider family in his closet that came out at the right time. Flip Mc McCarthy is a seasoned microbiologist but also a man of many trades. There seems to

be no topic one couldn't have a meaningful conversation with him. Racheal Howard has been my office mate for the last couple of years. She has quite a talent for organizing trips and I am thankful for the activities I took part in and the interesting people I met through her.

Carolyn Sheehan managed our lab groups. She did an outstanding job juggling the needs of 20+ people and was always willing to go above and beyond to accommodate our research needs. I cannot overemphasize how much easier life in the lab was and how much it facilitated my research.

Mark Hildebrand, who worked on the earlier parts of the bryostatin project, is an amazing molecular biologist. He knows any technique there is under the sun and he was willing to teach me patiently and with a fine sense of humor. This kept me trying again and again. I will always remember to boil my probe. Mark also is a talented singer and songwriter. His epic "*Tale of the B. neritina*" is included in the appendix. In the great tradition of ancient poets it shall conserve our bryostatin odyssey for generation come. May they be more careful what cloning to attempt.

Alex Worden was a post-doc one floor up. However for some strange reason (it might have to do with the nationality of her husband), she took care of me during the early years of grad school. This included dinner invitations, credit card assistance, play time with her amazing dogs but also serious discussions about character traits sometimes found in big shot scientist and how to avoid developing them.

Eric Schmidt has been an amazing collaborator in the *Prochloron* and trichamide projects. He was totally unselfish in agreeing to join forces; he taught me most of what I know about natural product chemistry. Eric is relentless when it comes to moving forward and publishing results, which has proven very beneficial for my publication record. He is also an adventurous hiker and this led to a memorable descent down a steep Utah mountain slope. This dissertation wouldn't have been possible without him.

John Faulkner encouraged me to pick up a second project and really cut me down when I suggested a couple, leaving only *Prochloron*. He clearly indicated that collaborating with his former student Eric Schmidt would be the road to success and unsurprisingly to me, this dissertation proves him right. John passed away in my second year at SIO, which prevented me from following in the footsteps of earlier Haygood students who have pursued projects in his lab.

Eddie Kisfaludy is an incredibly experienced diver. He took me on countless trips to collect my *Bugula*, but spiced them up by having me assist in catching sharks, knocking out and bagging fish and some other unusual underwater activities. I really enjoyed my time underwater.

Over the last couple of years a couple of talented undergraduate students joined our lab for a quarter to pursue research projects with me. Audra Budde, Lynn

Doan and Thu Mai Jan each volunteered their time to learn molecular biology as well as experience daily life in a research lab. I am grateful for their contributions, which are included in the bryostatin chapter. Amrish Patel was an exchange student from London who spent a couple of months working full time in the lab with me to learn molecular biology techniques. He is a talented scientist and had a significant impact in the bryostatin project. He also is a pleasant, patient, fun person. I am sure he could have a good career in science if he chooses to pursue it. Guido Bordignon entered the lab recently and spent some time with me learning how to grow bugs and work with their genes. His unique character and fun attitude towards life have been very entertaining and helped me to keep things in perspective during stressful times.

I also want to express my gratitude to my parents who have always been supportive of my education, even when it took me to the other end of the world. Their care and attention got me on the right start in school and were the basis of my later academic achievements. Their ongoing support allowed me to pursue a university degree in Germany.

Finally, but most importantly I want to thank my wife Lisa. Her gentleness and love kept me going when times were rough and made me reach for the stars when things were going well. She has truly turned my life around and things start to make sense. Everyday I have to remind myself that I am in fact not dreaming.

\*\*\*\*\*

Chapter 2: The bryostatin project was a long-running collaborative effort.

Laura Waggoner, Nicole Lopanik, Mark Hildebrand and Haibin Liu took part in the DNA sequencing and annotation effort as well as developing the biosynthetic hypothesis. Christine Anderson contributed all RNA data. David Sherman provided valuable input to the manuscript.

Chapter 3: Eric Schmidt collected all samples used in this study. He and his student Mohamed Donia provided helpful discussion. Ralph Lewin originally described *Prochloron*. He shared many of his interesting (and entertaining!) experiences with this bug.

Chapter 4: Kelly Roe and Kathy Barbeau provided advice on *Trichodesmium* culturing. Bioassays were run in the labs of Louis Barrow and Baldomero Olivero at the University of Utah and William Fenical and William Gerwick at Scripps. Chad Nelson at the University of Utah ran the mass spectrometry measurements. This chapter is a reprint of the material as it appears in Applied Environmental Microbiology 2006, 72: 4382-4387 with co-authors Margo G. Haygood, Diaa T. A. Youssef, and Eric W. Schmidt. The dissertation author was the primary investigator.

\*\*\*\*\*

This research was funded by the National Institute of Health, the National Science Foundation and the SIO Graduate Department. The SIO Grad Department and the Howard Hughes foundation have provided funding for attending a summer course in Microbial Diversity at the Marine Biological Laboratory in Woods Hole, MA.

## Vita

1975	Born in Mainz, Germany
1998	Teaching Assistant (Genetics), Colorado State University, Pueblo CO
1999	M.S. (Applied Natural Sciences) Colorado State University, Pueblo CO
1999-2000	Teaching Assistant (Zoology), Johannes-Gutenberg-University, Mainz Germany
2001	Thesis Research Fellowship (School of Medicine)
2001	Diploma (Biology), Johannes-Gutenberg-University, Mainz Germany
2001-2006	Graduate Student Researcher, University of California San Diego
2003	Teaching Assistant (Biochemistry), UC San Diego
2004	Microbial Diversity Summer Course, Marine Biological Laboratory, Woods Hole MA
2006	Ph.D. Marine Biology, UC San Diego

## Publications

Sudek S, Haygood MG, Youssef DT, Schmidt EW. 2006. Trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum* predicted from the genome sequence. *Appl Environ Microbiol* 72: 4382-4387.

Hildebrand M, Waggoner LE, Liu H, Sudek S, Allen S, Anderson C, Sherman DH, Haygood M. 2004. bryA: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. *Chem Biol* 11:1543-52.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J. 2005. 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* 102:7315-20.

Schmidt EW, Sudek S, Haygood MG. 2004. Genetic evidence supports secondary metabolic diversity in *Prochloron* spp., the cyanobacterial symbiont of a tropical ascidian. *J Nat Prod* 67:1341-5.

Perovic S, Schroder HC, Sudek S, Grebenjuk VA, Batel R, Stifanic M, Muller IM, Muller WE. 2003. Expression of one sponge Iroquois homeobox gene in primmorphs from *Suberites domuncula* during canal formation. *Evol Dev* 5:240-50.

Schroder HC, Sudek S, Caro S, Rosa S, Perovic S, Steffen R, Muller IM, Muller WE. 2002. Synthesis of the neurotoxin quinolinic acid in apoptotic tissue from *Suberites domuncula*: cell biological, molecular biological, and chemical analyses. Mar Biotechnol. 4:546-58.

Sudek S. Klonierung einer MET-verwandten Rezeptortyrosinkinase im marinen Schwamm *Suberites domuncula* (Cloning of a MET-class receptor tyrosine kinase gene from the marine sponge *Suberites domuncula*). Diploma thesis. Advisor: Dr. Werner E.G. Mueller. Johannes-Gutenberg University, Mainz, Germany

## **Fields of Study**

Major Field: Marine Biology

Studies in Microbiology

Prof. Margo Haygood

Prof. Doug Bartlett

Prof. Bianca Brahmsha

Studies in Natural Product Chemistry

Prof. William Fenical

Prof. Eric Schmidt

Studies in Phylogeny

Prof. Margo Haygood

Prof. John Huelsenbeck



## **ABSTRACT OF THE DISSERTATION**

Biosynthesis of natural products in marine bacteria: studies in molecular genetics,  
phylogeny and structural elucidation

By

Sebastian Sudek

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2006

Professor Margo G Haygood, Chair

Chapter 1 introduces marine natural products and discusses the symbiosis of bacteria and marine invertebrates as a source of these molecules. Indications for a bacterial origin of compounds isolated from invertebrates are presented and the major types of bacterial natural product biosynthesis pathways are introduced. The study of biosynthesis opens the possibility of solving the supply issue that often prevents the development of drugs from marine natural products. It also yields novel biochemical reactions and allows insight into the evolution of biosynthetic pathways.

Chapter 2 explores the biosynthetic genes responsible for the production of the anticancer compound bryostatin. Bryostatin was isolated from a marine bryozoan; here a putative biosynthetic pathway from a symbiotic bacterium via a modular polyketide synthase (PKS) mechanism is presented. Symbiont DNA was enriched using ultracentrifugation techniques and lambda phage library constructed. For homology screening initially a DNA fragment was used specific for a symbiont PKS derived in earlier work by degenerate PCR. Subsequently the ends of the isolated lambda fragments were used in a series of screenings. Short gaps were bridged by PCR. This chromosome walking approach yielded the entire 80 kb *bry* cluster. The cluster contains all the major elements needed for the production of bryostatin. A putative biosynthetic scheme is described.

Chapter 3 investigates the phylogeny of *Prochloron sp.*, a cyanobacterial symbiont of ascidians, based on 16S rRNA gene and 16S-23S internally transcribed spacer region sequences as well as denaturing gradient gel electrophoresis. *Prochloron* diversity is very low and there is no evidence for cospeciation with the ascidian host. This is in contrast to the vertical transmission of symbionts reported for some ascidians. *Prochloron* is the source of the patellamides, the relationship between phylogenetic and biosynthetic diversity is discussed.

Chapter 4 presents a novel cyclic peptide, trichamide, from the free-living cyanobacterium *Trichodesmium erythraeum*. A gene cluster related to the patellamide biosynthetic genes from *Prochloron sp.* was identified by homology in *T. erythraeum* sequence data available in Genbank. The structure of trichamide was proposed based on sequence information and could be verified by Fourier-transfer mass spectrometry.

# 1.

## **Introduction: Biosynthesis of natural products in (symbiotic) marine bacteria**

The dissertation explores several aspects of the biosynthesis of natural products in selected marine bacteria. This chapter introduces natural products from marine invertebrates and the bacterial symbionts often proposed to be involved in biosynthesis of these molecules. The term all-encompassing term “natural product” will be used in this chapter, these molecules are also often called “secondary” or “bioactive metabolites”.

### Marine Natural Products

Marine natural products are small organic molecules, typically with a molecular weight of 100s to 2000. They have diverse chemical structures and activities. Their ecological functions can be anti-competitor, anti-predation, trace metal scavenging and communication. This translates into pharmaceutically useful biological activities such as antibiotics, anticancer, antiparasitic, antiviral, and antifungal properties. Terrestrial environments have been and still are screened extensively for bioactive compounds. The marine realm is technically more challenging to explore; advances such as SCUBA diving and deep sea sampling

devices have made it accessible. Since the 1970s thousands of marine natural products have been described (for example, 869 structures are presented in a review covering only the year 2000 [14]).

Sessile marine invertebrates, especially sponges, but also ascidians and bryozoans have been particularly prolific sources. Like terrestrial plants, marine non-motile organisms have to find a way to counter predation. Physical defense mechanisms include shells, cuticles, external or internal spines and spicules. Chemical defense is the ability to produce toxic (or at least noxious) compounds that leave the animal unpalatable to potential predators. Often a mixture of both physical and chemical defense is employed. While motile, larvae of marine invertebrates are very vulnerable to predators and often the chemical defense is extended to or even mainly employed in the larval stages [35]. Natural products are also used to fend off competitors in the struggle for settling space on sublittoral substrates. Finally, marine invertebrates need a mechanism to prevent microbial/algal/fungal fouling of their surfaces, which can be achieved chemically.

#### Natural products and symbiosis

In the broadest definition symbiosis describes two species, which are associated persistently or at least for a major part of their life cycles. Often an evolutionary benefit for one or both of the partners can be demonstrated or hypothesized. Symbiosis between bacteria and marine invertebrates is common. There is a large diversity of lifestyles: the symbionts can be obligate or facultative, present as

biofilms on the host's surface or in body cavities, intercellular within the host tissue or even intracellular, in specialized host cells (bacteriocytes). In microbe-invertebrate symbiosis, the benefit for the host is often nutritional. For example, wood-boring shipworms (Teredinidae) are able to thrive on their diet because cellulose and lignin are degraded by symbiotic bacteria [9,50]. Many invertebrates in deep-sea vent community are dependent on chemoautotroph symbionts [49].

Symbiosis can also involve natural products. As described above, the ability to produce defensive compounds is a strong selective advantage for marine invertebrates to fight predation, compete for space and prevent fouling. Many animals lack the metabolic capabilities to synthesize potent compounds, so forming a partnership with a microbe is an effective strategy. Bacteria are ideally suited for this kind of symbiosis, because of their versatile biosynthetic capabilities. A classic example of chemical defense is found in the shrimp *Palaemon macrodactylus*. Isatin (Figure 1.1) is produced by an *Alteromonas* sp. persistently associated with the shrimp larvae and has been shown to prevent fungal infections [20]. When bacterial symbionts are uncultivated, it is difficult to obtain direct evidence for their involvement in the production of natural products isolated from marine invertebrates. Indications for a symbiotic origin are:

1. The natural product has previously been identified from a bacterium or at least has reminiscent structural features as in the case of ET-743 from the ascidian

*Ecteinascidia turbinata* and saframycin B from *Streptomyces lavendulae*

(Figure 1.1).

2. A similar compound has been isolated from an unrelated invertebrate. This is particularly appealing if the animals live in vastly different environments, which makes convergent evolution unlikely. Pederin from a beetle and onnamide A from a sponge (Figure 1.1) are discussed in more detail below.
3. The suggested mode of biosynthesis has previously only been found in bacteria as is the case with complex polyketides like bryostatin discussed in chapter 2.
4. Antibiotic treatment leads to a loss of compounds. *Bugula neritina* is the source of the anticancer compound bryostatin. When *B. neritina* larvae are treated with antibiotics, colonies developing from these larvae do not contain symbiotic bacteria or bryostatin [7, 37]. These experiments do not exclude the possibility that compounds are produced by the host in response to bacteria.
5. The compounds are localized in bacterial cells. In the sponge *Dysidea herbacea*, animal cells have been successfully separated from associated cyanobacterial *Oscillatoria spongelliae* filaments using a fluorescent activated cell sorter. 13-demethylisodysidenin and a brominated diphenyl ether were found only in the cyanobacteria, while two sesquiterpenes were located only in the sponge cells [47, 48]. In another instance, the bacterial population of the sponge *Theonella swinhoei* was separated by differential centrifugation. The cyclic peptide theopalauamide and the polyketide swinholide A were found in morphologically different bacteria [2]. These results should be interpreted with

caution. Natural products can be excreted by bacteria; they may diffuse or be transported around the host and be taken up by host cells. The opposite could also happen, i.e. host compounds could be excreted and taken up by symbiotic bacteria. Furthermore, certain nudibranchs have been shown to retain natural products derived from a nutritional source, green algae [21]. In the same way it seems possible that bacterial producers are digested by host cells, while their natural product are retained.

None of the indications listed is conclusive and multiple lines of evidence should be demonstrated before proposing a bacterial origin.

### Types of bacterial biosynthetic pathways

The *bry* cluster described in chapter 2 is an example of a modular polyketide synthase (PKS). Modular PKS are large polyfunctional megasynthases (up to 2000 kDa) that elongate and modify the nascent polyketide in an assembly-line fashion [30]. They are organized into functional modules that perform one specific elongation step by adding an acyl-CoA precursor and then transferring the polyketide chain to the next module. Each module has several catalytic domains (Figure 1.2). A canonical core module consists of a  $\beta$ -ketoacyl synthase (KS), an acyl transferase (AT) and an acyl carrier protein (ACP) domain. The AT selects and loads the extender unit (typically malonyl or methylmalonyl CoA) onto the ACP domain. In a decarboxylative step the KS condenses the nascent polyketide received from the ACP of the previous module



onto the acyl-CoA extender unit, effectively adding a minimum of two carbons to the polyketide. In most characterized PKSs, the AT functionality is integrated into each PKS module, but in several systems it resides on a discrete gene and there is no AT domain present in the PKS genes [4]. After the condensation step the ACP bridges the distance to the next module's KS domain with a phosphopantetheine arm to transfer the polyketide for the next extension step. Each module can contain different accessory domains. After extension, the  $\beta$ -keto group can be reduced to a hydroxyl group by a ketoreductase (KR) domain, to a trans-double bond by a dehydratase (DH) domain and to the fully saturated  $\beta$ -carbon by an enoyl reductase (ER) domain [10, 26].

Iterative PKS systems consist of complexes of mono-functional proteins [25]. The proteins contain the same functional domains as modular PKS but analogously to fatty acid synthesis are used iteratively once per elongation step. These PKS catalyse the formation of compounds that require aromatization and cyclization, but not extensive variable reduction or reduction/dehydration. A well characterized marine example is enterocin. Iterative PKS systems often catalyze elaborate C-C cleavage and re-cyclization reactions, but the biochemical details are not well understood. Type III PKS systems were previously mainly known in plants, but have recently been found in many bacterial genomes. They consist of monodimeric multifunctional proteins that perform iterative condensations often using larger precursor molecules. Type III PKS

typically produce monocyclic compounds. So far no marine natural products have been shown to be derived from Type III PKS.

Non-ribosomal peptide synthetases (NRPS) are functionally related to modular PKS. Instead of polyketides from acyl units, peptides are assembled from amino acids. NRPS are also organized in modules, a minimal NRPS module consists of an adenylation domain (A), condensation domain (C) and peptidyl carrier protein domain (PCP). The A domain activates a specific amino acid (analogous to a t-RNA) and transfers it to the PCP which holds on to the growing peptidyl as a thioester. The C domain forms a peptide bond between the next amino acyl and the peptidyl unit. There can be modifying domains for epimerization and heterocyclization [45] (Figure 1.2).

In both PKS and NRPS the product is released by a thioesterase. PKS and NRPS modules can mix in a single biosynthetic pathway (and even gene), this leads to hybrid polyketide/peptide compounds like onnamide A (Figure 1.1). Mixed PKS/NRPS are quite abundant in cyanobacteria [19].

Peptide natural products can also be synthesized in the standard ribosomal mode. This is the standard route to peptides in eukaryotes, for example conotoxins in cone snails [13], but also found in bacteria, exemplified by the patellamide/trichamide pathway discussed in chapter 4. A precursor gene is translated into a precursor peptide containing the amino acid sequence of the mature compound. Single amino acid

residues can be modified by dedicated enzymes and proteases cut the final product out of the precursor peptide.

### Examples of symbiotic natural products

The following examples highlight some well-studied symbiotic systems.

Onnamide A is polyketide/peptide hybrid from the sponge *Theonella swinhoei* with antiviral and antitumor activities [17]. The putative biosynthetic cluster has been characterized by its similarity to the pederin biosynthesis cluster [42]. Pederin is a defensive compound in the *Paederus* beetle family. It prevents predation of the beetle larvae by wolf spiders [32]. There is strong evidence from antibiotic treatment studies as well as structural considerations that pederin is produced by a  $\gamma$ -proteobacterial symbiont found in female *Paederus* beetles [31] through a polyketide synthase pathway. The putative onnamide cluster also consists of PKS/NRPS genes and surrounding genes are clearly bacterial [42]. This suggests the presence of a symbiotic bacterium in *Theonella swinhoei* that produces onnamide, although a candidate bacterium has not been identified and the ecological function of onnamide A is unclear at this point.

Bryostatin is a macrocyclic polyketide isolated from the marine bryozoan *Bugula neritina* [41], and the subject of chapter 2. It has been shown to be a feeding deterrent in *B. neritina* larvae [37]. Bryostatin has shown activity against several human cancers and over the last decade numerous clinical trials have been performed

[39]. Bryostatin specifically binds to protein kinase C and modulates the enzyme's activity. Recently, neurological effects of bryostatin have been documented. It enhanced long term memory in mollusks [1] and countered depression and dementia in rats [46]. Due to its chemical structure and the response to antibiotic treatment described above a bacterial origin of bryostatin is likely. A symbiotic  $\gamma$ -proteobacterium *Candidatus Endobugula sertula* (*E. sertula*) has been identified in *B. neritina*. *E. sertula* has been shown to be consistently associated with *B. neritina* [22], it is transferred vertically to a new bryozoan generation in a specialized groove on the *B. neritina* larva. A modular PKS gene *bryA* has been identified [24] and shown to originate in *E. sertula* by fluorescence in situ hybridization [6]. *BryA* contains four modules consistent with the first steps of bryostatin biosynthesis. The first module is the loading module. It contains DH and KR-like domains as well as the FkbH domain, which in other PKS system is proposed to form a methoxymalonate extender unit from a glycolytic pathway intermediate [3, 52]. The loading module is hypothesized to provide the unusual starter unit D-lactate. The three subsequent modules perform standard polyketide elongation reactions. The resulting nine carbon compound folds to form the southernmost part of bryostatin (Figure 1.3). Transcription of *bryA* has been shown by reverse transcriptase PCR on environmental RNA isolated from *B. neritina* samples [24]. *bryA* is part of a larger PKS gene cluster, the complete biosynthetic scheme is described in Chapter 2. *B. simplex*, a close relative of *B. neritina* also contains a specific symbiont, *Endobugula glebosa* and a PKS gene fragment homolog to the *bry* genes has been isolated [34]. There are indications for a bryostatin-like molecule in *B. simplex* that warrant further study.

The patellamides are cyclic peptides isolated from tropical didemnid ascidians. These ascidians have an obligate cyanobacterial symbiont *Prochloron* sp. An advantage of studying this system is the easy separation of symbiont and host. Mechanical pressure on the ascidian leads to the expulsion of highly enriched *Prochloron* sp. [33]. The patellamide biosynthesis cluster has been characterized from *Prochloron* sp. Biosynthesis is ribosomal via a microcin-like pathway. The amino acid sequence of mature patellamide could therefore be located in a small open reading frame after searching the translated version of the draft *Prochloron didemni* genome [44]. This precursor protein, PatE, is proposed to be modified by the products of surrounding genes, leading to the mature product patellamide. In this system there is direct evidence for bacterial production of the secondary metabolite: The pat gene cluster has been heterologously expressed in *Escherichia coli* and the production of authentic patellamide has been observed. The patellamides are mildly cytotoxic and are dispersed throughout the ascidian tunic, but patellamide C does not significantly reduce feeding by coral reef fishes [35], so the ecological function is unclear at present. Didemnid ascidians contain overlapping families of peptides (patellamides, mollamides, ulithiacyclamides and others [8, 16]). The ribosomal mode of biosynthesis suggests a straightforward mechanism to achieve this diversity. Point mutations in the precursor peptide gene *patE* would lead to altered amino acid sequences in the mature compound. So far, twenty-six versions of *patE*, differing only in the patellamide-coding region have been identified [11]. Homologs to the patellamide biosynthesis genes have been identified in phylogenetically diverse

bacteria (a *Streptomyces*, a chlorobium, a chlostridium and a *Rhizobium* species).

Another homolog from the cyanobacterium *Trichodesmium erythraeum* is described in chapter 4.

### Symbiosis and biosynthesis

Symbiotic systems are attractive for studying bacterial biosynthesis. The animals form a closed system in which the natural products accumulate into detectable quantities. Once a product is characterized, the finite (albeit in the case of sponges diverse) bacterial community can be searched for the biosynthetic source. It can be technically challenging to isolate the symbiotic DNA of interest from the host, if the bacteria are not abundant. Sometimes certain life stages of the host can contain higher symbiont level, differential centrifugation combined with competitive PCR has also been employed to enhance the symbiont content of DNA preparations [23]. Another option is a metagenomic approach. If the nature of the candidate biosynthetic pathway is known, the entire set of biosynthetic gene can be catalogued and scanned for potential candidates for the pathway of interest. This approach has been successful for onnamide A [42], but its pitfalls are evident in a failed attempt to isolate the discodermolide cluster from the sponge *Discodermia dissoluta*: no characteristic feature of the target cluster was known and the bacterial community was very diverse [43]. In the foreseeable future advances in DNA sequencing techniques (“454 synthesis-based sequencing”) should make obtaining the metagenome of all bacterial symbionts of an invertebrate a standard approach. The difficulty will then lie in

cataloguing all available clusters and determining which cluster is responsible for the natural product of interest.

In symbioses with vertical transmission, examining a group of related animals, their bacteria and bioactive metabolites allows insights into the evolution of biosynthetic pathways. Animals can readily be taxonomically classified and relationship between species assigned based on both morphological as well as genetic characters. In addition, some groups have a fossil record that can provide a timeline. The same relationship can then be proposed for the symbiotic bacteria as well as their biosynthetic pathways. This allows to elucidate evolutionary events affecting bioactive metabolites. The diversity of bacteria can be examined directly by genetic means (typically using the 16S rRNA gene). Comparing the phylogeny of bacteria and host animal is also instructive. Cospeciation (also called cophylogenetic descent) is the parallel and codependent speciation of host and symbiont. It leads to congruent trees of host and symbiont if the symbiont tree is superimposed on the host tree [27] (Figure 1.4). The congruence can be imperfect because of horizontal switching of host-specific symbionts between unrelated host lineages, symbiont duplication, i.e. speciation of symbionts within a non-speciating host lineage or sorting events, which include symbiont extinction and failure of a symbiont to colonize a new host lineage [28].

#### Significance of biosynthetic studies

Marine invertebrates often yield only miniscule amounts of a natural product. While a few mg of compounds are enough for modern structural elucidations,

typically 100s of mg are needed for bioactivity and early clinical studies. Gram quantities are required to supply an actual drug. This supply issue impedes the development of many compounds from marine invertebrates. Since direct routes like large-scale collection or aquaculture are typically not feasible or environmentally sound, indirect routes have been employed or envisioned. Total chemical synthesis has obvious advantages. Once developed it provides a secure and ideally scalable supply of compound. In the case of discodermolide multigram quantities of authentic compound were produced on an industrial scale [15]. The practicality of chemical synthesis may be problematic for complex natural products with multiple stereocenters. A possible alternative solution (that has yet to be successfully applied on a useful scale) is the production of compounds through heterologous expression of its biosynthetic pathway in an easily manageable host. This involves the identification of the biosynthetic gene cluster, the cloning into a suitable expression vector and expression in a host amenable to large-scale fermentation. As the technology becomes more widely available, *in-vitro* gene synthesis to accommodate host codon usage can be employed instead of cloning of the native genes.

