Investigating Natural Product Biosynthesis in Uncultivated Symbiotic Bacteria of the Marine Sponge *Theonella swinhoei*

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Agustinus Robert Uria

aus Tambarana, Zentral Sulawesi, Indonesien

Universität Bonn

2012

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Gutachter:

1. Prof. Dr. Jörn Piel

2. Prof. Dr. Uwe Deppenmeier

3. Prof. Dr. Albert Haas

4. Prof. Dr. Gabriele M. König

Tag der Promotion: 05.12.2012

Erscheinungsjahr: 2013

Dedicated to my parents with love and gratitude

Abstract

Marine sponges are a rich source of bioactive natural products with potent anticancer activities. Currently, the limited availability of most of these substances prohibits further drug development (Proksch *et al*., 2002). Highly complex consortia of bacterial symbionts associated with sponges have been frequently proposed to be the true producers of many secondary metabolites (Piel, 2004). However, the majority of these complex microbial assemblages are not amenable to cultivation (Amann *et al*., 1995; Hugenholtz *et al*., 1998; Friedrich *et al*., 2001; Webster & Hill, 2001), thereby hampering efforts to prove the symbiont hypothesis as well as to access their biosynthetic potential. Using metagenomic-based approaches, Piel and colleagues have previously provided the first genetic evidence for the bacterial origin of invertebrate-derived natural products by cloning the entire gene cluster for onnamide/theopederin from the Japanese sponge *Theonella swinhoei* (Piel *et al*., 2004a, 2004b).

The Japanese sponge *Theonella swinhoei* W

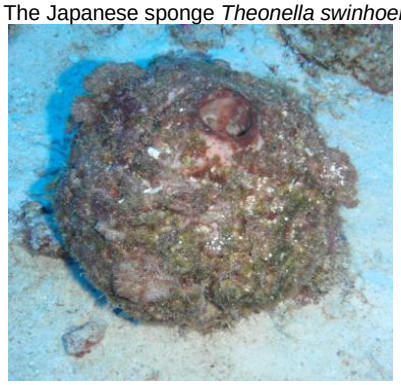


Photo by J. Tanaka

The Japanese sponge *Theonella swinhoei* Y

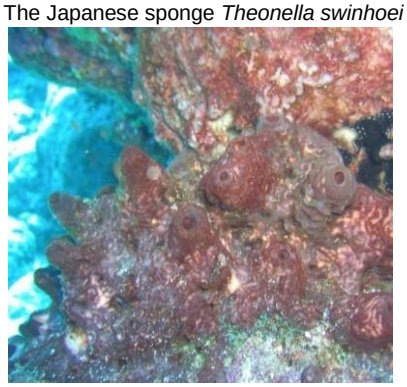
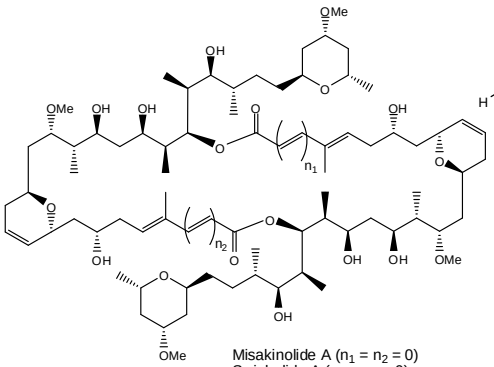


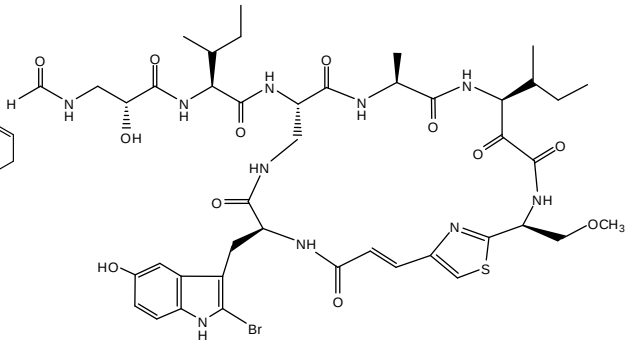
Photo by Y. Nakao

In this work, we investigated further natural product biosynthetic pathways from uncultivated symbiotic bacteria using Japanese *T. swinhoei* as a symbiotic assemblage model. The reasons for selecting this sponge are the wide variety of pharmaceutically important secondary metabolites isolated from this sponge as well as the high complexity of the associated bacteria (Fusetani & Matsunaga, 1993; Henstchel *et al*., 2002), which might play an important role in metabolite biosynthesis. Metagenome mining strategies that we applied and developed in this work have led to the cloning of two new biosynthetic pathways from this complex symbiosis model. Our bioinformatic analysis predicted that one pathway is responsible for the biosynthesis of misakinolide A, and another one for keramamide H. Interestingly, we found that the first pathway contains additional components that

match structures of swinholide A and hurghadolide A, potent actin polymerization inhibitors isolated from other sponges (Carmely & Kashman, 1985; Kitagawa *et al*., 1990; Doi *et al.* 1991; Youssef & Mooberry, 2006).



Misakinolide A ($n_1 = n_2 = 0$)
Swinholide A ($n_1 = n_2 = 0$)
Hurghadolide A ($n_1 = 1$, $n_2 = 0$)



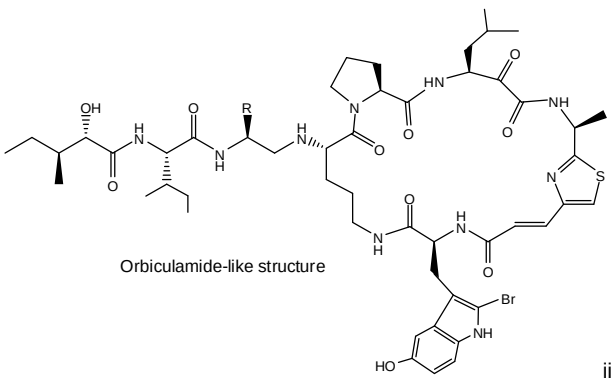
Keramamide H

Both biosynthetic pathways were encoded on two different gene clusters that exhibited typical bacterial gene features, strongly indicating that the producer of misakinolide A and keramamide H is a symbiotic bacterium. Since the screening system used to clone the gene clusters was based on the filamentous fraction dominated by "*Candidatus* Entotheonella sp.", a heterothropic delta-proteobacterium associated with *T. swinhoei*, we assumed that misakinolide A and keramamide H are produced by "Entotheonella sp." To confirm the taxonomic status of the bacterial producer, further analysis either by single cell studies or its combination with complete genome is currently underway.

Subsequent genome sequencing of another member of this as-yet uncultivable candidate genus from a different chemotype of *T. swinhoei* led to the identification of genes for the biosynthesis of orbiculamide-like structure, which is structurally related to keramamides. Therefore, the results in this work provide not only convincing proof for the microbial origin of marine natural products but also specific taxonomic information as well as the potential to sustainable supply of pharmacologically potential compounds.

"*Candidatus* Entotheonella sp."

Photo by T. Mori



Orbiculamide-like structure

Acknowledgements

First and foremost I would like to thank my supervisor, Prof. Dr. Jörn Piel for the great opportunity he has given to me to learn and perform research within his group. I am grateful for his insightful guidance, suggestions, ideas, conversations, supports, and encouragement. Thank you for editing this dissertation; your important advices and input significantly enhanced the content.

I would like to thank Prof. Dr. Grabriele M. König and Priv.-Doz. Dr. Thomas Kolter for evaluating my research progress and for providing the DAAD with recommendation letters that allowed the annual extension of my DAAD scholarship.

I would like to acknowledge the committee members Prof. Dr. Uwe Deppenmeier, Prof. Dr. Albert Haas, and Prof. Dr. Grabriele M. König for their critical reading, input, and helpful advice.

Many thanks go to all my colleagues in both the Piel and Gulder Groups for their friendship, humor, and help that provided a nice and fun environment for me to learn and perform research. My dissertation would not be possible without their support. My special thanks to Dr. Christian Gurgui for his kindness in sharing knowledge, protocols and experimental tricks, as well as for proof-reading my thesis draft. My sincere appreciation goes to him and his wife for their generosity and help especially when I found difficulties in my life. I am grateful to Silke Reiter and Bikram Pandey for their involvement and contributions in my PhD project.

I am thankful to Dr. Michael Freeman for sharing his incredible scientific knowledge and practical experiences, and for proof-reading my thesis draft. Many thanks to Dr. Brandon Morinaka for explaining me reaction mechanisms and proof-reading my thesis draft. I am grateful to Dr. Mike Wilson and Dr. Reiko Ueoka for sharing their scientific knowledge and having nice discussions. A big thank to Max Helf and Stefan Künne for their help especially when I have computer problems. I also want to thank Katja Jensen and Ursula Steffens for sharing valuable “yeast” protocols. Thanks too to Annette Kampa, Xiaofeng Cai, and Dr. Max Crüsemann for sharing their knowledge in protein expressions.

I am indebted to the past members of the Piel Group: Dr. Katja Fisch, Dr. Kathrin Reinhardt, Dr. Tu Ahn Nguyen, Dr. Sonia van der Sar, Dr. Jana Moldenhauer, Dr. Marija Avramović, Dr. Mina Eklund, Dr. Thomas Hochmuth, Dr. Katrin Zimmermann, Dr. Holger Niederkrueger, and Daniel Flachshaar. They have been supportive and helpful. Special thanks go to Dr. Katja Fisch for mentoring and teaching me with various laboratory techniques during the first year of my PhD research. I am especially thankful to Dr. Kathrin Reinhardt, Dr. Katrin Zimmermann and Dr. Holger Niederkrueger for their help, including matters regarding my visa extension.

I would like to thank Prof. S. Matsunaga (University of Tokyo), Dr. T. Mori, Prof. M. Takeyama (Waseda University), Dr. T. Wakimoto (University of Tokyo) for their collaboration in providing sponge and filamentous bacterial samples used in this work. I would like to thank Dr. Jörn Kalinowski and Dr. Christian Rückert (Bielefeld University) for their collaboration in genome sequencing. I am thankful to Dr. Vladimir Benes (EMBL-Heidelberg) for the collaboration in fosmid sequencing.

I am grateful to Prof. Dr. Inneke F.M. Rumengan (my former academic supervisor) for providing me with the foundation during my undergraduate studies for becoming a biotechnologist. Her support and encouragement were paramount in motivating me to get my scientific carrier started on the right path. I would like to thank Prof. Dr. Maggy Thenawidjaja Suhartono, Prof. Dr. Ekowati Chasanah, and Prof. Dr. Hari Eko Irianto for their encouragement and recommendation for me to pursue doctoral studies.

I am especially thankful to my parents and brothers for their prayers, love and encouragement. I am grateful to Edison Macusi and Dr. Jonattan Lassa for sharing brotherhood in Bonn. Thanks to friends and colleagues in Research Center for Marine and Fisheries Product Processing and Biotechnology, Jakarta for their friendship and support.

I would like to thank the German Academic Exchange Service (DAAD) for awarding me a scholarship that has enabled me to carry out doctoral studies in Germany. This PhD work was done in the Piel Group at Kekulé Institute of Organic Chemistry and Biochemistry, Rheinischen Friedrich-Wilhelms-Universität Bonn. This research project was financially supported by BMBF through GenBioCom 03155851 given to Prof. Dr. Jörn Piel.

Contents

List of Figures

List of Tables

Abbreviations

6-dEB	6-deoxyerythronolide
4`-PP	4`-phosphopantetheine
A domain	Adenylation domain
aaRSs	Aminoacyl-tRNA synthetases
Aboa	(*3S,4S,5E,7E*)-3-amino-4-hydroxy-6-methyl-8-(p-bromophenyl)-5,7-octadienoic acid
AIR carboxylases	1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate carboxylases
ACP	Acyl carrier protein
ACV	d-(L-α-aminoadipyl)-L-cysteinyl-D-valine
Adda	3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid
ARP	adhesion-related proteins
ArgA	Arginine-specific adenylation domain
Asn	Asparagine
AsnA	Asparagine-specific adenylation domain
AT	Acyltransferase
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BrTrp	Bromotryptophan
C domain	Condensation domain
CDPSs	Cyclodipeptide synthases
CoA	Coenzyme A
CR	Crotonase
Cys	Cysteine
DMSO	Dimethylsulfoxide
DEBS	6-deoxyerythronolide B synthase
CTAB	Cetyltrimethylammonium bromide
DGGE	Denaturing gradient gel electrophoresis
DH	Dehydratase domain
dH_2O	Distilled water
DNA	Deoxyribose nucleic acid
dNTPs	Deoxynucleoside triphosphates
dTTP	2'-deoxythymidine 5'-triphosphate
EB buffer	Elution buffer
EMA	European Medicines Agency
eDNA	Environmental DNA
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharides
FAD	Flavin adenine dinucleotide
$FADH_2$	Flavin adenine dinucleotide, Reduced
FASs	Fatty acid synthases
FDA	US Food and Drug Administration
FISH	Fluorescence *in situ* hybridization
FMN	Flavin mononucleotide
GNAT	GCN5-related *N*-acetyltransferase
GrsA1	Gramicidin Synthetase A1
HctF	A modular protein involved in hectoclorine biosynthesis
HMGS	3-hydroxy-3-methylglutaryl Coenzyme A synthase
HMW-DNA	High molecular weight DNA
HOCl	Hypochlorite
HPG	Homopropargylglycine
HPLC-MS	High-performance liquid chromatography-mass spectrometry
IC_{50}	Half maximal inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JamP	A modular protein involved in jamaicamide A biosynthesis
LMP agarose gel	Low melting point agarose gel
KAPP-loop domain	Kinase-associated protein phosphatase-loop domain
Kb	Kilobase pair

kDa	Kilo dalton
KR	Ketoreductase domain
KS	Ketosynthase domain
KS^{o}	Non-elongating ketosynthase domain
KtzN	A modular protein involved in kutzneride biosynthesis
LB medium	Luria broth medium
LC-MS	Liquid chromatography-mass spectrometry
MibH	A modular protein involved in microbisporicin biosynthesis
MIC	Minimum inhibitory concentration
MicA	A modular protein involved in microviridin biosynthesis
Min	Minute
MlnA	A modular protein involved in macrolactin biosynthesis
MOX	Monooxygenase
MT	Methyltransferase
NAD	Nicotinamide adenine dinucleotide
NFK	N-formylkynurenie
NpnB	A modular protein involved in Nostophycin biosynthesis
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthase
O-MT	Oxymethyltransferase
ORF	Open reading frame
OXY	Oxygenase
OX	Oxidoreductase
PDB buffer	Phosphate dilution buffer
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PEP synthase	Phosphoenolpyruvate synthase
PheA	Phenylalanine-specific adenylation domain
PK	Polyketide
PKS	Polyketide synthase
PPTase	Phosphopantetheinyl transferases
PS	Pyran synthase
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNAse	Ribonuclease
rRNA	Ribosomal RNA
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second (s)
SerA	Serine-specific adenylation domain
T domain	Thiolation domain
TDO	Tryptophan 2,3-dioxygenase
TE	Thioesterase
TEM	Transmission electron microscopy
T_m	Melting temperature
TP	Transposase
TPR	Tetratricopeptide repeat domain
Trp	Tryptophan
Trp-Hal	Tryptophan halogenase
TRP	Tetratricopeptide repeat domain-encoding proteins
Tris-HCl	2-Amino-2-(hydroxymethyl)-1,3-propanediol, hydrochloride
UV	Ultraviolet
X-Gal	5-Bromo-4-chloro-3-indolyl beta-D-galactopyranoside
WPD	World Porifera Database
ZmaA	A modular protein involved in Zwittermicin A biosynthesis

Chapter I

Introduction

Oceans account for over 70% of the earth`s surface area. The majority of these waters are at great depth, in which about 95% are more than 1000 m deep (Skropeta, 2008) or 60% are more than 2000 m deep (Bull *et al*., 2000). Fourteen of 35 animal phyla found in the sea are exclusively marine, whereas only 11 are terrestrial (Briggs, 1994; Ray, 1997; Sala and Knowlton, 2006). So far approximately 300,000 marine species have been described (Bouchet, 2006; Sala & Knowlton, 2006; Douvere and Laffoley, 2010), and the total number of eukaryote species in the world`s oceans is estimated at 2.21 million (Mora *et al*., 2001). The highest biodiversity in the ocean is found in coral reefs, where the number of species per unit area is extremely high and predicted to exceed that of tropical rainforests (Haefner, 2003; Sogin *et al*., 2006). Around 600,000 to more than 9 million species have been estimated in the coral reefs worldwide (Plaisance *et al*., 2011). The deep sea, on the other hand, is considered as high biodiversity repository due to the enormous surface area (Sala and Knowlton, 2006). Over 500 eukaryote species, encompassing 12 animal phyla and more than 150 new genera, have been described from deep vent sites in the last three decades (Thornburg *et al*., 2010). However, the marine biodiversity estimates assessed so far almost exclusively accounts for multicellular eukaryotes, mainly large- and well-studied macrofauna (Belwood and Hughes, 2001; Mellin *et al*., 2011; Plaisance *et al*., 2011). It does not account for single-cell eukaryotes, bacteria, archaea, and viruses (Sala & Knowlton, 2006), suggesting that the marine species diversity is much higher than previously thought (www.coml.org; Grassle and Maciolek, 1992; Committee on Biological Diversity in Marine Systems of the U.S. NRC, 1995).

Pioneering work by ZoBell and colleagues in the 1940s and 1950s showed the enormous number and diversity of true marine bacteria from the deep sea (up to 10 km deep). They described the multiplication, morphology, and physiological activities of deep sea bacteria able to grow both at normal pressures as well as those that prefer high hydrostatic pressures (up to 600 atmospheres) (ZoBell & Johnson, 1949; ZoBell and Oppenheimer, 1950; ZoBell and Morita, 1957). Recent genome shotgun sequencing conducted by Venter and coworkers showed the presence of at least 1800 new different genomic species of bacteria, including 148

previously unknown bacterial phylotypes, in just 1500 liters of seawater from the Sargasso Sea near Bermuda (Venter *et al*., 2004). A more recent ocean exploration (between 2000 to 2010) by the Census of Marine Life program involving 2,700 scientists from 80 nations revealed that the estimated number of species in the oceans are more than 1 million (excluding microbes). The diversity of marine microbes is estimated as high as 1 billion, with ~38,000 species and 5,000-9,000 species of microbial bacteria in a liter of seawater and a gram of sand, respectively (www.coml.org; Costello *et al*., 2010).

Fluorescent *in situ* hybridization studies by DeLong, Karners and colleagues showed that the global oceans harbor approximately 1.3×10^{28} archaeal cells, and 3.1×10^{28} bacterial cells. The pelagic archaea belonging to crenarchaeota accounts for approximately 20% of the global picoplankton cells, representing one of the most abundant cell types in the oceans (DeLong *et al*., 1999; Karner *et al*., 2001). The microbial biomass in the oceans is estimated at $3{,}6 \times 10^{29}$ microbial cells with a total cellular carbon content of 3×10^{17} grams (Sogin *et al*., 2006). The deep-ocean sediments (0-10 m) contain approximately 6.6×10^{29} marine prokaryotic (archaeal and bacterial) cells (Whitman *et al*., 1998). Recent molecular studies on symbiotic bacteria showed that the microbial consortia in many sponges can account for up to 60% of the total biomass and consist of hundreds to thousands of species with at least 18 bacterial and archaeal phyla identified from these hosts (Taylor *et al*., 2007). Some well-studied sponge specimens harbor bacterial cells in the density range of 6.4×10^{8} to 1.5×10^{9} bacterial cells per ml of sponge extract, which exceeds at of seawater by two to four orders of magnitude (Scheuermayer *et al*., 2006). It is therefore not surprising that marine microorganisms account for most of the oceanic biomass (Azam & Worden, 2004).

Due to the enourmous diversity and abundance of marine microorganisms, there has been an emerging interest among scientists to understand their bio-ecological roles. Significant roles have been attributed to marine microorganisms including their involvement in nearly all primary and secondary production as well as in all biogeochemical cycles in the oceans (Staley, 1985; Storm, 2008). In the deep-sea hydrothermal vents, extremophiles have the ability to adapt in extreme conditions such as high pressure, high temperature, low pH, low oxygen, toxic metals, and no sunlight (Skropeta, 2008). They play important roles in primary production in the vent communities via chemosynthesis (Thornburg *et al*., 2009). Recent genomic and

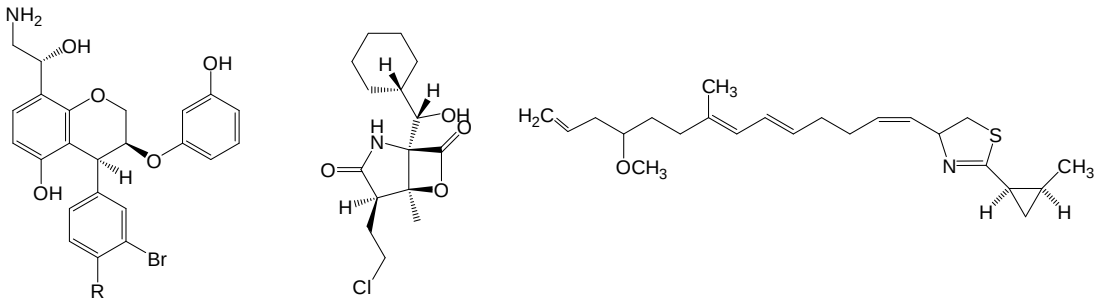
proteomic studies have revealed the remarkable diversity of metabolic pathways used by these chemosynthetic bacteria to obtain energy from their environments and feed their hosts (Dubilier *et al*., 2008). Symbiotic bacteria associated with marine animals have recently been shown to contribute to the survival and health of the host macro-organisms, including nutrition, chemical protection, localization cues, and developmental signals (Flachshaar & Piel, 2008, Woyke *et al.*, 2006).

A unique symbiotic role is exemplified by the luminescent symbionts of the Vibrionaceae group living in the light organs of squid of the genera *Uroteuthis*, *Loliolus* and *Euprymna.* These symbionts produce light to provide their host squid with a mechanism allowing them to evade predators by camouflage (Guerrero-Ferreira & Nishiguchi, 2007). Other unique symbiosis was observed between the organic compound-degrading Oceanospirillales and the marine worms of genus *Osedax* living on the carcasse of a grey whale at the depth of 2891 m in Monterey Canyon, off the coast of California (Goffredi *et al*., 2005). Recent shotgun sequencing by Dubilier and colleagues has revealed that two symbionts identified as *Olavius algarvensis* play roles in providing the worms with multiple sources of nutrition by fixating carbon into organic carbon via autotrophy as well as by synthesizing almost all amino acids and various vitamins (Woyke *et al*., 2005). It was found that these symbionts are also able to uptake and recycle the waste products of the worm, such as dicarboxylate, succinate, monocarboxylates acetate and propionate, explaining how the host lives without excretory system (Woyke *et al*., 2005). Further metaproteomic and metabolomic studies revealed unsual pathways possessed by the symbionts of *O. algarvensis* to cope with energy and nutrient limitation, such as the potential use of carbon mononoxide as an energy source and as yet undescribed energy-efficient steps in CO_2 fixation and SO_4^{2-} reduction (Kleiner *et al*., 2011). Interestingly, symbiotic microorganisms have been proposed to provide chemical protection or defense to most sessile marine invertebrates that lack mobility, such as sponges, bryozoans, and tunicates (Amsler *et al.*, 2001). To be able to accomplish all these bio-ecological roles as mentioned above, marine microorganisms appear to contain highly diverse and unique metabolic pathways, which can be explored for production of specific natural products that are useful for food, pharmaceutical, fuel and cosmetic industries. Therefore, biotechnological exploration of marine microorganisms undoubtedly offers exciting opportunities and challenges in discovery of intriguing biosynthetic pathways responsible for making compounds with promising pharmacological

activities. Further genetic manipulation of the biosynthetic pathways of interest could lead to the generation of structurally novel “unnatural” natural products with improved properties.

1.1. Pharmacological Potential of Marine Bacteria

Until 2009, approximately 569 compounds from marine bacteria have been reported (William *et al*., 2009), in which at least 15 marine bacterial natural products are currently being evaluated in clinical trials, especially as anticancer drugs (Gerwick *et al*., 2008). Approximately 390 novel natural products have been discovered from deep-sea (>50 m depth) organisms, including invertebrates, bacteria, archaea, and fungi (Skropeta, 2008). Interestingly a number of new compounds have recently been reported from bacteria living in the deep-sea hydrothermal vents, which is exemplified by the apoptosis-inducing agents ammonificins A (**1**) and B (**2**) isolated from the thermophilic anaerobic *ε*-Proteobacterium, *Thermovibrio ammonificans* (Andrianasolo *et al*., 2009) (Fig. 1.1.1).



Ammonificin A: R = OH (**1**)
Ammonificin B: R = Br (**2**)

Salinosporamide A (**3**)

Curacin A (**4**)

Figure 1.1.1 Examples of pharmacologically potential compounds discovered from marine bacteria: ammonificins A and B from *Thermovibrio ammonificans* (Andrianasolo *et al*., 2009), salinosporamide A from *Salinospora tropica* (Feling *et al*., 2003), and curacin A from *Lyngbya majuscule* (Gerwick *et al*., 1994).

The majority of the wide diversity of marine bacteria cannot be cultivated under laboratory conditions (Staley & Konopka, 1985; Iverson *et al*., 2012). However, even the few genera that have been cultured (Fenical, 1993) have enabled access to intriguing chemistry and unique biosynthetic pathways. Pioneering work of Fenical, Jensen, and coworkers has led to the discovery of at least 13 new genera of unique marine actinomycete bacteria from deep-sea sediments (Jensen *et al*., 1991; Fenical *et al*., 1993; Mincer *et al*., 2002). All of the isolated actinomycetes were cultured in a seawater-based medium and have been shown to produce unique metabolites (Jensen *et al*., 1991; Jensen & Fenical, 1994). Further chemical

investigation has shown that the obligate marine genus *Salinospora* is a particularly rich source of unprecedented secondary metabolites (Fenical & Jensen, 2006). One notable example is salinosporamide A (**3**), a very potent proteasome inhibitor (IC_{50} = 1.3 nM) (Feling *et al*., 2003; Fenical *et al*., 2009; Cragg *et al*., 2012). This potent cytotoxin is currently in phase I clinical trials and approaching phase II (Cragg *et al*., 2009; Gerwick & Moore, 2012). By genome sequencing of *S. tropica*, B. Moore and colleagues identified at least 17 biosynthetic gene clusters (Udwary *et al*, 2007). Interestingly, only three of them are responsible for the biosynthesis of known natural products, including salinosporamides, suggesting the remarkable biosynthetic capacity of this marine actinomycete (Udwary *et al*, 2007).

Marine cyanobacteria have attracted great attention as the producers of a large number of intriguing compounds with promising biological activities. These photosynthetic bacteria, also called blue green algae, are abundant under favorable environmental conditions in shallow tropical waters (Nunnery *et al*., 2010; Gerwick *et al*., 2008). Due to the profusive growth in nature, they can be collected in large enough quantities and cultivated under laboratory conditions (Gerwick *et al*., 2008). Starting with the pioneering work of R. E. Moore in the 1970s to the early 2000s, it was revealed that marine cyanobacteria are a significant source of pharmaceutically relevant secondary metabolites (Mynderse & Moore, 1978; Moore, 1996; Simmons *et al*., 2005; Gerwick *et al*., 2008; Cardellina II & Moore, 2010; Nunnery et *al*., 2010). Further investigation of marine cyanobacteria by Gerwick and coworkers has shown chemical structures, biosynthesis, and biological activities of these intriguing metabolites (Gerwick *et al*., 2008). Many of these cyanobacterial metabolites are structurally unique because they combine peptide and lipid features with additional functional modifications such halogen atoms to generate a wide variety of lipopeptides (Nunnery *et al*., 2010). A notable example, curacin A (**4**) from the truly marine cyanobacterium *Lyngbya majuscula* (Gerwick *et al*., 1994) is currently in preclinical trials as an anticancer agent (Newman & Crag, 2004; Simmons *et al*., 2005). This compound was found to contain unusual structural features, such as thiazole ring, cyclopropyl ring, and a terminal alkene moiety (Gerwick *et al*., 1994). Using a gene cloning strategy, the curacin A gene cluster was identified, including a gene cassette responsible for cyclopropyl ring formation (Chang *et al*., 2004). Further biosynthetic studies have revealed that the terminal alkene formation occurs through a unique mechanism involving hydrolysis, decarboxylation, and sulfate elimination (Gu *et al*., 2009; Gehret *et al*., 2011).

Recently symbiotic bacteria associated with marine invertebrates have gained increasing attention from a medical perspective because they are suspected as the true producers of many rare bioactive secondary metabolites. Marine invertebrates such as sponges, bryozoans, and tunicates are well known chemically rich sources of pharmaceutically important and structurally unique compounds (Faulkner, 1994; Hay & Fenical, 1996). A frequently observed phenomenon supporting the symbiosis hypothesis is that invertebrate-derived metabolites are structurally related to typical bacterial compounds (Bewley and Faulkner, 1998; Moore 1999; Piel, 2004). In addition, numerous marine invertebrates harbor complex polyketide and structurally modified peptide families that are common in bacteria, but rare in other animals (Piel, 2006; Uria & Piel, 2009). Due to the limited amount of the metabolites in the animal tissues, exploitation of the biosynthetic pathways of symbiotic bacteria could provide a sustainable supply of marine compounds for drug development (Piel, 2002). However, attempts to access these biosynthetic pathways have been hampered by the complexity of the microbial assemblages and the fact that they are not amenable to culture (Amann *et al*., Hugenholtz *et al*., 1998; Friedrich *et al*., 2001). For example, a recent study of the Great Barrier Reef sponge *Rhopaloeides odorabile* showed that the culturable percentage of symbiotic bacteria accounts for only 0.1-0.23% of the total bacterial community (Webster & Hill, 2001). Despite improvements in cultivation techniques, the majority of sponge-associated bacteria cannot be cultured in the laboratory media (Hentschel *et al*., 2002). A potential problem is that laboratory cultivation may destroy any cell-to-cell communication between organisms in the natural habitats that might be crucial for growth. In nature, symbiotic bacteria likely coexist with a host and/or other bacteria to obtain essential nutrients and substrates that are not available with standard microbiological approaches (Joint *et al*., 2010). Even if cultivation conditions can be identified, the desired compound might not be produced due to the absence of required environmental signals (Uria & Piel, 2009).

Recent advances in cultivation-independent techniques and DNA sequencing technology have enabled access to the biosynthetic pathways of pharmaceutically relevant natural products from complex microbial communities (Piel 2004; Schmidt 2008). Using metagenomics-based approaches, Piel and colleagues have provided the first genetic evidence for the bacterial origin of marine invertebrate-derived natural products by cloning the gene cluster for onnamides/theopederins (**5**) from the sponge *Theonella swinhoei* (Piel *et al*., 2004b). Subsequently, Piel and

colleagues cloned the gene cluster for psymberin (**6**) from the sponge *Psammocinia* aff. *bulbosa* (Fisch *et al*., 2009). More recently, Piel and colleagues isolated biosynthetic genes for polytheonamides A (**7**) and B (**8**) from *T. swinhoei* metagenome (Freeman *et al*., 2012). These giant toxins function as unimolecular ion channels in cell membranes (Iwamoto *et al*., 2010) with near femtomolar activity (Freeman *et al*., 2012). These highly cytotoxic 48-residue polypeptides contain 13 nonproteinogenic amino acids, and therefore, they were previously thought to have formed by a nonribosomal peptide synthetase (NRPS) (Fusetani & Matsunaga, 1993; Hamada *et al*., 2010; Itoh *et al*., 2012). However, recent metagenomic studies by Freeman *et al* (2012) have revealed a ribosomal origin of these exceptionally potent natural products with only six candidate enzymes involved in 48 posttranslational modifications (Freeman *et al*., 2012) (Fig. 1.1.2).

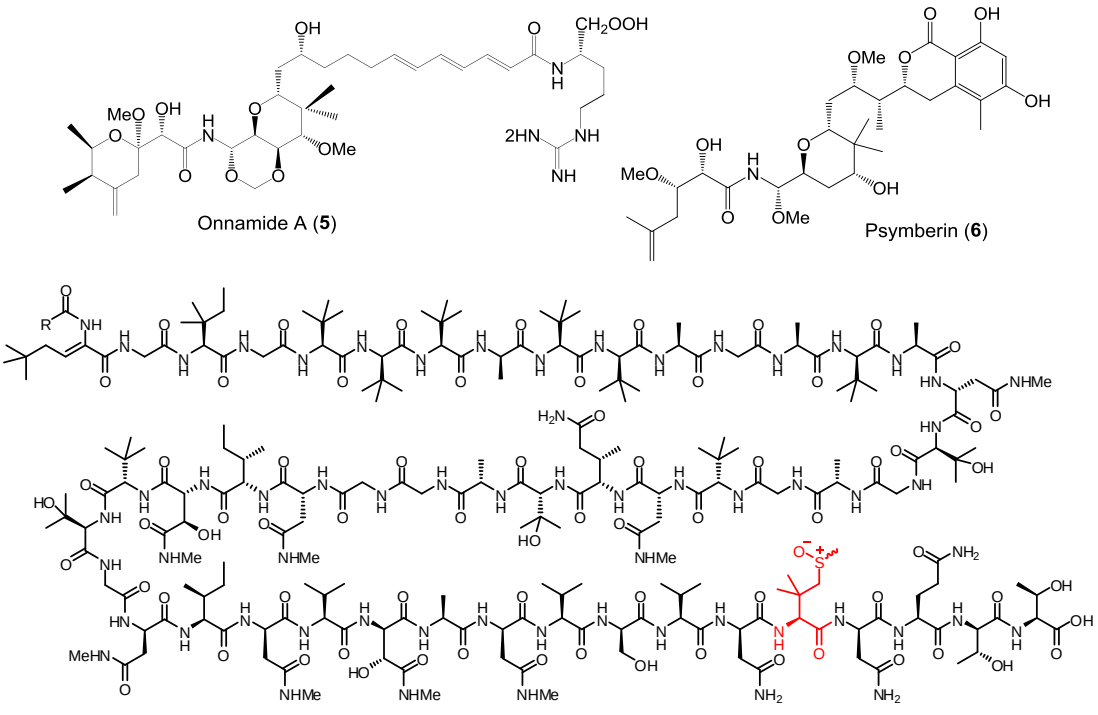


Figure 1.1.2 Pharmaceutically important natural products from marine sponges, in which their biosynthetic pathways have been characterized. These include onnamide A from the sponge *Theonella swinhoei* (Piel *et al*., 2004b), psymberin from the sponge *Psammocinia* aff. *bulbosa* (Fisch *et al*., 2009), and polytheonamides from *T. swinhoei* (Freeman *et al*., 2012). Polytheonamides A and B differ in the configuration of the sulfoxide moiety in the unusual residue β,β-dimethylmethionine sulfoxide (red color).

The biosynthetic pathways of a number of other natural products have been identified from the uncultivated bacterial symbionts of various other marine invertebrates, including the gene clusters coding for patellamides A (**9)** and C (**10**)

from a tunicate (Schmidt *et al*, 2005), bryostatin 0 (**11**) from a bryozoan, which is the common basis of the 20 known bryostatins including bryostatin 1 (**12**) (Sudek *et al*., 2007), and ET-583 (**13**), an intermediate involved in the biosynthesis of the chemotherapeutic ET-743 (**14**) from a tunicate (Rath *et al.*, 2011) (Fig. 1.1.3).

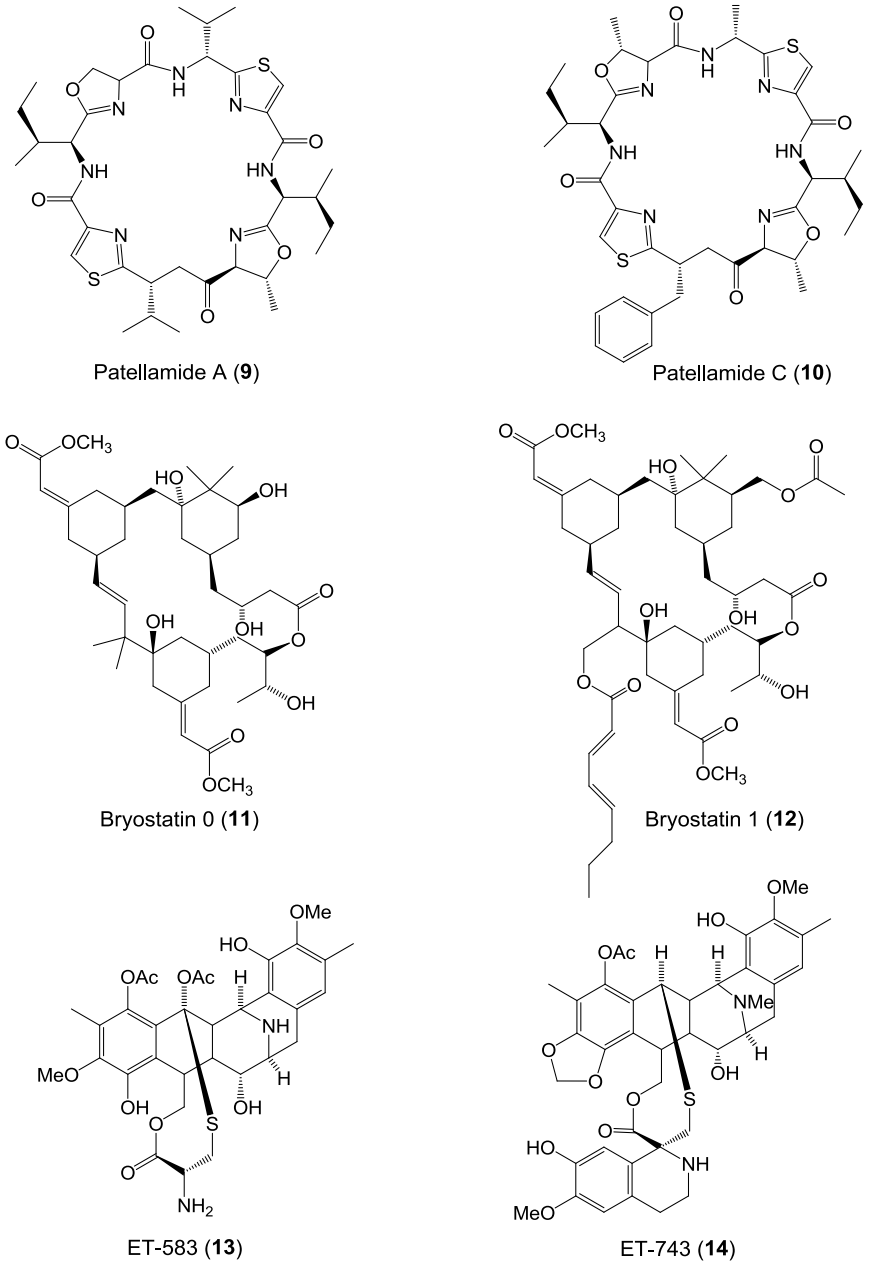


Figure 1.1.3 Pharmaceutically important natural products from various marine invertebrates, in which their biosynthetic pathways have been characterized. These include patellamides A and C from the cyanobacterial symbiont of tunicate *Lissoclinum patella* (Schmidt *et al*, 2005), bryostatin O that is a common basis of the 20 know bryostatins including bryostatin 1 from the bryzoan *Bugula neritina* (Sudek *et al*., 2007), and ET-583 that is an intermediate in the biosynthesis of ET-743 from the tunicate *Ecteinascidia turbinate* (Rath *et al*., 2011).

Discovery of the biosynthetic pathways for these complex compounds not only provides convincing proof for the microbial origin but also facilitates the sustainable supply of rare marine drug candidates. Since among marine invertebrates, sponges (Porifera) are the leading sources of bioactive natural products in terms of the number of bioactive compounds discovered (Proksch, 1994; Faulkner, 2000; Leal *et al*., 2012), therefore the next sections will highlight recent progress in sponge-symbiont studies and in exploring the biosynthetic potential of marine sponges for the sustainable supply of rare potent sponge-derived compounds.

1.2 Microbial Associations with Marine Sponges

Marine sponges (phylum *Porifera*) show a diverse range in structural, size, shape, and color, and are distributed in a wide range of marine habitats, including tropical, temperate, and antarctic oceans (Ruppert *et al.*, 2004). Currently, there are approximately 11,000 sponge species described in the World Porifera Database (WPD) of which around 8,500 are considered valid (Van Soest *et al*., 2012a; Van Soest *et al*., 2012b). Their multicellular body plan is relatively simple and specifically designed for filter feeding. The simplest form is a hollow cylinder that is attached to the substratum and surrounded by many small pores called ostia with a larger upper opening called osculum. The body surface is covered by an outer layer (pinacoderm), and the hollow interior (atrium) is lined by a monolayer of flagellated collar cells (choanoderm). The connective tissue layer located between the pinacoderm and the choanoderm is called the mesohyl (Ruppert *et al.*, 2004; Taylor *et al*., 2007; Eerkes-Medrano & Leys, 2006) (Fig. 1.2.1). Among many cells present in the mesohyl are archaeocytes, large amoeboid cells bearing a number of large lysosomes, which play a role in digestion and internal transport (Ruppert *et al.*, 2004). There are three cell types described in the pinacoderm, namely exopinacocytes that extend along the outer surfaces of the dermal membrane, ab-endopinacocytes that line the inner wall of the osculum and the excurrent canals, and pros-endopinacocytes that cover the subdermal space and the inner surface of the dermal membrane (Mergner, 1964). The pinacoderm is a monolayer composed of flatted cells. The flattened shape of the endopinacocytes is especially intended for rapid diffusion between the water channels and the mesenchyme (Bagby, 1970) (Fig. 1.2.1). As the name Porifera (pore bearers) implies, the sponge body is exceptionally porous which inhales seawater through the ostia throughout the body, then pass over the flagellated choanoderm to the atrium, and exits through the

osculum. This water circulatory system is called the aquiferous system (Ruppert *et al.*, 2004). The water volume pumped by a sponge is equal to its body volume that flows through the aquiferous system once every 5 seconds (Ruppert *et al.*, 2004) or up to 24,000 liters/kg its body in a single day (Vogel, 1977; Kennedy *et al*., 2007).

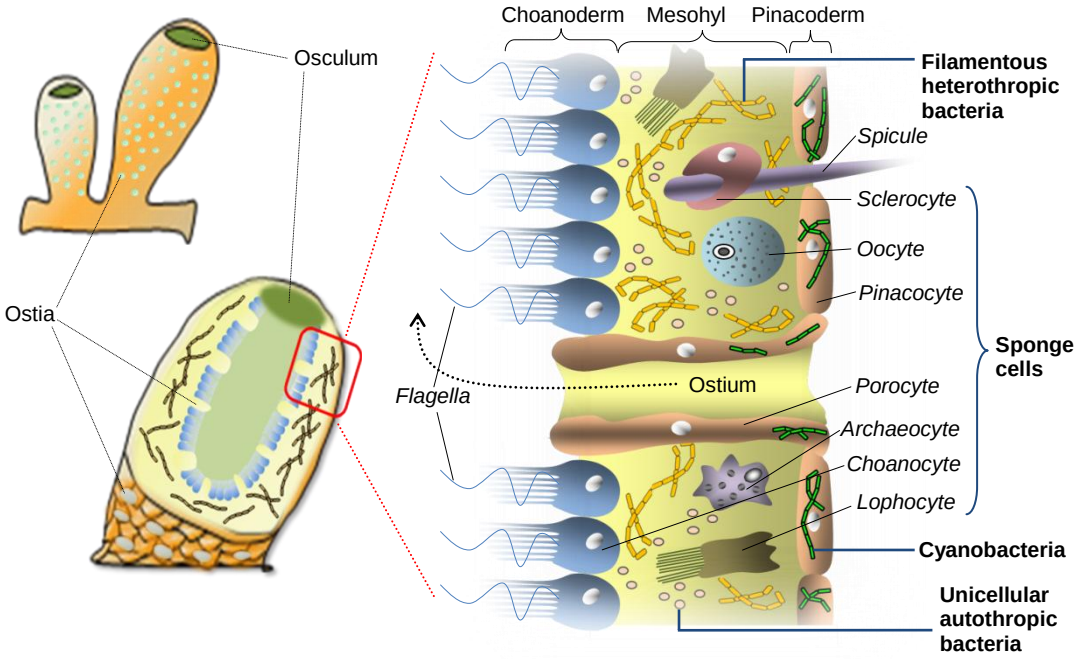


Figure 1.2.1 The sponge`s body plan, showing various sponge cell types and symbiotic bacterial cell types. Modified from Ruppert *et al* (2004).

The abundance and diversity of microbial communities in sponges were originally described by Wilkinson, Vacelet & Reisweg in the 1970s and 1980s (Reiswig, 1975; Wilkinson & Fay, 1979; Vacelet & Donadey, 1977). Based on electron microscopy and bacterial cultivation studies, they showed that the microbial biomass can account up to 40% of the total sponge volume. In particular the sponge specimens categorized as “bacteriosponges” are attributed to those with high numbers of bacteria (Vacelet and Donadey, 1977) with the densities of 10^8 to 10^{10} bacterial cells per gram of sponge body wet weight (Hentschel *et al*, 2006). It was found that sponge-specific microbial populations were particularly abundant in the sponge mesohyl (Wilkinson, 1978; Vacelet & Donadey, 1977). Recent developments in biomolecular techniques such as 16S rRNA gene analysis and denaturing gradient gel electrophoresis (DGGE) have enabled to obtain better insights into the diversity of symbionts in sponges. So far the diversity of microbes known in sponges encompass 14 recognized bacterial phyla, both major archaeal lineages, and assorted microbial eukaryotes. The most frequently isolated 16S rRNA gene sequences from sponges include those from Acidobacteria, Actinobacteria, and

Chloroflexi (Taylor *et al*., 2007). Hentschel and colleagues assessed the phylogenetic analysis of the mesohyl bacteria using fluorescence *in situ* hybridization (FISH) and showed that the vast majority of mesohyl bacteria belong to the domain with low GC content, which was mostly dominated by δ-proteobacteria, followed by γ-proteobacteria and representatives of the Bacteriodetes (Friedrich *et al*., 1999). In general, the internal mesohyl layer is primarily populated by heterotrophic and autotrophic bacteria. Whereas, the outer layer of the sponge exposed to sunlight is typically dominated by photosynthetic bacteria such as cyanobacteria (Kennedy *et al*., 2007) (Fig. 1.2.1).

By pumping larger volumes of seawater, marine sponges remove living microscopic particles such as bacteria, microalgae or detritus from the seawater column (Hentschel *et al*., 2006). The acquired bacteria cells entering the aquiferous system are subsequently transferred to the mesohyl layer. In the mesohyl layer, bacterial cells can be digested as nutrients through extracellular lysis and subsequent phagocytosis by sponge cells called archaeocytes (Vacelet & Donadey, 1977; Ribes *et al*., 1999; Vacelet & Duport, 2004). However, some bacteria can live and become established as part of the sponge-specific symbionts (Vacelet & Donadey, 1977). This uptake process of bacterial symbionts by the host sponges from the seawater environment is known as horizontal transmission. Interestingly, during horizontal transmission, sponges have the ability to distinguish between bacterial symbionts and seawater bacteria through a specific mechanism for recognition and rejection of self-particulate matters in the epidermal cells (Wilkinson *et al*., 1984).

The vertical transmission of symbionts from adult sponges to their offspring through reproductive stages was originally proposed by Lévi and Porte in the 1960s based on electron microscopy observation (Schmitt *et al*., 2008). This transmission is likely common and widespread among marine sponges in order to maintain stable symbiotic associations to each sponge generation (Schmitt *et al*., 2008; Lee *et al*., 2009). Recent molecular studies by Schmitt, Hentschel and colleagues on five different Carribbean sponge species harboring high complex microbial consortia have provided the first comprehensive phylogenetic analysis showing the collective vertical transmission of a complex adult microbial community through reproductive stages in the sponges, *Xestospongia muta*, *Ircina felix*, and *Corticium* sp (Schmitt *et al*., 2008). Subsequently they proposed the following transmission model for symbiotic associations of high microbial abundance sponges based on a combined

electron and fluorescence observation and phylogenetic studies. At first, the microbial symbionts in the adult mesohyl are vertically transmitted to larvae. Since larvae are unable to feed during the transmission, it is assumed that there is no microbial exchange with the environment. This non-feeding period results in the symbiont separation in different sponges that would lead to cospeciation. Subsequently when the larvae become juveniles, a filter-feeding period begins indicated by the ability to uptake seawater microbes. This uptake process may result in the horizontal transmission of microbial symbionts between the host sponges and the environment (Schmitt *et al*., 2008). This vertical transmission of bacterial symbionts from adult to embryo was supported by the work of Lee *et al* (2009) who applied DGGE and clone library analysis to demonstrate that the bacterial communities associated with the adults of the Caribbean sponge *Svenzea zeai* were highly similar to the embryos of the same sponge species, but they were completely different from those in the surrounding seawater (Lee *et al*., 2009).