Two studies in the didemnid-Prochloron system have used this approach [36, 44]: The patellamide biosynthesis gene cluster was cloned into *E. coli*. *E. coli* is easily manipulated and can be used in large-scale fermentation. The yields reported are low ( $\mu\text{g}$  of compound/liter of culture) and the patellamides presently have little clinical value, but a proof of concept has been achieved. Heterologous expression is much more difficult to achieve with PKS/NRPS pathways due to their size, the number of



catalytic steps and the need for accessory enzymes. In terrestrial *Streptomyces sp.* several compounds have been produced heterologously, a classical example is 6-deoxyerythronolide B, the aglycone part of the antibiotic erythromycin [29]. These experiments have been facilitated by the fact that closely related streptomyces species could be used as hosts, which may not be the case in some marine systems (for example, the putative bryostatin cluster comes from a  $\gamma$ -proteobacterium).

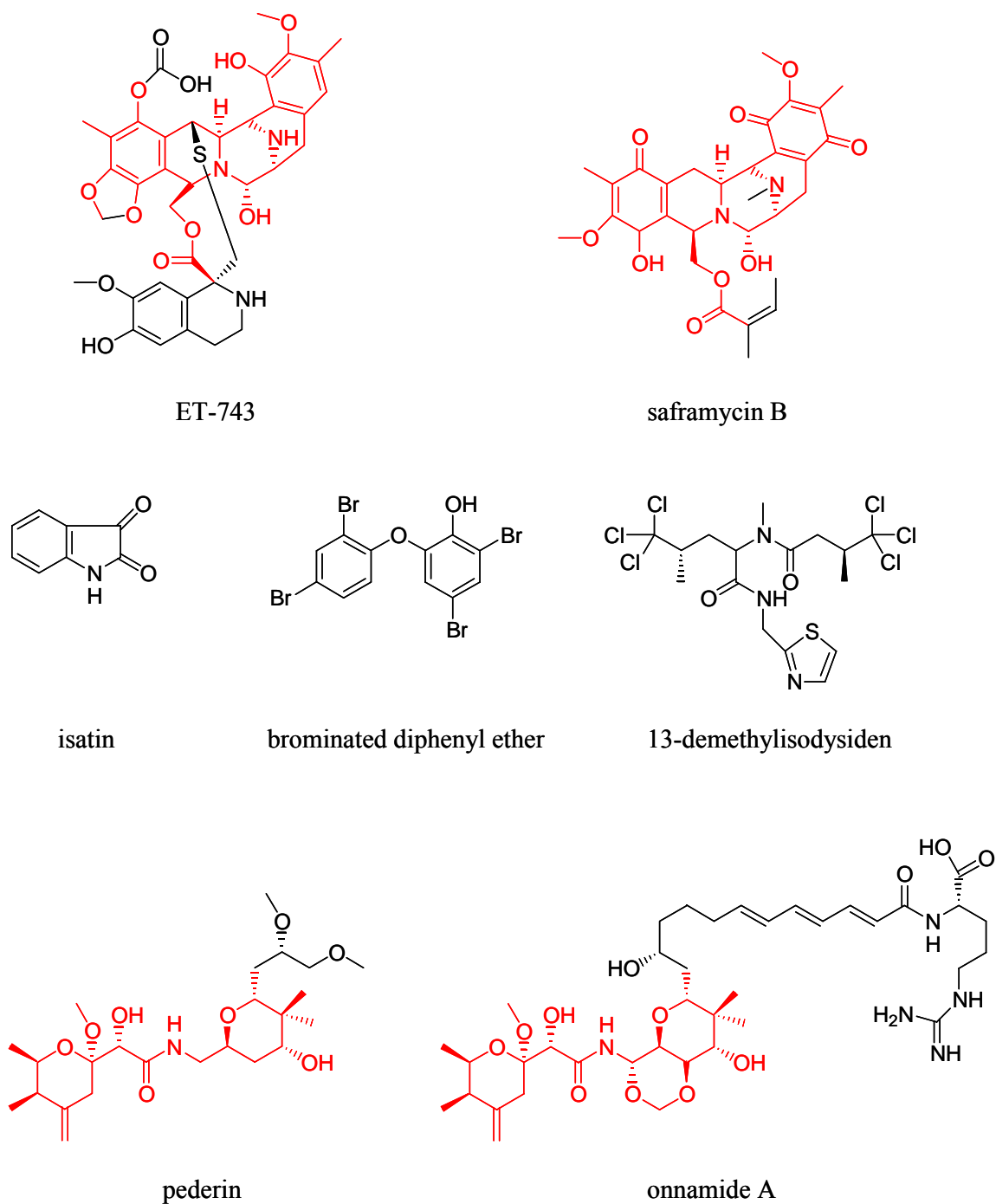
In addition to the practical application of supplying compounds, the study of biosynthesis pathways also gives valuable insight into novel biochemical reactions and enzymes. For example, halogenation is a common theme in marine natural products. Barbamide, a molluscicidal compound isolated from the filamentous cyanobacterium *Lyngbya majuscula* contains a unique trichloromethyl group [40]. This remarkable reaction is performed by two non-heme halogenases in the barbamide biosynthetic cluster. It requires oxygen,  $\alpha$ -keto glutarate and chloride and proceeds via a carbon radical created through hydrogen atom abstraction by a high-valent oxoiron species [18].

Biosynthesis studies also open the possibility of combinatorial approaches by dissecting and engineering PKS and NRPS pathway leading to “unnatural” natural products. This approach is more difficult than anticipated because inter-modular interactions are important for efficient catalysis and presently poorly understood [38]. Once advances have been made, presumably in the most studied terrestrial

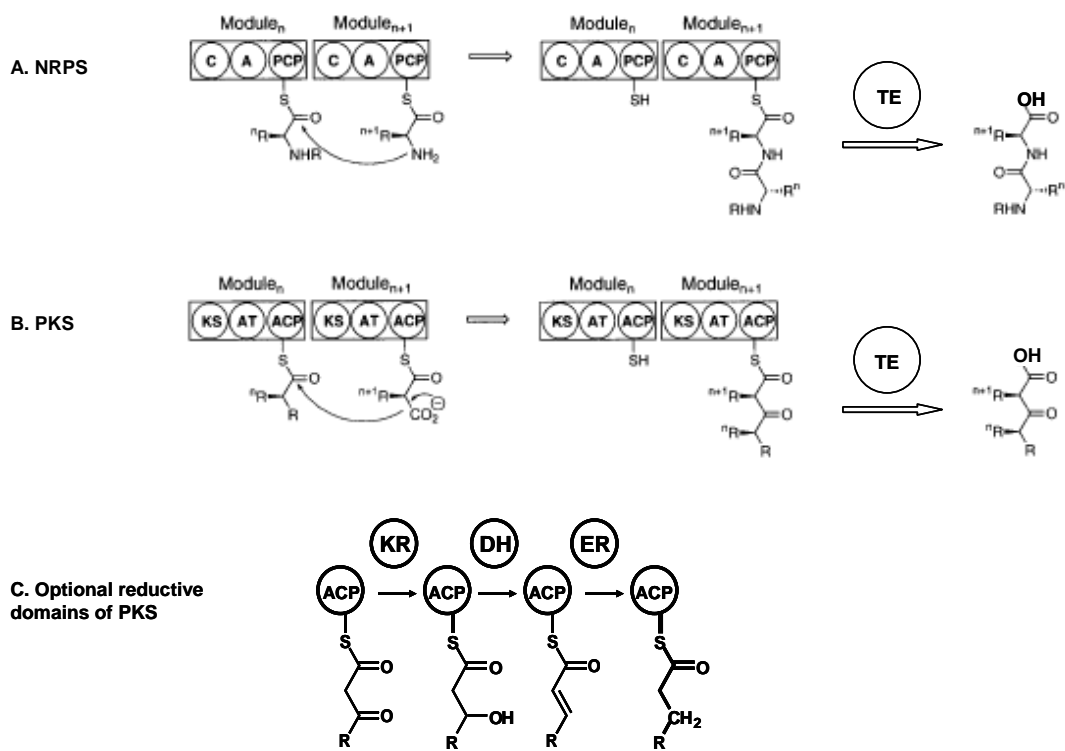
*Streptomyces sp.*, they may be applicable to more exotic biosynthetic genes from marine organisms.

Finally, phylogenetic analysis allows insights into the evolution of bioactive metabolite biosynthesis. Terrestrial actinomycetes are a prolific source of natural products. Continuing advances in the Fenical lab suggest that marine actinomycetes will become another treasure trove [5, 51]. Yet, the discovery of biosynthetic genes in diverse bacteria continues to blur the line between natural product producing versus non-producing bacteria. Clearly homologous clusters are found in diverse bacteria. For example one of the closest relative of the *bry* cluster from the  $\gamma$ -proteobacterium *E. sertula* is found in the gram-positive *Bacillus subtilis*. This suggests that horizontal gene transfer of natural product biosynthesis clusters is possible.

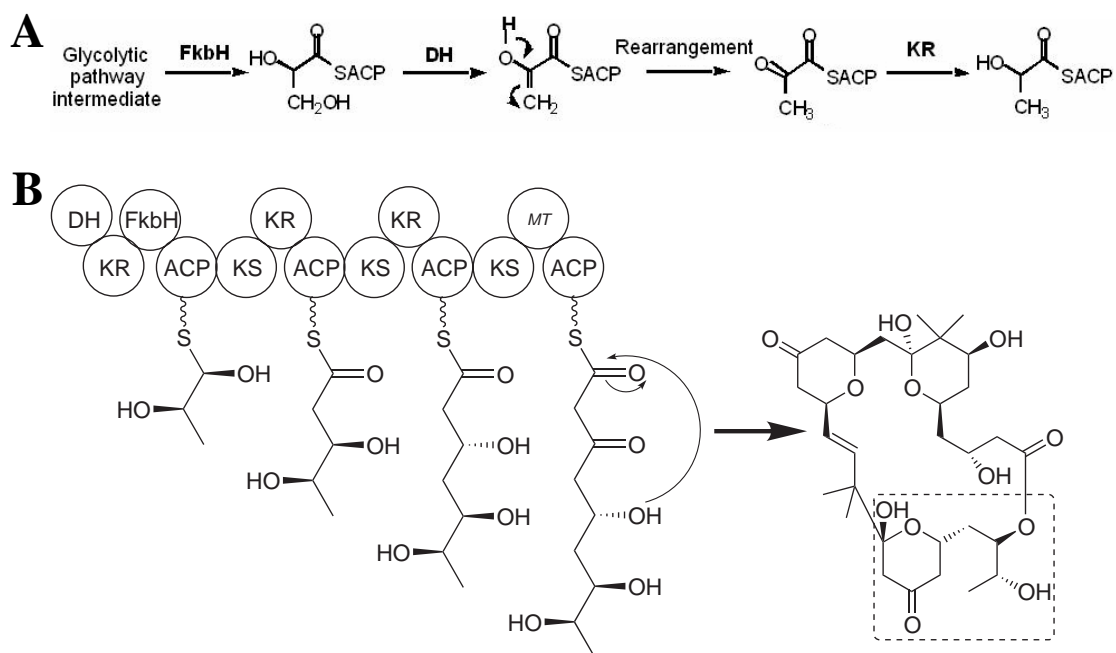
The research presented in this dissertation explores three aspects of natural product biosynthesis in three different systems (Figure 1.5). In chapter 2 the putative biosynthetic gene cluster for the production of the polyketide bryostatin by a symbiotic bacterium of the source organism *Bugula neritina* is presented. Chapter 3 is a phylogenetic analysis of *Prochloron sp.*, the cyanobacterial symbiont of tropical ascidians. *Prochloron* has been shown to produce a variety of peptides in different ascidians, but it proved to be genetically homogenous. Chapter 4 describes the discovery by genome mining and structural elucidation by mass spectrometry of a novel cyclic peptide in the free-living cyanobacterium *Trichodesmium erythraeum*.



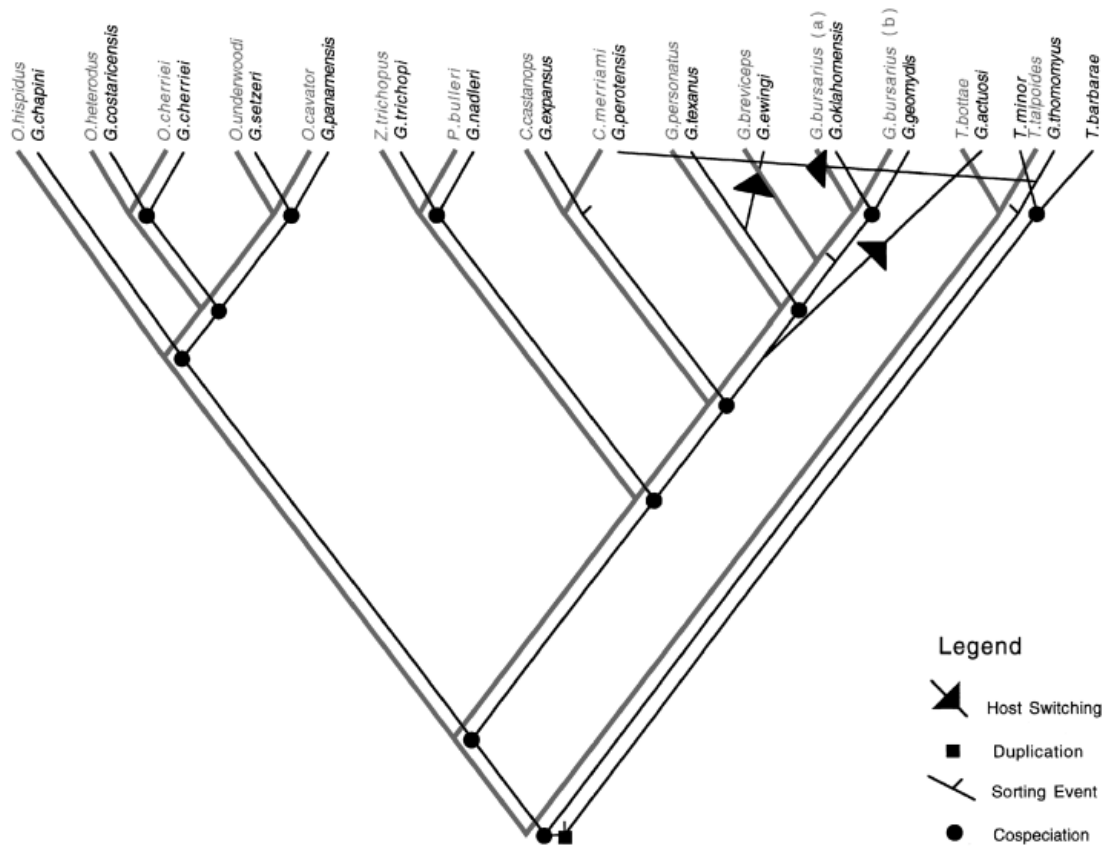
**Figure 1.1:** Structures of selected natural products. Conserved features between ET-743 and saframycin B as well as pederin and onnamide A are highlighted in red.



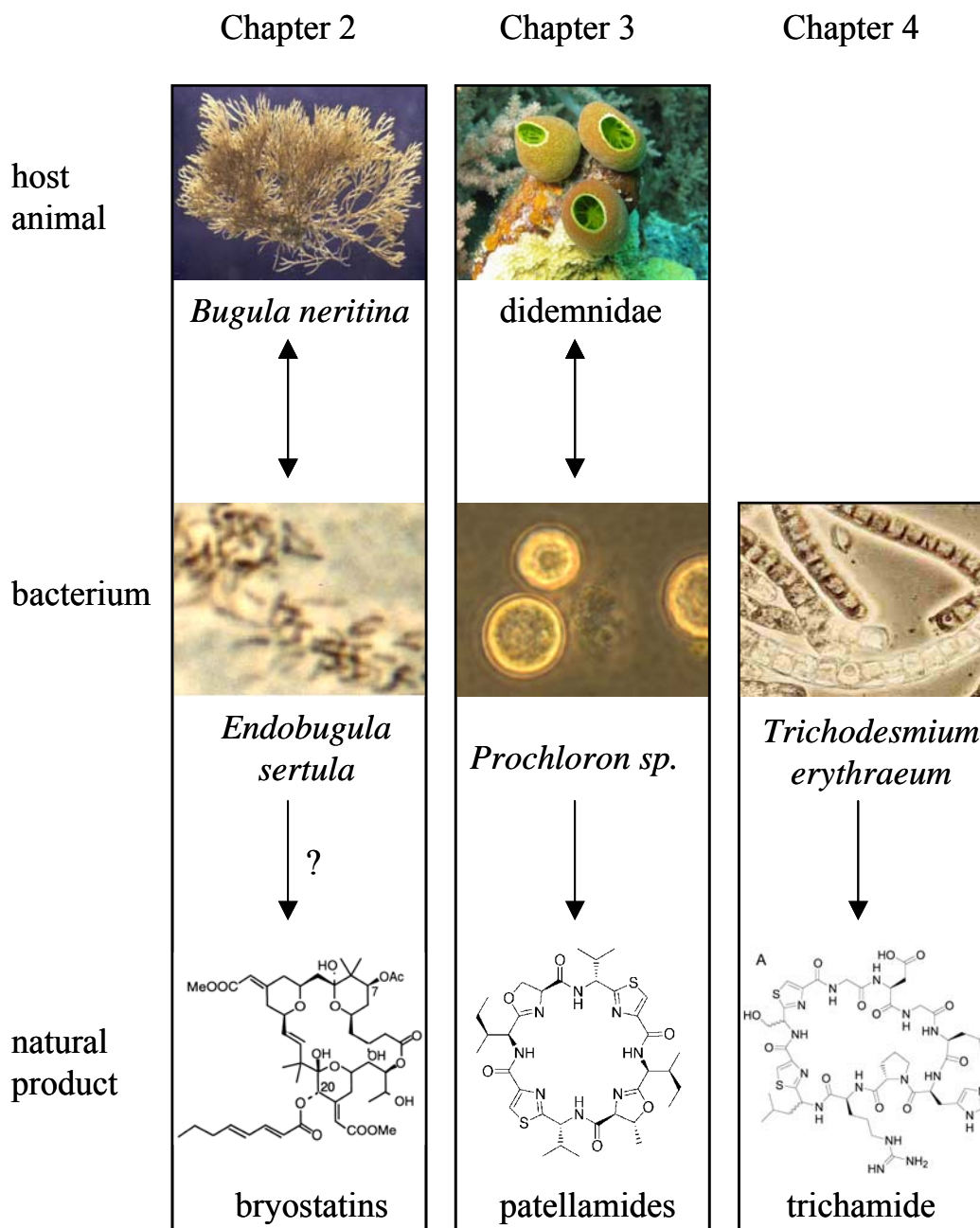
**Figure 1.2:** Modular organization of non-ribosomal peptide synthases A. and polyketide synthases B. PKS reductive steps are shown in C., each module does not have the full set of these domains leading to partially reduced  $\beta$ -keto groups. Figure modified from [12, 30].



**Figure 1.3:** The putative first steps of bryostatin biosynthesis, from [24]. (A) Proposed pathway for D-lactate formation by the loading module. (B) The proposed series of reactions catalyzed by BryA. Curved arrow indicates lactonization. The dashed box depicts the part of bryostatin that BryA synthesizes.



**Figure 1.4:** Classic example of congruent phylogenetic trees in the symbiotic association of gopher and their lice, from [28].



**Figure 1.5:** Overview of dissertation chapters. Photos: *B. neritina* by Koty Sharp, SIO. *E. sertula* from [6]. Didemnidae (*Didemnum molle*) by Chris Ireland, UU. *T. erythraeum* by Dave Caron, USC.

## References

1. Alkon, D. L., H. Epstein, A. Kuzirian, M. C. Bennett, T. J. Nelson. 2005. Protein synthesis required for long-term memory is induced by PKC activation on days before associative learning. *Proc Natl Acad Sci USA* 102: 16432-7.
2. Bewley, C. A., N. D. Holland, D. J. Faulkner. 1996. Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. *Experientia* 52: 716-722.
3. Carroll, B. J., S. J. Moss, L. Bai, Y. Kato, S. Toelzer, T. W. Yu, H. G. Floss. 2002. Identification of a set of genes involved in the formation of the substrate for the incorporation of the unusual "glycolate" chain extension unit in ansamitocin biosynthesis. *J Am Chem Soc* 124: 4176-7.
4. Cheng, Y.-Q., G.-L. Tang, B. Shen. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc Natl Acad Sci USA* 100: 3149-3154.
5. Cho, J. Y., H. C. Kwon, P. G. Williams, C. A. Kauffman, P. R. Jensen, W. Fenical. 2006. Actinofuranones A and B, polyketides from a marine-derived bacterium related to the genus streptomyces (actinomycetales). *J Nat Prod* 69: 425-428.
6. Davidson, S., S. W. Allen, G. E. Lim, C. M. Anderson, M. G. Haygood. 2001. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "*Candidatus Endobugula sertula*" of the bryozoan *Bugula neritina*. *Appl Env Microbiol* 67: 4531-4537.
7. Davidson, S. K. Biology of the Bryostatins in the marine bryozoan *Bugula neritina*: Symbiosis, cryptic speciation and chemical defense [PhD]. La Jolla: University of California San Diego; 1999.
8. Degnan, B. M., C. J. Hawkins, M. F. Lavin, E. J. McCaffrey, D. L. Parry, A. L. van den Brenk, D. J. Watters. 1989. New Cyclic Peptides with Cytotoxic Activity from the Ascidian *Lissoclinum patella*. *J Med Chem* 32: 1349-1354.
9. Distel, D. L., E. DeLong, J. B. Waterbury. 1991. Phylogenetic characterization and *in situ* localization of the bacteria symbiont of shipworms (Teredinidae: Bivalvia) by using 16S rRNA sequence analysis and oligonucleotide probe hybridization. *Appl Environ Microbiol* 57: 2376-2382.
10. Donadio, S., M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz. 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science* 259: 675-679.



11. Donia, M., S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel, E. W. Schmidt. 2006. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. submitted.
12. Du, L., C. Sanchez, B. Shen. 2001. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab Eng* 3: 78-95.
13. Duda, T. F., S. R. Palumbi. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci USA* 96: 6820-6823.
14. Faulkner, D. J. 2002. Marine Natural Products. *Nat. Prod. Rep.* 19: 1-49.
15. Freemantle, M. 2004. Scaled-up synthesis of discodermolide. *C & E news* 82: 33-35.
16. Fu, X., T. Do, F. J. Schmitz, V. Andrushevich, M. H. Engel. 1998. New cyclic peptides from the ascidian *Lissoclinum patella*. *J Nat Prod* 61: 1547-1551.
17. Fusetani, N., S. Matsunaga. 1993. Bioactive sponge peptides. *Chem Rev* 93: 1793-1806.
18. Galonic, D. P., F. H. Vaillancourt, C. T. Walsh. 2006. Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. *J Am Chem Soc* 128: 3900-1.
19. Gerwick, W. H., L. T. Tan, N. Sitachitta. 2001. Nitrogen-containing metabolites from marine cyanobacteria. *Alkaloids Chem Biol* 57: 75-184.
20. Gil-Turnes, M. S., M. E. Hay, W. Fenical. 1989. Symbiotic marine bacteria chemically defend crustacean embryos from a pathogenic fungus. *Science* 246: 116-118.
21. Hamann, M. T., C. S. Otto, P. J. Scheuer. 1996. Kahalalides: bioactive peptides from a marine mollusk *Elysia rufescens* and its algal diet *Bryopsis sp.* *J Org Chem* 61: 6594-6600.
22. Haygood, M., S. Davidson. 1997. 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.

23. Hildebrand, M., L. E. Waggoner, G. E. Lim, K. H. Sharp, C. P. Ridley, M. G. Haygood. 2004. Approaches to identify, clone and express symbiont bioactive metabolite genes. *Nat Prod Rep* 21: 122-142.
24. Hildebrand, M., L. E. Waggoner, H. Liu, S. Sudek, S. Allen, C. Anderson, D. H. Sherman, M. Haygood. 2004. *bryA*: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. *Chem Biol* 11: 1543-52.
25. Hopwood, D. A. 1997. Genetic contributions to understanding polyketide synthases. *Chem Rev* 97: 2465-2498.
26. Hopwood, D. A., D. H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* 24: 37-66.
27. Huelsenbeck, J. P., B. Rannala, Z. H. Yang. 1997. Statistical tests of host-parasite cospeciation. *Evolution* 51: 410-419.
28. Johnson, K. P., D. M. Drown, D. H. Clayton. 2000. A data based parsimony method of cophylogenetic analysis. *Zool scr* 30: 79-87.
29. Kao, C. M., L. Katz, C. Khosla. 1994. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* 265: 509-512.
30. Katz, L. 1997. Manipulation of modular polyketide synthases. *Chem Rev* 97: 2557-2575.
31. Kellner, R. L. 2002. Molecular identification of an endosymbiotic bacterium associated with pederin biosynthesis in *Paederus sabaesus* (Coleoptera : Staphylinidae). *Insect biochem mol biol* 32: 389-395.
32. Kellner, R. L., K. Dettner. 1996. Differential efficacy of toxic pederin in deterring potential arthropod predators of *Paederus* (Coleoptera: Staphylinidae) offspring. *Oecologia* 107: 293-300.
33. Lewin, R. A., L. Cheng. *Prochloron, a microbial enigma.*: Chapman & Hall; 1989.
34. Lim, G. E., M. G. Haygood. 2004. "*Candidatus Endobugula glebosa*", a specific symbiont of the marine bryozoan *Bugula simplex*. *Appl Env Microbiol* 70: 4921-4929.
35. Lindquist, N., M. E. Hay, W. Fenical. 1992. Defense of ascidians and their conspicuous larvae: adult vs. larval chemical defense. *Ecol Monog* 62: 547-568.

36. Long, P. F., W. C. Dunlap, C. N. Battershill, M. Jaspars. 2005. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. *Chembiochem* 6: 1760-1765.
37. Lopanik, N., N. Lindquist, N. Targett. 2004. Potent cytotoxins produced by a microbial symbiont protect host larvae from predation. *Oecologia* 139: 131-139.
38. Menzella, H. G., R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood, D. V. Santi. 2005. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* 23: 1083-1084.
39. Mutter, R., M. Wills. 2000. Chemistry and Clinical Biology of the Bryostatins. *Bioorg and Med Chem* 8: 1841-1860.
40. Orjala, J., W. H. Gerwick. 1996. Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 59: 427-430.
41. Pettit, G. R., C. L. Herald, D. L. Doubek, D. L. Herald. 1982. Isolation and structure of bryostatin 1. *J Am Chem Soc* 104: 6846-6847.
42. Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga. 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci USA* 101: 16222-16227.
43. Schirmer, A., R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, C. R. Hutchinson. 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl Environ Microbiol* 71: 4840-4849.
44. Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, J. Ravel. 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.
45. Schwarzer, D., R. Finking, M. A. Marahiel. 2003. Nonribosomal peptides: from genes to products. *Nat Prod Rep* 20: 275-287.
46. Sun, M. K., D. L. Alkon. 2005. Dual effects of bryostatin-1 on spatial memory and depression. *Eur J Pharmacol* 512: 43-51.

47. Unson, M. D., D. J. Faulkner. 1993. Cyanobacterial symbiont biosynthesis of chlorinated metabolites from *Dysidea herbacea* (Porifera). *Experientia* 49: 349-353.
48. Unson, M. D., N. D. Holland, D. J. Faulkner. 1994. A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar Biol* 119: 1-11.
49. Van Dover, C. L. The ecology of deep-sea hydrothermal vents. Princeton, NJ: Princeton Univ. Press; 2000.
50. Waterbury, J. B., C. B. Calloway, R. D. Turner. 1983. A cellulolytic nitrogen-fixing bacterium cultured from the Gland of Deshayes in shipworms (Bivalvia: Teredinidae). *Science* 221: 1401-1403.
51. Williams, P. G., G. O. Buchanan, R. H. Feling, C. A. Kauffman, P. R. Jensen, W. Fenical. 2005. New cytotoxic salinosporamides from the marine actinomycete *Salinispora tropica*. *J Org Chem* 70: 6196-6203.
52. Wu, K., L. Chung, W. P. Revill, L. Katz, C. D. Reeves. 2000. The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* 251: 81-90.

## 2.

### **Identification of the *bry* polyketide synthase gene cluster from *E. sertula*: the putative biosynthetic pathway for bryostatin**

#### **Abstract**

The bryostatins are protein kinase modulators with unique structural features and potential anti-cancer and neurological activities. They were isolated from the marine bryozoan *Bugula neritina*, but we suggest they are produced by a symbiotic bacterium, *Candidatus Endobugula sertula* (*E. sertula*). Here we present the putative biosynthetic genes: five modular polyketide synthase (PKS) genes, a discrete acyl transferase, a  $\beta$ -ketosynthase, a hydroxy-methyl-glutaryl (HMG) CoA synthase and an O-methyl transferase. The cluster was characterized in two *E. sertula* strains from different hosts. In one strain the gene cluster is contiguous, in the other strain it is split into two loci. One locus contains the PKS genes, the other the accessory genes. Here, we propose a hypothesis for the biosynthesis of the bryostatins. Using malonyl-CoA as the extender unit, thirteen PKS modules form the macrolactone ring. The methyl ester groups are added by an HMG gene cassette. The resulting hypothetical compound bryostatin 0 is the common basis for the 20 known bryostatins. As *E. sertula* is to date uncultured, the heterologous expression of this biosynthetic gene cluster represents the most viable method of producing the bioactive bryostatins in large enough quantities for development into a pharmaceutical.

## Introduction

Marine invertebrates, particularly sessile ones lacking physical defenses, are rich sources of bioactive compounds. The number of compounds characterized from marine sources has increased significantly in recent years [11]. Often, symbiotic microorganisms have been proposed as the true producers due to structural similarities with known microbial compounds and in a few cases convincing arguments for marine compounds produced by symbiotic bacteria have been made [5, 29, 31]. Many bioactive compounds have strong cytotoxic, anti-inflammatory, antimicrobial or antiviral activity, but the development into usable drugs is typically hampered by the scarcity of source material and/or the exceedingly low concentration of the compound. Large-scale isolation of compounds from natural sources is often environmentally detrimental and cost-prohibitive. One can envision four possible routes to provide sufficient amount of compounds: (i) aquaculture of the source animal, (ii) direct culturing of the symbiont, (iii) total chemical synthesis, and (iv) heterologous expression of biosynthetic genes in a suitable host. All four methods have potential benefits and drawbacks and the solution will presumably be different for different compounds.

The bryostatins are cyclic polyketides isolated from the marine bryozoan *Bugula neritina* [26, 27]. Their common feature is a 25-membered macrolactone ring with three component pyran rings (Figure 2.1). Bryostatin 1 has activity against a

variety of cancer cell lines ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)). It is in a number of phase II/III clinical trials alone and in combination with other drugs [3, 23, 25]. The effects of bryostatin 1, which modulates the activity of protein kinase C, is markedly different in different cell lines (reviewed in [25]). More recently bryostatin 1 has shown beneficial effects in countering depression and dementia in a rat model system [32] as well as enhancing long-term memory in a mollusk species [1]. One of the ecological roles of the bryostatins is to protect the host larvae from predators [19].

Complex polyketides such as the bryostatins are synthesized by modular polyketide synthases (PKS). PKSs are large multi-modular enzymes that elongate and modify the nascent polyketide in an assembly-line fashion. Each module performs one specific elongation step by adding an acyl-CoA precursor and then transferring the polyketide chain to the next module. Each module has several catalytic domains. A canonical core module consists of a  $\beta$ -ketoacyl synthase (KS), an acyl transferase (AT) and an acyl carrier protein (ACP) domain. The AT selects and loads the extender unit onto the ACP domain. The KS condenses the nascent polyketide received from the ACP of the previous module onto the acyl-CoA extender unit, effectively adding two carbons to the polyketide. In most characterized PKSs, the AT functionality is integrated into each PKS module, but in several systems it resides on a discrete and separate gene and there is no AT domain present in the PKS genes [4]. After the condensation step the ACP bridges the distance to the next module's KS domain with a phosphopantetheine arm to transfer the polyketide for the next

extension step. Each module can contain different accessory domains. After extension, the  $\beta$ -keto group can be reduced to a hydroxyl group by a ketoreductase (KR) domain, to a trans-double bond by a dehydratase (DH) domain and to the fully saturated  $\beta$ -carbon by an enoyl reductase (ER) domain [8, 18]. A number of other domains are occasionally found in PKS modules for instance methyl groups are added to the carbon backbone of yersiniabactin and epothilone by methyl transferase (MT) domains [22, 24]. While in most cases to date co-linearity between the order of modules on the gene and the order of biosynthetic reactions necessary was found, there are examples where domains or entire modules are skipped or used more than once [12, 38, 39].

Modular PKSs are typically only found in bacteria, making a bacterial symbiont of *B. neritina* a prime candidate for the true biosynthetic source of the bryostatins. The Haygood group has previously identified such a bacterium “*Candidatus Endobugula sertula*” (*E. sertula*) that is consistently associated with *B. neritina*, hasn’t been found elsewhere in the surrounding water column, and is transferred vertically between generations [14]. Elimination of *E. sertula* with antibiotics leads to *B. neritina* larvae in the next generation without bryostatin [5, 19].

There are three recognized sibling species of *Bugula neritina* [6, 20], “deep”, “shallow” and Northern Atlantic, each containing a strain of the symbiont *E. sertula*.



The species vary in their bryostatin composition. Of the 20 described bryostatins, most diversity is found in the C-7 and C-20 position (Figure 2.1). The “deep” species is normally found below ~10 meters and contains bryostatins with an octa-2,4-dienoate substituent at C-20 (among them the clinically relevant bryostatin 1), the “shallow” species, found above ~10 meters, does not contain bryostatins with this substituent. The third species found in the Northern Atlantic, has not been examined for bryostatin content. We hypothesize that a common precursor to all bryostatins is produced by a PKS in *E. sertula*.

A PKS gene fragment has been identified and localized to *E. sertula* cells [5]. More recently, an entire PKS gene, *bryA*, has been cloned from *E. sertula*. *BryA* has four PKS modules consistent with the early steps of bryostatin biosynthesis [16]. Here we present the entire ~80 kbp gene cluster proposed to code for the biosynthetic machinery to make a common precursor of the bryostatins.

## Materials and Methods

**Sample collection and genotype determination.** Samples of *Bugula neritina* were collected by SCUBA diving from three sites along the coast of San Diego CA, USA; Mission Bay (water depth ~5 m, sample not genotyped, but previous collections determined to be “shallow”), Scripps Pier (~ 7 m, CA “shallow”), Torrey Pines Artificial Reef II (~15 m, CA “deep”). A fourth sample was collected from

Radio Island Jetty in Morehead City, North Carolina (~7 m). The genotype of collected specimens was determined as described previously via a sequence polymorphism in the 16S gene [6].

**DNA isolation, cloning and sequencing.** All procedures to enrich *B. neritina* DNA preparations for *E. sertula* as well as the construction of the SuperCos I cosmid, lambda-ZAP and lambda-DASHII (Stratagene) libraries were described previously [5, 16]. Briefly, since DNA preparations from *B. neritina* contain only minor amounts of *E. sertula* DNA, different (ultra)centrifugation techniques are employed to enrich for the symbiont DNA. All libraries were constructed from partially digested genomic DNA according to manufacturer's protocols. Four SuperCos clones spanned most of the *bry* cluster, leaving two gaps, however due to stability issues, the entire cluster was re-cloned into lambda-DASHII (insert size ~15 kb) for CA "shallow" and lambda-ZAP (insert size ~5 kb) for CA "deep" and screened using cosmid fragments as probes. 32 lambda-DASHII and >100 lambda-ZAP clones were mapped to the *bry* cluster by end sequencing, overlapping clones were picked and fully sequenced by primer walking. Sequencing was performed on ABI3100 sequencers at the UCSD Center for AIDS research and through SeqXcel Inc., San Diego. The CA "shallow" cluster was only spot sequenced (~75% coverage); the CA "deep" cluster was sequenced to 2-4 fold coverage. Reads were assembled with Sequencher (Gene Codes Corp). Three long perfect repeats made assembly challenging. Each repeat was

covered by two overlapping clones, which contained part of the adjacent non-repeat region.

A fosmid library from NC populations of *B. neritina*/*E. sertula* was generated using CopyControl Fosmid Library Preparation kit (Epicentre). The library was probed with portions of KS genes that had been isolated from genomic DNA. Positive fosmids (2) were subcloned into pSMART-LC Kan sequencing vector (Lucigen Corp.) and end-sequenced. Because the whole cluster was not contained on these two fosmids, probes were generated from the 5' and 3' ends of the cluster and used to reprobe the library. The positive fosmids were subcloned and end-sequenced. Sequences were assembled using SeqMan (Lasergene). The fosmids were sequenced to 4-5 fold coverage.

Restriction mapping and Southern blotting were performed according to established protocols as described previously [16].

**RNA isolation and reverse transcription PCR (RT-PCR).** Enriched bacterial fraction RNA was isolated from Scripps Pier ("shallow") and Torrey Pines Artificial Reef ("deep") adult *B. neritina* as described previously [16]. RNA used for RT-PCR was purified on an RNeasy Mini Kit spin column (QIAGEN Inc.), treated

with DNase I (QIAGEN Inc.) as described by the manufacturer, and eluted with 30  $\mu$ l RNase-free water (QIAGEN Inc).

RT-PCR was performed as previously described [16]. Gene specific primers BryX-DH1R (GGC GTT GCC CAG GCA ATA TGT TGC) or BryPR (ACG TGA ATG AAA GGC AGC GC) were used to generate cDNA. PCR was performed on the cDNA with the corresponding primer pair, BryX-DH1F (GCT TTA CCC TGC TAT CCT TTT GCC) and BryX-DH1R or BryPF (GTG GGC AGG GTT CAC AGC AC) and BryPR. To confirm that RT-PCR products reflected RNA content rather than DNA contamination, control RT-PCRs without reverse transcriptase were conducted for each primer pair.

**Ribonuclease Protection Assay (RPA).** In order to make transcription templates for radio-labeled RNA probes, PCR was performed on “shallow” *B. neritina* DNA with the following *bry* cluster primer pairs: bryBF (GGT GAT GCC AAG ATG ACC ACC GC) and bryBR (CAC GAT CAT GAT TTA AAC GCT G), bryCF (CTA GAT GAT GAC GAT GAG ATT G) and bryXR (CAG ACT GCA ACA TGC GTA AGG C), bryX-DH1F and bryX-DH1R, bryX-KS3F (CCA GCC TTT TGA TCA GTT CAA GTC) and bryX-KS3R (GGG ATC TCC TAG TTT AGT CCC AGT), bryDF (TTC ATT TAT GTG CAG GAC ACA TAC) and bryDR (CCC AGA AGC CGA TTC CAG ATG CCC), and bryDendF (CAG ATA AAC CTA TAG AAG AGA TTG) and bryAR (GTT TTT TTC GGT ATT GTC GAA TGC). *bry*

cluster PCR products were purified from agarose gels with a Quiaex II Kit (QIAGEN Inc.) and cloned into pCR<sup>®</sup>4-TOPO<sup>®</sup> plasmids (Invitrogen) using a TOPO TA Cloning<sup>®</sup> Kit for Sequencing (Invitrogen). Following transformation into TOP10 One Shot Chemically Competent *E. coli* (Invitrogen), growth of transformants and isolation of plasmid DNA with a QIAprep Spin Miniprep Kit (QIAGEN Inc.), plasmids were digested with SpeI or NotI (New England Bio Labs), depending on insert orientation, so that transcription of the plasmid would terminate shortly after the insert region. The digested plasmid was then purified by standard phenol/chloroform extraction and ethanol precipitation. <sup>32</sup>P-labeled RNA probes complementary to *bry* cluster mRNA were transcribed with T3 RNA polymerase or T7 RNA polymerase using a MAXIscript<sup>™</sup> In Vitro Transcription Kit (Ambion) and [<sup>32</sup>P]UTP (Amersham Pharmacia Biotech). In order to maximize the portion of full-length transcripts in the RPA probe, the product of the transcription reaction was run on a denaturing polyacrylamide gel (6% acrylamide, 8M urea, 1X TBE), cut out of the gel and incubated in Probe Elution Buffer (Ambion) overnight at 37°C. 5 µg of enriched bacterial fraction RNA from “shallow” *B. neritina* and 8x10<sup>-4</sup> cpm of labeled RNA probe were used with a Hybspeed RPA Kit (Ambion) for each RPA. In the RPA, the RNA probe hybridized with its complement from the genomic RNA. Upon the addition of RNase A/T1, the single-stranded RNA was degraded leaving only the double-stranded probe-target hybrid, which is protected. The products were run on a denaturing polyacrylamide gel (6% acrylamide, 8M urea, 1X TBE) and exposed to X-ray film at -80°C.

## Results

**The bry gene cluster – Overview.** The construction and screening of “*E. sertula*” enriched *B. neritina* DNA libraries has previously been reported [16]. Briefly, a gene fragment from “*E. sertula*” amplified using degenerate primers for KS domains was used to identify four overlapping cosmids of a *B. neritina* DNA library. Due to instability upon propagation these cosmids could not be characterized. Subcloned fragments of the cosmid inserts were used as probes to repeatedly screen two libraries under high-stringency conditions: a lambda-ZAP library of California “deep, *E. sertula*-enriched” *B. neritina* and a North Carolina “shallow” *B. neritina* fosmid library. The clones identified were sequenced, assembled and annotated. The identified genes cluster with the modular PKS gene *bryA* we reported previously [16] and together constitute the entire putative bryostatin biosynthesis cluster. With one notable exception (see below) the arrangement of the genes is identical between “deep” and “shallow” strains of *E. sertula*. The sequence identity is high (~99%). A draft sequence was also obtained from CA “shallow” and proved to be virtually identical to NC “shallow” (data not shown).

The *bry* gene cluster has two parts: Five genes (*bryA-D* and *X*) totaling 71 kb coding for modular PKSs and four genes (*bryP-S*) totaling 6 kb with accessory functions: an acyl transferase, a discrete  $\beta$ -ketosynthase, a hydroxy-methyl-glutaryl CoA synthase and an O-methyl transferase (Figure 2.2 and table 2.1, function

proposed from BLAST searches). In CA/NC “shallow” *E. sertula* the accessory genes are directly upstream of the large PKS genes but in opposite orientation. In CA “deep” *E. sertula* a transposase lies directly upstream of the PKS, and although the accessory genes are present, their position relative to the PKS is unknown. They are not within ~30 kb upstream or ~5 kb downstream of the PKS genes as determined by draft quality sequencing of these regions (data not shown). There is a second transposase upstream of *bryP-R*. It is interesting to note, that the region directly upstream of *bryB* and *bryP* is extremely AT-rich (~75% over 200 nt) and hence susceptible to genomic rearrangement. Both parts of the cluster have coincident direction of transcription. The PKS genes have very small intergenic regions (0-6 bp) except for *bryX* to *bryA* (134 bp).

A peculiar feature of the PKS genes is the presence of three large (~4 kb) perfect repeats. The first repeat spans the first quarter of *bryB* and *C*, the second repeat is directly adjacent to the first in *bryB*, but about 6kb apart in *bryC*. The third repeat is found in the latter half of *bryD* and *A* (Figure 2.2). None of the repeat regions span genes. While the GC% is generally low at ~40%, it is notably higher in the repeat regions (~49%). The arrangement of genes has been confirmed by a series of Southern Blots and long-range PCRs (data not shown).

Downstream of the PKS cluster we found ~3 kb of truncated ORFs and hypothetical protein genes followed by a glutathione reductase gene, defining one end of the *bry* cluster. The gene density is notably lower in this area compared to the *bry* cluster. In the “deep” CA strain of *E. sertula*, there is a transposase directly upstream of the 5 large PKS genes *bryA-D+X* (Figure 2.2A). Upstream of the transposase there are a number of potential pseudogenes, primary metabolism genes and relatively long intergenic regions, but no homologies to *bryP-S* or the sequences surrounding these genes in the “shallow” *E. sertula* strains. In both strains of *E. sertula*, *bryS* is followed by a series of general oxidoreductase genes of unknown function (Figure 2.2A).

**Domain analysis.** The best-matching homologs for the entire *bry* PKS cluster are the *Bacillus subtilis* *pkzM* genes [17] and the *Paederus fuscipes* symbiont *ped* genes [28]. BlastP [2] analysis on the NCBI server ([www.ncbi.nlm.nih.gov/blast](http://www.ncbi.nlm.nih.gov/blast)) allowed the definition of catalytic domains and modules within the “*E. sertula*” PKS genes. *BryA* contains 4 modules thought to initiate bryostatin biosynthesis and has been described previously [16]. *BryB* and *C*, the first two of the 5 large PKS genes also contain four modules each (M4-11). *bryX* and *D* have two full modules each as well as a number of single domains (see table 2.1). We propose that the first KR-ACP (M11a) and the first complete modules of *bryD* (M12), as well as either the non-ribosomal peptide synthase domain (NRPS) condensation domain in *bryD* or the TE domain in *bryX* are used in bryostatin biosynthesis (Figure 2.1), while the other domains of *bryX* are inactive.



The first repeat in *bryBC* (Figure 2.2A) contains only a single KS domain. The second repeat contains a DH, MT, ACP and KS. The third repeat in *bryDA* contains the C-terminal region of a KS, DH, KR, ACP and the N-terminal region of another KS. Repeats are generally thought of as regions of increased genetic recombination and this domain content could allow “re-shuffling” of the PKS modules (see discussion).

The *bry* PKS genes lack integrated AT domains in each module. *BryP* could function as a discrete acyl transferase bringing a malonyl CoA extender unit to each of the 12 modules. This type of AT mechanism has been proposed in the closely related *ped* cluster and demonstrated biochemically in leinamycin biosynthesis [4]. *BryP* contains two distinct AT domains, the second domain has the complete active site residues GHS...R, in the first domain R is mutated to Q. The specificity motif defined for integrated ATs (YASH for methylmalonyl-CoA extension and HAFH for malonyl-CoA extension [7]) is only present partially in the first domain (HGFH), but closer to the expected extender unit.

***bry* cluster transcription.** Two very sensitive techniques, reverse transcription PCR (RT-PCR) and ribonuclease protection assays (RPAs) and were used to detect transcripts from the *bry* cluster in RNA preparations from *B. neritina*.

The presence of *bryX* transcripts in “shallow” *B. neritina* and *bryP* transcripts in “deep” *B. neritina* was demonstrated with RT-PCR (Figure 2.2B). *bryB*, *bryC*, *bryX*, *bryD* and *bryA* transcripts were detected in shallow type *B. neritina* RNA with RPAs (*bryC* and *X* in Figure 2.2C as examples). These results are consistent with transcription of the entire *bry* cluster in *B. neritina*.

## Discussion

**Bryostatin biosynthesis.** We have shown previously that the *bry* PKS cluster is the only large PKS cluster present in both shallow and deep strains of *E. sertula* [16]. The presence of such large open reading frames in an otherwise degenerate genome (as judged from ~35 kb of DNA sequence surrounding the PKS cluster) and the demonstration of transcription throughout the region suggest that these genes are functional. We propose the following model for the biosynthesis of a bryostatin precursor (Figure 2.3): Biosynthesis starts with *bryA* as described previously [16]. The next four modules (M4-7) come from *bryB*, the first PKS gene in the cluster extending the chain from 9 to 17 carbons. A methyl transferase domain in M4 introduces the first characteristic gem-dimethyl group. The successive use of a single MT domain has also been suggested in pederin [28] and yersiniabactin biosynthesis [22], in the latter it is supported by biochemical evidence. In addition to the MT M4 also contains a DH, which lacks the HxxxGxxxxP signature motif and therefore should be inactive. M9, where the second gem-dimethyl group is introduced (see below) has a similarly modified DH, while M3 and M7, which both also have MT

domains do not. Incidentally in our biosynthetic scheme, we assume that the MT in M3 and M7 are inactive. Therefore the M4 and M9 DH may be involved in the methyl transferase mechanism.

*BryC* also contains four modules (M8-11), which extend the chain to 25 carbons. M8 contains KS-DH-PS-KR-ACP, introducing a second double bond. The pyran synthase (PS) region has 26% identity over ~500 amino acids to a region in module 5 of *pedF* (where it is annotated as a dehydratase domain). Strikingly the dehydratase domain in module 5 of *pedF* and the PS region in module 8 of *bryC* are both in the correct position to catalyze the formation of the tetrahydro-pyran ring that pederin and bryostatin have in common. Analogous to Piel [28], we propose the bryostatin pyran ring from C11 to C15 is formed by a Michael type reaction catalyzed by the newly defined pyran synthase domain. The double bond between C11 and C12 is attacked by the nucleophilic oxygen on C15, leading to an intermediate state with a shifted double bond and a hydroxy group on C9, which can then rearrange to form the functionality found in bryostatin (see box Figure 2.3). In M9 a MT domain introduces the second gem-dimethyl group.

The next gene *bryX* is skipped; its function is unclear (see below). After module 11, the first incomplete module of *bryD* consisting of KR and ACP (M11a) would reduce the beta-keto group left untouched by module 11 proper. This “division

of labor” with one module doing the extension and the next doing the reductive step is very unusual, it could be achieved if there is an equilibrium between the ACP in module 11 and 11a, that allows the transfer of the polyketide directly from one ACP to the next without a keto-synthase step. The next complete module on *bryD* is proposed to do the final extension step.

We propose that the NRPS condensation domain in *bryD* together with the adjacent KS or the TE domain in *bryX* then cyclizes the molecule from C1 to C25. Two 5 carbon pyran rings form by the condensation of a keto and a hydroxyl group from C19 to C23 and C5 to C9. A further maturation step is the addition of two exomethylene groups on C13 and C21. The exomethylene groups are proposed to be synthesized by *bryR* and *S*. The HMG-CoA synthase *bryR* would condense acetate onto the keto groups at C13 and 21, after dehydration the carboxylic acid group would be methylated by the O-Methyl transferase activity of *bryS* (Figure 2.4). A similar function for HMG-CoA synthase homologs has been proposed in jamaicamide [9] and mupirocin [10] biosynthesis. It is interesting to note that during polyketide synthesis, when C13 and C21 carry the  $\beta$ -keto groups, a MT domain is present in the active modules 3 and 7. There is no methyl transfer needed in these steps, however the MT domain could activate the keto group to facilitate the HMG-CS reaction. Alternatively the MT in M3 and M7 could simply be inactive and the maturation steps could happen before or after the polyketide is released from the ACP. The resulting compound bryostatin 0 (Figure 2.4) is hypothetical, but is a plausible

common basis for the 20 known bryostatins. In all bryostatins (with the exception of bryostatin 3, 19 and 20 which have a further lactonization) further modifications are restricted to C7 and C20 (Figure 2.1). The mechanism (or even the physical location) of further maturation into these compounds remains to be investigated.

The model described here implies that the rules for co-linearity between genes and their proteins in PKS can be violated. The order in which biosynthesis happens is different from the gene order, namely *bryA*, the last gene in the cluster is the first gene in the biosynthetic scheme and *bryX* may be skipped. While the presence of an uninterrupted ORF of this size in both strains of *E. sertula* as well as our RT-PCR data shows that this gene is important, the domain content of *bryX* is enigmatic. A few functions required to produce complete bryostatins aren't apparent in the cluster, including the formation of C7/C20 substituents and phosphopantetheinylation but the domain structure of *bryX* does not appear to be appropriate for these functions (except for the TE domain as a potential longer-chain acyl transferase). Also *bryX* is identical in both "deep" and "shallow" *E. sertula*, which consistently have different suites of bryostatins.

The large repeats are unprecedented in PKS genes to our knowledge. They do suggest an attractive mechanism as to how variations in secondary metabolites could be achieved by recombination events. For example, if the second repeat were deleted,

the preceding KS of module 4 would be in the right position to interact with the now orphaned reductive domains of module 5. The other way around, if this repeat were inserted at the right position between a KS and an ACP in a hypothetical module, the newly fused gene would then contain two fully fledged modules, where the KS of the original module would need to interact with the repeat's ACP, and the repeat's KS would interact with the original ACP. The boundary of the third repeat is within a KS domain. Judging by the position of catalytic residues relative to the repeat's boundary deletion or insertion events would not lead to new functional KS domains.

One interesting aspect of the repeats found in the *bry* cluster is that KS domains are generally thought to be relatively specific to the chain length of their substrate. We propose that KSs identical in sequence perform in module 4/8 and module 5/10. It is possible that the formation of the tetrahydro pyran rings can mask the true chain length and allow identical KSs to catalyze reactions on such different substrates. Much further biochemical research is needed to test the biosynthesical hypotheses presented here.

Functional heterologous expression of the entire *bry* cluster would provide the hypothetical compound bryostatin 0 (Figure 2.4), differing from bryostatin 1 in C7 and C20 substituents. This compound contains all pharmacophore elements as proposed by Wender [35, 36]. C7 has been shown to be not relevant for activity. C20 substituents have recently been proposed as “tunable elements” to modify bryostatin

potency [33]. A chemical synthesis pathway for C20 modification of bryostatin 0 would presumably be difficult due to steric congestion leading to undesired side products [33]. It may be possible to simply overpower this problem with a cheap and basically unlimited supply of bryostatin 0 followed by separation techniques to isolate products with the correct bryostatin 1 substitution at C20. Alternatively, identification and functional expression of the additional biosynthetic genes from either host or symbiont would allow the direct production of authentic bryostatin 1.