There has been an emerging recognition about the mutualistic relationship between the bacterial symbionts and the host sponges themselves. In the surfaces or internal spaces of the sponge body, nutrients are much richer than in seawater and sediments, suggesting that the sponges offer nourishment and a safe habitat to their bacterial symbionts (Lee *et al*., 2001). The microbial symbionts contribute to the hosts by supporting the nutritional process through nitrogen fixation, nitrification and photosynthesis (Wilkinson & Fay, 1979; Hoffmann *et al*., 2009) and metabolic waste processing (Beer & Ilan, 1998). In particular, a nitrogen source is likely available in the form of ammonia as an endproduct of host metabolism (Scheuermayer *et al*., 2006). Providing nutrition to the host through nitrogen fixation is exemplified by symbiotic cyanobacteria, which are autotrophic symbionts of a variety of sponges (Wilkinson and Fay, 1979). Many sponges contain photosynthetic symbionts, more commonly cyanobacteria, in their tissue and obtain nutritional benefit from the photosynthate. Further investigation showed that cyanobacteria transfer excess photosynthate in the form of glycerol and a small organic phosphate through a mechanism similar to the glycerol 3-phosphate shuttle functioning between chloroplasts and cytoplasm in plant cells (Arillo *et al*. 1993). The excess photosynthate is utilized by the sponges (Ruppert *et al.*, 2004).

Recently, Hentschel, Piel and coworkers have revealed the functional properties and lifestyle of a member of the possibly ancient symbiont phylum, “*Candidatus*

Poribacteria". This symbiont was isolated as individual cells from the marine sponge, *A. aerophoba* by FACS. Whole genome amplification and sequencing indicated that the "Poribacteria" were mixotrophic bacteria with the capacities of autotrophic CO_2 fixation via the Wood-Ljungdahl pathway. Interestingly, two intermediate enzymes in the anaerobic respiratory chain, namely nitrite reductase and nitric oxide reductase were found. This suggests the ability of this symbiont to maintain a respiratory chain under oxygen-limiting conditions, such as during the sponge`s non-pumping periods in which the meshopyl turns anaerobic. In addition, several putative proteins in the sequenced genome might be involved in the sponge-microbe symbiosis, such as adhesins, adhesion-related proteins (ARP) and tetratricopeptide repeat domain-encoding proteins (TPR) (Siegl *et al*., 2011).

Another ecological role of microbial symbionts that have attracted special attention from a medical perspective is the ability to produce toxic metabolites (Faulkner *et al*, 1994). There is growing evidence that symbiotic bacteria produce toxic compounds in order to provide the chemical defense system to the hosts, because sponges lack the complex immune system of higher animals (Kennedy *et al*., 2007; Lee *et al*., 2001). In particular the toxic compounds released by the symbionts may help their sponge hosts to deter predators, keep competitors away, paralyze their prey, prevent settlement of other organisms on their body`s surfaces (biofouling), or compete for space with other organisms (Haefner, 2003). The substances playing a role as chemical defense have been the main focus of interest in recent years for being developed as clinically used medicines. Because these toxic substances are in general available in extremely limited amounts in the sponge tissue, efforts to study the involvement of bacterial symbionts in producing sponge-derived metabolites could lead to a sustainable supply of marine drug candidates.

1.3 Sponge-derived Metabolites and Symbiosis Hypothesis

One of the most interesting aspects of marine sponges is the presence of bioactive compounds in their tissue, which may play a role to chemically defend sponges against environmental pressure factors such as predation, overgrowth and settlement of other organisms, or competition for space (Proksch, 1994; Haefner, 2003). Marine sponges have been an interesting target for extensive studies on pharmaceutically valuable compounds since the discovery of two cytotoxic nucleosides from the marine sponge *Cryptotethia crypta* by Bergmann and

coworkers in the late 1940s and the early 1950s (Bergmann & Feeney, 1950, 1951; Bergmann & Swift, 1951). These nucleosides, named spongothymidine (**15**) and spongouridine (**16**), subsequently provided the basis for the synthesis of the antiviral agent vidarabine (Ara-A) (**17**), the anticancer agent cytarabine (Ara-C) (**18**) (Proksch *et al*., 2002), and the antiviral agent zidovudine (AZT) (**19**) (Bowling *et al*., 2007) (Fig. 1.3.1). Ara-C (**18**) became the first marine-derived anticancer drug that is currently used in the routine treatment of patients with leukemia and lymphoma (Proksch *et al*., 2002).

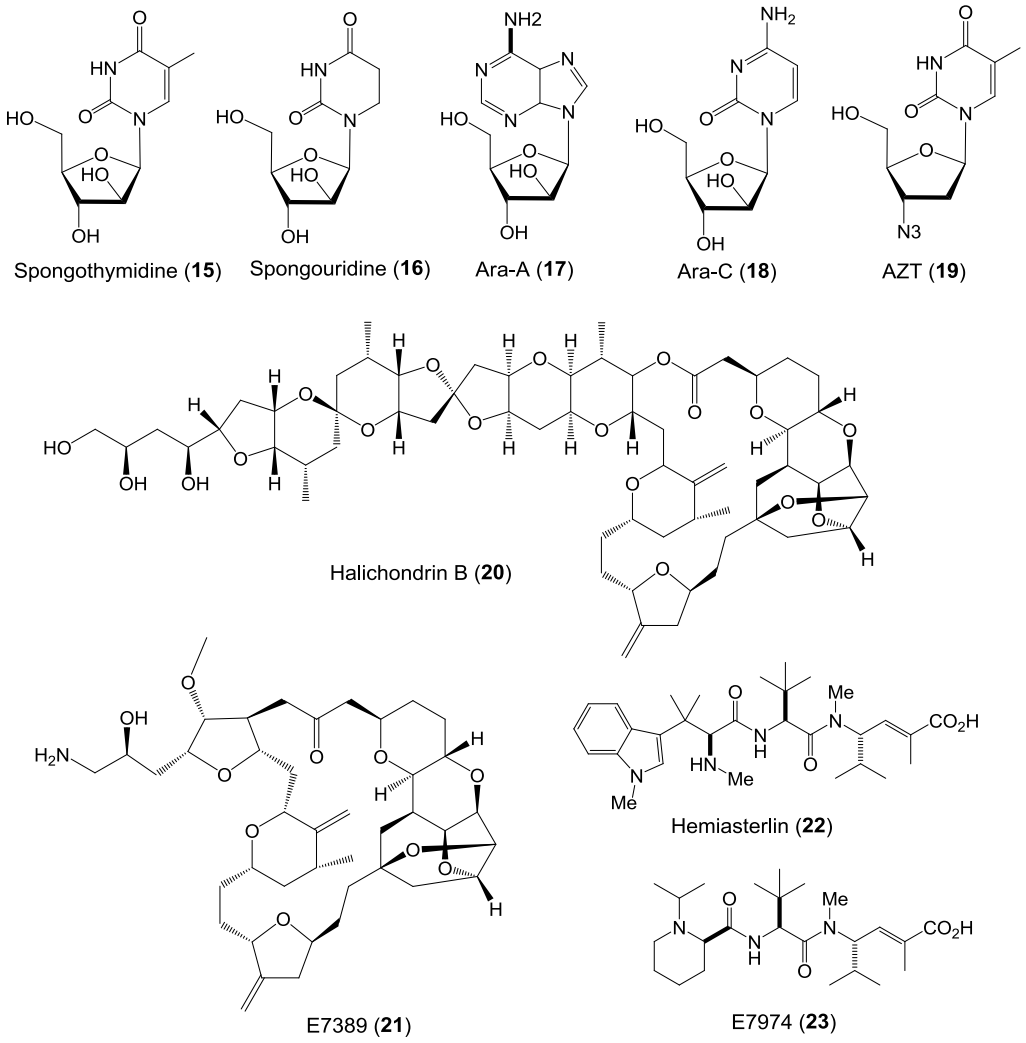


Figure 1.3.1 Sponge-derived natural products and their analogs that are approved by FDA and/or EMA as clinically used agents or in clinical trial. Spongothymidine and spongouridine inspired the development of the well-known medicines Ara-A, Ara-C, and AZT. Halichondrin B became a basis for the synthesis of the newly approved anticancer drug Halaven® (E7389). Hemiasterlin isolated from several marine sponge species inspired the synthesis of E7974, an anticancer drug candidate currently in human clinical trials.

In more recent year, a synthetic analog of halicondrin B (**20**), termed Eribulin Mesylate (E7389, Halaven®) (**21**), has been approved as a new chemotherapy

agent for metastatic breast cancer patients by US Food and Drug Administration (Donoghue *et al*., 2012) as well as by the European Medicines Agency (Menis and Twelves, 2011; Gourmelon *et al*., 2012; Ortega & Cortés, 2012). It inhibits microtubule polymerization and aggregates soluble tubulin in nonreprodutive form (Gourmelon *et al*., 2012). Halichondrin B (**20**) is a large macrolide isolated from the Japanese sponge, *Halicondria okadai* (Uemura *et al*., 1985; Hirata *et al*., 1986). Total synthesis of this tubulin-binding agent was performed by Kishi and colleagues (Aicher *et al*., 1992). In addition, a synthetic analogue of the marine sponge natural product Hemiasterlin (**22**), termed E7974 (**23**) (Kowalczyk *et al*., 2005), is in current phase I clinical trial as an anticancer agent (Rocha-Lima *et al*., 2012; Gerwick & Moore, 2012). It acts through an antimitotic mechanism inhibiting microtubule polymerization (Kuznetsov *et al*., 2009). Hemiasterlin, a modified liniear tripeptide, was isolated from the marine sponges *Hemiasterella minor* (Talpir *et al.,* 1994), *Cymbastela* sp. (Coleman *et al*., 1995), and *Auletta* sp. and *Siphonochalina* sp. (Gamble *et al*., 1999) (Fig. 1.3.1).

Among marine sponges, the order Lithistida (Demospongiae) is a renowned source of a diverse array of structurally complex biologically active metabolites. Over 300 natural products have been reported from these lithistid sponges and structurally include sterols, macrolides, acetogenins, alkaloids, and peptides (Fusetani & Matsunaga, 1993; Bewley & Faulkner, 1998; Wright *et al*., 2010). Within the Lithistida, the chemistry of *Theonella* and *Disocdermia* species has been studied extensively. Early work by Djerassi, Kashman, and coworkers on *T. swinhoei* and *T. conica* in the 1980s led to the isolation of the sterols, conicasterol and theonellasterol (Kho *et al*., 1981). Furher work by Carmely and Kashman on the Red Sea *T. swinhoei* led to the isolation of the polyketide swinholide A. This cytotoxic compound was first identified as a monomeric 22-membered macrolide (Carmely & Kashman, 1985). Misakinolide A was first identified from *Theonella* sp. as a monomeric macrolide structurally very similar to swinholide A, but lacked an additional double bond in the lactone ring (Sakai *et al*., 1986). Subsequently, two macrolides, bistheonellide A and B, were discovered from a *Theonella* sp. collected at Hachijo-Jima island, Japan. NMR data, spectroscopic and chemical studies led to the conclusion that bistheonellide A was identical to misakinolide A (**24**) and the chemical structure proposed for both was dimeric instead of monomeric (Kato *et al*., 1987). This allowed the previous monomeric swinholide A (**25**) structure to be revised to be a symmetric cyclic 44-membered dimer using X-ray single crystal

analysis, as reported for the Okinawan *T. swinhoei* (Kitagawa *et al*., 1990; Doi *et al.* 1991) (Fig. 1.3.2). Recent studies by Youssef and Mooberry (2006) on the Red Sea *T. swinhoei* resulted in the isolation of hurghadolide A (**26**), an unprecedented 42-membered dilactone moiety that is structurally related to swinholide A. It is asymmetric due to the absence of an olefin in the lactone ring of only in one of the monomeric lactone. Compounds closely related to compounds **24-26** have been reported from remarkably diverse sources, including luminaolide (**27**) from algae (Kitamura *et al*., 2009), scytophycin C (**28**) from the filamentous cyanobacterium *Scytonema pseudohofmanni* (Ishibashi *et al*., 1986), and lobophorolide (**29**) from seaweed (Kubanek *et al*., 2003) (Fig. 1.3.2). This suggests that misakinolide A and its analogs are produced by microorganisms associated with the sponges. This symbiont hypothesis was somewhat supported by scanning electron micrograms for the Okinawan *T. swinhoei* containing swinholide A (Kitigawa *et al*., 1990) and the *T. swinhoei* from Hachijo-jima Island, showing the abundant presence of filamentous bacteria in the sponge tissues (Fusetani & Matsunaga, 1993).

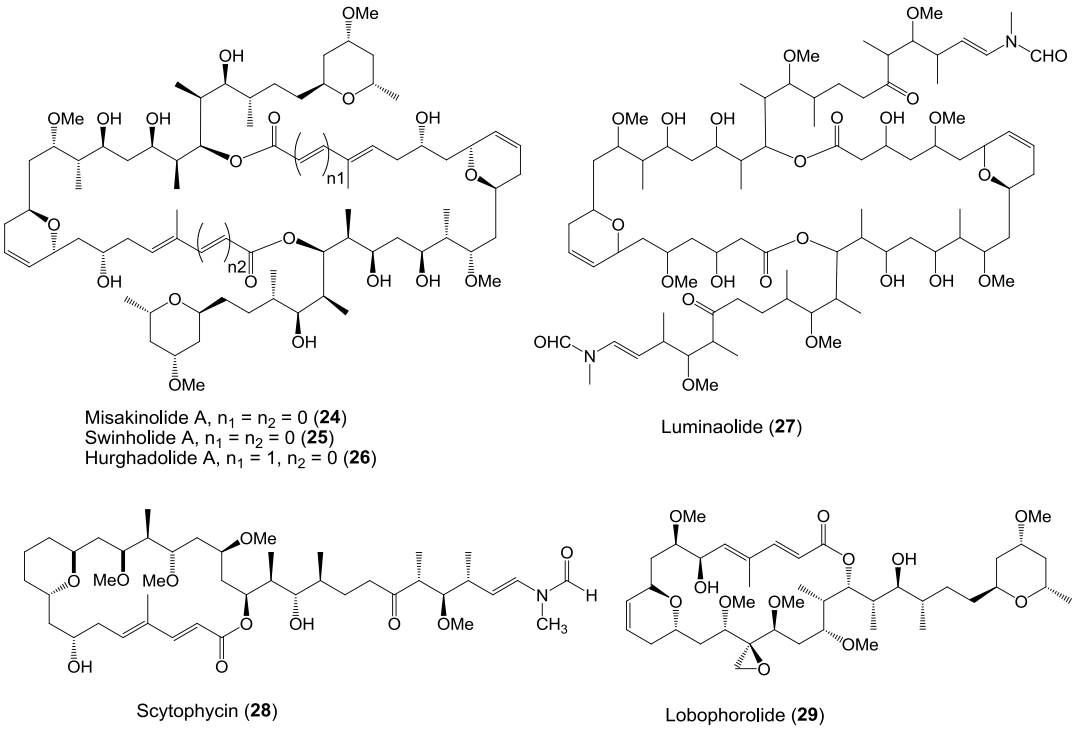
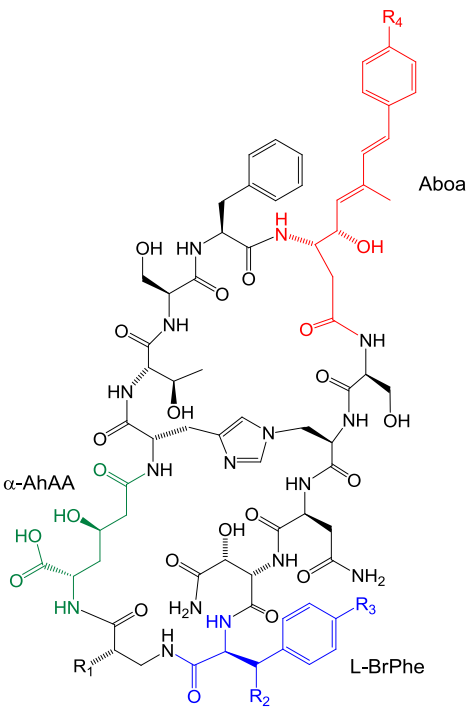


Figure 1.3.2 Macrolides closely related to misakinolide A reported from diverse sources: swinholide A and hurghadolide A from the sponge *T. swinhoei* (Carmely & Kashman, 1985; Kitagawa *et al*., 1990; Doi *et al.* 1991; Youssef and Mooberry, 2006), luminaolide from the crustose coralline algae *Hydrolithonreinboldii* (Kitamura *et al*., 2009), lobophorolide from the seaweed Lobophora variegate (Kubanek et al., 2003), and scytophycin C from the filamentous cyanobacterium *Scytonema pseudohofmanni* (Ishibashi *et al*., 1986).

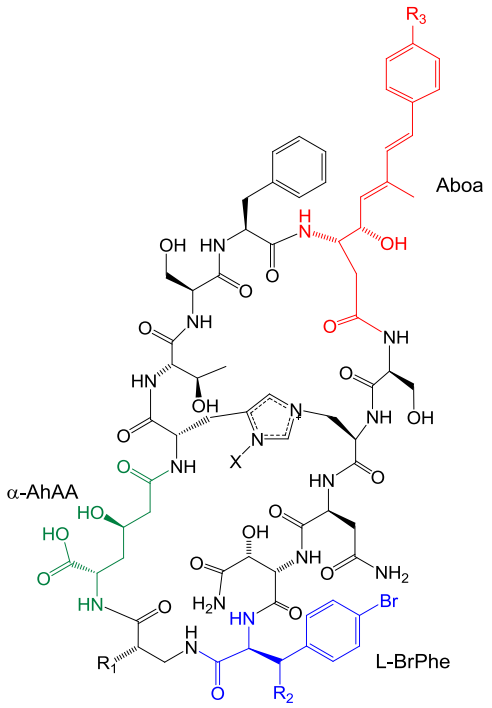
Although all three compounds (swinholide A, misakinolide A, and hurghadolide A) are potent actin polymerization inhibitors (Carmely & Kashman, 1985; Kitagawa *et*

al., 1990; Youssef & Mooberry, 2006), they show different mechanisms of activity. Misakinolide caps the barbed end of filaments while swinholide A severs actin filaments (Terry *et al*., 1997). Hurghadolide A (**26**) was 10 times more potent than swinholide A at disrupting microfilaments (Youssef & Mooberry, 2006). Compound **26** showed *in vitro* cytotoxicity against human colon adenocarcinoma (HCT-116) with IC_{50} values of 5.6 and 365 nM (Youssef & Mooberry, 2006).

Further work by Fusetani, Matsunaga and coworkers on the *T. swinhoei* collected in Hachijo-jima Island, Japan has led to the discovery of the complex bicyclic peptide theonellamide F (**30**) (Matsunaga *et al*., 1989) and theonellamides A-E (**31-35**) (Matsunaga and Fusetani, 1995). Theonellamide F (**30**) showed antifungal activity that inhibited the growth of *Candida albicans* with an MIC 6.3 µg/ml (Fusetani and Matsunaga, 1993). Subsequently, Faulkner and co-workers isolated theonegramide (**36**) and theopalaumide (**37**) from *T. swinhoei* collected in the Philippines (Bewley and Faulkner, 1994) and Palau (Schmidt *et al*., 1998) (Fig. 1.3.3).



Theonellamide F: R_1 = H, R_2 = H, R_3 = Br, R_4 = Br (**30**)
Theonellamide B: R_1 = OH, R_2 = Me, R_3 = Br, R_4 = H (**31**)
Theonellamide C: R_1 = H, R_2 = H, R_3 = H, R_4 = Br (**32**)



Theonellamide A: R_1 = OH, R_2 = H, R_3 = H, X = β-D-Gal (**33**)
Theonellamide D: R_1 = H, R_2 = H, R_3 = Br, X = β-L-Ara (**34**)
Theonellamide E: R_1 = H, R_2 = H, R_3 = H, X = β-D-Gal (**35**)
Theonegramide : R_1 = OH, R_2 = Me, R_3 = H, X = β-D-Ara (**36**)
Theopalaumide : R_1 = OH, R_2 = Me, R_3 = H, X = β-D-Gal (**37**)

Figure 1.3.3 The antifungal complex cyclic peptide theonellamides F and A-E from Japanese *T. swinhoei* (Matsunaga *et al*., 1989; Matsunaga and Fusetani, 1995), which is closely related to theonegramide from Philippines *T. swinhoei* (Bewley & Faulkner, 1994) and theopalaumide from Palauan *T. swinhoei* (Schmidt *et al*., 1998). The unusual amino acids are indicated with red color for Aboa, green color for (2*S*,3*R*)-3-hydroxyasparagine (*α*-AhAA), and blue color for L-ρ-bromophenylalanine (L-BrPhe).

Theonegramide and theopalaumide are structurally similar to theonellamide F, but differ by the addition of a sugar moeity to the histidoalanine-bridge, in which theonegramide harbors D-arabinose whereas theopalaumide contains D-galactose. Interestingly, these three antifungal peptides have the unusual amino acid residue (*3S,4S,5E,7E*)-3-amino-4-hydroxy-6-methyl-8-(p-bromophenyl)-5,7-octadienoic acid (Aboa), which is structurally similar to 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid (Adda) present in microcystins associated with the bloom of cyanobacterium *Microcystis* spp. (Namikoshi *et al*., 1992), suggesting the possible symbiotic bacterial origin of theonellamides and its analogs.

Experimental support of the microbial origin of sponge-derived secondary metabolites was initially provided by Faulkner and coworkers who investigated the cellular localization of swinholide A and the complex cyclic peptide theopalauamide in the Palauan *T*. *swinhoei* (Bewley *et al*. 1996). In their experiments, three different types of bacterial cells were initially separated from the sponge cells by differential centrifugation: (i) the filamentous heterotrophic bacteria (found in the endosome part), (ii) the unicellular cyanobacteria (identified as *Aphanocapsa feldmanni*) (present only in the ectosome), and (iii) the unicellular heterotropic bacteria (distributed throughout the sponge tissue, both endosome and ectosome). Chemical analysis showed that swinholide A was accumulated in the mixed unicellular bacterial fraction, whereas the antifungal theopalaumide was located mainly in the filamentous bacteria (Bewley *et al*., 1996). In contrast, work of Gerwick and coworkers on free-living cyanobacteria collected from two different locations led to the isolation of swinholide A (Andrianasolo *et al*., 2005). At this point, it was unclear which one the actual producer of this compound was, the sponge-associated cyanobacteria or the unicellular heterotrophic bacteria (Andrianasolo *et al*., 2005; König *et al*., 2006). Another possibility is accumulation of swinholide A in a certain type of bacterial population due to the transport of such compound across cellular membranes and whole tissues (Uria & Piel, 2009).

Identifying the taxonomy affiliation of the filamentous heterotrophic bacterial fraction containing theopalaumide was subsequently pursued (Schmidt *et al*., 2000). Although over 100 efforts to cultivate these filamentous symbionts, no growth was observed under any condition (Schmidt *et al*., 2000). Using J agar with sodium thiosulfate, sponge extract, and sodium silicate, Schmidt *et al* (2000) observed colonies containing some cells identified as "*Candidatus* Entotheonella sp." cells together with other bacteria, suggesting that *E. palauensis* is probably an obligate

symbiont requiring specific conditions present inside the host *T. swinhoei*. Transmission electron microscopy (TEM) analysis showed that the cell integrity was maintained during the separation procedure and has confirmed that the cyanobacterium *Aphanocapsa feldmanni* was limited only to the ectosome (Bewley *et al.* 1996). In a separate study, Brück *et al* (2008) reported also the presence of the symbiotic “Entotheonella sp.“ in the mesohyl of the Caribbean Sea marine sponge, *Discodermia dissoluta*. This sponge is well known as the source of discodermolide, a natural product with potential anticancer properties. TEM analysis on its thin tissue section showed a large number of large filamentous microbes in the mesohyl as well as other unicellular microorganisms. The filamentous microbial population was subsequently obtained by differential centrifugation. Further 16S rDNA analysis of the filamentous microbe-containing fractions indicated the sequence similarity to uncultured “Entotheonella sp.“ This result was supported by FISH probing showing that the filamentous structures were bound and illuminated by the probes targeting “Entotheonella sp.” (Brück, *et al.*, 2008).

Further work by Fusetani, Matsunaga, and coworkers on *T. swinhoei* with the bright yellow inner body, designated as chemotype yellow, led to the isolation of the cytotoxic cyclic peptides cyclotheonamides (Fusetani, *et al*., 1990) and orbiculamide A (**38**) (Fusetani *et al*., 1991) (Fig. 1.3.4). Interestingly, orbiculamide A (**38**) contains three unusual amino acids, namely brominated hydroxytryptophan, theonalanine, and theoleucine (Fusetani *et al*., 1991), suggesting the involvement of non-ribosomal peptide synthase in the biosynthesis of orbiculamide A. Closely related peptides to orbiculamide A include keramamides B-D (**39-41**) from Okinawan *T. swinhoei* (Kobayashi *et al*., 1991), and discobahamins A (**42**) and B (**43**) from Bahamian *Discodermia* sp. (Gunasekera *et al*., 1994). All of these peptides contain 5-hydroxytryptophane and thiazole moieties (Fig. 1.3.4). The structural characteristic of orbiculamide A and keramamides B-D is the addition of a bromide atom to the hydroxytryptophan (Fusetani *et al*., 1991; Kobayashi *et al*., 1991), which is absent in discobahamins A and B.