The sequence data presented is the first step towards solving the bryostatin supply problem using a biosynthetic route. There are substantial technical hurdles to overcome: The *bry* PKS cluster are enormous in size, which makes cloning challenging and functional expression of the entire cluster daunting. Also, the low %GC of the *bry* cluster could be challenging for a heterologous host to translate, as well as making long stretches of DNA unstable and prone to recombination. Finally as discussed above, some genes in bryostatin biosynthesis have not been identified. In spite of these potential difficulties expressing the biosynthetic genes in a heterologous host would be very cost effective to provide a virtually unlimited supply of authentic bryostatin. The *bry* genes also have the potential to be used in a combinatorial biosynthesis approach. By deleting or modifying modules or domains variations of bryostatin could be produced and complement the chemically synthesized “bryologs” currently being developed. This compares favorably with the other potential routes to bryostatin: Aquaculture of *B. neritina* to produce bryostatins while technically

feasible has proven to be not cost-effective [21]. Direct culturing of obligate-symbiotic bacteria is notoriously difficult because typically the bacteria have a hard time adapting to conditions outside of the narrow range provided by the host. No exception to this rule, *E. sertula* has proven unculturable so far. While total chemical synthesis of bryostatin is possible [13] its practical use is limited because of the >70 steps needed. In our opinion the synthesis of simpler analogues based on the pharmacologically active part of the molecule [37] is the other promising route. It has already lead to a number of compounds with interesting activities [34]. A potential drawback is, that each new analogue (chemically synthesized or derived from combinatorial biosynthesis) presumably has to be evaluated again in clinical trials since it is not authentic bryostatin 1.

**Genomic rearrangement.** The major difference between the two *bry* clusters presented in this study is the separation of *bryABCDX* from *bryPQRS* in “deep” *E.sertula*. The transposase genes found directly upstream of these two gene sets strongly suggests transposons as the cause of rearrangement. Since the two transposases upstream of *bryB* and *bryP* are not identical, multiple transposition events must have taken place. The acyl transferase activity of *bryP* is integral for PKS function, so we assume the gene arrangement found in “shallow” *E. sertula* with all *bry* genes in one locus is ancestral. The fragmentation of a PKS pathway is not unprecedented. The *ped* genes are found in three loci in the *Paederus fuscipes* symbiont genome [29, 30], while the closely related *onn* genes [29] are clustered. The



*ped* loci are also bordered by transposase-like sequences. A symbiotic lifestyle generally leads to genome degradation and an increased percentage of transposases. It is tempting to speculate, that the strictly symbiotic lifestyle of *E. sertula* has also led to the fragmentation of the putative bryostatin biosynthesis pathway.

**Conclusions.** The *bry* cluster described in this paper is an example of a PKS cluster found in a marine invertebrate that is implicated in the production of a bioactive compound originally isolated from the host animal. While the organization of the genes and lack of appropriately sized PKS genes elsewhere in the genome makes us reasonably sure that the *bry* cluster is responsible for bryostatin production, there is no direct evidence. This would require either culturing and mutational studies of *E. sertula* or heterologous expression of all (or at least significant parts) of the *bry* genes. As described above, both of these tasks are formidable. The *bry* genes share a surprising number of features with the pederin/onamide clusters: domains from these clusters are usually the best hits in blastP searches after the *B. subtilis* *pks* genes, both clusters have methyl transferases, discrete ATs and in the case of onamide HMG-CoA synthases. They also share a domain that presumably is responsible for the formation of a tetrahydro pyran ring via a Michael type reaction mechanism. It is interesting to note that pederin/onnamide is a mixed NRPS/PKS system, while bryostatin is pure PKS suggesting that nature is able to mix and match the two systems easily. There are three long perfect repeats in the *bry* cluster. While their borders do not exactly coincide with PKS modules, the recombination events

commonly associated with repeated stretches of DNA hint at a mechanism how modular PKS genes can quickly evolve. In conclusion, the *bry* cluster has some unique and interesting features and we provide some testable hypotheses for its function opening a rich field of further study.

\*\*\*\*\*

The bryostatin project was a long-running collaborative effort. Laura Waggoner, Nicole Lopanik, Mark Hildebrand and Haibin Liu took part in the DNA sequencing and annotation effort as well as developing the biosynthetic hypothesis. Christine Anderson contributed all RNA data. David Sherman provided valuable input to the manuscript.

**Table 2.1:** Genes and PKS domains in the *bry* cluster.

Gene	Size (amino acids)	Predicted catalytic domains and modules or function	homologs (blastP)
<i>bryA</i>	4888	L: DHh-KRh-FkbH-ACP	<i>B. subtilis</i> PKS, <i>Paederus fuscipes</i> symbiont <i>ped</i> cluster
<i>bryB</i>	5511	1: KS-KR-ACP	
		2: KS-KR-ACP	
<i>bryC</i>	5381	3: KS-( <i>MT</i> )-ACP	
		4: KS-DH*-MT-ACP	
		5: KS-DH-KR-ACP	
		6: KS-DH*-KR-ACP	
		7: KS-( <i>MT</i> )-ACP	
		8: KS-DH-PS-KR-ACP	
<i>bryX</i>	3476	9: KS-DH*-MT-ACP	
		10: KS-KR-ACP	
		11: KS-ACP	<i>B. subtilis</i> AT (pksC) <i>B. subtilis</i> pksF <i>B. subtilis</i> HMG-CS (pksG) <i>Microbulbifer</i> sp. SAM-dependent methyl transferase
<i>bryD</i>	4396	( <i>MT</i> -ACP)	
		( <i>KS</i> -DH-ACP)	
		( <i>KS</i> -ACP)	
		TE-KS	
		11a: KR-ACP	
		12: KS-DH*-KR-ACP	
<i>bryP</i>	632	KS-ACP-NRPS	
<i>bryQ</i>	405		
<i>bryR</i>	419	AT-AT	
<i>bryS</i>	364	KS	
		HMG-CS	
		O-MT	

\* active site mutated, proposed to be inactive

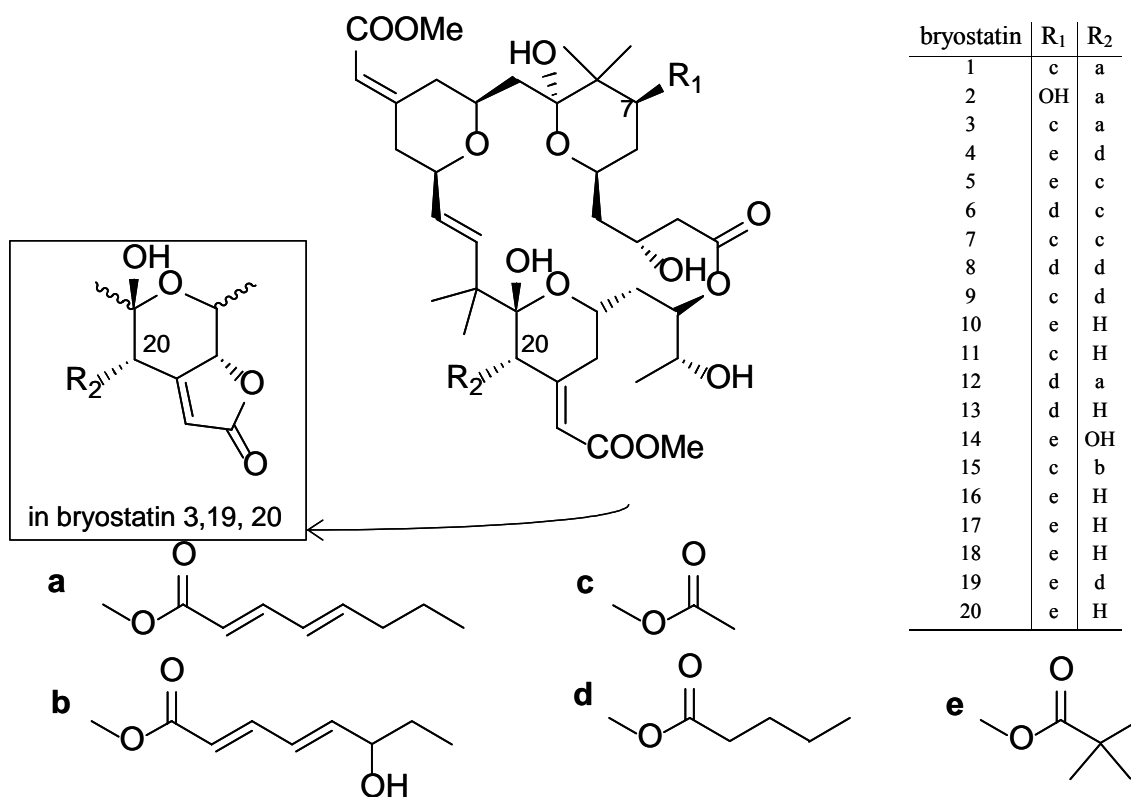
(*italics*), role in bryostatin biosynthesis unclear

KS = beta-keto synthase, KR = ketoreductase, DH = dehydratase,

ACP = acyl carrier protein, MT = methyl transferase, TE = thioesterase,

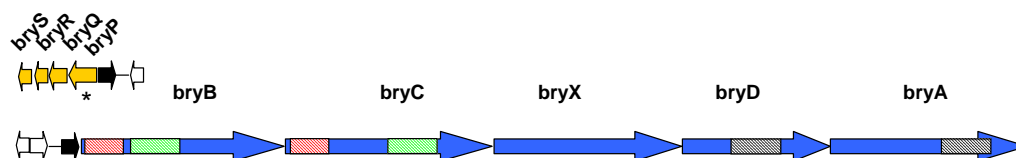
AT = acyl transferase, HMG-CS = hydroxymethyl glutaryl CoA synthase

The loading module (L) has some unique domains, DHh and KRh are DH and KR-like sequences, together with the FkbH domain they are proposed to catalyze the formation of the unusual D-lactate starter unit as described previously [16]. Module 8 contains a unique domain, proposed to form the tetrahydro pyran ring from C11-C15.

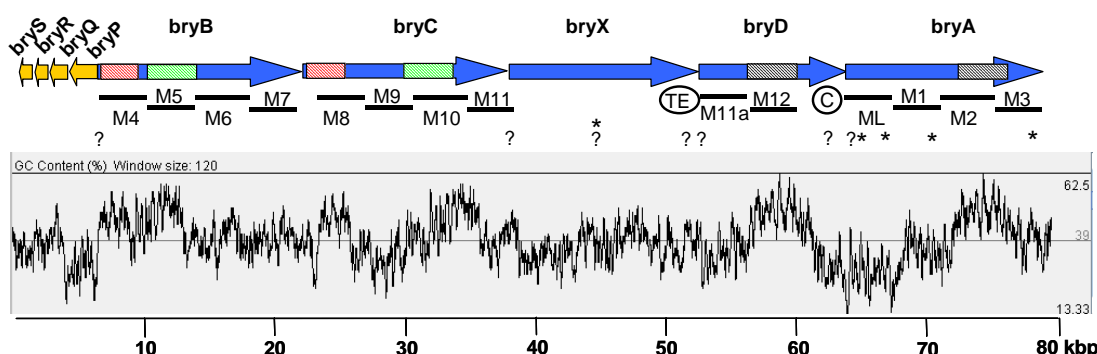


**Figure 2.1:** Structure of bryostatins, modified from [5]. Bryostatin 10 has been shown to be a feeding deterrent in *B. neritina* larvae and probably is the ecologically significant compound. Bryostatin 1 is currently in clinical trials for a variety of cancer treatments.

## DEEP

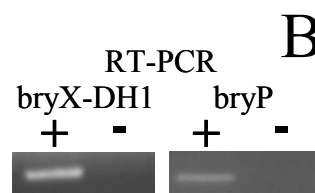


## SHALLOW

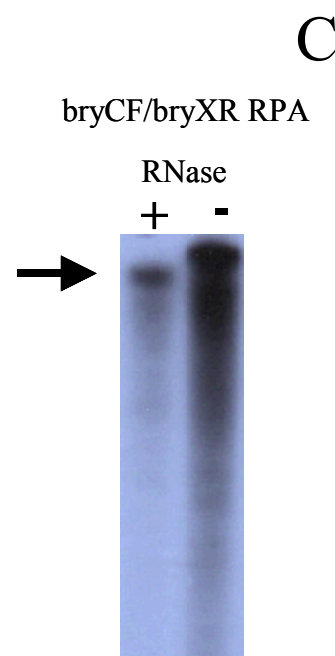


A

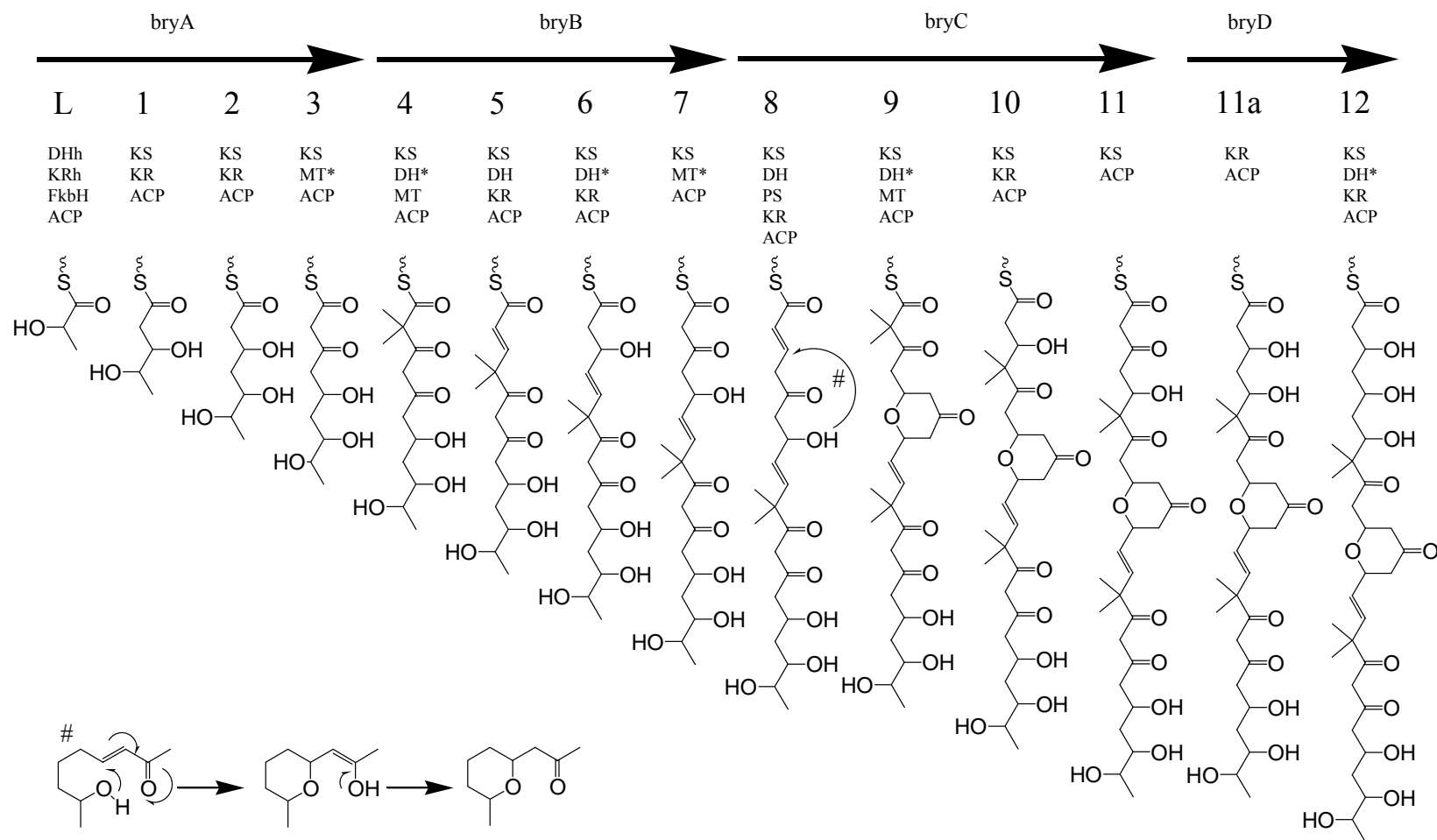
**Figure 2.2:** (A) The *bry* cluster in “deep” and “shallow” *E. sertula*. Modular polyketide synthase genes in blue, genes with accessory function in yellow and transposases in black. Genes with primary metabolism homologs in open arrows. The red, green and black striped regions are perfect repeats. Black bars approximate PKS modules involved in bryostatin biosynthesis. The two possible off-loading domains, a thioesterase in *bryX* and a NRPS condensation domain in *bryD* are circled (not to scale). Transcripts were detected by RT-PCR (\*) or RPAs (↑) in native *B. neritina* RNA samples. The %GC plot shows higher values in repeats. (B) RT-PCR products from *bryX* (“shallow” *B. neritina* DNA) and *bryP* (“deep” *B. neritina* DNA) in which reverse transcriptase was included (+), showing the presence of transcript, or omitted (-), demonstrating the absence of DNA contamination, in the RT-PCR protocol. (C) BryCF/BryXR RPA products in which RNase A/T1 was included (+), demonstrating the presence of transcript, or omitted (-), showing the full length of the labeled RNA probe. The RPA products of the other genes are omitted.



B



C



**Figure 2.3:** Proposed pathway for bryostatin biosynthesis. Functional arrangement of domains and the intermediate polyketides are shown. Arrows denote genes, PKS modules are designated by L and 1-12, and their domains are listed. Box details the formation of the pyran ring via Michael reaction. Abbreviations see Table 2.1.



## References

1. Alkon, D. L., H. Epstein, A. Kuzirian, M. C. Bennett, T. J. Nelson. 2005. Protein synthesis required for long-term memory is induced by PKC activation on days before associative learning. *Proc Natl Acad Sci USA* 102: 16432-16437.
2. Altschul, S. F., W. Gish, W. Miller, E. W. Wyers, D. J. Lipman. 1990. Basic Local Alignment Search Tool. *J Mol Biol* 215: 403-410.
3. Blackhall, F. H., M. Ranson, J. A. Radford, B. W. Hancock, M. Soukop, A. T. McGown, A. Robbins, G. Halbert, G. C. Jayson. 2001. A phase II trial of bryostatin 1 in patients with non-Hodgkin's lymphoma. *Br J Cancer* 84: 465-469.
4. Cheng, Y.-Q., G.-L. Tang, B. Shen. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc Natl Acad Sci USA* 100: 3149-3154.
5. Davidson, S., S. W. Allen, G. E. Lim, C. M. Anderson, M. G. Haygood. 2001. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "*Candidatus* Endobugula sertula" of the bryozoan *Bugula neritina*. *Appl Env Microbiol* 67: 4531-4537.
6. Davidson, S. K., M. G. Haygood. 1999. 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.
7. Del Vecchio, F., H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortes, B. A. Rudd, J. Staunton, P. F. Leadlay. 2003. Active-site residue, domain and module swaps in modular polyketide synthases. *J Ind Microbiol Biotechnol* 30: 489-94.
8. Donadio, S., M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz. 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science* 259: 675-679.
9. Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, W. H. Gerwick. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* 11: 817-833.



10. El-Sayed, A. K., J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas. 2003. Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem Biol* 10: 419-430.
11. Faulkner, D. J. 2002. Marine Natural Products. *Nat Prod Rep* 19: 1-49.
12. Gaitatzis, N., B. Silakowski, B. Kunze, G. Nordsiek, H. Blocker, G. Hofle, R. Muller. 2002. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J Biol Chem* 277: 13082-13090.
13. Hale, K. J., M. G. Hummersone, S. Manaviazar, M. Frigerio. 2002. The chemistry and biology of the bryostatin antitumour macrolides. *Nat Prod Rep* 19: 413-453.
14. Haygood, M., S. Davidson. 1997. 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.
15. Hildebrand, M., L. E. Waggoner, G. E. Lim, K. H. Sharp, C. P. Ridley, M. G. Haygood. 2004. Approaches to identify, clone and express symbiont bioactive metabolite genes. *Nat Prod Rep* 21: 122-142.
16. Hildebrand, M., L. E. Waggoner, H. Liu, S. Sudek, S. Allen, C. Anderson, D. H. Sherman, M. Haygood. 2004. *bryA*: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. *Chem Biol* 11: 1543-1552.
17. Hofemeister, J., B. Conrad, B. Adler, B. Hofemeister, J. Feesche, N. Kucheryava, G. Steinborn, P. Franke, N. Grammel, A. Zwintscher, F. Leenders, G. Hitzeroth, J. Vater. 2004. Genetic analysis of the biosynthesis of non-ribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* A1/3. *Mol Genetics Genomics* 272: 363-378.
18. Hopwood, D. A., D. H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* 24: 37-66.
19. Lopanik, N., N. Lindquist, N. Targett. 2004. Potent cytotoxins produced by a microbial symbiont protect host larvae from predation. *Oecologia* 139: 131-139.
20. McGovern, T., M. Hellberg. 2003. Cryptic species, cryptic endosymbionts, and geographic variation in chemical defenses in the bryozoan *Bugula neritina*. *Mol. Ecol.* 12: 1207-1215.

21. Mendola, D. 2003. Aquaculture of three phyla of marine invertebrates to yield bioactive metabolites: process developments and economics. *Biomol Eng* 20: 441-458.
22. Miller, D. A., L. Luo, N. Hillson, T. A. Keating, C. T. Walsh. 2002. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem Biol* 9: 333-44.
23. Mohammad, R. M., N. V. Adsay, P. A. Philip, G. R. Pettit, V. K. Vaitkevicius, F. H. Sarkar. 2001. Bryostatin 1 induces differentiation and potentiates the antitumor effect of Auristatin PE in a human pancreatic tumor (PANC-1) xenograft model. *Anti-Cancer Drugs* 12: 735-740.
24. Molnar, I., T. Schupp, M. Ono, R. E. Zirkle, M. Milnamow, B. Nowak-Thompson, N. Engel, C. Toupet, A. Stratmann, D. D. Cyr, J. Gorlach, J. M. Mayo, A. Hu, S. Goff, J. Schmid, J. M. Ligon. 2000. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem Biol* 7: 97-109.
25. Mutter, R., M. Wills. 2000. Chemistry and clinical biology of the bryostatins. *Bioorg Med Chem* 8: 1841-1860.
26. Pettit, G. R. 1991. The bryostatins. *Prog Chem Org Nat Prod* 57: 153-195.
27. Pettit, G. R., C. L. Herald, D. L. Doubek, D. L. Herald. 1982. Isolation and structure of bryostatin 1. *J Am Chem Soc* 104: 6846-6847.
28. Piel, J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc Natl Acad Sci USA* 99: 14002-14007.
29. Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga. 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci USA* 101: 16222-7.
30. Piel, J., G. Wen, M. Platzer, D. Hui. 2004. Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *Chembiochem* 5: 93-8.

31. Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, J. Ravel. 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: 7315-20.
32. Sun, M. K., D. L. Alkon. 2005. Dual effects of bryostatin-1 on spatial memory and depression. *Eur J Pharmacol* 512: 43-51.
33. Wender, P. A., J. L. Baryza. 2005. Identification of a tunable site in bryostatin analogs: C20 Bryologs through late stage diversification. *Org Lett* 7: 1177-80.
34. Wender, P. A., M. O. Clarke, J. C. Horan. 2005. Role of the A-ring of bryostatin analogues in PKC binding: synthesis and initial biological evaluation of new A-ring-modified bryologs. *Org Lett* 7: 1995-8.
35. Wender, P. A., C. M. Cribbs, K. F. Koehler, N. A. Sharkey, C. L. Herald, Y. Kamano, G. R. Pettit, P. M. Blumberg. 1988. Modeling of the bryostatins to the phorbol ester pharmacophore on protein kinase C. *Proc Natl Acad Sci USA* 85: 7197-7201.
36. Wender, P. A., J. DeBrabander, P. G. Harran, J. M. Jimenez, M. F. Koehler, B. Lippa, C. M. Park, C. Siedenbiedel, G. R. Pettit. 1998. The design, computer modeling, solution structure, and biological evaluation of synthetic analogs of bryostatin 1. *Proc Natl Acad Sci USA* 95: 6624-6629.
37. Wender, P. A., K. W. Hinkle, M. F. Koehler, B. Lippa. 1999. The rational design of potential chemotherapeutic agents: synthesis of bryostatin analogues. *Med Res Rev* 19: 388-407.
38. Wenzel, S. C., R. Muller. 2005. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. *Curr Opin Chem Biol* 9: 447-458.
39. Xue, Y., D. Wilson, D. H. Sherman. 2000. Genetic architecture of the polyketide synthases for methymycin and pikromycin series macrolides. *Gene* 245: 203-211.

### 3.

## **Diversity of the symbiotic cyanobacterium *Prochloron sp.* in didemnid ascidians**

### **Abstract**

*Prochloron sp.* is the cyanobacterial symbiont of a tropical ascidian family, the didemnids. By 16S rRNA gene and 16S-23S internally transcribed spacer region sequencing and denaturing gradient gel electrophoresis the diversity of *Prochloron sp.* from different didemnid species was explored. Multiple *Prochloron* strains were identified in a single didemnid colony, which accounts for the mixture of natural products found in the ascidians. *Prochloron* showed surprising genetic homogeneity (>99%) over considerable geographical distances as well as between host species suggesting continuous horizontal exchange of symbionts despite of reports of vertical transmission.

## Introduction

Ascidians are sessile, filter-feeding, marine invertebrates. The ascidian tunic, a structural support and protective layer consisting of cellulose, is a unique adaptation in the animal kingdom. Traditionally ascidians, cephalochordata and vertebrata make up the phylum chordata, although typical chordate signatures like notochord, dorsal neural tube and muscular tail are only found in the larval stages of ascidians [29]. Recent molecular studies showing a high degree of divergence between the ascidians and the other two members of the phylum, lead to the proposal of a new phylum “tunicata” separate from the chordates [31]. Ascidians in the family didemnidae are colonial; thousands of individual animals (zooids) share a common tunic. The didemnidae are globally distributed, although most species diversity is found in tropical waters [12], where they form a major component of the filter feeding communities in sublittoral locations [13]. 500 species in eight genera have been described [13, 14]. Four genera, *Didemnum*, *Lissoclinum*, *Trididemnum* and *Diplosoma* contain species that bear the bacterial symbiont *Prochloron sp.*

*Prochloron sp.* is a cyanobacterium with unusual pigment composition [15]. It contains chlorophyll A and B, but none of the phycobilins typical for cyanobacteria. Cells are spherical and fairly large (~15 µm, Figure 3.1C). *Prochloron sp.* is typically found in tropical didemnids, where it makes up a significant part of the biomass (Figure 3.1) [16]. It has also been reported to be associated with sponges and other ascidians, often as an epibiont (Ralph Lewin, personal communication). Recently, *Prochloron* has been identified in stromatolites from Western Australia [3].

*Prochloron* cells have occasionally been reported in the water column [16]. Despite prolonged efforts, *Prochloron* remains uncultivated [16]. Full genome sequencing of *Prochloron* sp. has recently been completed, genome annotation is currently in progress (a collaborative effort of The Institute of Genomic Research: Jacques Ravel, M.J. Rosovitz and Jonathan Eisen; University of Utah: Mohamed Donia and Eric Schmidt; Scripps Institution of Oceanography: Christine Anderson, Sebastian Sudek and Margo Haygood; Station Biologique Roscoff, France: Frederic Partensky).

The nature of the symbiosis between *Prochloron* and the host didemnid is not completely resolved. Potential advantages for *Prochloron* are shelter, supply of ammonia and UV-protection. One advantages for the host is clearly nutritional. *Prochloron* sp. has been shown to contribute a significant portion of the reduced carbon used in host respiration (30-56% in *L. patella* under high-light conditions) [2] and the growth of didemnid colonies increases under elevated light conditions [20]. The symbiont may also provide a chemical defense through the synthesis of natural products (see below). Contrary to an earlier report [11], *Prochloron* is probably not involved in nitrogen fixation as the characteristic *nif* genes are absent from the genome (unpublished results, The Prochloron Genome Team).

The location and distribution of *Prochloron* sp. differs depending on the host (Figure 3.2). It has been found extracellular in the tunic, lining the cloacal cavity, or even intracellular (in *L. punctatum*) [10, 16]. In the genus *Diplosoma*, vertical transmission between ascidian generations has been shown. The larvae have a

specialized organ, the rastrum, to collect some *Prochloron* from the parental colony into a special pouch [9]. In the other genera, patches of *Prochloron* can be associated with the exterior of the larval tail [30]. Both mechanisms ensure that the daughter colony has a starting culture of the *Prochloron* symbiont.

A diverse group of amino acid-derived compounds has been isolated from didemnids [27]. One family, the patellamides are cyclic peptides. They can be rationalized as containing two elements, each having the sequence thiazole-nonpolar amino acid-oxazoline-non polar amino acid. The elements can be identical as in patellamide A or different (while following the described template) as in patellamide C (Figure 3.3, [8]). The patellamides are mildly cytotoxic and some have been reported to reverse cancer multidrug resistance [8]. Patellamide C has no significant effect as a chemical defense against predation by coral reef fish [18], but this does not exclude the possibility that other patellamides are more active or active against different kinds of predators. While the patellamides are evenly distributed throughout the didemnid colony [25], they have been shown to originate in *Prochloron sp.* [26]. Biosynthesis of the patellamides is ribosomal via a microcin-like pathway. A precursor protein PatE, containing the sequence for two compounds (see chapter 4, Figure 3.2) is modified by dedicated enzymes leading to the mature products [26]. The diversity of patellamides is achieved by a remarkable mechanism: point mutations in the precursor protein gene were found specifically in the region where the sequence for the mature product is located, while the rest of the pathway is unchanged (Donia et al., unpublished results).

This project investigated the phylogenetic diversity of *Prochloron* and how it is related to chemical diversity and host speciation. Based on literature reports suggesting a vertical mode of transmission, we hypothesized that cospeciation (i.e. parallel and co-dependent speciation indicated by congruent phylogenetic trees) between the didemnids and *Prochloron* sp. would be detectable.

### **Materials and Methods**

**Sample collection.** Didemnid samples (table 3.1) were collected in Palau (2002) and Papua New Guinea (2004) and stored in RNAlater by Eric Schmidt (University of Utah). *Lissoclinum patella* and *Didemnum molle* could be identified by their characteristic morphology in the field (Figure 3.1), sample 05-49 was identified as *Diplosoma virens* by 18S rRNA gene sequencing in the Schmidt lab, four other didemnids could not be identified by either method and are listed as unknown. These samples (05-31, 42, 45 and 46) are not *L. patella* or *D. molle* based on morphology. An unidentified sponge containing *Prochloron* sp. was also collected (sample 05-38).

**DNA isolation.** The *Prochloron*-containing part of the tunic (as determined by bright green coloration) was dissected out, and ~300 mg was minced using a razor blade. DNA was isolated using the DNeasy Tissue Kit (Qiagen) according to manufacturer's protocol with the following modifications: The amount of resuspension and lysis buffers was doubled and the lysis step was run overnight. The



presence of purified high-molecular weight DNA was confirmed by agarose gel electrophoresis. The yield varied between 50 ng and 1 µg.

**Denaturing Gradient Gel Electrophoresis (DGGE).** DGGE was performed following established protocols [7] using the Bio-Rad DCODE system. Briefly, a 16S gene fragment was PCR amplified with universal primers 1055f and 1392r with GC clamp. PCR products were loaded onto the gradient gels with 30-70% denaturant. The gel was stained with SYBR Gold (Molecular Probes Inc.), and bands were visualized with UV illumination and excised from the gel. The gel matrix was mechanically disrupted and the DNA extracted into 20 µl of ddH<sub>2</sub>O in an overnight incubation at 37 °C. One microliter was used as the template to reamplify the bands with the same primers for DNA sequencing.

**16S rRNA gene and 16S-23S internal transcribed spacer (ITS) analysis.** The Prochloron 16S rRNA gene sequence was determined by two protocols. DNA was amplified (Roche, Taq polymerase) with general eubacterial primers 27f and 1492r, TOPO-TA (Invitrogen) cloned according to the manufacturer's protocol and sequenced (SeqXcel Inc.). In some samples a significant percentage of clones turned out to contain 16S rRNA genes of unrelated bacteria, presumably originating from seawater in the ascidian body cavities. In these cases, a colony PCR was run with the Prochloron-specific primers pr16Sf and pr16Sr. Multiple positive clones per sample were selectively grown in liquid media, the plasmid re-isolated (QIAGEN) and sequenced with the general primers.

To skip the time-consuming cloning step, in the second protocol two separate PCR reactions were run for each sample using the primer pairs 27f + pr16Sr and pr16Sf + 1492r (Figure 3.4). The overlapping products were sequenced using the same primers and sequences assembled to re-constitute the quasi-full length 16S gene. In most samples double peaks in the sequencing chromatograms indicated single nucleotide polymorphisms. All 16S rRNA PCR reactions were run using Taq polymerase under the following conditions: initial denaturation at 95°C for 5 minutes, 35 cycles of 95°C for 30 sec, 52°C for 30 sec and 72°C for 1 min, and final elongation at 72°C for 7 min with the following template DNA dilutions: 1x, 1/10x, 1/100x.

The 16S-23S internal transcribed spacer region (ITS) was amplified using the primer pair pr16Sf and 23S115r (Figure 3.4). The product was sequenced using the same primer and 1055f as an internal primer. The ITS PCR reactions were run as described for 16S with the following modifications: annealing temperature 54C and elongation time 2 min.

Sequence reads were assembled with Sequencher (Gene Codes Corp.)

**Phylogenetic analysis.** Sequences were aligned using ClustalX [28]. The ends of the alignment were trimmed to avoid bias by longer sequences having unaligned parts. A total of 1209 characters were used for 16S and 462 for ITS. Phylogenetic trees were calculated using PAUP (Sinauer Associates) for maximum likelihood under the

general time reversible (GTR) model with gamma distribution rates and MrBayes [24] for Bayesian analysis under the GTR model with inverse gamma distribution rates. Bootstrapping for maximum likelihood was performed with 1400 replicates.

## Results

The diversity of the two Palauan samples (see table 3.1) was explored using denaturing gradient gel electrophoresis (DGGE). PCR products obtained with universal bacterial primers were run on a denaturing gel, with which separation depends on the nucleotide sequence of the PCR product (Figure 3.5). In both samples, a single major band was present and identified as *Prochloron sp.* through sequencing (identical to Genbank accession number X63141). Some weak bands were also present in the reef sample, and their sequence was shown to be identical to the major band. These bands are probably caused by incomplete denaturation of the PCR product. The DGGE results indicate, that *Prochloron* is overwhelmingly dominant in the Palau DNA preparations. This is consistent with light microscopy studies, where only the characteristic *Prochloron* cell type was observed. At the level of DGGE, there seemed to be no diversity of *Prochloron* strains.

To further study this question, *Prochloron* 16S rRNA gene sequences were amplified using general and *Prochloron*-specific primers from the Papua New Guinea samples. A total of 13 samples from seven different didemnid species and one sponge (table 3.1) were found to be >99.5% identical (Figure 3.5 and 3.6). To explore the diversity within a single host, the 16S rRNA gene of sample PNG05-019 was

amplified with the general primers 27f and 1492r and cloned into the TOPO vector. Eight *Prochloron* 16S rRNA genes were sequenced leading to 4 different sequences (>99.5% identity). It is interesting to note that in pairwise alignment, the sequences from PNG05-019 are not closer to each other than to the other sequences from different hosts (Figure 3.6).

The 23S-16S internally transcribed region should be mutating more rapidly than the 16S rRNA gene. This region is cut out of the rRNA precursor to yield mature 16S and 23S rRNA and so there are fewer functional constraints compared to the 16S rRNA, which needs to fold properly for a functioning ribosome. Using pr16Sf with the general primer 23S119r, the ITS of five samples was amplified and sequenced. The *Prochloron* ITS is 462 nucleotides long. It contains a single tRNA gene (Figure 3.4). Sequence diversity between the *Prochloron* samples was again found to be low. (overall >97% identity). Surprisingly, there were almost no differences between most of the samples (3 mutations over 462 nucleotides) except for the sequences obtained from sample 05-031 (an additional 13 mutations). These variations clustered around nucleotide 210 and 305 (Figure 3.7).

The low diversity found in all *Prochloron* sp. and for both gene regions raises the question if all *Prochloron* sp. are really identical and the variation is caused by nucleotide misincorporation during PCR, especially considering that the environmental DNA preparations may contain small molecules interfering with the polymerase reaction (DNA sequencing errors can be discounted because samples were

sequenced with long overlaps and often multiple times, confirming the variations). While the occurrence of PCR misincorporations cannot be excluded for all sequence variations, some variations are certainly natural mutations based on two arguments: sometimes a particular nucleotide was different in multiple samples, suggesting a variable position in the gene (Figure 3.7). Furthermore, a certain part of the ITS region (Figure 3.8) had a high occurrence of variations. One would expect PCR misincorporations to be randomly distributed.

Initially a phylogenetic tree of the 16S sequences was constructed using the neighbor-joining algorithm. The tree grouped the Prochloron 16S sequences from *L. patella* (and one *D. molle*) together, the “unknown host” (which are not *L. patella* or *D. molle*) formed a second clade and the sequences from *D. molle* a third. The bootstrap support for this tree was too low to be convincing (data not shown). A second phylogenetic tree of the 16S sequences was constructed using Bayesian analysis with *Synechocystis* clone 29P18 as an outgroup. All taxa are unresolved with the exception of 05-025 and 05-034, both from *D. molle*, which clade together with good posterior probability support (Figure 3.9). Similar results were obtained using the maximum likelihood algorithm. The high degree of identity (17 informative sites out of 1209) doesn't give enough resolution to establish separate clades. The ITS sequences (which also include about ~one third of the 16S rRNA gene, see methods section) also show polytomy. Only the sequences from sample PNG05-31 form a separate clade (Figure 3.10).

## Discussion

All *Prochloron* 16S rRNA sequences known to date are ~99% identical. This includes sequences obtained in this study from Papua New Guinea and Palau didemnids, and sequences in Genbank from a Palau didemnid and Australian stromatolites (the stromatolite sequences are only ~550 bp long). This homogeneity is surprising considering they come from different locations within the Indo-Pacific, different host ascidians (and a sponge) and even from stromatolites. All efforts to construct a phylogeny have given inconclusive results due to the low number of informative characters. Surprisingly the 16S-23S ITS did not show higher divergence and does not yield higher phylogenetic resolution. A previous study based on DNA-dependent RNA polymerase sequences found divergence between five *Prochloron sp.* samples [21] (see below).

Working on the symbiosis of vesicomid clams with chemoautotrophic proteobacteria Peek et al propose cospeciation based on congruent phylogenetic trees as well as the timing of cladogenic events. The amount of divergence they find within symbionts (3.9%) equals to ~100 million years as an independent lineage (they assume a divergence rate of 2% per 50 million years). This coincides with the age of the oldest fossilized vesicomid clams [22].

Hirose et al. have recently published an extensive analysis of didemnid phylogeny based on 18S rRNA sequences [30]. Based on their phylogenetic trees, the

didemnid ascidians have speciated between ~175-350 million years ago, assuming a mutation rate of 1-2% per 50 million years in the 18S gene [19]. There is no fossil record of the didemnid ascidians, but the earliest ascidian fossils are believed to be 520 million years old [4]. The bacterial 16S rRNA gene is thought to mutate at the same rate as 18S rRNA gene [19], the amount of variation in the 16S rRNA genes found in this study (<0.5%) is therefore much less than what should be expected if the didemnids and *Prochloron* had cospeciated. A possible explanation for our data is that the mutation rate of the *Prochloron* 16S rRNA gene is much lower and therefore the variation is too little to show cospeciation. A more convincing possibility is that there is no cospeciation. *Prochloron* would be continually exchanged between didemnids (and stromatolites!) through horizontal transfer. One would expect there to be a pool of free-living *Prochloron*, so that genetic exchange is rapid enough to explain the homogeneity over large geographical distances. In this context it would be instructive to study Caribbean *Prochloron* sp., since current research is focused on the Indo-Pacific.

Single *Prochloron* cells have been found in the water column around ascidians, but have never been reported in open ocean water, despite of efforts to catalogue microbial diversity using molecular biology techniques. It is hard to estimate how rare (and therefore overlooked) *Prochloron* could be while still maintaining genetic homogeneity at least across the Indo-Pacific. A possible explanation is the existence of a *Prochloron* resting stage that is not characteristic under the microscope, maybe even

resistant to common DNA extraction methods (there is no precedence in cyanobacteria to our knowledge).

Cospeciation of bacteria and host is usually found in systems with vertical transmission (for example vesicomysid clams described above), whereas in horizontal transmission systems little bacterial divergence is apparent. For example, vestiminiferan tubeworms have symbiont-free juvenile stages. The adults are in an obligate symbiosis with chemoautotrophic  $\gamma$ -proteobacteria. Feldman *et al.* have shown that there is little divergence of symbionts in hot vent vestiminiferans, even among different tubeworm species and over long geographic distances based on 16S rRNA gene sequences [6]. This situation is clearly reminiscent of the findings in this study. Considering the evidence for vertical transmission of *Prochloron sp.*, the following scenario is possible: Vertical and horizontal transmission are not mutually exclusive. All (or most) didemnids supply their larvae with a starter culture of photosynthetic *Prochloron sp.*, clearly a nutritional advantage. Yet, *Prochloron* cells are also recruited (and expelled) in later life stages without specificity for didemnid species. Some form of understanding between host and symbiont *in spe* is needed as roaming *Prochloron* has to pass the digestive tract to arrive in the cloacal cavity, where the symbiont resides in most didemnids (Figure 3.2). If a *Prochloron* cell gets detached from the cloacal cavity wall it may be expelled with the water flow. Occasionally these wandering *Prochloron sp.* successfully colonize other surfaces like sea cucumbers and sponges as epibionts.



While there are indication for cospeciation of extracellular symbionts in the bryozoan genus *Bugula sp.* [17] and the sponge order Dictyoceratida [23], in many bacteria-animal symbiosis with cospeciation, the bacteria are maintained intracellularly (i.e in bacteriocytes): for example vesicomylid clams and proteobacteria [22], aphids and *Buchnera sp.* [5], tsetse flies and  $\gamma$ -proteobacteria [1]. *Prochloron* is an extracellular symbiont except in the genus *Diplosoma*. In this study, only one *Diplosoma* sample was characterized, it seems possible that cospeciation within this one genus would occur. This could be tested by obtaining a wider range of *Diplosoma* samples.

Despite the high similarity of 16S rRNA gene and 16S-23S ITS sequences, different *Prochloron* strains are distinguishable in the ascidians. The single nucleotide polymorphisms found in the 16S rRNA sequencing suggest that even within a single host animal multiple strains of *Prochloron* are present. By cloning the PCR product pool, four slightly different 16S rRNA genes could be identified in a single ascidian sample, but the search was not exhaustive. The number of *Prochloron* strain per host didemnid is unclear presently. The *Prochloron sp.* diversity coincides with multiple versions of the patellamide biosynthesis precursor protein gene *patE*. Multiple versions of *patE*, differing in the patellamide-coding sequence have been identified in a number of didemnid samples (Donia et al. unpublished results). Quantitative PCR experiments have shown that the ratios of the different *patEs* can be up to 1:70. (Donia et al. unpublished results). This strongly suggests, that the different *patE* versions do

not reside in a single genome but that there are multiple, closely related strains of *Prochloron* in different abundances, each with a different *patE*. This array of *Prochloron sp.* then leads to the well-documented mixture of patellamides found in a single didemnid.

The sequencing of the *Prochloron* genome indicates that there are minimal differences between *Prochloron* strains in a single ascidian with the exception of *patE* (and hence the ability to produce certain patellamides). This suggests that the chemical diversity has ecological significance beyond a possible general function of all patellamides as feeding deterrents against the didemnid's predators. One could speculate that there is an arms race between the *Prochloron* strains. If the patellamides were toxic for *Prochloron* and required a resistance mechanism specific for each patellamide, acquiring the biosynthetic/resistance potential for a new patellamide would be a selective advantage. This could lead to different *Prochloron* strains with different patellamides as observed. Its resistance to culturing efforts impedes studies on *Prochloron*, we hope that the genome sequencing will help answer these intriguing questions.

The didemnid – *Prochloron sp.* symbiosis is summarized in Figure 3.11. This study shows that rigorous phylogenetic analysis is instructive. It has established the presence of multiple strains of *Prochloron* in different as well as in a single host didemnid. There is no indication of cospeciation between *Prochloron* and the didemnids. A larger sample pool including more *Diplosoma*-derived as well as

Caribbean samples would be a profitable direction of future research. In contrast to this study, Palenik *et al.* [21] have shown that there is limited diversity between *Prochloron sp.* based on the DNA-dependent RNA polymerase C gene (*rpoC*) sequences of six ascidians (up to 2.6%). This suggests that *rpoC* is a better phylogenetic marker for *Prochloron*. Certainly the most complete approach would be a phylogeny based on the sequencing of whole *Prochloron* genomes from multiple geographical locations and different host species.

\*\*\*\*\*

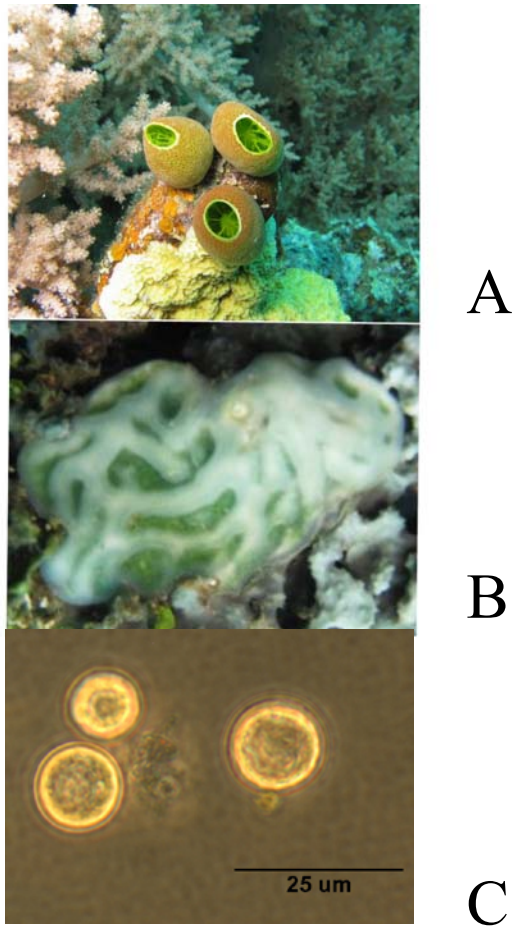
Eric Schmidt collected all samples used in this study. He and his student Mohamed Donia provided helpful discussion. Ralph Lewin originally described *Prochloron* and shared many of his interesting (and entertaining!) experiences with this bug.

**Table 3.1:** *Prochloron* samples analyzed in this study. 16S rRNA genes and internally transcribed spacer region (ITS) were amplified in samples indicated by +. Both of the Palauan 16S rRNA gene sequences were found to be identical to the sequence deposited in Genbank previously (accession number: X63141). In samples 05-19, 05-20 and 05-31 multiple strains within one host were identified.

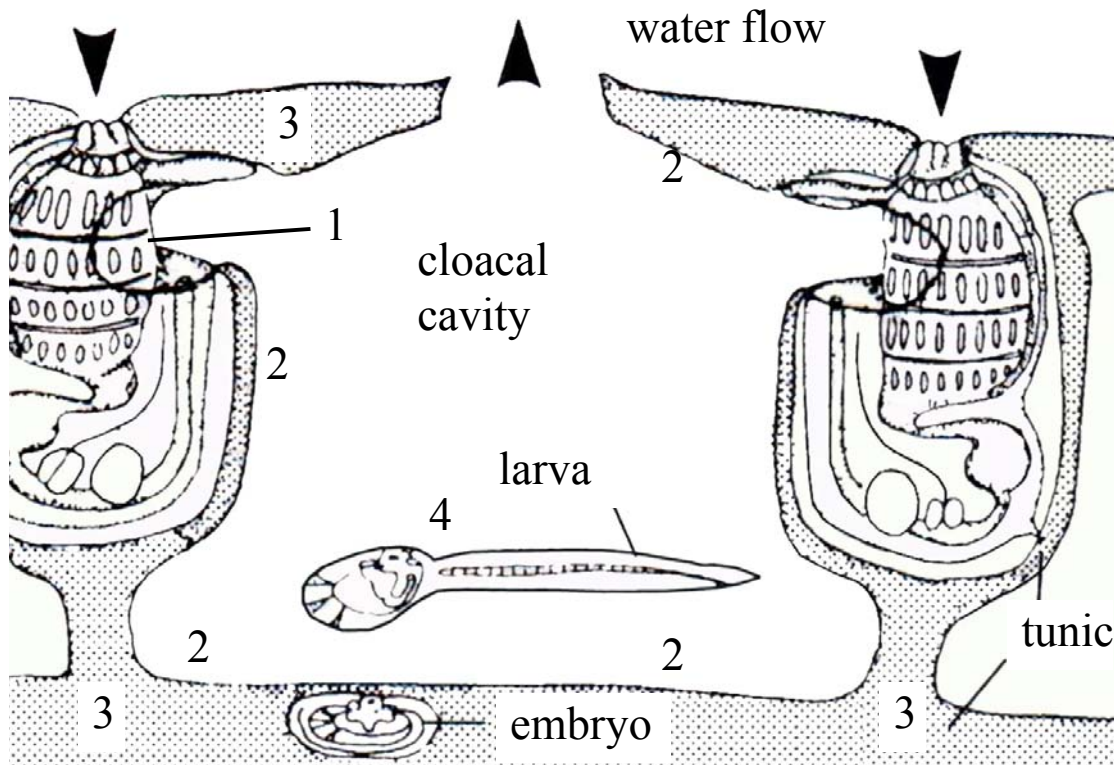
sample	host	16S	ITS
Palau reef	<i>L. patella</i>	*	
Palau seagrass	<i>L. patella</i>	*	
PNG05-19	<i>L. patella</i>	+	+
PNG05-20	<i>L. patella</i>	+	+
PNG05-25	<i>D. molle</i>	+	
PNG05-27	<i>D. molle</i>		+
PNG05-28	<i>D. molle</i>	+	
PNG05-31	unknown A	+	+
PNG05-33	<i>L. patella</i>	+	
PNG05-34	<i>D. molle</i>	+	
PNG05-38	sponge		+
PNG05-42	unknown B	+	
PNG05-45	unknown C	+	
PNG05-46	unknown D	+	
PNG05-49	<i>Dip. virens</i>	+	

**Table 3.2:** Oligonucleotide primers used in this study.

primer	sequence
27f	aga gtt tga tcm tgg ctc ag
1055f	atg gct gtc gtc agc t
1392r	(g <sub>x</sub> c <sub>y</sub> ) <sub>40</sub> acg ggc ggt gtg tac
1492r	tac ggy tac ctt gtt acg act t
pr16Sf	agg gtc cgc agg tgg cga a
pr16Sr	ctt cgg cac agc tcg ggt c
23S115r	ggg ttb ccc cat tcg g

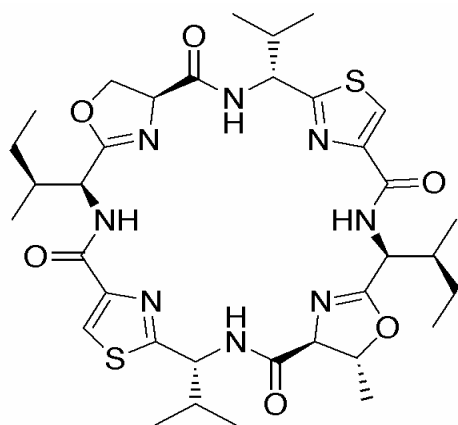


**Figure 3.1:** *Didemnum molle*, *Lissoclinum patella* and symbiotic *Prochloron sp.* A. *D. molle* and B. *L. patella* represent two genera of didemnid ascidians. Green coloration is caused by the symbiotic cyanobacterium *Prochloron sp.* Photos by Chris Ireland (University of Utah). C. *Prochloron sp.* in phase contrast light microscopy

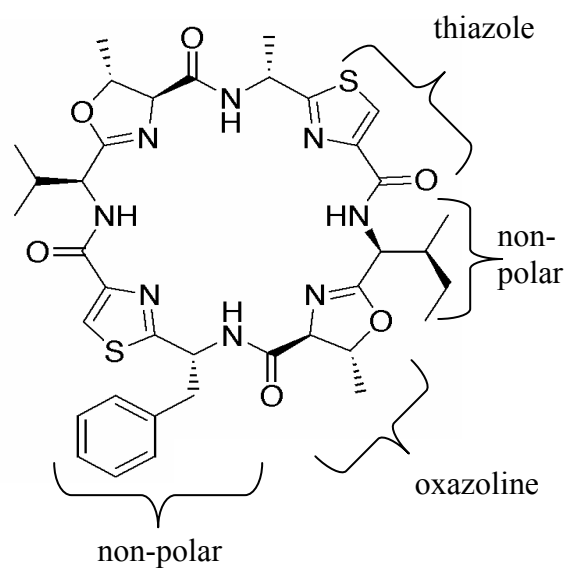


**Figure 3.2:** Schematic of a didemnid ascidian, modified from [29]. Locations of *Prochloron* sp. in different didemnids indicated by numbers; species are only examples, *Prochloron* location may not be strictly correlated with species: 1. lining the peribranchial space in *Diplosoma similis* 2. lining cloacal cavity in *Lissoclinum patella* and *Didemnum molle*. 3. in the tunic, extracellular in *Diplosoma virens*, intracellular in mesenchymal cells in *Lissoclinum bistratum*. 4. associated with the larvae, in a specialized pouch in *Diplosoma* sp. or attached to the larval tail *L. patella*

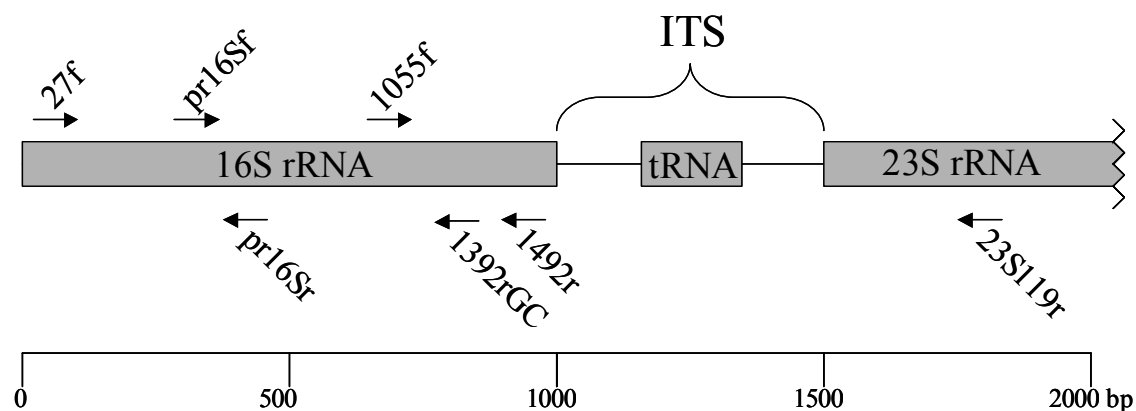
patellamide A



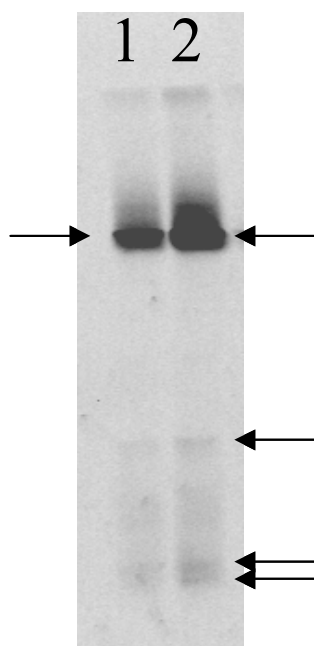
patellamide C



**Figure 3.3:** Patellamides A and C, cyclic peptides biosynthesized by *Prochloron* sp. They consist of elements with the substructure: thiazole, non-polar, oxazoline, non-polar amino acid.