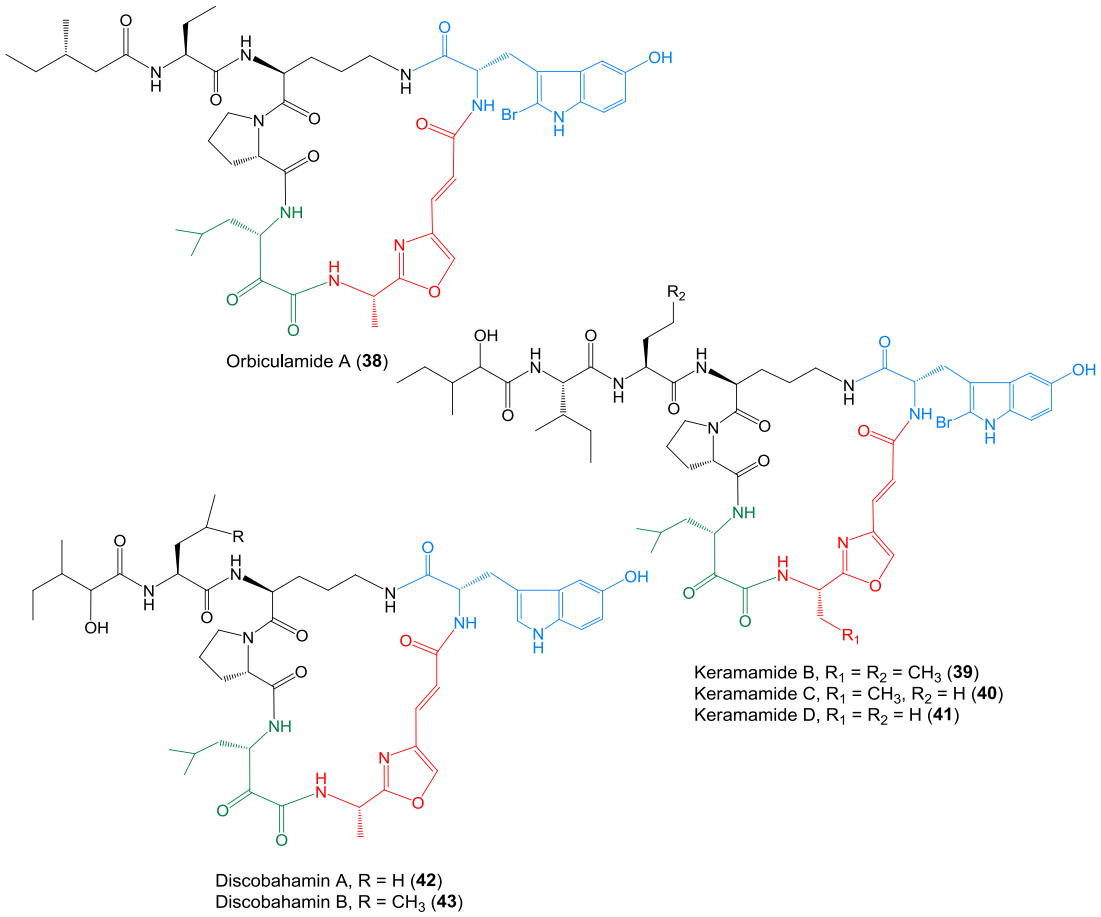


Figure 1.3.4 Oxazole-containing cyclic peptides isolated from marine sponges: orbiculamide A (Fusetani *et al*., 1991), keramamides B-D (Kobayashi *et al*., 1991), discobahamins A and B (Gunasekera *et al*., 1994). They contain unusual amino acid residues including hydroxytryptophane (blue color), theonalanine (red), and theoleucine (green).

Orbiculamide A is also closely related to other reported cyclic peptides from Okinawan *T. Swinhoei*, which include keramamides F (**44**) (Itagaki *et al*., 1992), keramamides E, G (**45**), H (**46**), and J (**47**) (Kobayashi *et al*., 1995), keramamides K and L (Uemoto *et al*., 1998), keramamides M and N (Tsuda *et al*., 1999), and calyxmides A (**48)** and B (**49**) recently reported by Wakimoto, Abe, and colleagues for the sponge *Discodermia calyx* collected near Shikine-jima Island (Kimura *et al*., 2012). It is proposed that the oxazole rings in the theonalanine of compounds **33**-**43** are biosynthetically derived from serine, whereas the thiazole rings present in **44**-**49** are presumably derived from cysteine. It is also of interest to understand how the α-keto group in the theoleucine is biosynthetically formed.

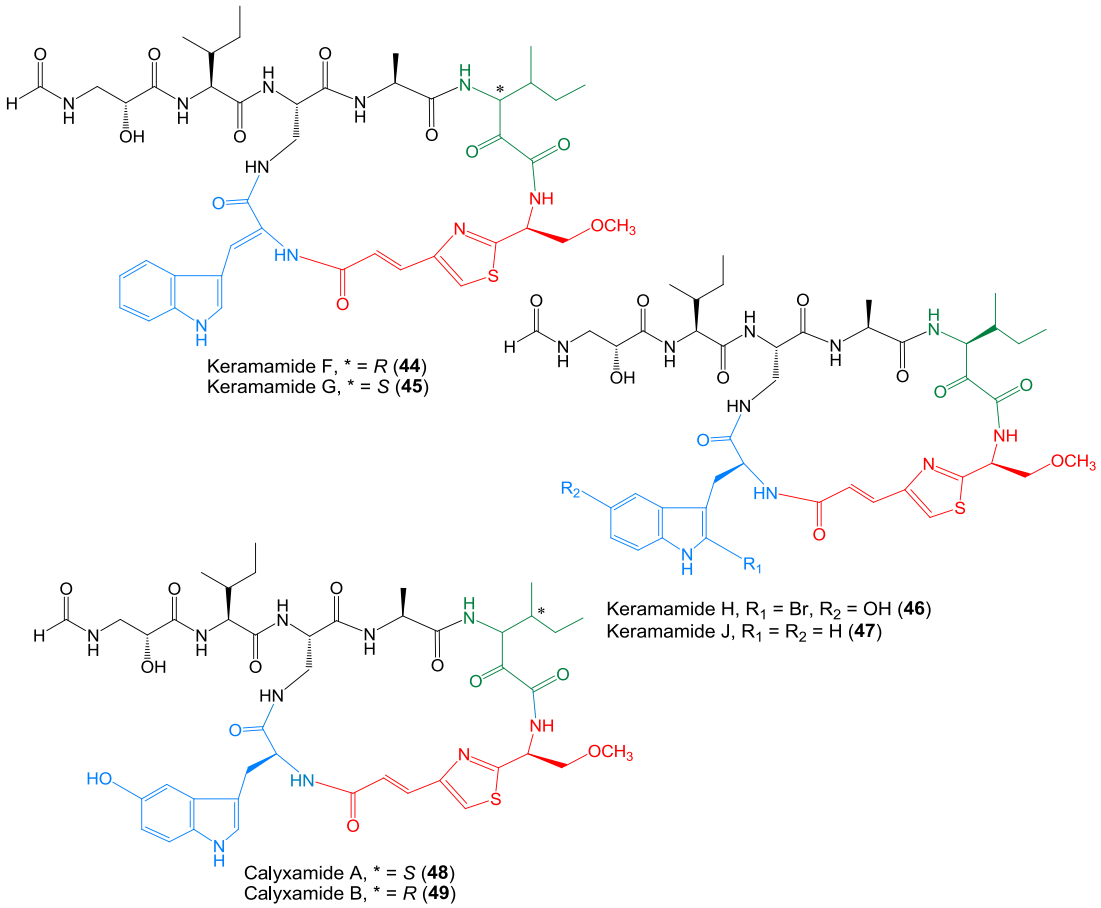


Figure 1.3.5 Thiazole-containing cyclic peptides isolated from marine sponges: keramamides F (Itagaki *et al*., 1992), keramamides G, H, and J (Kobayashi *et al*., 1995), and calyxamides A and B (Kimura *et al*., 2012). They contain unusual amino acid residues including tryptophane (blue color), theoleucine (green), and oxymethylated theonalanine (red). Keramamides F and G differ from keramamide J for the presence of an α,β-*Z*-double bond in the Δ-Trp.

Other closely related peptides include keramamide A (**50**), and konbamide (**51**) isolated from the Okinawan *Theonella* sp. (Kobayashi *et al*., 1991) (Fig. 1.3.6). These two peptides are lack of neither theoleucine nor theonalamine. However, they are unique peptides due to the presence of a halogenated hydroxytryptophan and an unusual ureido bond between two amino acid monomers (Kobayashi *et al*., 1991). Interestingly, in separate studies by Harada *et al* (1995) and Murakami *et al* (1997), anabaenopeptin B (**52**) (Fig. 1.3.6), closely related to **50** and **51** in structure, was reported for the cyanobacteria, *Anabaena flos-aquae* NRC 525-17 and *Oscillatoria agardhii* NIES-204, respectively. This suggests that symbiotic cyanobacteria might be the actual producers of compounds **50** and **51** (Murakami *et al*., 1997).

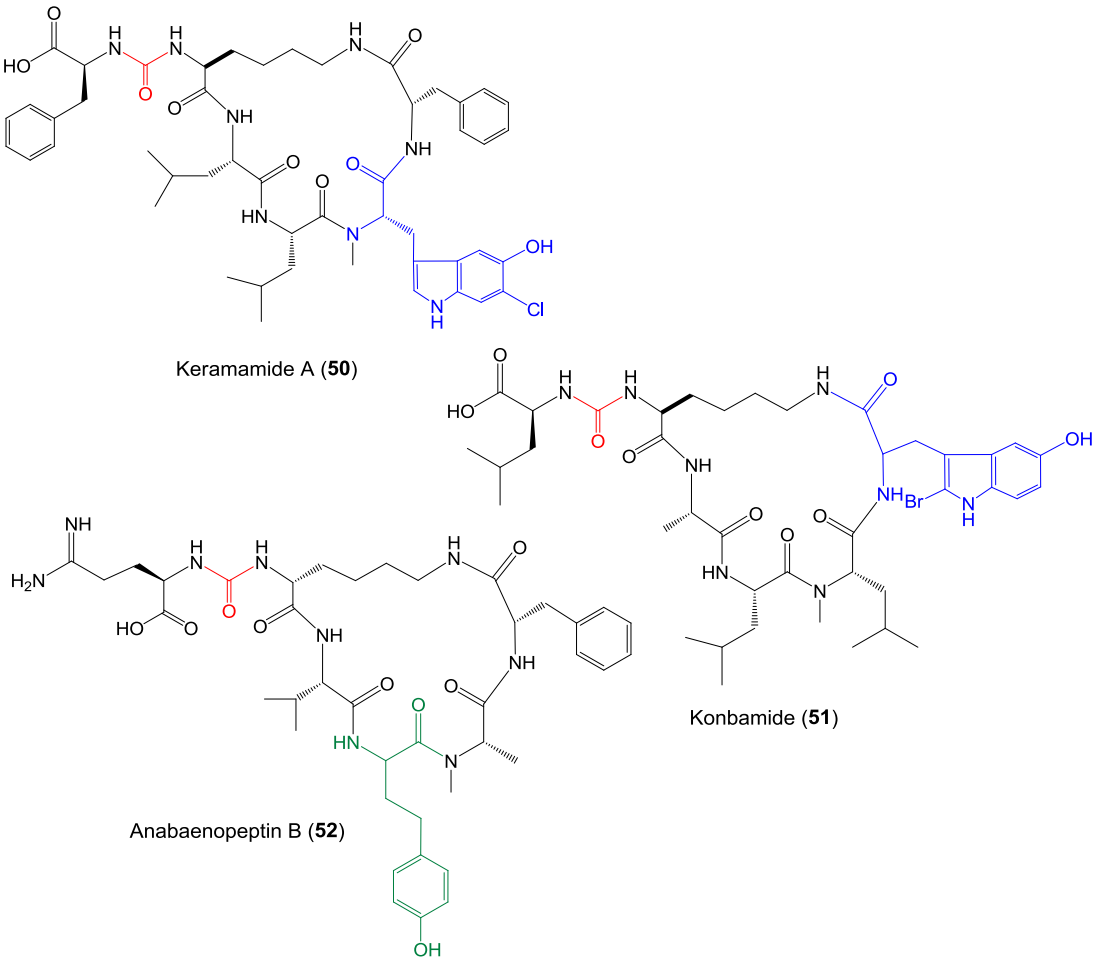


Figure 1.3.6 Ureido bond-containing cytotoxic cyclic peptides. Keramamide A and konbamide was isolated from the marine sponge *Theonella* sp. (Kobayashi *et al*., 1991), and anabaenopetin B was from the cyanobacteria, *Anabaena flos-aquae* NRC 525-17 (Harada *et al*., 1995) and *Oscillatoria agardhii* NIES-204 (Murakami *et al*., 1996). Besides ureido moiety (red color), they contain other unusual amino acid residues including halogenated hydroxytryptophan (blue) or homotyrosine (green).

Other complex compounds discovered from the *T. swinhoei* are polyketides of the onnamide and theopederin series. Onnamide A (**5**) was initially discovered from Okinawan *Theonella* sp. by Higa and coworkers (Sakemi *et al*., 1988). This compound showed potent antiviral *in vitro* activity against herpes simplex virus type-1, vesicular stomatitis virus, and coronavirus A-59 (Sakemi *et al*., 1988). Subsequently, Fusetani, Matsunaga, and Nakao isolated a series of additional compounds closely related to onnamide A from the *T. swinhoei* collected in Hachijo Island, which included dihydroonnamide A (**53**), onnamides B-E (**54**-**57**) (Matsunaga *et al*., 1992), and theopederins A-E (**58-61**) (Fusetani *et al*., 1992) (Fig. 1.3.7). Theopederins showed stronger cytotoxic activity than onnamides. In particular, theopederins A and B exhibited promising antitumor activity against P388 cells (Fusetani, 1996).

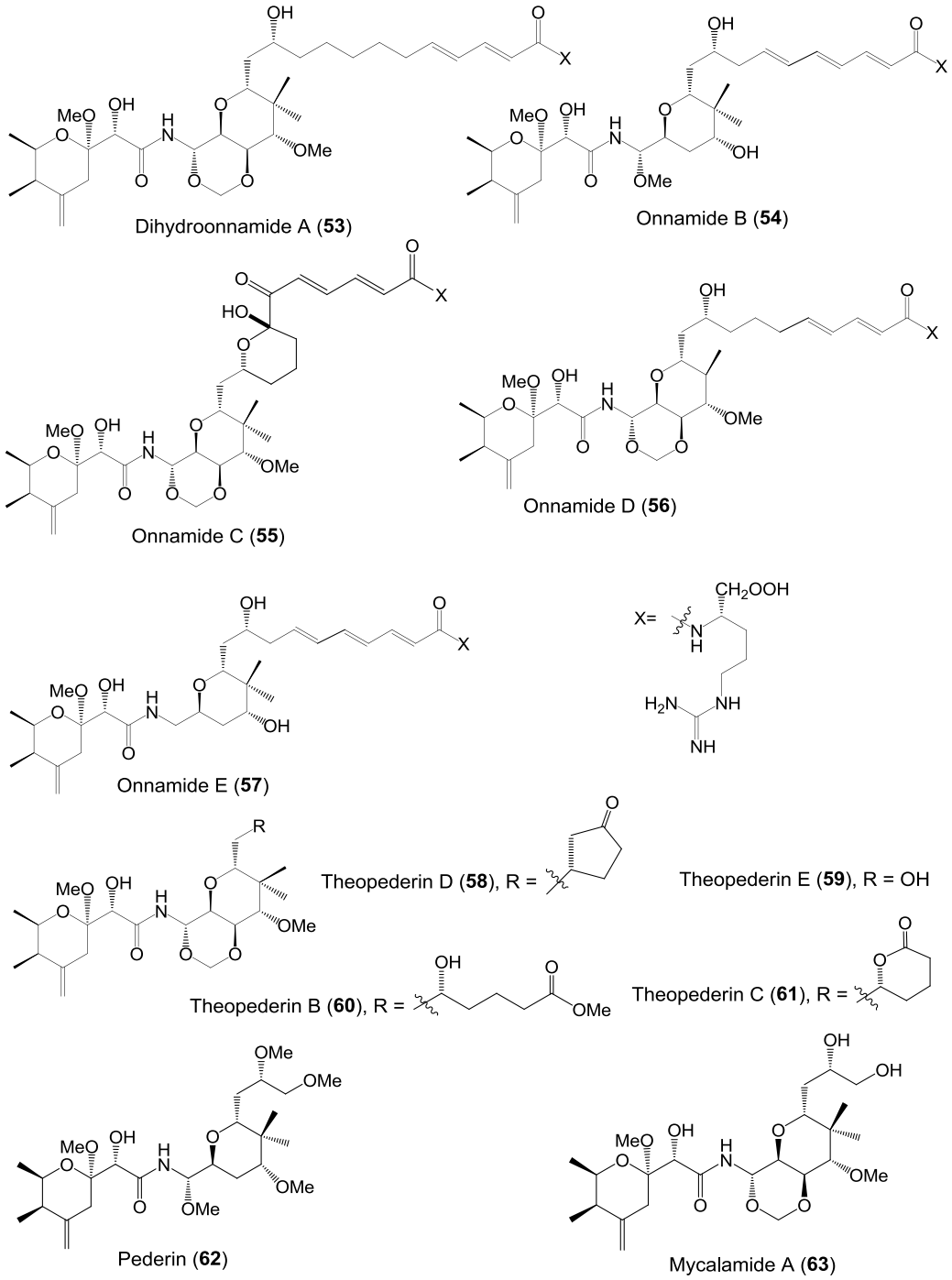


Figure 1.3.7 Onnamide-type compounds from various sources: dihydroonnamide A, onnamides B-E (Matsunaga *et al*., 1992), theopederins A-E (Fusetani *et al*., 1992) from Hachijo-Jima *T. swinhoei,* pederin from *Paederus* and *Paederidus* (Kellner and Detter, 1996), mycalamide A from New Zealand *Mycale hentscheli* (Perry *et al*., 1990), and psymberin from *Psammocinia* aff. *bulbosa* (Cichewicz *et al*., 2004).

Onnamide and theopederin series are strikingly similar to pederin (**62**), a highly antitumor toxin of rove beetles of the genera *Paederus* and *Paederidus* (Kellner & Detter, 1996), mycalamide A (**63**) from the New Zealand sponge *Mycale hentscheli* (Perry *et al*., 1990), and psymberin (**5**) from the sponge *Psammocinia* aff. *bulbosa*

(Cichewicz *et al*., 2004) also known as irciniastatin A from the Indo-Pacific sponge *Ircinia ramosa* (Pettit *et al*., 2004) (Fig. 1.3.7). The occurrence of these closely related compounds in different sponges and animals suggested a microbial origin in their biosynthesis. They all show similar cytototoxic activities, and only pederin and mycalamide A are highly dermatoxic (Piel, 2010). Psymberin showed highly selective antitumor activity (Cichewicz *et al*., 2004), and is therefore the most promising to be developed as a drug candidate. Due to the rarity of this compound, it took over 11 years (1990-2001) to process approximately 600 *Psammocinia* fractions for structural elucidation (Cichewicz *et al*., 2004).

Onnamide-type natural products and many other complex sponge-derived natural products including those mentioned above belong to the class of polyketides or hybrid polyketide/non-ribosomal peptides. To obtain insights into the biosynthesis and biosynthetic origin of complex polyketides of pharmaceutical relevance from marine sponges, Piel and colleagues have developed metagenomics-based strategies using onnamide-type compounds as a polyketide model (Piel *et al*., 2004a; Piel *et al*, 2004b, 2007; Fisch *et al*., 2009). These onnamide-type metabolites were considered as an ideal model for such studies because in the previous work, the entire gene cluster for the biosynthesis of the related compound pederin was successfully cloned from the beetle *Paederus fuscipus* (Piel, 2002). Discovery of the pederin gene cluster provided guidance to isolate the biosynthetic pathways of sponge-derived onnamide-type compounds (Piel *et al*., 2004b). Piel and coworkers have developed metagenomic strategies that feature screening tools to rapidly identify biosynthetic genes for sponge polyketides in highly complex sponge assemblages (Piel *et al*., 2004b; Nguyen *et al*., 2008; Fisch *et al*., 2009). In addition, another exciting strategy developed by the Piel group was to generate methods that enable one not only to catch representative genomes from the wide diversity of microbial symbionts in nature, but also to facilitate rapid screening of the biosynthetic pathways of interest among many diverse non-target pathways present in a sponge symbiotic system (Hvartin and Piel, 2007). With these strategies, Piel group have successfully isolated the entire gene clusters encoding the biosynthesis of onnamide A, theopederin and psymberin from complex sponge systems, thereby providing the first genetic evidence of the bacterial origin of marine natural products (Piel *et al*., 2004b, 2007; Fisch *et al*., 2009).

1.4 Biosynthetic Insights into Microbial Secondary Metabolites

Many potent antitumor secondary metabolites isolated from marine sponges are complex polyketides (PKs) and hybrids with non-ribosomal peptides (NRPs) (Piel, 2004; Fusetani, 2010). This sub-section highlights the biosynthetic insights of microbial secondary metabolites that belong to group of PKs, NRPs, and PK/NRP hybrids. Subsequently metagenomic strategies to isolate biosynthetic gene clusters will briefly be described.

1.4.1 Polyketide biosynthesis

Polyketides are a large family of structurally diverse natural products, which include many important pharmaceuticals, agrochemicals, and veterinary agents. These natural products are used clinically as antitumor agents (epothilone and doxorubicin), the antihelminthic agent (avermectin B_1), antibacterial agents (erythromycin A, tetracycline, and nystatin) (Cane *et al*., 1998), and others. Interestingly, the enormous diversity of polyketides are built from simple carboxylic acid monomers through a number of programmed events catalyzed by multifunctional proteins termed polyketide synthases (PKSs) (Cane *et al*., 1998; Llewellyn & Spencer, 2007; Hertweck, 2009). These programmed catalytic events may include selection of starter and extender units, carbon chain length, folding, degree of reduction, termination, and tailoring (Moore & Hertweck, 2001).

In general, PKS-catalyzed processes are similar to those of fatty acid synthases (FASs) that use simple precursors such as acetyl-CoA and malonyl-CoA monomers to build fatty acid chains through repetitive decarboxylative condensations (Hopwood, 1997; Hertweck, 2009). These enzymatic processes are mediated by the repeated groups of protein domains called modules. Each module usually catalyzes one cycle of chain elongation by incorporating one building block and passes the elongated product on to a downstream module (Cane *et al*., 1998; Piel 2010). Each module is subdivided into domains, each of which is responsible for a single chemical manipulation. A ketoacyl synthase (KS) domain catalyzes the chain elongation, an acyltransferase (AT) domain selects the correct acyl-CoA extender, and an acyl carrier protein (ACP) domain serves as anchor to the polyketide chain. Various functional modifications on the β-position of the polyketide backbones are performed by optional domains, such as a ketoreductase (KR) that converts the *β*-keto into a hydroxyl group, a dehydratase (DH) that generates double bonds, and an

enoylreductase (ER) that reduces the double bonds to a saturated intermediate (Rawlings, 2001; Calderone *et al*., 2007; Piel*.,* 2010) (Fig. 1.4.1). The most significant difference between PKSs and FASs are the reductive steps in each elongation. In PKSs, the reductive steps are optional, an intermediate can be partly or fully reduced before being transferred to the next round of elongation. This results in a much more structurally diverse natural product. In contrast, FASs typically fully reduce units after individual elongations (Rawlings, 1998; Rawlings, 2001; Hertweck, 2009).