**Figure 3.4:** rRNA operon of *Prochloron* sp. Position of primers used in this study are indicated by arrows (not to scale).



**Figure 3.5:** Denaturing gradient gel electrophoresis of *Prochloron* sp. DNA preparations (30-70% gradient). Lane 1: Palau seagrass sample; Lane 2: Palau reef sample. Arrows indicate the bands excised and sequenced.



	1	2	3	4	5	6	7	8	9	10	11	12	13
1 05-019B-Lp													
2 05-019C-Lp	0.085												
3 05-019D-Lp	0.170	0.085											
4 05-020-Lp	0.085	0.000	0.085										
5 05-025-Dm	0.255	0.170	0.255	0.170									
6 05-028-Dm	0.085	0.000	0.085	0.000	0.170								
7 05-031-xx	0.085	0.000	0.085	0.000	0.170	0.000							
8 05-033-Lp	0.085	0.000	0.085	0.000	0.170	0.000	0.000						
9 05-034-Dm	0.255	0.170	0.255	0.170	0.000	0.170	0.170	0.170					
10 05-042-xx	0.085	0.000	0.085	0.000	0.170	0.000	0.000	0.000	0.170				
11 05-045-xx	0.341	0.255	0.341	0.255	0.426	0.255	0.255	0.255	0.426	0.255			
12 05-046-xx	0.341	0.255	0.341	0.255	0.426	0.255	0.255	0.255	0.426	0.255	0.341		
13 05-049-Dipl	0.170	0.085	0.170	0.085	0.255	0.085	0.085	0.085	0.255	0.085	0.341	0.341	
14 03-Palau	0.085	0.000	0.085	0.000	0.170	0.000	0.000	0.000	0.170	0.000	0.255	0.255	0.085

**Figure 3.6:** Pairwise distance matrix for *Prochloron* sp. 16S rRNA sequences in percent. Lp = *Lissoclinum patella*, Dm = *Didemnum molle*, Diplo = *Diplosoma virilens*, Palau = *L. patella* from Palau (all other samples are from Papua New Guinea), xx = unknown host

```

                                GCTAGTTGGAAGGGTAAAAGCCTACCAAGGCGATGATCGGTAGCTGGTCT
05-019A-Lp : .....
05-034-Dm : .....G.....A.....
05-025-Dm : .....A.....
05-020-Lp : .....
05-033-Lp : .....
05-019D-Lp : .....
05-28-Dm : .....
05-019C-Lp : .....C.....
03-Palau : .....C.....
05-019B-Lp : .....C.....
05-31-xx : .....
05-049-Dip : .....G.....A.....
05-042-xx : .....
05-045-xx : .....G.....
05-046-xx : .....C.....
syn.cystis : .T.....TG.....C.C.....

```

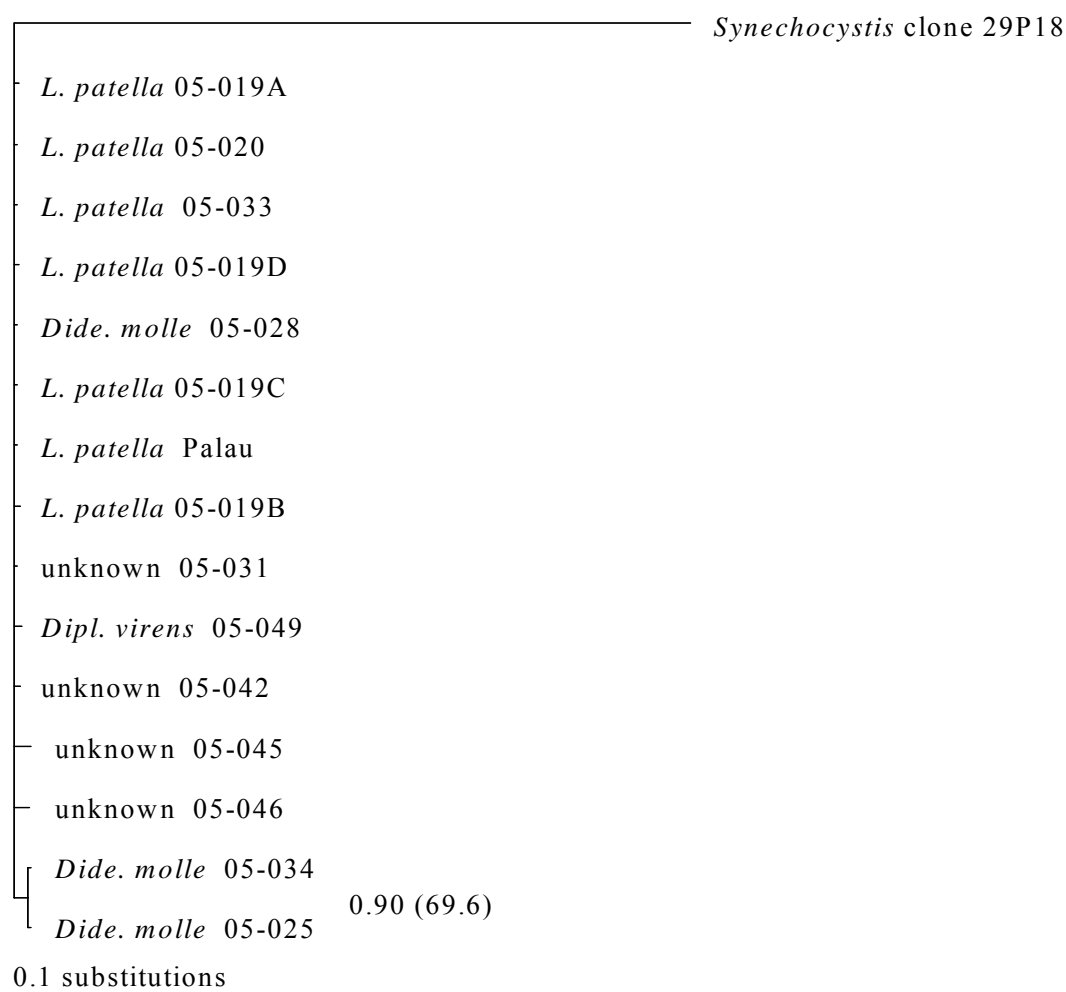
**Figure 3.7:** Partial 16S alignment of the most diverse part of the 16S rRNA gene. Syn.cystis = *Synechocystis* clone 29P18, other samples see Figure 3.9.

```

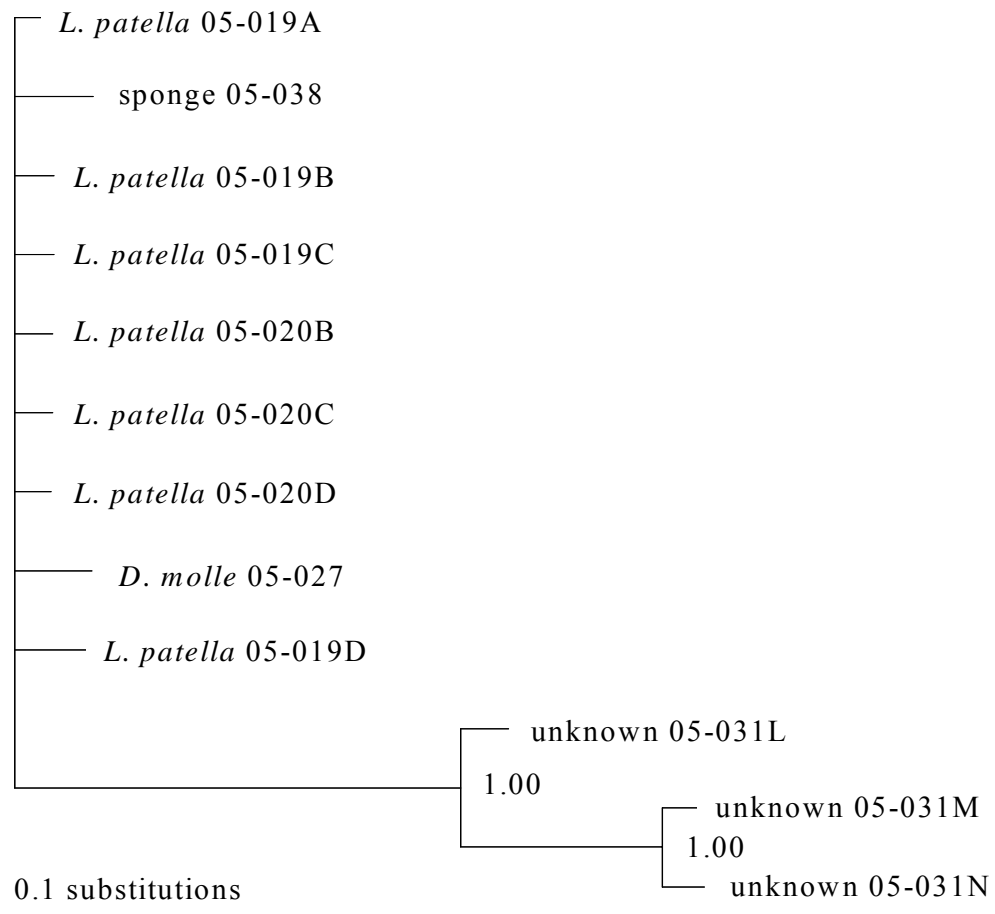
05-019A  AGTCCAGGAT  GGCCACCTA  AAATACACAG  GCATCCTGCC  TGTGCCACAA
05-038   .....
05-019B   .....
05-019C   .....
05-020B   .....
05-020C   .....
05-020D   .....
05-027    .....
05-019D   .....
05-031M   .....GGA.....
05-031N   .....GGA.....
05-031L   .....GGA.....

```

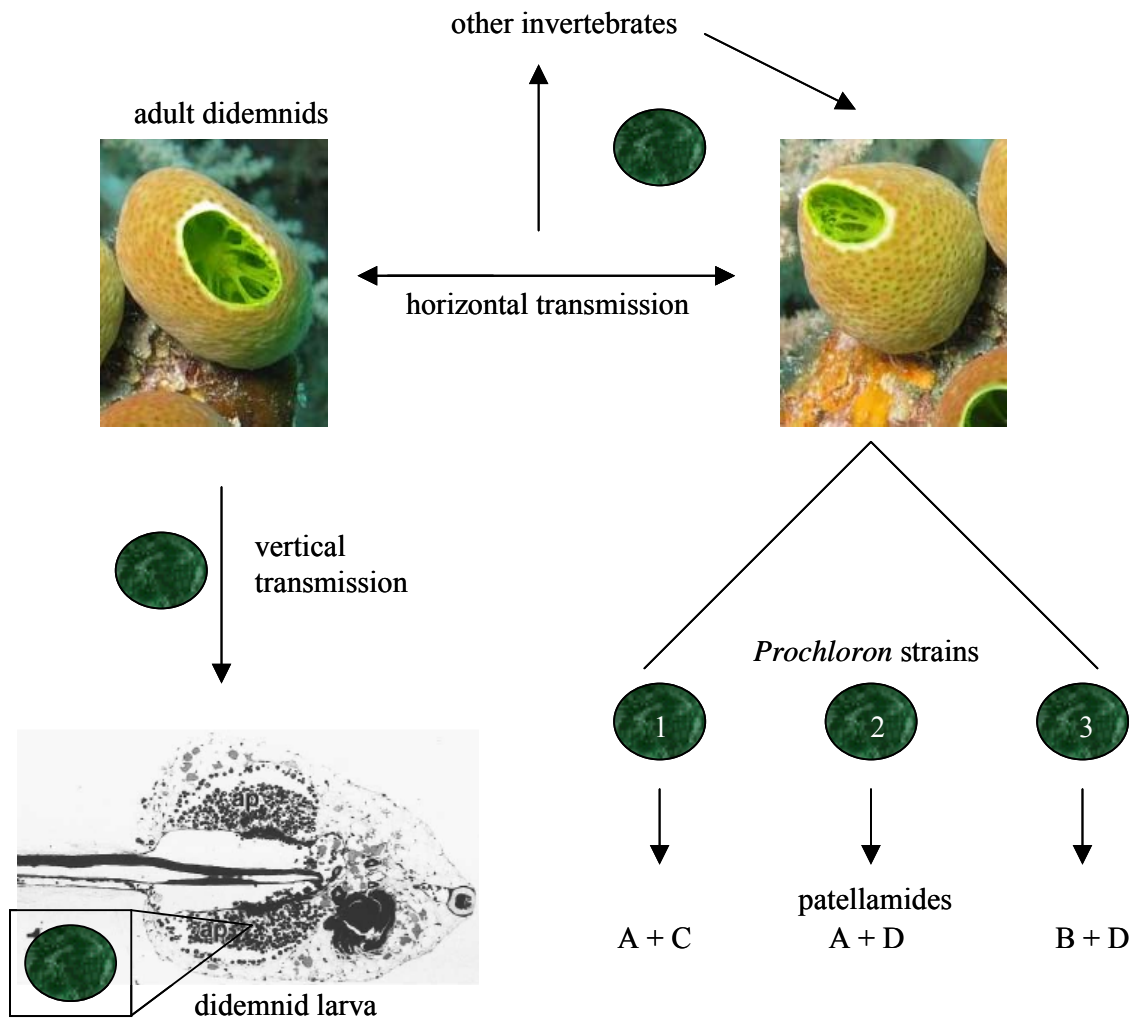
**Figure 3.8:** Partial 16S-23S ITS alignment showing clustered mutations of sample 05-031.



**Figure 3.9:** Phylogenetic tree of *Prochloron* spp. based on the 16S rRNA gene using *Synechocystis* clone 29P18 as an outgroup estimated using MrBayes. A maximum likelihood estimation with PAUP gives similar results. Bayesian posterior probability and (maximum likelihood bootstrap value) is shown for the *Dide. molle* clade.



**Figure 3.10:** Phylogenetic tree of *Prochloron* spp. based on the 16S-23S ITS estimated using MrBayes. Bayesian posterior probability is shown on nodes.



**Figure 3.11:** Overview of the didemnid – *Prochloron* sp. symbiosis. *Prochloron* sp. is transmitted both vertically and horizontally. Occasionally it ends up on other invertebrates. Multiple strains of *Prochloron* producing different patellamides are present within a single didemnid colony and may be competing against each other employing patellamides.

Bottom left: transmission electron micrograph of the frontal section of a *Diplosoma similis* tadpole larvae from [9]. The location of *Prochloron* in the “algal pouch” is indicated.

## References

1. Aksoy, S. 1995. Molecular analysis of the endosymbionts of tsetse flies: 16S rDNA locus and over-expression of a chaperonin. *Insect Mol Biol* 4: 23-9.
2. Alberte, R. S., L. Cheng, R. A. Lewin. 1987. Characteristics of Prochloron/ascidian symbioses. II. Photosynthesis-irradiance relationships and carbon balance of associations from Palau, Micronesia. *Symbiosis* 4: 147-170.
3. Burns, B. P., F. Goh, M. Allen, B. A. Neilan. 2004. Microbial diversity of extant stromatolites in the hypersaline marine environment of Shark Bay, Australia. *Environ Microbiol* 6: 1096-101.
4. Chen, J. Y., D. Y. Huang, Q. Q. Peng, H. M. Chi, X. Q. Wang, M. Feng. 2003. The first tunicate from the Early Cambrian of South China. *Proc Natl Acad Sci USA* 100: 8314-8.
5. Clark, M. A., N. A. Moran, P. Baumann, J. J. Wernegreen. 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution Int J Org Evolution* 54: 517-25.
6. Feldman, R. A., M. B. Black, S. C. Cary, R. A. Lutz, R. C. Vrijenhoek. 1997. Molecular phylogenetics of bacterial endosymbionts and their vestimentiferan hosts. *Mol Mar Biol Biotechnol* 6: 268-277.
7. Ferris, M. J., G. Muyzer, D. M. Ward. 1996. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Appl Environ Microbiol* 62: 340-346.
8. Fu, X., T. Do, F. J. Schmitz, V. Andrushevich, M. H. Engel. 1998. New cyclic peptides from the ascidian *Lissoclinum patella*. *J Nat Prod* 61: 1547-1551.
9. Hirose, E. 2000. Plant rake and algal pouch of the larvae in the tropical ascidian *Diplosoma similis*: An adaptation for vertical transmission of photosynthetic symbionts *Prochloron* sp. *Zool. Science (Tokyo)* 17: 233-240.
10. Hirose, E., T. Maruyama, L. Cheng, R. A. Lewin. 1996. Intracellular symbiosis of a photosynthetic prokaryote, *Prochloron* sp., in a colonial ascidian. *Invertebr Biol* 115: 343-348.

11. Kline, T. C., R. A. Lewin. 1999. Natural  $^{15}\text{N}/^{14}\text{N}$  abundance as evidence of  $\text{N}_2$  fixation by *Prochloron* (Prochlorophyta) endosymbiotic with didemnid ascidians. *Symbiosis* 26: 193-198.
12. Kott, P. Replication in the ascidiacea: an adaptive strategy in the coral reef environment. In: *Proceedings of the Fourth International Coral Reef Symposium Manila 1981*; 1982: University of the Philippines, Quezon City; 1982. p. 725-733.
13. Kott, P. 2001. The Australian Ascidiacea Part 4, Aplousobranchia (3), Didemnidae. *Memoirs of the Queensland Museum*: 1-407.
14. Kott, P. 2004. New and little-known species of didemnidae (ascidiacea, tunicata) from Australia (part I). *J Natl Hist* 38: 731-774.
15. Lewin, R. A. 1976. Prochlorophyta as a proposed new division of algae. *Nature* 261: 697-698.
16. Lewin, R. A., L. Cheng. *Prochloron, a microbial enigma.*: Chapman & Hall; 1989.
17. Lim, G. E. *Bugula* (bryozoa) and their bacterial symbionts: a study in symbiosis, molecular phylogenetics and secondary metabolism [Ph.D.]. La Jolla CA: University of California, San Diego; 2004.
18. Lindquist, N., M. E. Hay, W. Fenical. 1992. Defense of ascidians and their conspicuous larvae: adult vs. larval chemical defense. *Ecol Monog* 62: 547-568.
19. Ochman, H., A. C. Wilson. 1987. Evolution in bacteria: evidence for universal substitution rate in cellular genomes. *J Molec Evol* 26: 74-86.
20. Olsen, R. R. 1986. Light-enhanced growth of the ascidian *Didemnum molle*/*Prochloron sp.* symbiosis. *Mar Biol* 93: 437-442.
21. Palenik, B., D. G. Swift. 1996. Cyanobacterial evolution and prochlorophyte diversity as seen in DNA-dependent RNA polymerase sequences. *J Phycol* 32: 638-646.
22. Peek, A. S., R. A. Feldman, R. A. Lutz, R. C. Vrijenhoek. 1998. Cospeciation of chemoautotrophic bacteria and deep sea clams. *Proc Natl Acad Sci USA* 95: 9962-9966.

23. Ridley, C. P., P. R. Bergquist, M. K. Harper, D. J. Faulkner, J. N. Hooper, M. G. Haygood. 2005. Speciation and biosynthetic variation in four dictyoceratid sponges and their cyanobacterial symbiont, *Oscillatoria spongelliae*. *Chem Biol* 12: 397-406.
24. Ronquist, F., J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
25. Salomon, C. E., D. J. Faulkner. 2002. Localization Studies of Bioactive Cyclic Peptides in the Ascidian *Lissoclinum patella*. *J Nat Prod* 65: 689-692.
26. Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, J. Ravel. 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.
27. Sings, H. L., K. L. Rinehart. 1996. Compounds produced from potential tunicate-blue-green algal symbiosis: a review. *J Ind Microbiol Biotechnol* 17: 385-396.
28. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucleic Acids Res* 25: 4876-4882.
29. Westheide, W., R. Rieger. *Spezielle Zoologie: Einzeller und Wirbellose Tiere*: Gustav Fischer Verlag; 1996.
30. Yokobori, S. I., A. Kurabayashi, B. A. Neilan, T. Maruyama, E. Hirose. 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.
31. Zeng, L., B. J. Swalla. 2005. Molecular phylogeny of the protochordates: chordate evolution. *Can J Zool* 83: 24-33.



## 4.

### **Trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum* predicted from the genome sequence**

#### **Abstract**

A gene cluster for the biosynthesis of a new small cyclic peptide, dubbed trichamide, was discovered in the genome of the global, bloom-forming marine cyanobacterium *Trichodesmium erythraeum* ISM101 because of striking similarities to the previously characterized patellamide biosynthesis cluster. The *tri* cluster consists of a precursor peptide gene containing the amino acid sequence for mature trichamide, a putative heterocyclization gene, an oxidase, two proteases and hypothetical genes. Based upon detailed sequence analysis, a structure was predicted for trichamide and confirmed by Fourier-transform mass spectrometry. Trichamide consists of 11 amino acids, including two cysteine-derived thiazole groups, and is cyclized by an N-C terminal amide bond. As the first natural product reported from *T. erythraeum*, trichamide shows the power of genome mining in the prediction and discovery of new natural products.

## Introduction

*Trichodesmium* is a genus of marine diazotrophic, non-heterocysteous cyanobacteria. It occurs throughout the open waters of oligotrophic tropical and subtropical oceans and forms filaments (trichomes) of 20-200 cells that can further aggregate into colonies several millimeters across. *Trichodesmium* can form blooms in excess of 100,000 km<sup>2</sup> [12], which are most commonly composed of *T. erythraeum* and *T. thiebautii*. *Trichodesmium* spp. have been the subject of intense research mainly for two reasons. First, they contribute a significant portion (40% or more, [12]) to global oceanic nitrogen fixation, thereby directly affecting the biogeochemical carbon flux in tropical oceans with implications for the world's climate [1]. Second, massive coastal *Trichodesmium* blooms have been reported to have toxic effects, both directly on invertebrates [8, 10] and on humans ("Trichodesmium or Tamandare fever", [21]) as well as indirectly by inducing blooms of other organisms [3, 14] that can be potentially harmful. While cyanobacteria are a prolific source of diverse natural products and toxins [2, 6, 18], to our knowledge a toxic compound (or any natural product) has not been isolated from a *Trichodesmium* species despite some efforts [9].

While most small peptides found in cyanobacteria are biosynthesized by non-ribosomal peptide synthases [17], we have recently reported a microcin-like pathway for the biosynthesis of a family of cyclic peptides, the patellamides, from *Prochloron*

*didemni*, a cyanobacterial symbiont of tropical ascidians (accession number: AY986476) [23]. The patellamides are moderately cytotoxic and composed of a pseudosymmetrical, cyclic dimer, with each substructure having the sequence thiazole-nonpolar amino acid-oxazoline-nonpolar amino acid. Despite these unusual features, patellamide biosynthesis is ribosomal. The *pat* gene cluster consists of a precursor peptide gene, which codes for the patellamide amino acid sequence, and a number of genes with protease and other peptide modifying homologies [23]. BLAST searches in GenBank with the *pat* genes revealed homologs in *T. erythraeum* IMS101. This led us to investigate the presence of a potential patellamide-like biosynthesis cluster as well as its product, a small cyclic peptide, dubbed trichamide in *T. erythraeum*.

## Materials and Methods

**Bioinformatics.** Most of the *T. erythraeum* IMS101 genome was shotgun sequenced by the Joint Genome Institute (JGI) and is available in GenBank (www.ncbi.nlm.nih.gov). The contig with accession number NZ\_AABK04000003 contains the *pat* homologs listed before [23]. Nucleotides 785,500 to 803,500 of this contig were downloaded and manually annotated in Artemis (Sanger Institute). Predicted ORFs were compared to the JGI auto-annotation and putative functions assigned by BLASTP on GenBank.