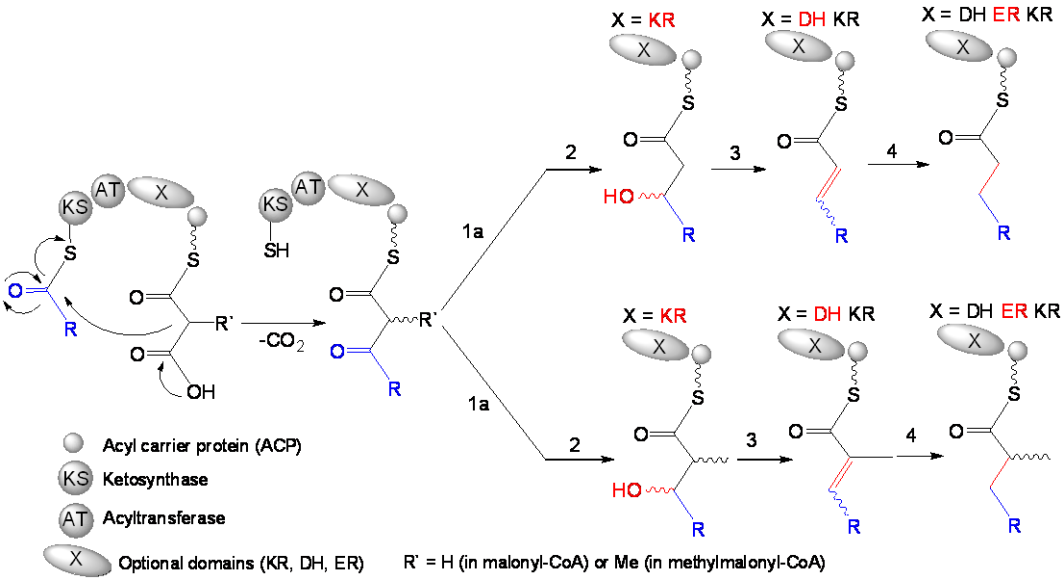


Figure 1.4.1 General biosynthetic steps of a bacterial modular PKS. In a *cis*-AT PKS system, the AT domain is integrated in each module. Notes: x, optional domain(s); 1a, extension with malonyl-CoA; 1b, extension with methylmalonyl-CoA; 2, reduction; 3, dehydration; 4, reduction. Abbreviations: KR, β-ketoacyl reductase; DH, dehydratase; and ER, enoyl reductase.

Bacterial modular PKSs usually employ malonyl-CoA or methylmalonyl-CoA monomers in chain extension. When methylmalonyl-CoA is used as extender units, the resulting intermediates harbor methyl branches at the α-position. Therefore, polyketide structures can be predicted from PKS domain sequences or *vice versa*, because the module architecture and function of PKSs corresponds to the resulting intermediates. This correspondence is known as the colinearity principles (Hertweck, 2009; Piel 2010). These canonical rules are usually applied for a subclass of modular type I PKSs called "*cis*-AT PKSs", in which the AT domain is present in each module. Type I refers to linearly arranged and covalently fused catalytic domains within large multifunctional enzymes (Hertweck, 2009). The

modular *cis*-AT PKS is best exemplified by the well-studied 6-deoxyerythronolide B synthase (DEBS) (Fig. 1.4.2).

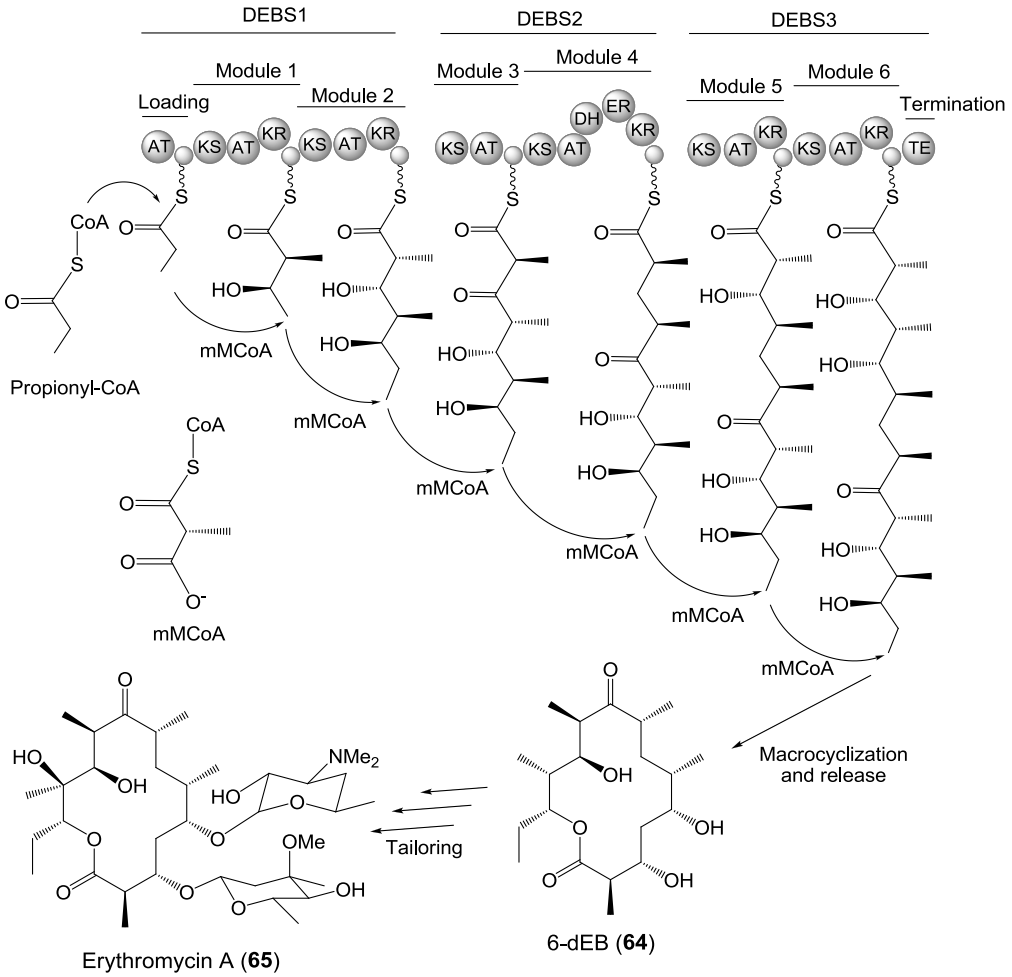


Figure 1.4.2 The well-studied AT PKS, deoxyerythronolide B synthase (DEBS). The loading module containing AT_L loads propionyl-CoA starter unit and transfers the activated propionyl grout to the adjancent ACP. Subsequently the ACP-bound propionyl group is assembled with six (2*S*)-methylmalonyl-CoA extender units catalyzed by a set of core domains (KS, AT, and ACP). The assembly line process is initiated by the KS domain that receives the growing polyketidyl chain from the ACP domain of the preceding module. The AT domain transfers a CoA-linked carboxyacyl extender unit to the Ppant arm of the ACP. The KS then catalyzes decarboxylative condensation between the growing chain and the extender unit to generate a β-ketoacyl thioester-ACP intermediate. Each intermediate undergoes β-keto processing events catalyzed by up to three optional/auxiliary domains (KR, DH, and ER). The resultant final intermediate undergoes macrocyclization and release from the DEBS, resulting in 6-dEB (Hertweck, 2009; Yuzawa *et al*., 2011). This 6-dEB undergoes further enzymatic elaboration by a series of tailoring enzymes such as glycosyl transferases, methyl transferases, and regiospecific hydroxylases, leading to the formation of erythromycin A (Staunton & Wilkinson, 1997). Abbreviations: AT, acyl transferase; ACP, acyl carrier protein; KS, β-ketoacyl synthase; KR, β-ketoacyl reductase; ER, enoyl reductase; DH, dehydratase; TE, thioesterase; and mMCoA, methylmalonyl-CoA.

This PKS system assembles one unit of propionyl-CoA and six units of methylmalonyl-CoA into 6-deoxyerythronolide B (6-dEB) (**64**) scaffold of erythromycin A (Staunton & Wilkinson, 1997, Rawlings, 2001). Erythromycin A

(Ilotycin, Lilly) (**65**) has been well known in clinical medicine to fight against Gram-positive bacterial infections as well as pulmonary infections (Staunton, 1997). It was isolated in 1952 from *Sacharopolyspora erythraea* (formerly *Streptomyces erythreus*) (McGuire *et al.*, 1952), and five years after its discovery, the chemical structure was elucidated (Wiley *et al*., 1957).

Cloning of the three giant genes coding for DEB biosynthesis was simultaneously reported by Leadlay and coworkers (Cortes *et al*., 1990) and Katz and coworkers (Tuan *et al*., 1990). This subsequently has allowed the detailed biochemical and structural studies of this multifunctional megasynthase (Khosla *et al*., 2007) by heterologous expression of the entire DEBS in *Streptomyces coelicolor* (Kao *et al*., 1994) and in *E. coli* (Pfeifer *et al*., 2001), *in vitro* enzymatic reconstitution of 6-dEB (Pieper *et al*., 1995), elucidation of the 3-D structural model (Staunton *et al*., 1996), and genetic engineering of DEBS proteins to generate diverse polyketide structures with certain size and functional groups (Caffrey *et al*., 1992; Roberts *et al*., 1993; Brown *et al*., 1995; Kao *et al*., 1995; Kao *et al*., 1996; McDaniel *et al*., 1999; Xue *et al*., 1999). Due to most of current knowledge about modular PKS derived from DEBS studies, DEBS has recently emerged as the prototype of modular PKS systems (Cane, 2010).

DEBS is a giant multifunctional megasynthase (>3000 amino acids) consisting of three large polypeptides, designated as DEBS1, DEBS2 and DEBS3 (Cortés *et al*., 1990; Donadio *et al*., 1991). Each harbors two modules that are responsible for chain elongation and modification of the newly generated β-ketone. Each module contains a set of catalytic domains (100-400 amino acids each) that controls the incorporation of each precursor into the polyketidyl chain (Khosla *et al*., 1999; Chan *et al*., 2008). The first module, called loading module, harbors an AT domain (designated as AT_L) that preferentially loads and transfers propionyl-CoA to the loading ACP (ACP_L) (Fig. 1.4.3A), whereas the AT domains in the extender modules specifically load (2S)-methlymalonyl-CoA (Marsden *et al*., 1994; Lau *et al*., 1999). In DEBS, the resulting activated propionyl starter unit undergoes repetitive Claisen condensations with six (2*S*)-methylmalonyl-CoA extender units (Fig. 1.4.3A). During erythromycin biosynthesis, all seven DEBS modules containing 28 domains performs 21 steps to make a linear polyketide chain that is subsequently released by TE through direct lactonization, resulting in 6-dEB (Cortes *et al*., 1995; Kao *et al*., 1995; Cane *et al*., 1998; Lau *et al*., 2000). The number of the modules is equivalent

to the number of precursors incorporated into the assembly line. Therefore, the number of modules and the precursors dictate the chain length, while the domain composition in each module determines the structure of each elongation unit (Yuzawa *et al*., 2011).

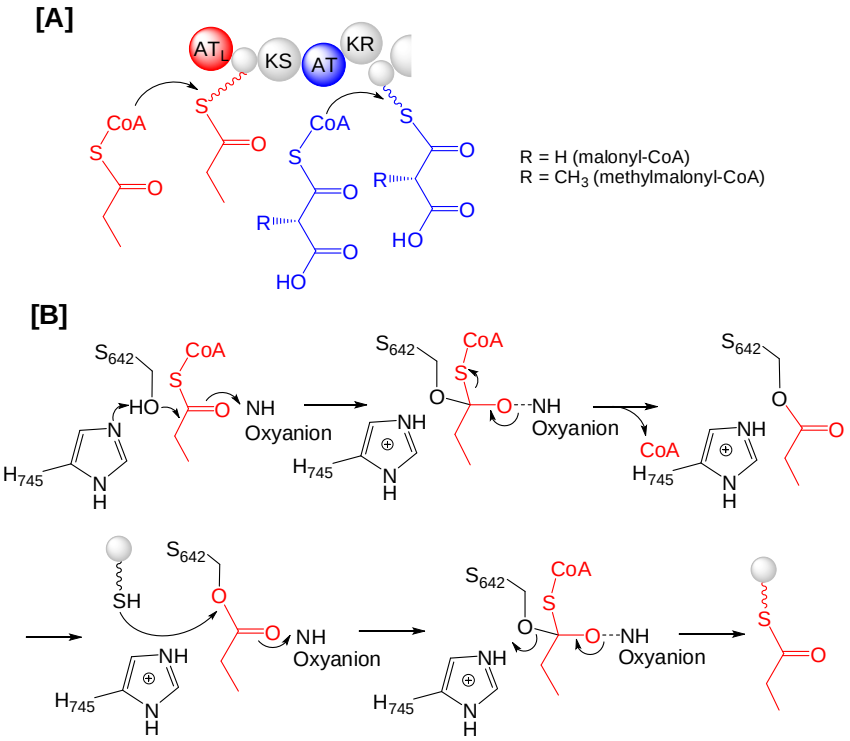


Figure 1.4.3 Basic mechanisms of initiation by AT domain in type I PKS. **[A]** The initiation module may harbor no KS domain at the N-terminus of the assembly line, as reported for deoxyerythronolide B (6-dEB) biosynthesis. DEBS initiation module consists of the didomain AT_L fused with an ACP. In the case, the AT_L selects propionyl-CoA and transfers the activated propionyl group to the P-pant of the adjancent ACP. Subsequently this ACP donates the propionyl group to the next linear modules (Keating & Walsh, 1999; Khosla, 1997). [**B**] Proposed mechanism of the AT domain in DEBS biosynthesis. The active-site Ser (S_{642} in DEBS AT5) in the highly conserved GHSXG motif forms a canonical catalytic dyad by hydrogen bonding to the conserved active His (H_{745} in DEBS AT5), leading to the activation of nucleophilic attack by the active site Ser hydroxyl group on the methylmalonyl thioester carbonyl group. This enhanced nucleophilicity of the Ser then results in the formation of a tetrahedral intermediate, in which the negative charge is stabilized by an oxyanion hole. The active site His protonates the existing CoA moiety, resulting in the replacement of the CoA moiety with ACP. Through a nucleophilic attack on the ester carbonyl group activated by the cysteamine thiol of the ACP phophopantetheine, the second tetrahedral intermediate is generated. This intermediate is resolved again by the active site His, followed by reprotonation of the active site Ser, resulting in the release of the acyl-ACP product (Khosla *et al*., 2007; Smith & Tsai, 2007; Tsai & Ames, 2009).

Using mutantional and computational approaches, key residues within AT domains that correlate with specificity have been successfully identified (Reeves *et al*., 2001; Smith & Tsai, 2007). There are four conserved motifs identified to be responsible for the AT specificity. The first is the RVDVVQ motif located 30 residues upstream of

the catalytic Ser (Haydock *et al*., 1995). Further AT domain analysis by Yadav *et al*. (2003) determined more detailed motifs: [RQSED]V[DE]VVQ for methylmalonyl-AT and ZTX$[AT][QE] for malonyl-AT, in which Z is a hydrophilic residue and $ is an aromatic residue (Smith & Tsai, 2007). Secondly, the GHSXG motif around the catalytic Ser was also identified to contribute to specificity. The X residue is a branched hydrophobic amino acid, in which it is glutamine (Q_{643} in DEBS AT5) in methylmalonyl-specific ATs, whereas in malonyl-specific ATs, the X residue is either valine or isoleucine (Haydock *et al*., 1995; Smith & Tsai, 2007). Third, the YASH motif located 100 residues downstream of the active site Ser were found in methylmalonyl-specific extender ATs in DEBS, while the HAFH motif was identified in malonyl-specific ATs (Haydock *et al*., 1995; Reeves *et al*., 2001; Smith & Tsai, 2007). Fourth, starter AT domain in DEBS has the following specificity motifs: RVEVVQ instead of RVDVVQ, GHSIG instead of GHSQG, and MAAH instead of YASH, suggesting the motif difference between extender and starter ATs (Khosla *et al*., 1999; Smith & Tsai, 2007). The transfer of acyl moieties between CoA and ACP by the AT domains occurs through a ping-pong bi-bi catalytic mechanism using a Ser-His catalytic dyad on the basis of studies of the reactions catalyzed by the type II malonyl-CoA:ACP transacylase from *E. coli* (Fig 1.4.3B) (Ruch & Vagelos, 1973; Smith & Tsai, 2007; Khosla *et al*., 2007; Tsai & Ames, 2009).

The KSs in modular PKSs employ a Cys-His-His triad at the active center to form a new carbon-carbon bond through a catalytic mechanism that is similar to that of porcine FAS KS domains (Witkowski *et al*., 2002; Zhang *et al*., 2006; Smith & Tsai, 2007) (Fig. 1.4.4). In principle, the KS catalytic mechanism begins with the transfer of the upstream acyl group to the KS active site, followed by the KS-catalyzed decarboxylation of the downstream (methyl)malonyl-*S*-ACP, resulting in a carbanion that elongates and translocates the intermediate to the downstream ACP (Keating & Walsh, 1999).

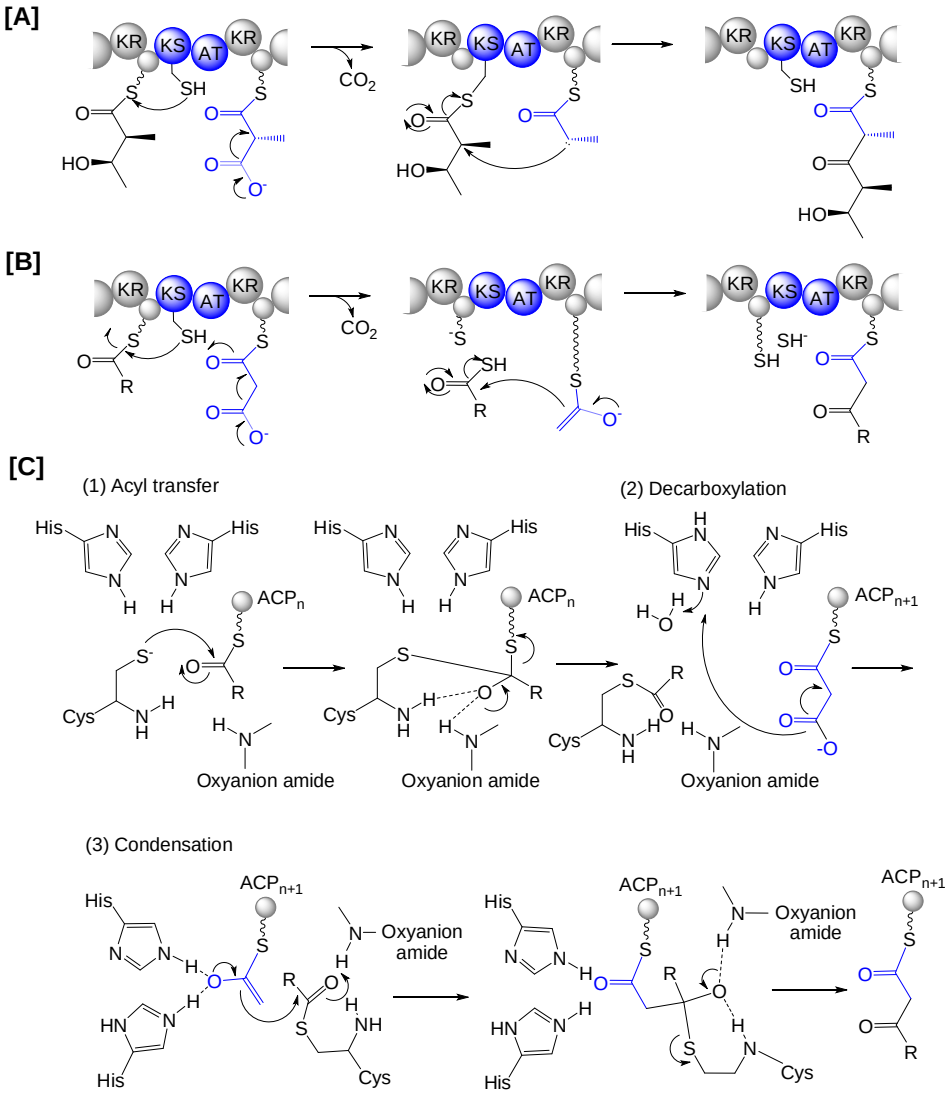


Figure 1.4.4 Basic mechanisms of elongation by KS domain in type I PKS. **[A]** In 6-Deoxyerythronolide B (6-dEB) biosynthesis, chain extention occurs through a series of sequential condensation of one unit of propionyl CoA and six units of methylmalonyl-CoA units. The KS domain accepts the growing chain from the upstream ACP onto the thiol group (-SH). Then the KS decarboxylates the downstream ACP-attached methylmalonly thioester to create a carbanion that attacks the upstream acyl-*S*-ACP thioester in a Claisen condensation, resulting in a C-C bond. **[B]** During chain extension with malonyl-CoA units, decarboxylation of the downstream malonyl-*S*-ACP results in a nucleophilic thioester enolate that subsequently attack the upstream acyl-*S*-ACP thioester. This leads to the formation of a C-C bond through a Claisen condensation (Keating & Walsh, 1999; Fischbach & Walsh, 2006; Chan *et al*., 2008). **[C]** In acyl transfer step, the negative charge on the thioester carbonyl is neutralized by the oxyanion hole generated by the backbone amides of the active site cysteine and an amino acid residue near the C-terminus, resulting in the stabilization of the tetrahedral intermediate.The free ACP is released and the chain extender (malonyl- or methylmalonyl-*S*-ACP) binds to the enzyme. In decarboxylation step, the malonyl C_3 carboxylate is attacked by an active-site water molecule likely activated by a conserved histidine, resulting in the release of bicarbonate. The second conserved histidine likely facilitates the formation of the carbanion at C_2, thereby stabilizing the enol intermediate. In the condensation step, the tetrahedral intermediate is generated by the attack of the carbanion on the acyl-thioester. This is stabilized initially by the oxyanion hole that then breaks down to form the β-ketoacyl product (Smith and Tsai, 2007).

The type I KRs belong to the short-chain dehydrogenase family (Persson *et al*., 2003). Their structure harbours two principal subdomains: the N-terminal structural subdomain and the C-terminal catalytic subdomain (Khosla *et al*., 2007; Smith & Tsai, 2007). The catalytic subdomain has the typical SDR motifs that include the TGxxxGxG motif, the D_{63} and NNAG motifs, the active-site tetrad Asn-Ser-Lys-Tyr, and the PG motif (Persson *et al*., 2003). The active-site tetrad in the *act* KR is composed of N_{114}-S_{144}-Y_{157}-K_{161}. In DEBS KR1 and Tyl KR1, the positions of the Asn and Lys are switched (K_{1776}-S_{1800}-Y_{1813}-N_{1817}) (Keatinge-Clay, 2007; Keatinge-Clay and Stroud, 2006). The active-site tetrads of these KR domains relies on the proton-relay mechanism as proposed for the *E. coli* FabG in the reduction of the ketone to the alcohol (Fig. 1.4.5A) (Zhang *et al*., 2003; Price *et al*., 2001; Price *et al*., 2004; Smith & Tsai, 2007; Tsai & Ames, 2009). Active KR domains are indicated by the presence of the conserved Rossmann fold motif (GGXGXXG) as well as the conserved catalytic motif (K-S-Y) (Reid *et al*., 2003). The stereochemistry of a β-hydroxyl group formed by the modular PKS KR domains can be predicted based on the presence or absence of LDD and PxxxN motifs. The presence of these motifs predicts the 3*R* stereomer (or D-stereomer) whereas the absence of these motifs predicts the 3*S* stereomer (or L-stereomer) (Fig. 1.4.5B) (Caffrey, 2003; Reid *et al*., 2003; Baerga-Ortiz, 2006). A functional DH domain is usually recognizable by the presence of the conserved HxxxGxxxxP motif near to its N-terminus (Bevitt *et al*., 1992), in which the histidine and gylcine (G) form the catalytic dyad for the activity (Cortes *et al*., 1995; Tsai & Ames, 2009; Valenzano *et al*., 2010). In addition, a functional DH domain usually contains the conserved Dxxx(Q/H) motif, in which the catalytic aspartic acid (D) is essential for the activity (Fig. 1.4.5C) (Keatinge-Clay, 2008; Valenzano *et al*., 2010).

The ER active site may reside in a cleft containing the NADPH-binding site and other active site residues. The NADPH-binding site is recognized by the GxGxxG/AxxxG/A motif as has been confirmed by mutagenesis of the animal FAS KR (Witkowski *et al.,* 2004). Cofactor binding is assumed to be stabilized by hydrophilic residues in the active site via hydrogen bond interactions with the hydroxyl groups on the nicotinamide ribose. In particular the active site Tyr and Lys should be important for the cofactor binding and catalytic mechanism, as proposed for type II FAS ER domain (Smith & Tsai, 2007).