**Culturing.** *T. erythraeum* IMS101 [20] was obtained from John Waterbury, WHOI via Katherine Barbeau at SIO. The culture is non-axenic, i.e. does contain other heterotrophic bacteria (K. Barbeau, personal communication). Cultures were grown in R medium at 25°C under 12 hour light-dark photocycle with slow stirring as well as daily inversion of the culture flasks. R medium (modified from John Waterbury's recipe): 25% ddH<sub>2</sub>O and 75% natural sea water from Scripps pier are mixed and amended with 8 µM KH<sub>2</sub>PO<sub>4</sub>, 2.5 µM EDTA, 0.1 µM ferric citrate, 0.1 µM MnCl<sub>2</sub>, 10 nM Na<sub>2</sub>MoO<sub>4</sub>, 10 nM ZnSO<sub>4</sub>, 0.1 nM CoCl<sub>2</sub>, 0.1 nM NiCl<sub>2</sub>, and 0.1 nM Na<sub>2</sub>SeO<sub>4</sub>. All components are 0.2 µm filter-sterilized. *T. erythraeum* requires a 10% inoculum to start cultures; accordingly, 800 ml of culture were used in 8 liters of R medium. After 12-14 days, the culture was vacuum filtered through a 5 µm polycarbonate filter to retain *T. erythraeum* colonies and remove free-living bacteria. The cell material was rinsed off the filters into a 50 ml Falcon tube with ddH<sub>2</sub>O, immediately frozen at -80°C and later lyophilized. The average yield was ~10 mg dried cells per liter culture volume.

**Extraction and purification.** Lyophilized cyanobacterial pellets were extracted 3x with a ~100-fold excess of methanol. The methanolic extract was dried, yielding a crude extract that was used for initial electron spray ionization mass spectrometry (ESI-MS). For Fourier-transform MS (FT-MS), the crude extract was purified with a C<sub>18</sub> ZipTip (Millipore).

A portion of the crude methanolic extract (23 mg) was further purified by partitioning between ethyl acetate and water. The aqueous part was fractionated over a HP20SS column with 25, 50, 75 and 100% acetone. As determined by ESI-MS, the 25 and 50% acetone (aq.) fractions contained the 1099 peak and were combined. This combined fraction was run on a HPLC Phenomenex C<sub>18</sub> analytical column with the following protocol (all solvents contained 0.01% trifluoroacetic acid): 5 min of water, 5-35 min gradient from 0-100% acetonitrile, 10 min of 100% acetonitrile. Fractions were collected in minute intervals. Only fractions eluting at 16-17 and 17-18 minutes contained a 1099 peak as determined by ESI-MS. These fractions did not contain a single compound, since additional peaks beside 1099 were present in the MS. The amount of material in the two HPLC fractions was too low to measure.

In an improved procedure, a methanolic extract (57 mg) was partially purified by step gradient on a column containing 7 g C<sub>18</sub>, using solvents containing 0.01% trifluoroacetic acid. Fractions were eluted with water, followed by 25%, 50% and 100% acetonitrile (aq). The 100% elution fraction was further purified on a Phenomenex C<sub>18</sub> column as described before. A single peak with the correct diode array profile cleanly eluted at 16.6 min. By ESI-MS analysis, this HPLC-peak contained the 1099 ion. The concentration of trichamide was below a measurable limit and was thus estimated by comparison of the diode array absorbance at 240 nm with those for standards of ulithiacyclamide at varying concentrations. This intensity

should depend mainly upon the concentration of thiazole, since both ulithiacyclamide and trichamide have no other chromophores at this wavelength. By this method, the total amount of trichamide isolated was estimated to be 25-50 µg.

**Mass Spectrometry.** Crude extracts and partially purified fractions were monitored by ESI-MS and by FT-MS on a ThermoFinnigan LTQ-FT at 100,000 resolution (i.e. mass 400). FT-MS/MS experiments were run with collision-induced dissociation (CID) and infrared multiphoton dissociation (IRMPD) techniques. Predicted masses were calculated using the following values: C=12, H=1.007825, N=14.003074, O=15.994914, S=31.97207.

## Results and Discussion

**Biosynthetic genes.** Using genomic data available from GenBank, we have annotated a 12.5 kb gene cluster proposed to be responsible for the biosynthesis of trichamide (hence named *tri* cluster). The %GC is 40, higher than the average %GC of *T. erythraeum* at 34. On both sides it is bordered by tRNA-synthetase genes, potentially implicating horizontal gene transfer. The *T. erythraeum* genome is not closed, currently residing in 52 contigs at GenBank. The contig containing the *tri* genes (GenBank accession: NZ\_AABK04000003) is 842 kb long and also contains a number of ribosomal proteins. A BLAST analysis of the ribosomal proteins finds

similarities in other cyanobacteria, so it is assumed that this contig is indeed from *T. erythraeum* and not from a possible contamination by heterotrophic bacteria.

The *tri* cluster contains 11 ORFs designated *triA-K* (Figure 4.1 and Table 4.1). Four of these (*triBCEF*) are short and have sequence identity only to conserved hypothetical proteins, while *triI* is only hypothetical with no significant sequence identities. Some of these ORFs may not be actively transcribed. This paper will focus on the remaining six genes, for which function may be inferred.

The product of *triG* is the putative precursor protein. It was identified by two 5 amino acid motifs (GPGPS...SYDGD) that closely resemble the proposed cyclization signal found before and after the patellamide A and C sequences in the precursor protein of patellamide biosynthesis, PatE (Figure 4.2 and [23]). Analogous to patellamide biosynthesis, these motifs would define the borders of the eleven amino acid peptide, GDGLHPRLCSC. TriG also contains a leader sequence of 43 amino acids without similarities in GenBank except that five of the first six amino acids are identical to those of PatE.

*TriA* has high similarity to *patD*, which is proposed to be involved in heterocyclization of cysteine and/or threonine/serine into thiazoline and oxazoline

rings. The putative function was assigned on the basis of low sequence identity to previously characterized proteins: for the N-terminal part the adenylating enzyme MccB from microcin biosynthesis [7], for the C-terminal part a possible hydrolase, SagD from *Streptomyces iniae* [5].

TriD has high similarity to the N-terminal part of PatG and to oxidases. Previously, we predicted that this part of PatG would oxidize the intermediate thiazoline rings into thiazoles [23].

BLASTP analysis of TriH and K gives homology to subtilisin-like proteases. They have high similarity to PatA and the C-terminal part of PatG. We predict that these proteases would be involved in the maturation of PatE by cleaving the product from leader and trailer sequence and assume the same function in trichamide biosynthesis. It is interesting to note that TriH and TriK have 48% identity to each other.

TriJ has 50/72% similarity/identity to PatB. There is no other homolog to either of the two proteins in GenBank. PatB is not required for biosynthesis but seems to improve patellamide yield in heterologous expression experiments with the *pat* cluster (Eric Schmidt, unpublished results). The high identity between TriJ and PatB



over their entire length and presence in both clusters does suggest that they serve a role in peptide biosynthesis.

There are few differences between the *pat* and *tri* clusters: PatG has two domains: one for oxidation and one for proteolytic cleavage. In *T. erythraeum* these functionalities are separated into two proteins, TriD and TriH, respectively. The only *pat* gene without a homolog in the *tri* cluster (excluding very short putative ORFs) is *patF*, which has no significant homologies in GenBank. Overall, the *pat* and *tri* clusters have striking similarities. The biosynthetic genes have between 45-60% identity, and both gene clusters consist of a heterocyclization enzyme, an oxidase, two proteases and *patB/triJ*, a gene of unknown functionality. Also, while there is variability in the length of the precursor protein, both in terms of the leader sequence as well as in product sequence (8 amino acids for patellamide, 11 for trichamide), the five amino acid cyclization signals before and after the peptide are highly conserved.

Based upon these similarities in biosynthesis genes, we predicted the presence of a patellamide-like compound, trichamide, a cyclic, thiazole-containing peptide in *T. erythraeum*. Depending on the pattern of cyclization of the peptide and/or heterocyclization of serine and cysteine moieties, the possible molecular weight of the compound would be between 1079 and 1157.

**Mass spectrometry.** Initial screening of a crude extract of *T. erythraeum* with MALDI-TOF MS revealed the presence of a major peak at 1099. A molecule with this mass can be constructed from the precursor peptide sequence GDGLHPRLCSC by heterocyclization and oxidation of two of the three possible amino acids - cysteine, serine, cysteine - to thiazoles or oxazoles and cyclization of the entire peptide. Alternatively, this mass is also consistent with heterocyclization of the remaining amino acid to a thiazoline or oxazoline moiety in a linear peptide.

A high-resolution experiment on a Fourier-Transform MS/MS system gave a molecular ion at  $(M+H)^+ 550.23166$ , only 0.022 ppm different from the theoretical value of  $(M+H)^+ 550.231648$  for the predicted structure, validating the presence of a molecule containing the trichamide molecular formula  $C_{46}H_{66}N_{16}O_{12}S_2$  (supporting material).

Further MS/MS fragmentation experiments of mass 550.2 using collision-induced dissociation (CID) and infrared multiphoton dissociation (IRMPD) techniques revealed fragmentation patterns in congruence only with a cyclic peptide (Table 4.2 and Figures 4.6 and 4.7). With the exception of ion A, all masses are within ~3 ppm of their predicted values. This leaves three possible heterocyclization patterns that have identical mass: 1) thiazole-serine-thiazole; 2) thiazole-oxazole-cysteine; and 3) cysteine-oxazole-thiazole. The data are consistent with 1) on the

basis of three arguments: First, heterocyclization of adjacent amino acids has no precedent in the patellamide structural literature; in fact when two cysteine residues are adjacent in the patellamide family as in the ulithiacyclamides [4], only one is cyclized. Second, the patellamide class of compounds does not contain oxazoles, but only oxazolines. Third, it is highly unlikely that an enzyme would specifically modify one cysteine but not the other.

Because of the ribosomal mode of synthesis and in accordance with the patellamides, all of the amino acids in this molecule should adopt the L-configuration. Exceptions to this rule may be serine and leucine 2, which are adjacent to thiazole. These stereocenters readily undergo epimerization, and they are often found in either the D or L form in patellamides. We are currently examining the patellamide biosynthetic proteins to determine whether or not this epimerization is enzymatic. The proposed structure of trichamide is shown in Figure 4.3.

**Biosynthetic Pathway.** Closely paralleling patellamide biosynthesis, we predict the following pathway for trichamide biosynthesis (Figure 4.4). TriG is the precursor protein and forms the substrate for posttranslational modification by TriA, D, H and K. First, TriA modifies the cysteine moieties of TriG to form thiazoline groups. This could be an ATP consuming process as in microcin heterocyclization [16], needing the ATP hydrolysing functionality of the N-terminal part of TriA, while the reaction itself would be catalyzed by the uncharacterized C-terminal part. Next, TriD oxidizes thiazolines to thiazoles. TriA and TriK cleave the propeptide guided by

the conserved motifs GXXXS and XYDG. We propose that one protease cleaves the peptide bond after the header sequence leading to a free amide group. The other protease would cut the back end and catalyze a transpeptidation reaction between the two ends of the peptide leading to the mature cyclic form in a mechanism similar to the well characterized peptidoglycan cyclization by a serine protease, penicillin binding protein (reviewed in [22]). It is possible that the significant similarities between the two proteases allow them to form a dimer, which catalyzes both the hydrolysis of two peptide bonds and the cyclization in concert. It is interesting to note that the biosynthetic cluster of the linear peptide goadsporin [19] does not contain the two subtilisin-like proteases found in the *tri* and *pat* clusters, in agreement with an involvement of TriHK in cyclization. Recently Milne et al. published a computational study in which preorganization of patellamides were predicted to lead to cyclization and an enzyme would thus not be required [15]. The differences in size and sequence and the maintenance of dedicated proteases in patellamides and trichamide argue against this possibility. Finally, the absence of a PatF homolog in *T. erythraeum* and the requirement of PatF in patellamide biosynthesis (Eric Schmidt, unpublished data) implicate PatF in oxazoline formation, which is not part of the trichamide pathway.

The biosynthetic scheme presented here is in accordance with all available data. Due to the ribosomal mode of biosynthesis, i.e. the coding of the trichamide amino acids in *triG*, the link between the *tri* genes and trichamide is direct. Still,

ultimate proof of function could only be obtained by heterologous expression or knock-out experiments with the *tri* genes.

Patellamide and trichamide biosynthesis may be examples of a more common pathway to small peptides. Besides the aforementioned goadsporin from *Streptomyces* *sp.* TP-A0584, at the time of this writing clustered ORFs with 35-40% identity to TriA and D are present in the genomes of phylogenetically distant bacteria: plut\_0880 and 0878 in *Pelodictyon luteolum*, Chlorobia (GenBank accession: CP000096), swolDRAFT\_1502 and 1501 in *Syntrophomonas wolfei*, Chlostridia (GenBank accession: NZ\_AAJG01000002), and blr4538 and 4539 in *Bradyrhizobium japonicum*, Rhizobiales [11] (GenBank accession: BA000040).

**Trichamide function.** Trichamide is hydrophilic, partitioning to the aqueous fraction relative to ethyl acetate. In addition, it is found only in the cells and is not excreted in significant quantities to the growth medium (data not shown). These properties suggest an antipredation defense function, rather than anticompetitor or communication functions. To test biological activities, *T. erythraeum* crude methanolic extracts were tested for general cytotoxicity (HCT-116 at 10 µg/ml and CEM-TART at 5 and 50 µg/ml) and anti-HIV (1 and 10 µg/ml), antifungal (*Candida albicans* at 10 µg/ml) or antimicrobial (*Staphylococcus aureus* and *Enterococcus faecium* at 10 µg/ml) effects. No significant activity was found in these assays (data

not shown). A number of algal blooms have neurotoxic effects [2] and neurotoxicity of environmental *Trichodesmium* sp. in mice has previously been reported [9]. The crude methanolic extract of *T. erythraeum* IMS101 also exhibited neurotoxicity in a mouse assay, but purified trichamide was not the active component (data not shown). Guo and Tester have found that healthy *Trichodesmium* sp. cells do not affect the copepod *Acartia tonsa*, while aged or lysed *Trichodesmium* cells are toxic [8]. This result is consistent with the properties of trichamide, which suggest that the compound is maintained inside healthy cells, but would be released into seawater from lysed cells. Testing the effect of trichamide on health and feeding behavior of copepods and other grazers of *Trichodesmium* might reveal the ecological function of trichamide. The presence of trichamide should be examined in *Trichodesmium* bloom waters. It would also be interesting to determine if *T. thiebautii*, the other major *Trichodesmium* species, which in one report has proven more toxic than *T. erythraeum* [10], contains trichamide (or a related compound) and the necessary biosynthetic capabilities.

**Conclusion.** The ongoing exponential growth of DNA sequence data will lead to the discovery of many natural product biosynthesis pathways for which no actual product has been characterized. Careful study of these pathways can lead to discovery of novel products. Challis et al. identified a non-ribosomal peptide gene (NRPS) cluster in the genome of a *Streptomyces* and discovered a novel compound, coelichelin [13]. NRPS pathways are well-known routes to natural products and the

genus *Streptomyces* is a prolific source. The genus *Trichodesmium* was not previously known to produce natural products and trichamide is only the second example, after the *pat* cluster of a cyclic peptide biosynthesized this way. Depending on the type of pathway, genomic mining should encompass careful curation. While the auto-annotation of the *T. erythraeum* genome identified most of the *tri* genes as ORFs of unknown function, the most essential part of the cluster, the precursor peptide gene, was discovered manually.

This study is also an example for the power of general and timely access to genomic data. Even though at the time of this writing the *T. erythraeum* IMS101 genome is not completely sequenced or published, the public availability of the draft sequence data allowed assignment of function to the *tri* gene cluster and the discovery of a novel cyclic peptide.

\*\*\*\*\*

Kelly Roe and Kathy Barbeau provided advice on *Trichodesmium* culturing. Bioassays were run in the labs of Louis Barrow and Baldomero Olivero at the University of Utah and William Fenical and William Gerwick at Scripps. Chad Nelson at the University of Utah ran the mass spectrometry measurements. This

chapter is a reprint of the material as it appears in *Applied Environmental Microbiology* 2006, 72: 4382-4387 with co-authors Margo G. Haygood, Diaa T. A. Youssef, and Eric W. Schmidt. The dissertation author was the primary investigator.



**Table 4.1:** the *tri* cluster proteins and their homologs.

protein	length (aa)	Homolog, GenBank accession	%identity/%similarity	Predicted function	GenBank accession
TriA	769	PatD, AAY21153	57/70	Adenylation/heterocyclization	ZP_00672901
TriB	112	Conserved hypothetical, NP_942321	53/70	-	ZP_00672900
TriC	124	Conserved hypothetical, BAB73591	60/78	-	ZP_00672899
TriD	475	PatG, N-terminal, AAY21156	45/59	oxidase	ZP_00672897
TriE	106	transposase-like, ZP_00345329	79/85	-	ZP_00672896
TriF	112	Conserved hypothetical, ZP_00675293	78/91	-	ZP_00672895
TriG	67	None	-	Precursor protein	794178..381 of NZ_AABK04000003
TriH	666	PatA, AAY21150	60/72	Subtilisin-like protease	ZP_00672894
TriI	72	-	-	-	ZP_00672893
TriJ	71	PatB, AAY21151	52/70	-	ZP_00672892
TriK	702	PatG, C-terminal, AAY21156	48/64	Subtilisin-like protease	ZP_00672891

**Table 4.2:** Mass spectrometry of trichamide and fragments. Proposed peptide structure are in one-letter amino acid code, Thia = cysteine modified to thiazole, FT-MS = Fourier-transform mass spectrometry, CID = collision induced dissociation, IRMPD = infrared multiphoton dissociation. Artifacts intrinsic to the machine and visible in other spectra of unrelated molecules constitute the other major peaks in the spectrum and are not tabulated here.

	ion	proposed structure	Observed mass	Theoretical mass	difference (ppm)
<b>FT-MS</b>	<b>I</b>	M+H*2, GDGLHPRL-Thia-S-Thia	550.23166	550.231648	0.022
	<b>II</b>	M+H*2, <sup>34</sup> S	551.22845	551.22955	2.0
	<b>III</b>	M+H, <sup>13</sup> C <sub>2</sub>	551.23627	551.2350025	2.3
<b>CID-MS/MS of 550.2</b>	<b>A</b>	M+H*2 of ion F	481.20212	481.19883	6.8
	<b>B</b>	M+H*2 of parent ion minus C=O	536.23520	536.23520	0.0
	<b>C</b>	M+H, PRL-Thia-S-Thia-GD	792.29387	792.29213	2.2
	<b>D</b>	M+H, PRL-Thia-S-Thia-GDG	849.31559	849.31359	2.4
	<b>E</b>	M+H, PRL-Thia-S-Thia-GDGL	962.39975	962.39766	2.2
	<b>F</b>	M+H, HPRL-Thia-S-Thia-GDG	986.37429	986.37250	1.8
<b>IRMPD-MS/MS of 550.2</b>	<b>G</b>	M+H*4 of parent ion	275.11674	275.115824	3.3
	<b>H</b>	M+H, G-L-H	308.17152	308.17220	2.2
	<b>I</b>	M+H*2 of parent ion	550.23035	550.231648	2.4
	<b>J</b>	M+H, RL-Thia-S-Thia-GD	695.23870	695.239367	1.0
	<b>K</b>	as ion C	792.29121	792.29213	1.2
	<b>L</b>	M+H, L-Thia-S-Thia-GDGLH	846.30172	846.302694	1.2
	<b>M</b>	as ion E	962.39676	962.39766	0.9



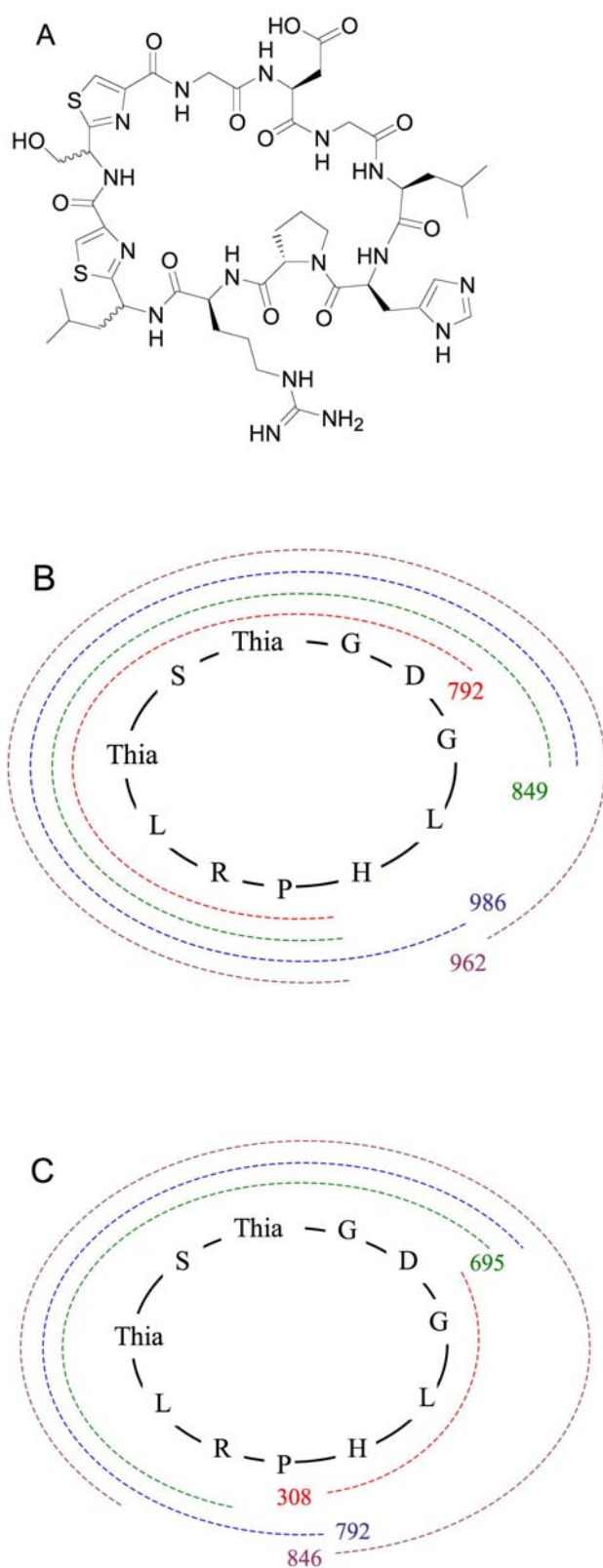
**Figure 4.1:** The *tri* gene cluster. Arrows denote ORFs and their direction, black ORFs are tRNA synthetases, white ORFs are conserved hypothetical without homolog in the pat cluster, green ORFs are pat homologs, the precursor peptide gene is in orange.

```

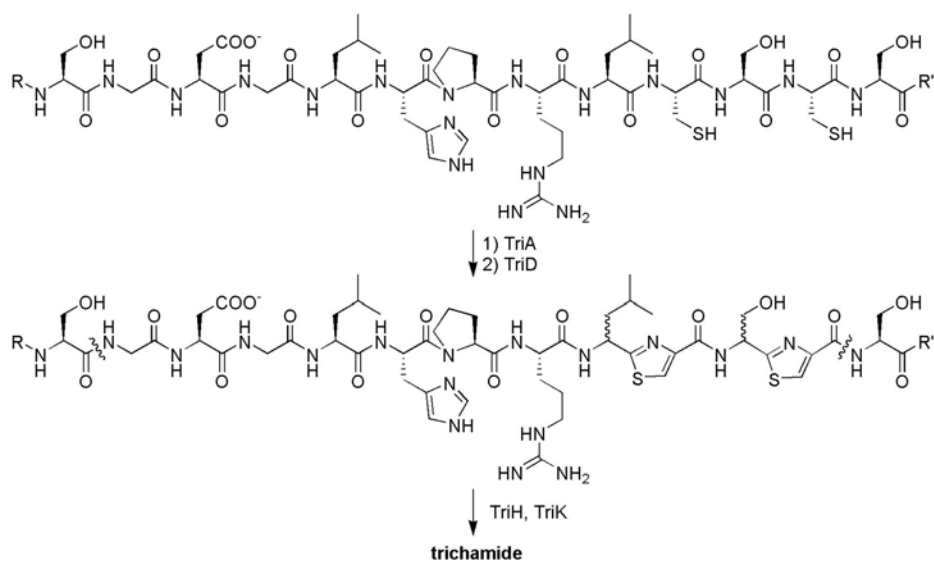
>PatE  MNKKNI-----LPQQGQPVIRTAGQLSSQLAELSEEALGDAGGLEASVTACITFC---AYD
                                             GVEPSITVCISVC---AYDGE
>TriG  MGKKNIQPNSSQPVFRSLVARPALEELREENLTEGNQGHGPLANGGPGPSGDGLHPRLCSCSYDGDE

```

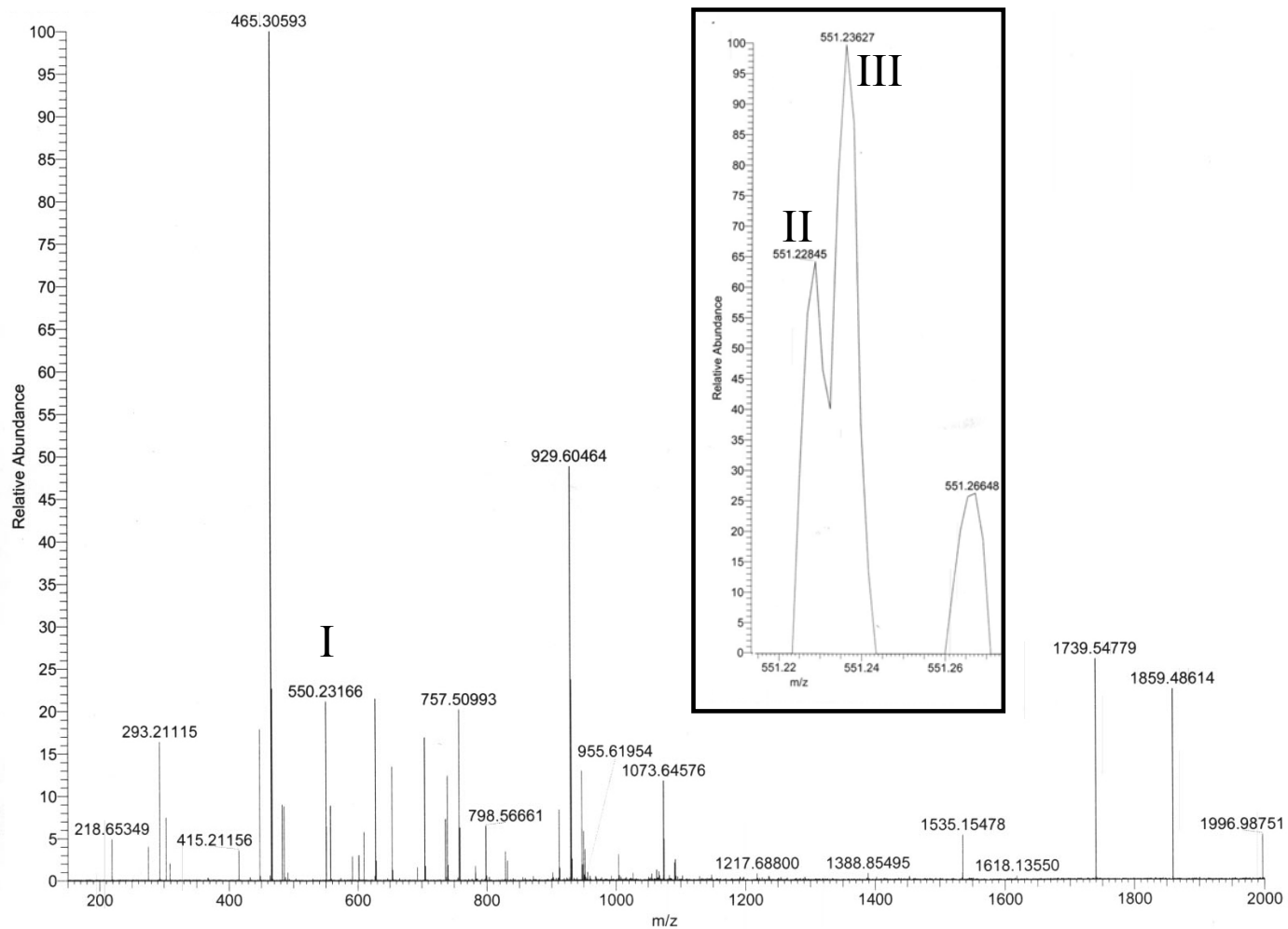
**Figure 4.2:** Alignment of the precursor peptides PatE and TriG. The sequence encoding patellamide C, patellamide A and trichamide (top to bottom) is underlined, proposed cyclization signal is in bold.



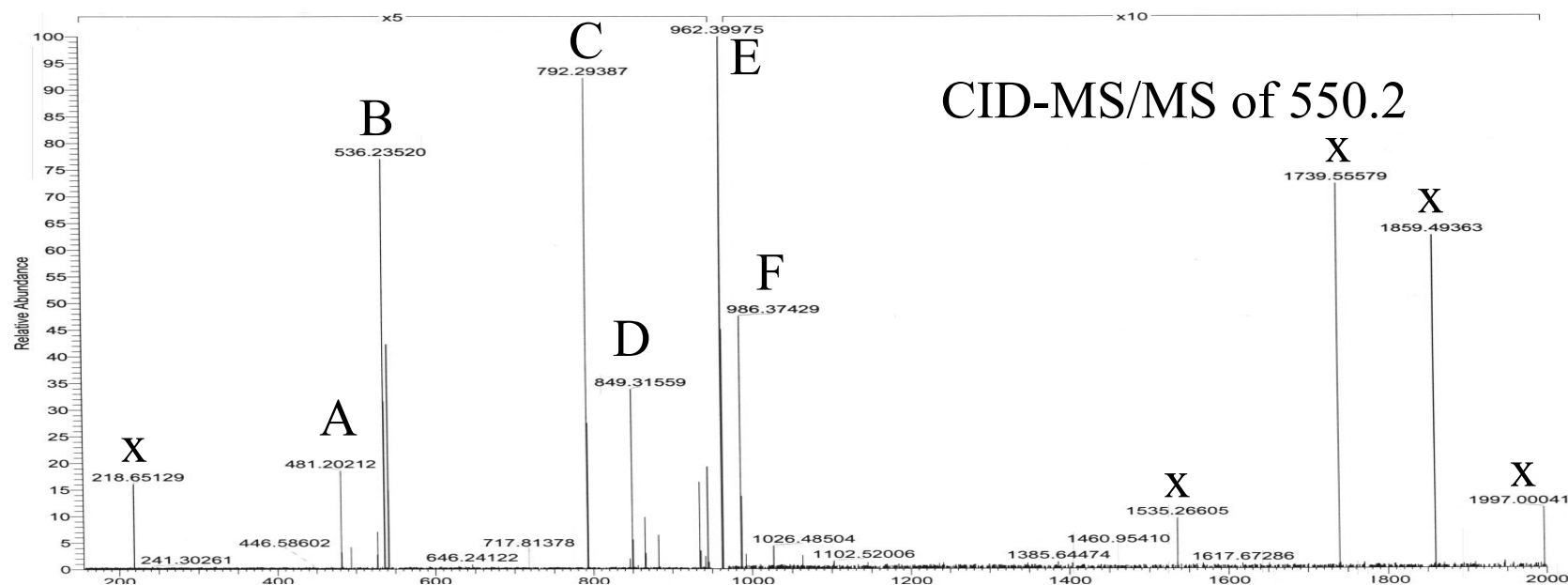
**Figure 4.3:** A. Structure of trichamide. Stereochemistry is inferred, not determined experimentally, as described in the text. B. Assignment of CID-MS fragments from Table 4.2 to the trichamide structure. C. Assignment of IRMPD-MS fragments.



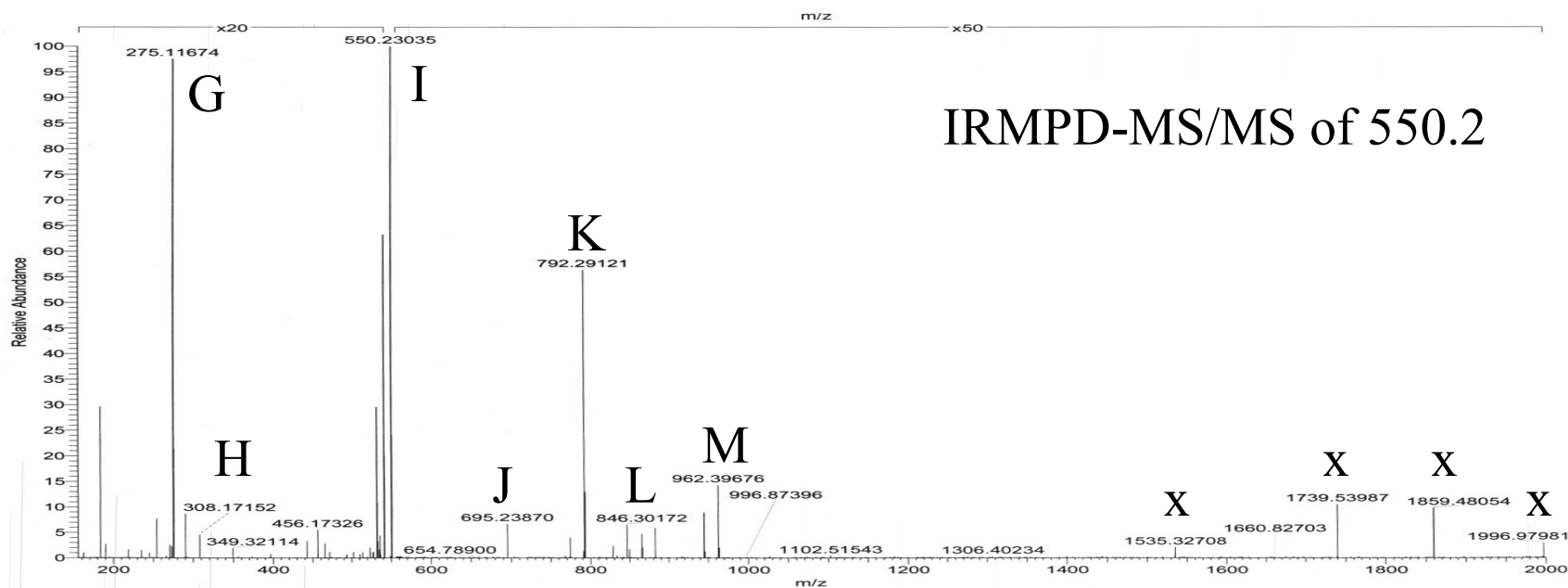
**Figure 4.4:** Proposed biosynthetic pathway to trichamide.



**Figure 4.5:** Fourier transfer –mass spec of crude *Trichodesmium* extract. Peaks are present for the trichamide parent ion (I), the  $^{34}\text{S}$  isotope (II) and the  $^{13}\text{C}_2$  isotope (III).



**Figure 4.6:** MS fragmentation patterns under CID (collision-induced dissociation) of ion 550.2. Peaks labeled “x” are artifacts of the instrument, present in all spectra measured no matter the sample type or source. All other ions can be accounted for as in Table 4.2 and Figure 4.3.



**Figure 4.7:** MS fragmentation patterns under IRMPD (infrared multiphoton dissociation) of ion 550.2. Peaks labeled “x” are artifacts of the instrument, present in all spectra measured no matter the sample type or source. All other ions can be accounted for as in Table 4.2 and Figure 4.3.



## References

1. Capone, D. G., J. P. Zehr, H. W. Paerl, B. Bergmann, E. J. Carpenter. 1997. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* 276: 1221-1229.
2. Carmichael, W. W. 1992. Cyanobacteria secondary metabolites - the cyanotoxins. *J Appl Bacteriol* 72: 445-459.
3. Devassy, V. P., P. M. Bhattathiri, S. Z. Qasim. 1979. Succession of organisms following *Trichodesmium* phenomenon. *Indian J Mar Sci* 8: 88-93.
4. Fu, X., T. Do, F. J. Schmitz, V. Andrushevich, M. H. Engel. 1998. New cyclic peptides from the ascidian *Lissoclinum patella*. *J Nat Prod* 61: 1547-1551.
5. Fuller, J. D., A. C. Camus, C. L. Duncan, V. Nizet, D. J. Bast, R. L. Thune, D. E. Low, J. C. De Azavedo. 2002. Identification of a streptolysin S-associated gene cluster and its role in the pathogenesis of *Streptococcus iniae* disease. *Infect Immun* 70: 5730-9.
6. Gerwick, W. H., L. T. Tan, N. Sitachitta. 2001. Nitrogen-containing metabolites from marine cyanobacteria. *Alkaloids Chem Biol* 57: 75-184.
7. Gonzalez-Pastor, J. E., J. L. San Millan, M. A. Castilla, F. Moreno. 1995. Structure and organization of plasmid genes required to produce the translation inhibitor microcin C7. *J Bacteriol* 177: 7131-40.
8. Guo, C., P. A. Tester. 1994. Toxic effect of the bloom-forming *Trichodesmium* sp. (cyanophyta) to the copepod *Acartia tonsa*. *Nat Toxins* 2: 222-7.
9. Hawser, S. P., E. J. Carpenter, G. A. Codd, D. G. Capone. 1991. A neurotoxic factor associated with the bloom-forming cyanobacterium *Trichodesmium*. *Toxicon* 29: 277-278.
10. Hawser, S. P., J. M. O'Neil, M. R. Roman, G. A. Codd. 1992. Toxicity of blooms of the cyanobacterium *Trichodesmium* to zooplankton. *J Appl Phycol* 4: 79-86.

11. Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada, S. Tabata. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 9: 189-197.
12. Karl, D., A. Michaels, B. Bergman, D. Capone, E. Carpenter, R. Letelier, F. Lipschultz, H. Paerl, D. Sigman, L. Stal. 2002. Dinitrogen fixation in the world's oceans. *Biogeochemistry* 57-58: 47-98.
13. Lautru, S., R. J. Deeth, L. M. Bailey, G. L. Challis. 2005. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol* 1: 265-269.
14. Lenes, J. M., B. P. Darrow, C. Cattrall, C. A. Heil, M. Callahan, G. A. Vargo, R. H. Byrne, J. M. Prospero, D. E. Bates, K. A. Fanning, J. J. Walsh. 2001. Iron fertilization and the *Trichodesmium* response on the West Florida shelf. *Limnol Oceanogr* 46: 1261-1277.
15. Milne, B. F., P. F. Long, A. Starcevic, D. Hranueli, M. Jaspars. 2006. Spontaneity in the patellamide biosynthetic pathway. *Org Biomol Chem* 4: 631-638.
16. Milne, J. C., A. C. Eliot, N. L. Kelleher, C. T. Walsh. 1998. ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. *Biochemistry* 37: 13250-13261.
17. Moore, B. S. 2005. Biosynthesis of marine natural products: microorganisms (Part A). *Nat Prod Rep* 22: 580-593.
18. Namikoshi, M., K. L. Rinehart. 1996. Bioactive compounds produced by cyanobacteria. *J Ind Microbiol* 17: 373-384.
19. Onaka, H., M. Nakaho, K. Hayashi, Y. Igarashi, T. Furumai. 2005. Cloning and characterization of the goadsporin biosynthetic gene cluster from *Streptomyces* sp. TP-A0584. *Microbiology* 151: 3923-33.
20. Prufert-Bebout, L., H. W. Paerl, C. Lassen. 1993. Growth, nitrogen fixation, and spectral attenuation in cultivated *Trichodesmium* species. *Appl Environ Microbiol* 59: 1367-1375.
21. Sato, S., M. N. Paranangua, E. Eskinazi. 1963. On the mechanism of red tide of *Trichodesmium* in Recife north eastern Brazil, with some considerations of the relation to human disease Tamandare Fever. *Trab do Instit Oceanogr Univ Fed de Pernambuco Recife* 5/6: 7-50.

22. Scheffer, D., M. G. Pinho. 2005. Bacterial cell wall synthesis: new insights from localization studies. *Microbiol. Mol Biol Rev* 69: 585-607.
23. Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, J. Ravel. 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: 7315-20.

## Appendix

### The Tale of the *B. neritina*

By Mark Hildebrand, Fall 2003  
(to the tune of “The Wreck of the Edmund Fitzgerald” by Gordon Lightfoot)

Well, some time ago, a PI named Margo  
Had a dream about curing the cancer,  
Well she looked to the sea, and what did she see  
That *B. neritina* was the answer  
Well it wasn't the host that attracted her most  
But a tiny bacteria inside it,  
As we all know, Wollacot's larvae did show  
That they lived in the pallial sinus

Seana started one day, and got DNA  
From larvae in different locations  
She amplified, and began to surmise  
16S permutations  
Two populations there'd be, both shallow and deep  
Of gamma proteobacteria  
In situ's were done, and it was then named  
*Endobugula sertula*

More evidence was needed, to show that indeed  
Bryostatin was from this bug  
If we could cultivate, it sure would be great  
An endless supply of the drug  
Or clone out the genes, what a novel scheme  
And express them in another host  
Bryopyran rings, and other things  
Man that surely would be the most

Scott Allen came to town, by PCR bound  
To clone out the KSa  
Degenerate primers and touchdown designer  
Eventually got it to stay  
But he decided, to join the other side  
And got lucrative offer  
“My lack of finance, won't let me skip the chance  
To fill up my personal coffers”

Several interacted, to write a new chapter  
 In the Endobugula caper  
 PCR surveys, and other forays  
 The supporting evidence paper  
 Antibiotic, gave *E. sertula* a lick  
 Its levels fell to the basement  
 Bryo activity also dropped, but it didn't stop  
 As judged by phorbol displacement

Mark's grant had run out, when Scott left town  
 The timing seem predestined  
 Margo then explored, if he'd like to come on board  
 He said yes, thanks for asking the question  
 That's awfully kind, 40% time  
 My schedule I'll be rearranging  
 Little did he know, what lay in store  
 Or how his life would be changing

All started OK as he isolated DNA  
 Of sufficient molecular weight  
 Got into a groove, and then did remove  
 The pigment by cesium gradient  
 Partials then performed, they came quite normally  
 Just a little indigestion  
 Sucrose gradients came next, from which you select  
 The fragments for your ligation

The pigment by the way, was of interest I say  
 Margarita showed it absorbed in UV  
 The bad rays will go, inhibit the glow  
 Applied to the face liberally  
 Consider the fact that it had bryostatin  
 And was brown in color  
 Self-tanning yes, it will be the best  
 Way to prevent skin cancer

Our research to speed, we found that we needed  
 A way to grind that was groovy  
 Well come on lets move on, buy a Polytron  
 And let's make Bugula smoothies  
 They splattered and spilled, but the grinding was a thrill  
 The whirl of our success  
 Bacteria intact, but the host was shellacked  
 My goodness what a mess

Differential spins, host debris did win  
 The race to become the pellet  
 The bacteria did float, along with the foam  
 A way for them to embellish  
 Check enrichment now, competitor is how  
 Another of Scott's contributions  
 For every problem we faced, a step we'd take  
 Towards coming up with the solutions

The pace went so well, there's really no telling  
 How fast we'll make bryostatin  
 These thoughts were arising, but on the horizon  
 A sinister thing did happen  
 The cosmids you see, were high copy  
 And subject to deletions  
 They generated a gap, which started to tap  
 Our confidence into submission

Filling in the gaps, was one of the traps  
 That soon began taking its toll  
 They swallowed you see, time and energy  
 Like a supermassive black hole  
 But we toiled away, by night and by day  
 It became an incredible feat  
 When we were finally done, we found more than one  
 Enormous, humongous repeat

Along the way we discovered one day  
 The end of the PKS cluster  
 When we saw the last, four amino acids  
 Wouldn't you just go and figure!  
 Histidine glutamic, then double leucine  
 A word this sequence did spell  
 It seemed most fitting, this devilish fiend's  
 Sequence ended in HELL

We realized one day, we had cloned every way  
 Possibly known to man  
 There was ZAP, and DASH, cosmids and plasmids  
 All became part of the plan  
 They all fell short, I have to report  
 None of them got us to there  
 After all this work, there aren't many perks  
 It seems like we're getting nowhere

Summers were spent, cloning in BACs  
 Or pulsed field separations  
 Success was absent; time would've been better spent  
 On extended four month vacations  
 Mark said with a frown, Bugula's wearing me down  
 I'm losing my sense of calm  
 But what kept him going, in background was knowing  
 He'd work on his diatoms

When he tried to construct, ORF5 in a pUC vector  
 It was inherently unstable  
 He'd meet with Margo, and tell her so  
 These things just aren't normal  
 When last he was seen, he had turned to yeast  
 For ORF5 expression  
 If these experiments fail, I can at least brew some ale  
 And drink away all my frustrations

When Laura came on board, we felt reassured  
 That the project really was growing  
 From worms did she come to the neritina fun  
 She got into without really knowing  
 What lie ahead, it could never be said  
 She knew what was in her future  
 But the curse wouldn't hide, it would only decide  
 On the particular brand of her torture

She extracted DNA and enriched it one day  
 By Hoechst dye centrifugation  
 Tiny fractions she collected, and read them in the spec  
 'till sufficient accumulation  
 She pooled all those, and by PCR showed  
 Where the KSa was located  
 Once they were combined, she next spent her time  
 On Southernns she was dedicated

Well the Southernns looked great, we all were elate  
 At the beautiful signals appearing  
 She straightened out the map, we began to clap  
 And some of us started cheering  
 But the curse is so tough; it said "that is enough  
 I'll turn their joy into sorrow  
 Happy now they'll be, but just wait and see  
 What's in store for them tomorrow"

In the summer of '01, we collected a ton  
 Of the B. neritina larvae  
 Our goal to ensure, that the DNA'd be pure  
 As we strove to make our libraries  
 Laura did the preps, we felt sure they'd be the best  
 For any such use we could come by  
 For Southernns and clones, and even genomes  
 Our spirits were soaring high



She collected the brown, larvae and froze them down  
Preparing them for the extraction  
The DNA came out, and enough no doubt  
For cesium centrifugation  
When the gradients came down, our smiles turned to frowns  
The DNA had disappeared  
Explanations we had none, except for one  
The Bermuda Triangle moved westward

To further her pains, she expressed some domains  
For antibody preparation  
For first the amount, you could not count  
It was very low expression  
That problem was solved and she got more involved  
And worked with sincere devotion  
Six-His purified, the curse then told a lie  
That these really were her proteins

We later found out, after months of doubt  
That E. coli had harbored a sleeper  
Activated by the curse, it made things worse  
Our despair became even deeper  
The protein she'd cleaned, was just part of the scheme  
Of the curse for our disgrace  
Some literature search confirmed the worst  
It was superoxide dismutase

Then there was Carolyn, who led a harrowing  
Existence running the show  
She kept us organized, and well supplied  
And assisted PI, Margo  
When Bugula we desired, an assistant by her side  
They gathered it off of the pier  
I'm sure that their dive, brought some excitement  
Compared with her duties here

She recently was honored, a special award  
 For her work with EH and S  
 The plaque that they gave her, I think that it said  
 She really can clean up a mess  
 To Carole Kent she'd say, we'll do it your way  
 But it was said with a wink  
 Carole's gone let's pollute, and ethanol dilute  
 Before pouring it down the sink

As all of us know, we had video  
 That was held on a regular schedule  
 Where we sat in a room, sort of a temple of doom  
 And our every movement was so visible  
 It was such a fright, to sit under the lights  
 Radiating a shine from our faces  
 There for each one to see, without flattery  
 We all could have used some makeup

At least the room was tip top, but Minnesota's was not  
 Their microphones, they didn't matter  
 The images we received, well they made us grieve  
 'Twas like looking through fathoms of water  
 "We can't see a thing in your cloning scheme  
 But your stories really compel us  
 Just download the images from the Discus  
 We follow along as you tell us"

Some good can be said for the regular way  
 We gathered together in earnest  
 The camaraderie was good, and we knew we would  
 Hear from David Sherman  
 He's a very busy man, with lots of plans  
 And his mind, it is always moving  
 In barely a wink, 'fore the meeting he'd think  
 "Now just exactly what are we doing?"

At one of these things Haibin gave a scheme  
 Explaining how all fit together  
 It was quite creative, we were all elated  
 It seemed to avert disaster  
 Unconventional it was, but so was the cluster  
 It didn't have any ATs  
 A module to load, was part of the code  
 And NRPS replaced the TE

It jumped around, until it found  
 A partner domain  
 Loading module lactate, it really was great  
 A test had come up in his brain  
 One way to see, was with DEBS-TE  
 Experiments that he started  
 We all were encouraged, but then got discouraged,  
 'Twas then that Haibin departed

Deanne spent part time, chopping and grinding  
 Bugula ad infinitum  
 Several spins, then Percoll she'd begin  
 In the hopes of purification  
 Her ultimate goal, was to culture the sole  
 Bacteria of our interest  
 What she found instead, was *E. sertula* was dead  
 Using different staining treatments

Christine came along for a summer SURFF  
 To work in the lab for a few weeks  
 In situ's she did, they were publish'ed  
 She thought graduate school would be a breeze  
 She applied and came in, and thought she would win  
 Her degree in a short/sweet duration  
 Little did she know, from way down below  
 The curse had just set its hooks in

It was weeks, it was days, doing RPAs  
 To test if the cluster was active  
 They'd work, then they didn't, and on infinite  
 These experiments were an awful disaster  
 RNA's a fickle thing, just to begin  
 And mighty tricky to handle  
 But on top of it all, the curse came to call  
 Making Christine's life all scrambled

So she searched for a scheme, to fulfill her dreams  
 And thought about genomics  
 "If we sequence it all, the mysteries will fall  
 And I will complete my thesis"  
 She and Liz gave a go, to do some flow  
 Cytometric genomic comparisons  
 The genome was small, but then came the fall  
 For the curse had just started its harassment

The grant didn't come through, what else is new  
 The challenge wasn't that simple  
 Other bugs for sure, can be isolated pure  
 But not so for our dear *E. sertula*  
 This bug is unclean, and is always seen  
 With a host of contamination  
 When you think about it, purity will never be writ  
 On the tomb of this small abomination

Now Grace tread with care, she didn't dare  
 Fully expose herself to the danger  
 She did some trees and activities  
 But otherwise remained a stranger  
 To the neritina curse, avoiding the worst  
 She was smarter than most of us  
 Simplex and turrida, but not neritina  
 Are species that she did trust

When simplex gave KSa, it was quite a day  
 You simply must go ahead and clone it  
 She said "hmmm; let me think I'm right on the brink  
 Of finishing up my thesis  
 Well I just won't have time, it's awfully kind  
 Of you to make that suggestion  
 But I'll be leaving soon, and some poor buffoon  
 Will have to tackle that question"

Then one day there came, Sebastian's the name  
 A native German national  
 A hard working lad, he jumped into the lab  
 Thinking that research was rational  
 My poor boy don't you see, this is Bugulee  
 You're bound to become quite vexed  
 There's a curse you will find, it's a matter of time  
 'fore you won't know what's coming next

But he went for the deep, right into the breach  
 To clone the PKS cluster  
 But deep Bugula was worse, it had its own curse  
 And few E. sertula it could muster  
 He spent quite a long time, collecting and grind  
 -ing all the material he wanted  
 Yields were pitifully spare, and he was not aware  
 That this project was certainly haunted

I'll take a short cut, and try my luck  
 By doing long range PCR  
 In just a few days I'll go all the way  
 And eclipse the sequence of Mark's  
 Well the PCR's worked, with one little quirk  
 T'was a wonderful schema  
 You think so, OK but the DNA  
 Had more mutations than Hiroshima

He then relegated to clone, set out on his own  
 With guidance from shallow sequence  
 As the years slipped by, he said with a sigh  
 I'm losing my sense of reason  
 An example of the trap, he couldn't fill a gap  
 By partial digestion coverage  
 Well these two Sau3A's, were 4 kb away,  
 That's 16 times farther than average

Well Sebastian we've heard, will be traveling abroad  
 To deal with his Germanic visa  
 Listen to me man, get out while you can  
 From the curse of the B. neritina  
 Use your trip overseas, to set sail on a fresh breeze  
 And search for a new horizon  
 A new project you should find, if you want to keep your mind  
 Don't return to your third floor dungeon

Jen took a dare to repeat Russell Kerr's  
 Experiments in bryo labeling  
 Label and extract, and then rotovap  
 To increase the concentration  
 Then count and see, and do TLC  
 To find out if something happened  
 After a few months, of coming up skunked  
 She ran clear out of the building

Now quite recently, Koty gave us to see  
 Her confocal images were awesome  
 E. sertula abounded, within and around  
 Like some cosmic fluorescent blossom  
 Was it Cy 5 or sci-fi, I couldn't rely  
 On experience to make it clear  
 At any minute, I expected a ship  
 Like the Enterprise to appear

The pictures were stunning, my mind started running  
How can the curse let this be?  
Then I realized in the blink of an eye  
The curse had selectivity  
I'll tell you this much, it says look but don't touch  
With your molecular tools  
Touch my DNA, and I'll blow you away  
And make you all look like fools

Progress the years and all of our fears  
Will fade into the past  
But I can't say, that they'll ever go away  
Such horrors generally last  
As we lay in our beds, thoughts will come to our heads  
Of our days on La Jolla Shores  
We'll awake with a scream, it was only a dream  
Of being engulfed by lophophores

When we're about to collapse, from failed synapses  
We'll tell our grand – sons and daughters  
To them we'll regale, this tremendous tale  
Of our sailing on treacherous waters  
With lots more research, we'll iron out the quirks  
And finally get to the answer  
They'll be lots of Bryo, and we'll use it to show  
How we can cure our own cancer