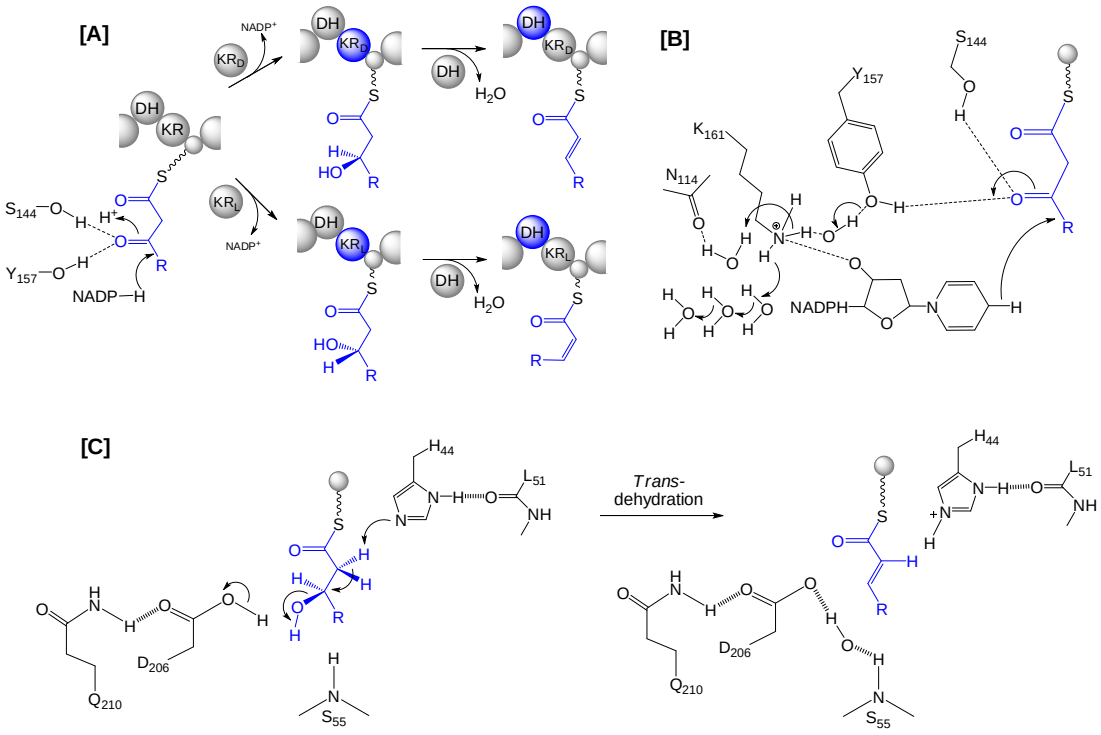


Figure 1.4.5 Proposed mechanisms of modifications by KR and DH domains in modular type I PKS. **[A]** The proposed stereospecific reduction of KR domain based on SDR homology modeling. The substrate is rotated about the axis of the reduced carbonyl bond to orient the substrate differently in the two types of KRs. The relative positions of the catalytic residues and the NADPH are maintained for the two different stereochemical products, and only single active site geometry is required for each type of KRs (Reid *et al*., 2003). **[B]** The proton-relay mechanism of KR domain in the reduction of the ketone to the alcohol. Tyr (K_{161}) and Lys (Y_{157}) form hydrogen bonds with the NADH ribose and nicotinamide ring. Four crystalline water molecules form hydrogen bonds with N_{114} and K_{161}. These waters form a proton-relay network as proposed by Price *et al* (2004) for *E. coli* FabG-$NADP^+$ (Korman *et al*., 2004; Tsai and Ames, 2009). **[C]** The proposed mechanism of DH4 of DEBS. The conserved Gly as a part of the HxxxGxxxxP motif serves to make a turn, enabling van der Waals interactions between H_{44} and P_{53}. The catalytic D_{206} as a part of the Dxxx(Q/H) motif makes hydrogen bonds with the side chain of Q_{210} (Tsai and Ames, 2009).

Once the resultant DEB intermediate reaches the final carrier domain at the last module, the intermediate is subsequently released by a specialized thioesterase (TE) domain. In principle, there are two general mechanisms for chain release from by TE domains: (i) intramolecular capture by a nucleophile leading to macrolactonization with concomitant release, as has been shown for erythromycin and DEBS3) (Fig. 1.4.6), and (ii) intermolecular hydrolysis that releases the free acid (Keating & Walsh, 1999; Staunton & Wilkinson, 1997); Tsai *et al*., 2002; Khosla *et al*., 2007; Du & Lou, 2009) as shown by linear PKs such as thailandamide (Nguyen *et al*., 2008) and bacillaene (Moldenhauer *et al*., 2007; Butcher *et al*., 2007).

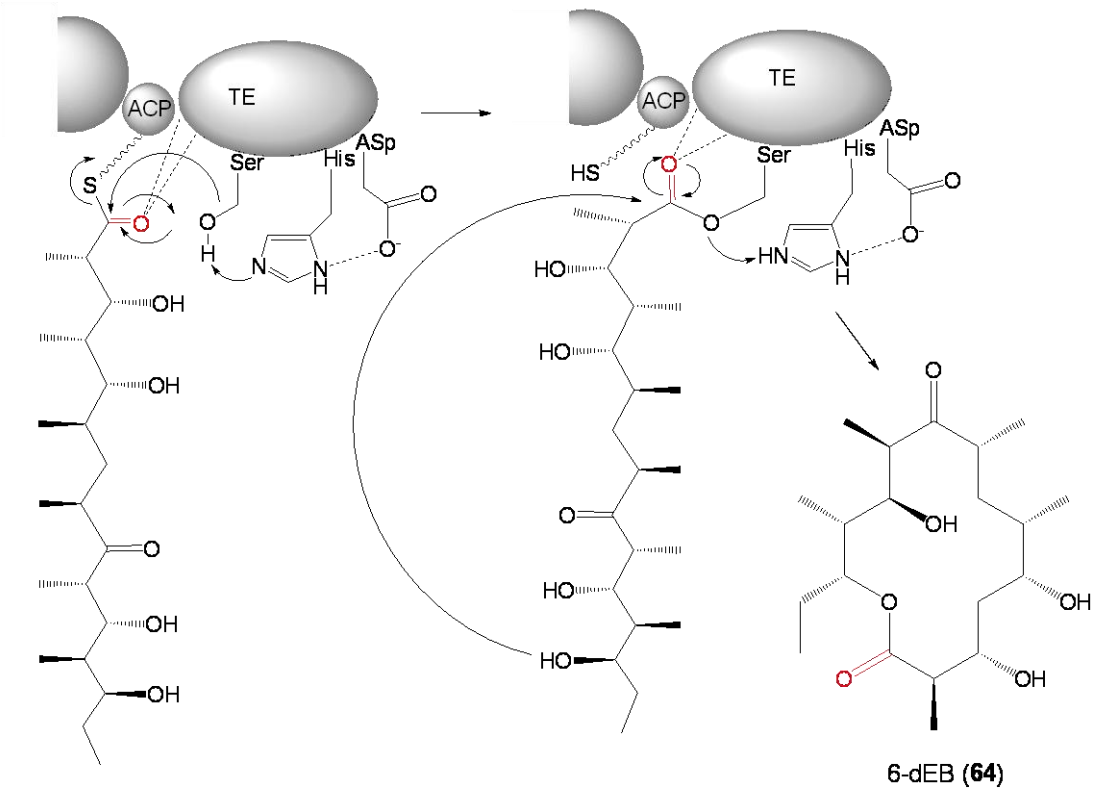


Figure 1.4.6 Basic mechanisms of termination in 6-deoxyerythronolide B (6-dEB) biosynthesis.Termination in DEBS system occurs via a TE-mediated release mechanism, resulting in the direct lactonization of the final intermediate in DEBS. The active site of the TE domain contains a catalytic triad consisting of Ser_{142}, His_{259}, and Asp_{169}. It was proposed that the His_{259} residue stimulates nucleophilic attack by Ser_{142} on the polyketide acyl-ACP-TE, resulting in transfer of the polyketide acyl chain from the thiol of the neigbouring ACP to the OH group of the Ser_{142} on the TE. The resulting oxygen-ester on the acyl-*O*-Ser intermediate is then cleaved by nucleophilic attack of the OH group at C-13 to release the product DEB. Adapted from Staunton and Wilkinson (1997), Tsai *et al* (2002), Khosla *et al* (2007), and Du & Lou (2009).

The classical colinearity principles do not always apply for bacterial modular PKSs. In some cases, the module architecture does not fit to the resulting polyketide structures. The PKS systems that deviate from the coliniearity rules are often characterized by the lack of AT domains in each module. Instead, a free-standing AT domain is responsible for loading acyl building blocks. Due to the absence of integrated AT domains in individual modules, as only one type of building block is loaded, usually malonyl-CoA. In these PKS systems termed "*trans*-AT PKSs" (Fig. 1.4.7), methyl branches are introduced at the α-positions of growing intermediate through *S*-adenosylmethionine (SAM)-dependent α-methylation by methyltransferase (MT) domains present in the PKS modules (Piel, 2010).

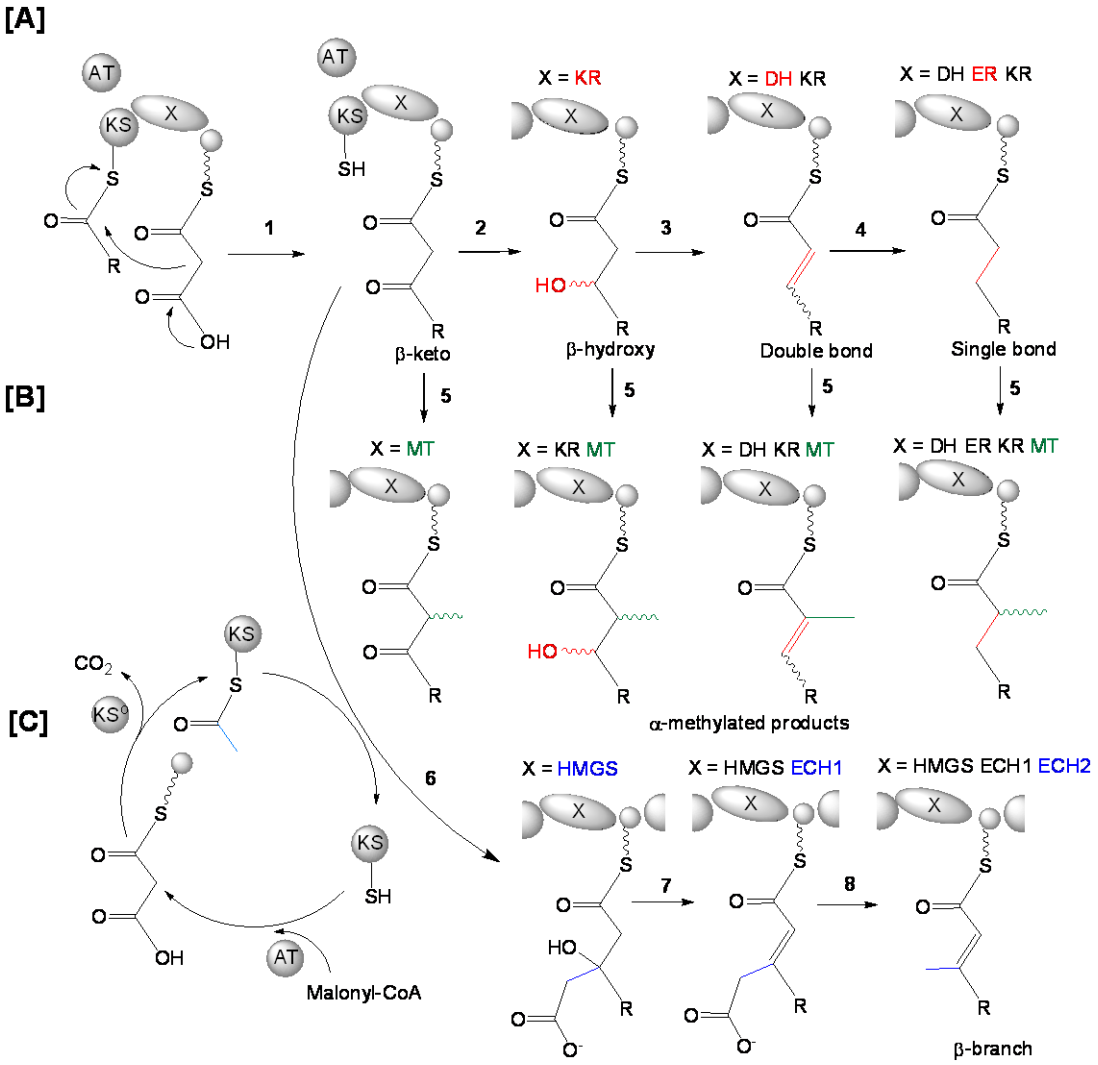


Figure 1.4.7 Basic mechanisms of *trans*-AT PKS, showing precursor loading and elongation with various reductive modifications at β-positions **[A]**, α-methylation by methyltransferase **[B]**, and β-methylation that involves HMGS-mediated aldol addition of a free standing acetyl-ACP with β-ketoacyl ACP, and further ECH1-catalysed dehydration and ECH2-catalysed decarboxylation **[C]**. Notes: 1, condensation; 2, reduction; 3, dehydration; 4, reduction; 5, methylation; 6, aldo addition; 7, dehydration; 8, decarboxylation; and x, optional domains. Adapted from Nguyen *et al* (2008), Gu *et al* (2009), and Piel (2010).

Interestingly many *trans*-AT PKS systems are able to perform unusual branching at the β-position of acyl-ACP intermediates (Piel, 2002; El-Sayed *et al*., 2003; Edwards *et al*., 2004; Piel *et al*., 2004; Calderone *et al*., 2006; Chang *et al*., 2007; Fisch *et al*., 2009) in a process resembling isoprenoid-like methylation. The β-branch incorporation occurs through a KS^{0}-mediated decarboxylation of malonyl-ACP to acetyl-ACP. This is followed by an HMGS-mediated aldol addition of acetyl-ACP to β-ketoacyl ACP and subsequent ECH1-catalysed dehydration and ECH2-catalysed decarboxylation (Calderone *et al*., 2006). Further work by Piel and coworkers on relocating the terminal TE domain to the up-stream modules deduced the timing of

β-branching in the bacillaene PKS, which occurs during elongation (Moldenhauer *et al*., 2007). The presence and biochemistry of HMGSs, ECHs, and other enzymes may determine the β-branching types in polyketides, including methyl-, ethyl-, methylene, ester, or heterocyclic moieties. Variation in these β-branching components can generate remarkably diverse polyketide moieties (Gulder *et al*., 2011).

1.4.2 Nonribosomal peptide biosynthesis

Many microorganisms produce peptide antibiotics and other biologically active substances through biosynthetic pathways catalyzed by large multimodular protein templates, termed nonribosomal peptide synthetase (NRPS) (Finking & Marahiel, 2004; Marahiel & Essen, 2009). Many of the nonribosomal peptides are clinically used drugs. Notable examples are cyclosporine A, penicillin, and vancomycin (Kleinkauf & Döhren, 1990). The remarkable structural and functional diversity in nonribosomal peptides arises from 20 proteinogenic amino acids and a larger variety of nonproteinogenic amino and aryl acids as the building blocks during chain elongation and post assembly (Fischbach and Walsh, 2006). Nonproteinogenic amino aryl acids incorporated into the NRPS assembly lines includes D-isomers, carboxy acids, N-methylated residues, heterocyclic rings and fatty acids. Common transformations that further diversify NRP structures during chain elongation include α-ketone reduction, epimerization, and heterocyclization (Walsh and Fischbach, 2010). Post modifications also contribute to the diversity of NRP structures, including glycosiylation and oxidative cross-linking (Sieber and Marahiel, 2003).

NRPSs apparently use a similar strategy as PKSs for peptide biosynthesis (Du *et al*., 2001). It has a modular structure that is composed of a series of modules. Each module is responsible for the incorporation and modification of an amino acid unit into the growing peptide chain. Therefore the order and number of the modules in a NRPS system indicate the sequence and number of building blocks in the resulting peptide product (Cane, 1997; Du *et al*., 2001). A typical NRPS module harbors at least three domains: an adenylation (A) domain responsible for amino acid activation and incorporation, a thiolation (T) domain for thioesterification of the activated amino acid residue, and a condensation (C) domain for elongation of the growing peptide chain thorugh transpeptidation (Cane *et al*., 1998; Cane & Walsh, 1999; Konz & Marahiel, 1999; Du *et al*., 2001; Schwarzer & Marahiel, 2001).

In NRPS, selection and activation of the substrate amino acids are performed by adenlyation (A) domains to generate the corresponding aminoacyl adenylate. The resulting adenylated products are then attacked by the nucleophilic thiol group of a 4`-phosphopentethein covalently bond with a peptide carrier protein (T) (Fig 1.4.8A). Resolving the 3-D structure of the phenylalanine activating A-domain of Gramicidin Synthetase A1 (PheA-GrsA1 in a complex with AMP and phenylalanine has enabled the identification of their amino acid residues generally responsible for the coordination of ATP and substrate (Conti, *et al*., 1997; Schwarzer and Marahiel, 2001; Schwarzer, *et al*., 2003). These residues are then shown to be highly conserved, forming ten core motifs (A_1-A_{10}) based on the alignment with other A-domains (Schwarzer, *et al*., 2003). Subsequently, general rules, also known as nonribosomal codes, for the substrate specificity of A domains have recently been established, thereby offering exciting opportunities for engineered production of new peptide analogs (Stachelhaus *et al.,* 1999; Challis *et al*., 2000).

There are various other strategies of monomer activation for the formation of peptide bond in secondary metabolites. One of them has recently been reported by Walsh and coworkers for a peptide bond-forming enzyme in the biosynthesis of the antibiotic pacidamycin. This enzyme, called tRNA-dependent aminoacyl transferase (PacB), uses tRNA-bound amino acids as substrates to generate peptide bonds (Fig 1.4.8B) (Zhang *et al*., 2011; Walsh & Zhang, 2011). Another strategy for substrate activation is based the acyl-AMP ligase family as reported for the biosynthesis of the tripeptide antibiotics dapdiamides (Hollenhorst *et al*., 2009; Dawlaty *et al*., 2010). This enzyme mediates the formation of a peptide bond between the activated acyl-AMP intermediate and an amino group of a free cosubstrate (Fig. 1.4.8C) (Kadi *et al*., 2007; Giessen & Marahiel, 2012). Another substrate activation strategy relied on the ability of cyclodipeptide synthases (CDPSs) to use two tRNA-bound amino acids as substrates for the direct formation of two new peptide bonds. These tRNA-bound amino acids are generated by the activity of aminoacyl-tRNA synthetases (aaRSs) (Fig 1.4.8D). This strategy was reported in the biosynthesis of a diketopiperazine (Vetting *et al*., 2010; Bonnefond *et al*., 2011) and albonoursin (Lautru *et al*., 2010). Another monomer activation strategy involves the activity of ATP-grasp ligases containing an atypical ATP-binding site, called the ATP-grasp fold, as reported for the DdaG of dapdiamide synthetase. This enzyme family catalyzes the attachment of a phosphate group on the substrate carboxylate to form an acylphosphate intermediate through the hydrolysis of ATP into ADP. The resultant activated

carboxylate is subsequently attacked by an amino group of a free substrate leading to the formation of a new peptide bond (Fig. 1.4.8E) (Fawaz *et al*., 2011).

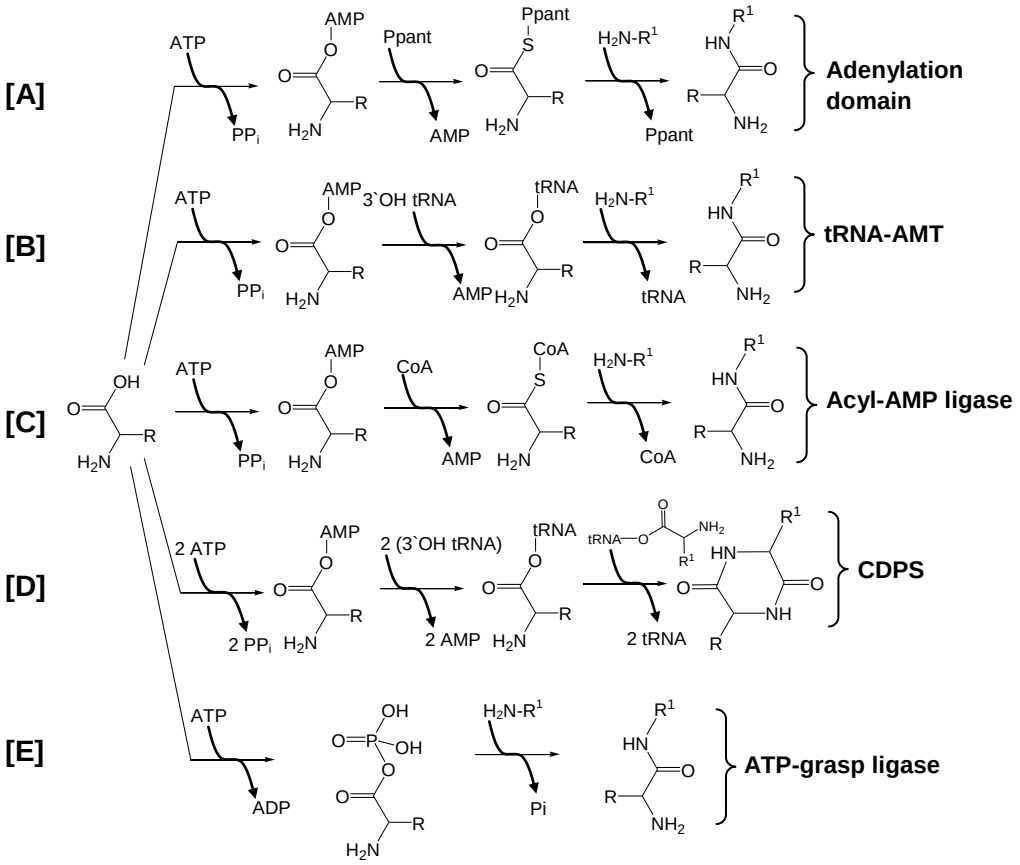


Figure 1.4.8 Various strategies of amino acid activiation to form peptide bonds in secondary metabolite biosynthesis. **[A]** In NRPS, selection and activation of building blocks are performed by adenlyation (A) domains to form adenylated products. Then the aminoacyl moiety of the adenylate is covalently tethered to a prosthetic phosphopantetheinyl (Ppant) group of a T domain. When a peptidyl-S-Ppant donor is attacked by a monomeric aminoacyl-*S*-Ppant nucleophile, peptide-bond formation and chain translocation occur (Eppelmann *et al*., 2002; Cane & Walsh, 1999). **[B]** tRNA-bound amino acids are initially generated by aminoacyl-tRNA synthetases (aaRSs). The the tRNA-dependent aminoacyl transferase uses a tRNA-bound amino acids as donor and delivers it to a T-loaded nonribosomal dipeptide as acceptor, thereby forming a peptide bond (Zhang *et al*., 2011). **[C]** The substrate carboxylate as an adenylate is initially activated by an acyl-AMP ligase. The resultant activated carboxylate reacts with CoA to form the CoA-thioester that is subsequently attacked by an amino group of free cosubstrate, thus resulting in a new peptide linkage (Kadi *et al*., 2007; Dawlaty *et al*., 2010; Koetsier *et al*., 2011; Giessen & Marahiel, 2012). **[D]** CDPs activate carboxylates for nucleophilic attack by donor amino groups to bring together monomers in the assembly (Hollenhorst *et al*., 2009). **[E]** ATP-grasp ligase activates carboxylate substrates via acyl-phosphate formation with distinct ATP cleavage patterns to form tandem peptide bonds (Fawaz *et al*., 2011).

In the incorporation of an amino acid into a growing peptidyl chain, the activated amino acid (aminoacyl-adenylate) is attacked by the enzyme-attached thiol moiety 4`-phosphopantetheine (4`-PP) results in an aminoacyl-*S*-Pant thioester. Subsequently acyl-*S*-Pant thioesters on two adjancent PCP domains facilitate amide

formation. This occurs through the transfer of the carboxyl group of one amino residue to the adjancent amino group of the next amino acid (Marahiel *et al*., 1997) (Fig. 1.4.9A).

Recently there have increasing numbers of secondary metabolites, in which their biosynthesis are encoded by NRPS-PKS hybrid assembly lines combining both NRPS and PKS modules. Some examples are surfactin (Cosmina *et al*., 1993), rapamycin (Aparicion *et al*., 1996), yersiniabactin (Gehring *et al*., 1998), mycosubtilin (Duitman *et al*., 1999), bleomycin (Du *et al*., 2000), nostopeptolide A (Hoffmann *et al*., 2003), and nostocyclopetide (Becker *et al*., 2003). An interesting aspect of these hybrid assembly lines is the mechanisms underlying the elongation of a PKS-bound growing polyketidyl chain with an NRPS module or vice versa. This hybrid chain elongation occurs through the formation of either C-N or C-C bonds in NRPS/PKS interfaces. At the PKS/NRPS interface, a ketide chain formed by an upstream PKS module is accepted by the C domain of the downstream NRPS module. The C domain subsequently mediates a C-N bond formation, allowing the elongation of the polyketide chain with an amino acid (Du *et al*, 2001; Fischbach & Walsh, 2006) (Fig 1.4.9B). At the NRPS/PKS interface, the KS domain accepts an upstream peptide chain rather than a ketide chain and catalyzes the formation of a C-C bond leading to the elongation of the peptide chain with a short carboxylic acid (Fig 1.4.9C) (Du *et al*., 2001; Fischbach & Walsh, 2006). The C-N or C-C bond formation in PKS/NRPS hybrid assembly lines are best exemplified by the HMWP2 and HMWP1 proteins of yersiniabactin synthetase (Miller *et al*., 2002) and the first NRPS subunit of FK520 synthase (Gatto *et al*., 2005).

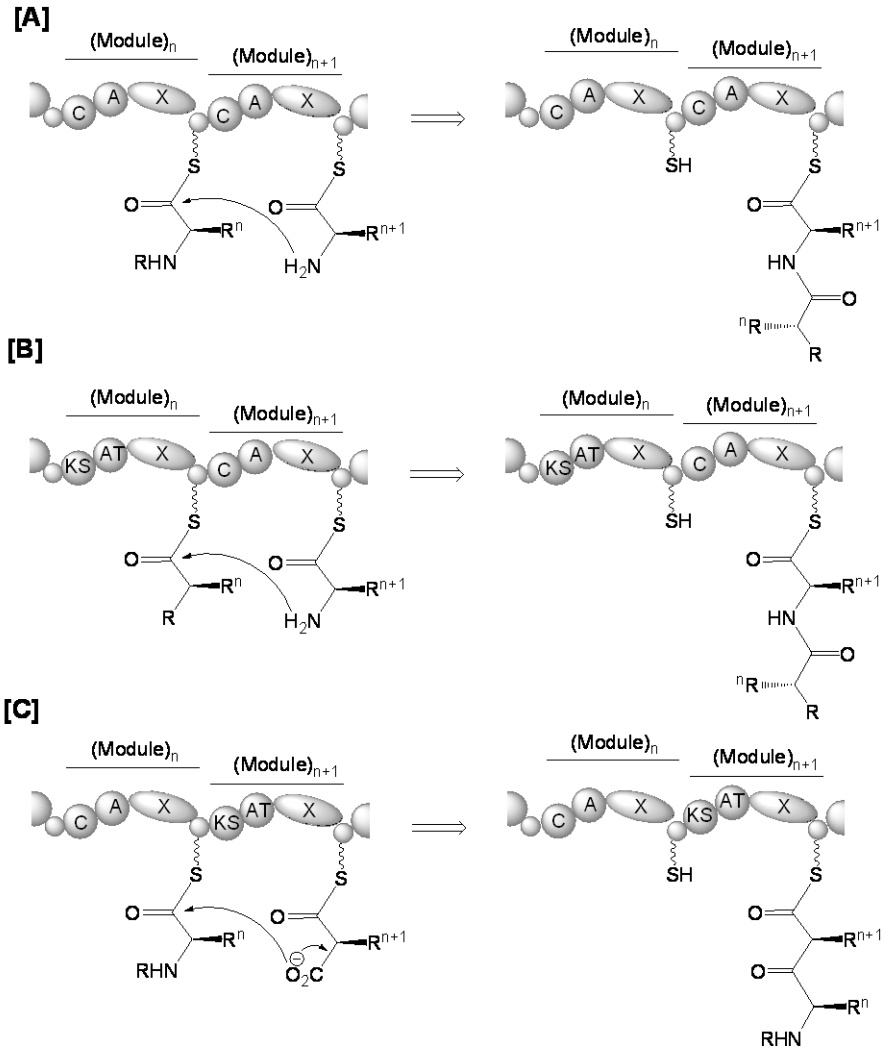


Figure 1.4.9 Bond formation in nonribosomal peptides and hybrids with polyketides. **[A]** C-N bond formation in the NRPS system. The C domain catalyzes nucleophilic attack of the amino group of an amino acyl-*S*-T donor on the acyl group of an upstream electrophilic acyl- or peptidyl acyl-*S*-T chain, resulting in the formation of peptide bond (Cane and Walsh, 1999). **[B]** C-N bond formation in the PKS/NRPS hybrid system is mediated by the C domain. In particular the C domain catalyzes the nucleophilic substitution between the acyl group of the growing polyketidyl chain of acyl-*S*-ACP from the upstream PKS module and the amino group of the downstream aminoacyl-*S*-T, resulting in the integration of the polyketide chain with an amino acid via C-N bond formation (Du *et al*. 2001) **[C]** C-C bond formation in the NRPS/PKS hybrid system. This is catalyzed by the C domain by transferring the growing peptide chain of peptidyl-*S*-T from the upstream NRPS module to the KS active site Cys to generate a peptidyl-*S*-KS. Subsequent decarboxylative condensation with its cognate malonyl-S-ACP leads to the extension of the peptide chain with a short carboxylic acid (Du *et al*. 2001; Fischbach and Walsh, 2006; Giessen and Marahiel, 2012). Note: x, optional domain(s).

1.4.3 Metagenomic strategies to isolate biosynthetic pathways

Searching for biosynthetic pathways of interest in complex symbiotic systems such as a sponge is extremely challenging due to the high microbial complexity. Sponge symbiotic systems, for example, can harbor a high density range of bacterial cells between 6.4 x 10^8 to 1.5 x 10^9 bacterial cells per ml of sponge extract (Scheuermayer *et al*., 2006) and may contains hundreds to thousands of microbial species (Taylor *et al*., 2007). The high microbial complexity of a sponge system suggests that the symbiotic assemblage contain numerous homologous genes from diverse pathways (Piel *et al*., 2004a).

One of the strategies to access biosynthetic pathways of pharmaceutical relevance from a complex microbial consortium is based on phenotypic detection of growing *E. coli* clones on plates, such as antibacterial activity, color production, and signaling phenomena (Courtois *et al*., 2003; Banik & Brady, 2006; Brady, 2007; Brady *et al*., 2009). Using top agar overlay assays, Brady and Clardy screened a cosmid library for antibacterially active *E. coli* clones that led to the characterization of various new long-chain N-acylated amines with antibacterial activities (**66**) (Brady and Clardy, 2000) and a new antibiotic isonitrile functionalized indole (**67**) (Brady and Clardy, 2005). By observing color production, Brady *et al* (2001) identified some blue-colored clones in a cosmid library constructed from soil eDNA. They found that one of these clones produced the broad-spectrum antibacterial pigment violacein (**68**) (Brady *et al*., 2001). Based on visual screening of pigmented clones, Gillespie *et al* (2002) isolated brown-colored clones, in which one of them produced the antibiotics turbomycin A (**69**) and turbomycin B (**70**) (Gillespie *et al*., 2002) (Figure 1.4.10).

Screening of library could also be performed by combining direct cloning approach and chemical analysis to detect the biosynthetic products, as has successfully been shown for cyanobactin family by Jaspars group (Long *et al.,* 2005) and Schmidt group (Schmidt *et al*., 2005). Cyanobactins such as patellamide A (**9**), ulithiacyclamide (**71**), and trunkamide (**72**) were isolated from the ascidian *Lissoclinum patella* (Degnan *et al*., 1989) (Figure 1.4.10). Through cell separation studies, it was suggested that the ascidian`s symbiont *Prochloron* as cyanobactin producer (Degnan *et al*., 1989).

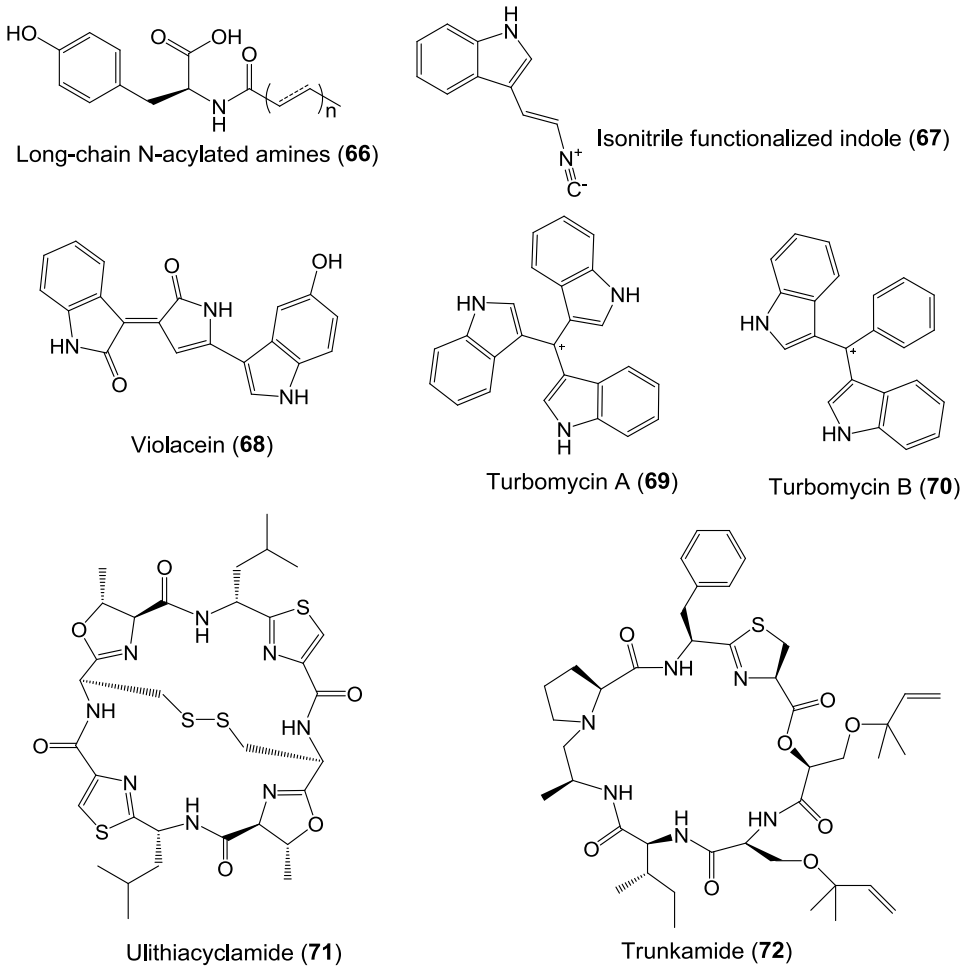


Fig. 1.4.10 Some examples of molecules from the metagenomic expression of biosynthetic pathways. Compounds **66**-**70** were discovered by Clardy lab and Brady lab on the basis of phenotypic detection of growing *E. coli* clones on plates, such as Long-chain *N*-acylated amines (antibacterial activity), color production, and signaling phenomena (Brady and Clardy, 2000; Brady *et al*., 2001; Brady and Clardy, 2005; Gillespie *et al*., 2002). Molecules **9**, **10**, **71**, and **72** of which the biosynthetic pathways have been cloned from enriched fractions of the ascidian symbiotic bacterium *Prochloron* spp. by Jaspars group (Long *et al.,* 2005) and by Schmidt group (Schmidt *et al*., 2005).

Jaspars and coworkers generated a large-insert library from *Prochloron*-enriched DNA using the *λ*Red system, which was subsequently stored in 96-well microtitter plates. Clones from each plate were combined and fermented in broth medium. The cells in the culture were removed by centrifugation, leaving free-cell broth. To remove HPLC interferences, the broth was concentrated by solid-phase extraction and eluted with methanol. The dried residue was cleaned by water/didchloromethane partitioning and subsequently analyzed by HPLC-MS. Once plates containing patellamide D and ascidiacyclamide were located, combined clones from each plate column were cultured and analyzed by LC-MS. Subsequently individual clones from each column containing the desired metabolites were fermented separately for patellamide clone selection (Long *et al.,* 2005). In the

same year in 2005, Schmidt and coworkers reported the identification the biosynthetic gene clusters for ribosomal precursor peptides encoding cyanobactins through shotgun sequencing of *Prochloron* mategenome. Subsequent expression of the gene clusters in *E. coli* led to the production of patellamides A (**9**) and C (**10**) (Schmidt *et al*., 2005; Schmidt & Donia, 2009).

Alternatively, libraries can be screened by directly blotting colonies from agar plates onto Nylon membranes, which are then analyzed by hybridization as described by Ginolhac *et al*. (2004). However, libraries on agar plates can not be stored for a longer time. As the alternative, clones on a plate can be combined into pools that can be added with glycerol and stored at -80^{o}C for a long time. All clone pools can be screened by combining pool dilutions and PCR-based approach (Piel, 2002). This strategy enabled cloning of a *trans*-AT PKS gene cluster from the metagenomic DNA of the beetle *Paederus fuscipes*. Bioinformatic analysis predicted that the isolated gene cluster code for the biosynthesis of pederin (Piel, 2002, Piel *et al*., 2004a), which were subsequently supported by functional studies (Zimmermann *et al*., 2009). This discovery marked the first report of an entire biosynthetic gene cluster of an uncultivated symbiotic bacterial origin.

Discovery of the pederin gene cluster subsequently provided guidance to isolate onnamide-type compounds from much more complex symbiotic systems, marine sponges. By phylogenetic analysis, Piel *et al* (2004b) identified *trans*-AT PKS genes among various diverse PKS types obtained from the metagenome of the Japanese sponge *T. swinhoei*. Based on the information obtained from the isolated *trans*-AT PKSs, the metagenomic library was constructed from total *T. swinhoei* DNA and then screened. This led to the cloning of a large part of onnamide gene cluster (Piel *et al*., 2004b). In the studies of onnamide-type compounds, Piel group developed a method that enabled to rapidly identify or target the biosynthetic genes in highly complex sponge symbiont assemblages (Hrvatin & Piel, 2007; Fisch *et al*., 2009). The method termed a liquid gel pool analysis involves the construction of a metagenomic library in a 3-D format, allowing well-separated colonies grow in the average density of 1000 clones/ml. This is followed by rapid screening by whole-cell PCR. In PCR-screening step, clone pools in the same rows were combined into super-pools and screened by whole-cell PCR. Positive clone pools were amplified at lower densities and screened further until individual positive clones were detected. This approach enables one not only to catch representative genomes from the wide

diversity of microbial symbionts in nature, but also to facilitate rapid screening of the biosynthetic pathways of interest among many diverse non-target pathways present in a sponge symbiotic system. The advantages of employing this screening method are that the whole procedure from transfection to the pure clone was completed in four days by a single person, being highly economic in terms of labor, consumables and freezer space (Hrvatin & Piel, 2007). With these screening strategies, Piel group have successfully isolated the entire gene clusters encoding the biosynthesis of onnamide A, theopederin and psymberin from complex sponge systems. The architectures of the cloned biosynthetic gene clusters indicated bacterial origin, thereby providing the first genetic evidence of the bacterial origin of marine invertebrate-derived natural products (Piel *et al*, 2004b, 2007; Fisch *et al*., 2009).

Recently, Piel and coworkers have developed a novel predictive approach to enable the accurate prediction of the structures of polyketides encoded on *trans*-AT PKS genes (Nguyen et al., 2008). The approach was developed on the basis of functional and phylogenetic studies (Nguyen *et al*., 2008), in which its utility has been demonstrated in the genome mining of thailandamide and elansolid polyketides (Nguyen *et al*., 2008; Teta *et al*., 2010) and the metagenome mining of psymberin (Fisch *et al*., 2009). Therefore, the approach is especially very useful for targeting biosynthetic pathways of pharmacological significant from complex microbial consortia due to the high complexity of genomes coding for many diverse biosynthetic pathways. In addition, this approach is particularly very useful for natural product discovery research from marine sponge, due to the fact that only small percentage of sponge PKS genes were found to encode complex polyketide (Fieseler *et al*., 2007; Hochmuth & Piel, 2009). Subsequent expression of biosynthetic gene clusters in culturable microbes not only reveals fundamental insights into the chemistry and ecology of uncultured symbionts but also provide access to sustainable production systems. Moreover, with the genes in hand, pathways could be genetically altered to produce structurally novel analogs with improved pharmacological profiles (Uria and Piel, 2009).

Chapter 2

Research Goals

Highly complex consortia of bacterial symbionts associated with sponges have been frequently proposed to be the true source of many sponge-derived metabolites. However, the majority (>99%) of these symbionts are uncultivable (Amann *et al*., 1995; Hugenholtz *et al*., 1998; Friedrich *et al*., 2001; Webster and Hill, 2001), thereby hampering efforts to test the symbiont hypothesis and to study biotechnological applications. Using metagenomic-based approaches, Piel and coworkers have verified the symbiont hypothesis by cloning entire gene clusters from bacterial symbionts responsible for the production of the antitumor polyketides, onnamide A (**5**) from Japanese specimens of *Theonella swinhoei* (Piel *et al*., 2004b; Hrvatin & Piel, 2007; Nguyen *et al*., 2008) and psymberin (**6**) from the sponge *Psammocinia* aff. *bulbosa* (Fisch *et al*., 2009).

[A]

Photo by Y. Nakao

[B]

Photo by J. Tanaka

Figure 2.1 Two forms of *T. swinhoei*: chemotype Y [**A**], chemotype W **[B]**.

In the Piel group, the sponge *T. swinhoei* has been chosen as a symbiotic assemblage model to study natural product biosynthetic pathways. This is due to the wide variety of pharmaceutically important secondary metabolites isolated from this sponge as well as the high complexity of the associated bacteria, which might play an important role in metabolite biosynthesis (Matsunaga and Fusetani, 2003; Matsunaga *et al*., 1989; Matsunaga *et al*., 1992). The *T. swinhoei* specimens occur in two chemotypes that grow side-by-side at Hachijo Island, Japan. The chemotype with the bright yellow inner body (Y) (Fig. 2.1A) especially contains onnamides (**5**, **52**-**56**) and theopederins (**57**-**60**) (Matsunaga *et al*., 1992), aurantosides A (**73**), B

(**74**) (Matsunaga *et al*., 1991), nazumamide A (**75**) (Fusetani *et al*., 1991), cyclotheonamides A-B (**76**-**77**) (Fusetani, *et al*., 1990), orbiculamide A (**38**) (Fusetani *et al*., 1991), and pseudotheonamides (**78**, **79**) (Nakao *et al*., 1999). The chemotype with white interior (W) (Fig. 2.1B) contains theonellamides (**30**-**35**) (Matsunaga *et al*., 1989), misakinolide A (also known as bistheonellide A) (**24**) (Kato, *et al*., 1987), and keramamides (**39**-**41**, **44**-**47**) (Itagaki *et al*., 1992) (Fig. 2.2).

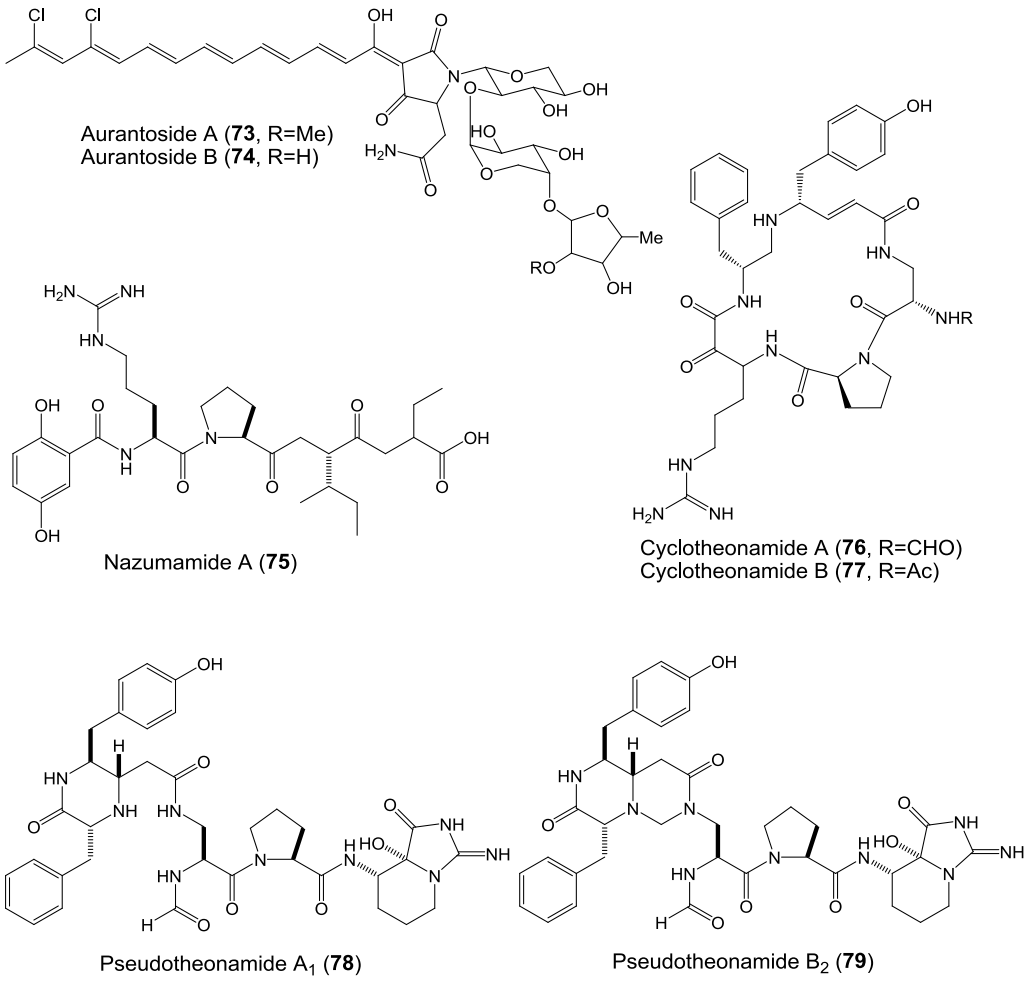


Figure 2.2 Structures of some bioactive secondary metabolites isolated from the chemotype Y of the Japanese *T. swinhoei* , including aurantosides A and B (Matsunaga *et al*., 1991), nazumamide A (Fusetani *et al*., 1991), cyclotheonamides A and B (Fusetani, *et al*., 1990), orbiculamide A (Fusetani *et al*., 1991), and pseudotheonamide B_2 (Nakao *et al*., 1999).

The first goal of this work was to clone at least one new biosynthetic pathway of pharmacological relevance from the Japanese *T. swinhoei* by employing and developing metagenomic strategies. If possible, information on modified peptides should be gained, since sources of these compounds had previously not been examined in sponges.

Although metagenomic methods are able to distinguish between prokaryotic and eukaryotic origins, these approaches do not usually provide specific taxonomic information. Therefore, the second goal of this work was to provide taxonomic information of the uncultured bacterial symbionts harboring the biosynthetic pathway(s) of interest. A special interest was directed to "*Candidatus* Entotheonella spp.", filamentous heterothropic δ-proteobacteria associated with *T. swinhoei* of both chemotypes (Y and W) (J. Piel, S. Matsunaga, unpublished observations). These symbiotic bacteria are considered as a good model to study marine natural product biosynthesis, because they are uncultivable as of yet, easy to isolate by mechanical separation, and loosely related to myxobacteria, an important group of free-living polyketide producers. In addition, these symbionts occur in different sponges and feature large unusual morphology (Schmidt *et al*., 2000; Brück *et al*., 2008). The outcomes of this work were expected not only to further test the symbiont hypothesis, but also to provide specific taxonomic information as well as to facilitate the sustainable supply of rare marine drug candidates.

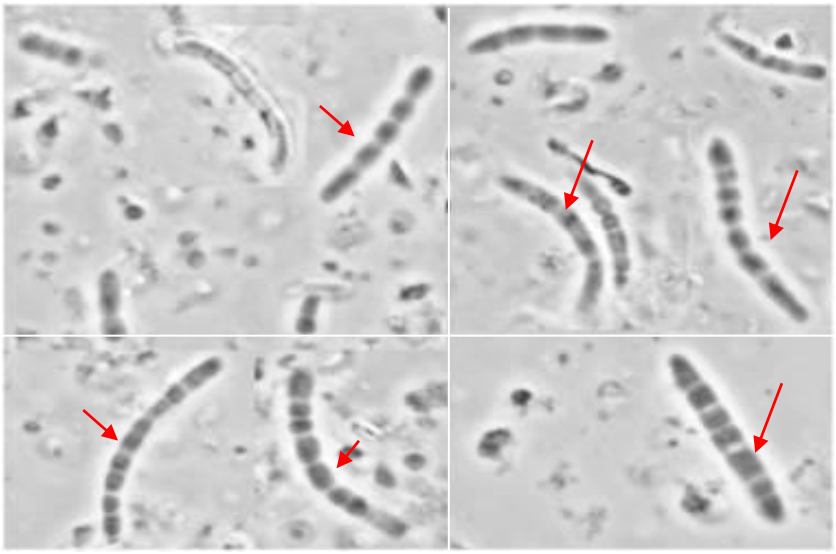


Photo by C. Gurgui

Figure 2.3 The symbiotic "*Candidatus* Entotheonella spp." in the homogenized sponge sample of *T. swinhoei* chemotype W (indicated by red arrows).

Chapter 3

Results and Discussion

Finding a biosynthetic pathway of interest in a sponge system is highly challenging due to the high complexity of the associated symbiotic assemblage. The fact that the majority (>99%) of the symbiotic bacteria present in the sponge are as-yet uncultivable (Amann *et al*., 1995; Hugenholtz *et al*., 1998; Friedrich *et al*., 2001; Webster and Hill, 2001) has hampered attempts in accessing biosynthetic pathways by cultivation-based approaches. Even if cultivation conditions can be met for individual symbionts, the target compound might not be produced due to the lack of required environmental signals (Uria and Piel, 2009). To tackle the challenge, the Piel group has recently developed efficient cultivation-independent approaches to access the biosynthetic pathways of secondary metabolites from a complex symbiotic system. These molecular approaches, called metagenome mining, have enabled to clone the biosynthetic pathways of pharmaceutically valuable secondary metabolites from the complex sponge symbiotic assemblage (Piel *et al*., 2004b; Hrvatin & Piel, 2007; Nguyen *et al*., 2008; Fisch *et al*., 2009).

Because all of these pharmaceutically important compounds mentioned above belong to polyketides and non-ribosomal peptides, the metagenomic approaches developed in this work have focused on the biosynthetic systems that generate these two compound classes, namely modular polyketide synthetases type I (PKS) and nonribosomal peptide synthases (NRPS), respectively. The experimental results discussed in this chapter will encompass (i) molecular detection of the key biosynthetic genes in the symbiotic assemblage, (ii) isolation and analysis of complex PKS/NRPS gene clusters, and (iii) preliminary biosynthesis insights from the genome sequencing of an uncultivated bacterial symbiont.

3.1 Metagenomic Survey of Biosynthesis Genes

To obtain insights into the diversity of biosynthetic genes present in the *T. swinhoei* chemotype W, we conducted metagenomic survey of PKS/NRPS genes by amplifying, cloning, and sequencing of KS and A domain fragments. The genes encoding KS domains were used as the diagnostic target, since their sequences are most conserved sequences among PKS domains (Aparicio *et al*., 2000). In addition,

genes coding for A-domains were chosen as the cloning target for NRPS detection due to the highly conserved amino acid sequences throughout this domain in comparison with other NRPS domains (Turgay *et al*., 1992). Due to the huge diversity of homologous genes from various pathways present in the total DNA of sponges, cultivation-independent strategies were developed, which combined techniques for simplifying the symbiotic complexity and designing effective diagnostic primers. By performing the PCR technology along with gene cloning, sequencing, and bioinformatic analysis, important information on potential PKS and NRPS genes from this chemically rich sponge was obtained, which will be discussed in this sub-section.

3.1.1 PCR cloning of biosynthesis genes from sponge metagenome

PCR amplification of KS domain-containing genes was based on using oligonucleotides designed on two highly conserved motifs in KS domains: DPQQ and HGTGT (Piel, 2002) (methodology section Table 4.3.1). In this work, at first we prepared the metagenomic DNA from the Japanese sponge *T. swinhoei* chemotype W using a CTAB-based DNA isolation (Piel *et al*., 2004b; Hrvatin and Piel, 2009; Gurgui and Piel, 2010). The extracted total DNA was separated by electrophoresis and purified from the agarose gel. Using the purified sponge DNA as the template, we detected the PCR products of the expected size (~700 bp). The PCR products were then purified from the gel and cloned into *E. coli* (Fig. 3.1.1). Some of the clones were sequenced and analysed by the BlastX program (Altschul *et al*., 1990).

The BlastX results showed that all of the five clones (supplementary S3.1.1) exhibited high similarities (88 to 95%) to KS domains of the module SupA identified from both the sponges *Theonella swinhoei* and *Aplysina aerophoba* (Fieseler *et al*., 2007). SupA contains a single complete PKS module (KS-AT-DH-MT-ER-KR-ACP) with additional KS and inactive AT domains in the C-terminus. This domain architecture is typical for animal FASs (Schweizer & Weissman, 2001), in which the presence of KR-DH-ER-MT domains fully reduces the product backbone, resulting in a methyl-branched fatty acid (Fieseler *et al*., 2007). To determine the phylogenetic affiliation of the isolated KS sequences, their translated proteins were aligned with three protein homologs from FAS (SupA, SA1_PKSA, SA1_PKSB) and 19 known KS domains of both *cis*-AT and *trans*-AT PKSs retrieved from the GenBank database. The *cis*-AT PKSs included nostophycin synthase from the

cyanobacterium *Nostoc* sp. Strain 152 (Fewer *et al*., 2011), cryptophycin synthase from the cyanobacterium *Nostoc* sp. (Magarvey *et al*., 2006), and nodularin synthetase from the cyanobacterium *Nodularia spumigen*a (Moffitt & Neilan, 2003). The *trans*-AT PKSs included onnamide A synthetase from *T. swinhoei* (Piel *et al*., 2004), difficidin synthase from *Bacillus amyloliquefaciens* FZB42 (Chen *et al*., 2007), and bacillaene from *Bacillus amyloliquefaciens* FZB42 (Chen *et al*., 2007). The multiple alignments were conducted in BioEdit (Hall 1999) using ClustalW program (Thompson *et al*., 1994). The alignment result was transferred to MEGA4 software (Tamura *et al*., 2007). The phylogenetic analyses were performed using the Neighbor-Joining method (Saitou & Nei, 1987) and evaluated based on Felsenstein`s bootstrap test (Felsenstein, 1985).

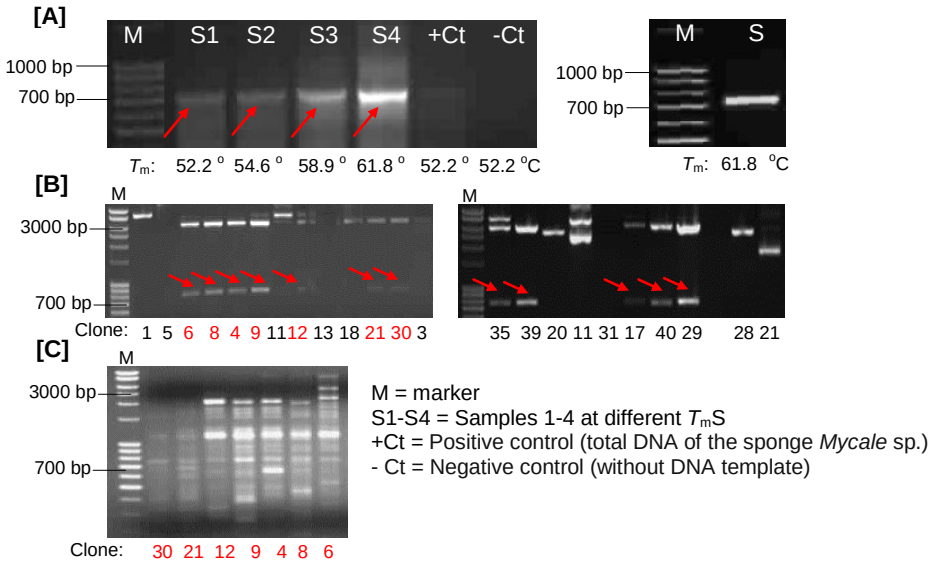


Fig. 3.1.1 Isolation of KS domain-encoding genes from the metagenome of *T. swinhoei* chemotype white (TSW). **[A]** PCR amplification was carried out at four different annealing temperatures (T_ms). **[B]** The ~700 bp PCR product obtained at the optimal T_m (61.8°C) was purified from the agarose gel and cloned into *E. coli* using pBlueScript II SK(-). Some clones were picked up based on blue-white screening and subjected to *Eco*RI-*Hind*III restriction to check the presence of inserts (indicated by red arrows). **[C]** The positive clones were further subjected to *Rsa*I-*Hha*I digestion, and sequenced.

The resulting phylogenetic tree shown in Figure 3.1.2 revealed that the protein sequences of the five isolated KS clones (marked with red color) did not group into the *trans*-AT and *cis*-AT PKS clades, indicating that none of them are responsible for the biosynthesis of complex polyketides. Instead, they fell into the same clade as KS domains of SupA, suggesting their possible roles in the lipid or methyl-branched fatty acid biosynthesis. Similar results were previously reported by Piel, Henstchel and coworkers who detected a large number of clones harbored *sup* (sponge ubiquitous PKS) genes involved in the fatty acid biosynthesis (Fieseler *et al*., 2007).

The extremely low probability in obtaining KS sequences of pharmacologically relevant PKS from the sponge metagenome is due to the huge amount of fatty acids in sponges that can account for an average of 86% of all KS sequences in average based on the study on 20 different sponge species, as reported by Fieseler *et al* (2007).

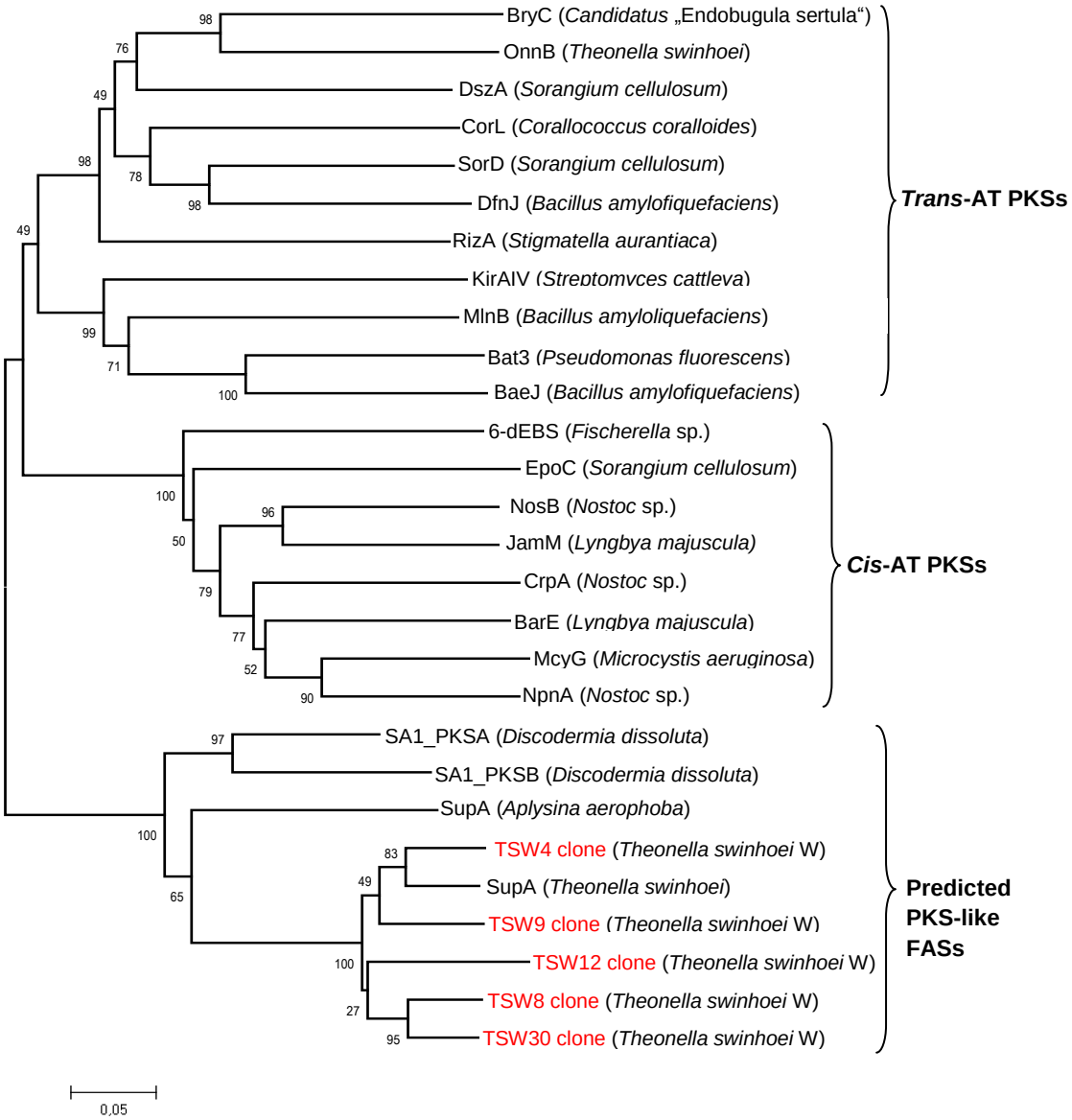


Fig. 3.1.2 Phylogenetic analysis of partial KS amplicons (TSW4, TSW8, TSW9, TSW12, TSW30) cloned from the metagenomic DNA of *T. swinhoei* chemotype W. This phylogenetic analyses were performed in MEGA4 program (Tamura, 2007) using the Neighbor-Joining method (Saitou and Nei, 1987). The bootstrap percentages are shown next to the branches. The boostrap consensus tree was inferred from 500 replicates. McyG of microcystis (YP_001658877) JerC of jerangolid (ABK32289), DszA of disorazoles (AAY32964), CorL of corallopyronin A (ADI59534), SorD of sorangicin (ADN68479), Bat3 of batumin (ADD82941) DszA of disorazoles (AAY32964), StiG of stigmatellin (CAD19091), JamM of jamaicamides (AAS98784), 6-dEBS of 6-deoxyerythronolide-B synthase (ZP_08988403.1), NpnA (AEU11005.1), CrpA (ABM21569.1), BarE (AAN32979.1), NosB (AF204805_2), BryC (ABM63528.1).

3.1.2 PCR cloning of biosynthesis genes from uncultivated bacteria

To tackle the problem, we used an uncultured bacterial cell fraction prepared from *T. swinhoei* chemotype W. In collaboration with Prof. S. Matsunaga (University of Tokyo), Prof. J. Piel performed differential centrifugation to separate bacterial cell types from sponge cells (Fig. 3.1.3). Two types of the bacterial populations were obtained (Piel *et al*., 2004b): (i) a filamentous bacterial fraction and (ii) a unicellular heterothrophic bacterial fraction. Based on microscopic observation, the filamentous bacterial fraction obtained was mainly dominated by “Entotheonella sp.“ (J. Piel, S. Matsunaga, unpublished observation).

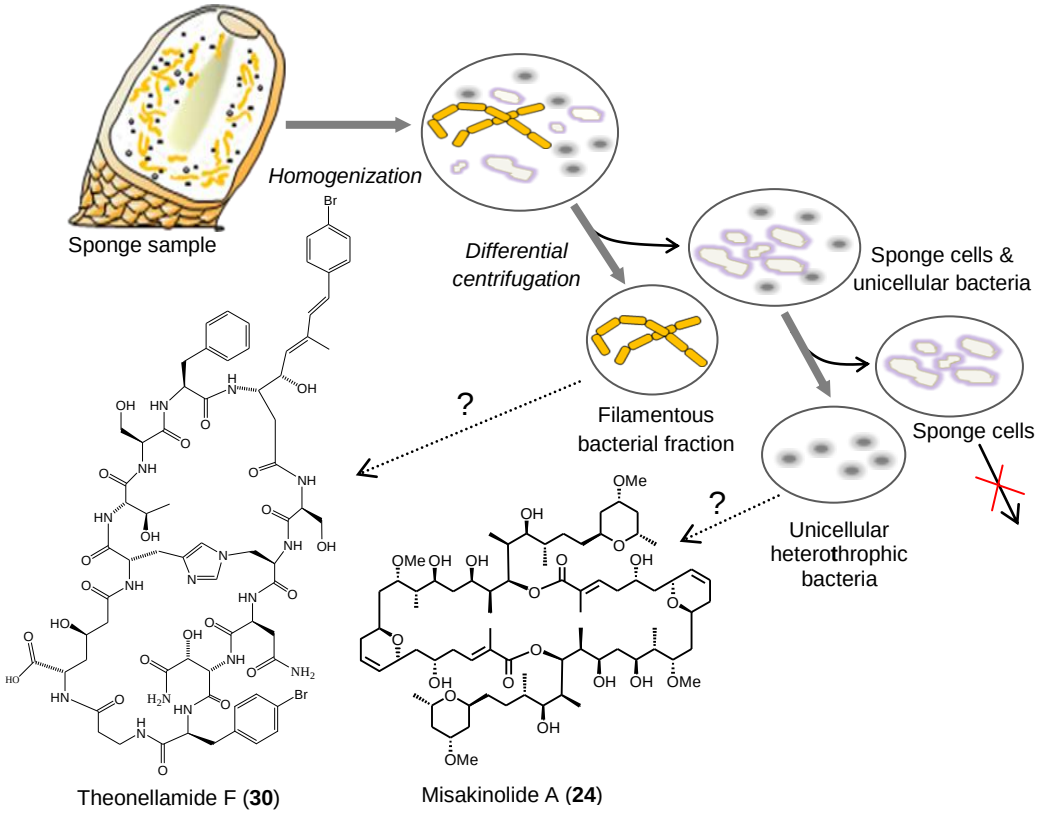


Figure 3.1.3 Compound localization in the Japanese *T. swinhoei* chemotype W. The sponge sample was homogenized in calcium/magnesium-free artificial seawater (CMF-ASW) and the suspension was filtered through a nylon sieve to remove larger particles. The filtrate was subjected to differential centrifugation at increasing speed, thereby resulting in the separation of cell types with different size and densities. Based on earlier studies of Bewley *et al* (1996), it was proposed that theonellamide F and misakinolide A are located in the filamentous bacteria and unicellular heterothropic bacteria, respectively.

This mechanical separation was not only intended to obtain an enriched filamentous fraction, but also to reduce the complexity of the sponge symbiotic assemblage, thereby simplifying the access to potential biosynthetic pathways present in the sponge. Based on the earlier localization studies by Bewley *et al* (1996), we

hypothesized that theonellamide F reported from this sponge (structurally similar to theopalaumide, see Introduction) was located in the filamentous bacteria identified as "*Candidatus* Entotheonella sp.", whereas misakinolide A (structurally similar to swinholide A) was accumulated in unicellular heterotrophic bacterial fraction. "Entotheonella sp." is an especially interesting model to study the biosynthetic capacity of an uncultivated bacterial symbiont. Besides potentially producing the promising antifungal compound theonellamide F, it has large unusual morphology and occurs in different sponges. In addition, it is easy to isolate by mechanical separation, and taxonomically distant from known cultivated bacteria (Schmidt *et al*., 2000; Brück *et al*., 2008).

Because "Entotheonella sp." was proposed as the true producer of theopalauamide based on compound localization studies (Schmidt *et al*., 2000), our initial studies on theonellamide F were focused on gaining access to the biosynthetic pathway of this antifungal cyclic peptide from the filamentous fraction. For this purpose, the biogenesis of some unusual building blocks of theonellamide F was theoretically predicted in order to determine whether or not a PKS system and other enzymatic modifications are involved in the theonellamide biosynthesis. This information was useful not only to confirm if the existing KS-based primers could be used, but also to get insights about which other parts of the pathway could be targeted by PCR.

One of the unusual amino acids forming theonellamide F is the L-p-bromophenylalanyl residue. As shown in Fig. 3.1.4, this brominated residue is presumably formed via L-phenylalanine bromination catalyzed by a halogenase, suggesting that diagnostic PCR-primers could be designed based on the halogenase enzyme, especially for a phenylalanine or tyrosine halogenase. Another unusual building block is (2*S*)-2-amino-4-hydroxyadipic acid that could arise from the hydroxylation of L-α-aminoadipic acid, which is an intermediate in the lysine biosynthetic pathway (Velasco *et al*., 2002). The NRPS system that involves loading L-α-aminoadipic acid was reported for the biosynthetic pathway of the tripeptide d-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV), a common intermediate in β-lactam antibiotic families, the penicillins and cephalosporins (Martín *et al*., 2000). This suggests the possibility to design screening primers based on the adenylation domain specific for aminoadipic acid.

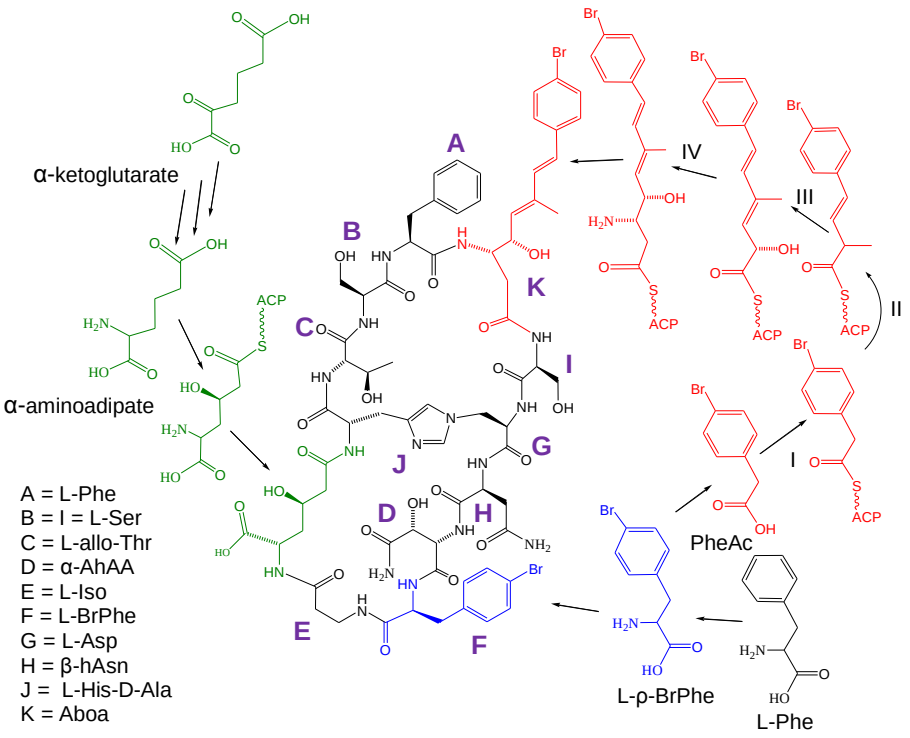


Figure 3.1.4 The proposed biogenesis of some building blocks in theonellamide F. The incorporated brominated phenylalanine (BrPhe) is presumably from L-Phe via halogenation (blue color). The proposed Aboa biosynthesis (red color) starts from the L-Phe-derived PheAc, which is loaded by an NRPS (I) and processed through a PKS assembly line involving *a* double bond shift by β,γ-dehydration and α-methylation (II), α-hydroxylation (III), and aminotransferase (IV). The α-AhAA probably arises from α-aminoadipic acid, which is derived from α-ketoglutarate in lysine biosynthesis (green color). Abbreviations: α-AhAA is (2*S*,3*R*)-3-hydroxyasparagine, (*2S,4R*)-2-amino-4-hydroxyadipic acid, *L*-His-D-Ala is *Z*-*L*-histidino-D-alanine, L-p-bromophenylalanine, and Aboa is (3*S*,4*S*,5*E*,7*E*)-3-amino-4-hydroxy-6-methyl-8-(*p*-bromophenyl)-5,7-octadienoic acid.

Another unusual amino acid is (*3S,4S,5E,7E*)-3-amino-4-hydroxy-6-methyl-8-(ρ-bromophenyl)-5,7-octadecanoic acid (Aboa), which is a close structural analog of 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid (Adda) present in microcystins (Namikoshi *et al*., 1992). The Adda biosynthetic pathway involves loading phenylacetate (PheAc) and a subsequent extension reaction by a series of PKS modules (Tillet *et al*., 2000). For Aboa, the biosynthetic pathway presumably arises from L-phenylalanine that is converted into PheAc through transamination and oxidative carboxylation. Subsequent decarboxylative condensations of PheAc with malonyl-CoA-derived extender units catalyzed by a series of PKS modules involving a double bond shift, α-methylation and α-hydroxylation would lead to the formation of the Aboa residue, as described in Fig. 3.1.4. Therefore, all of the building blocks, except Aboa, were assumed to be enzymatically incorporated into theonellamide F by NRPS modules. Because the biosynthesis of theonellamide F most likely involves a PKS system for Aboa residue formation, we initially targeted the PKS genes using the existing KS-based primers.

By using the KS-based primers with the filamentous bacterial fraction as the PCR template, we attempted to detect of PKS type I genes by performing a whole cell PCR technique. A gradient PCR was done at different T_ms (52, 55.5, 60 ^oC) to know the optimal annealing temperature (Fig 3.1.5). The target band of around 700 bp was observed at the T_m of 60 ^oC, which was used further as the optimal T_m to get more amplicons. The resulting PCR product was subsequently purified from the gel and cloned into the host, *E. coli* XL1 Blue MRF. White clones that appeared on the solid media (based on blue-white screening) were picked, cultivated, and checked for the presence of correct-sized inserts. The positive clones were then sequenced and analyzed to predict the functional properties.

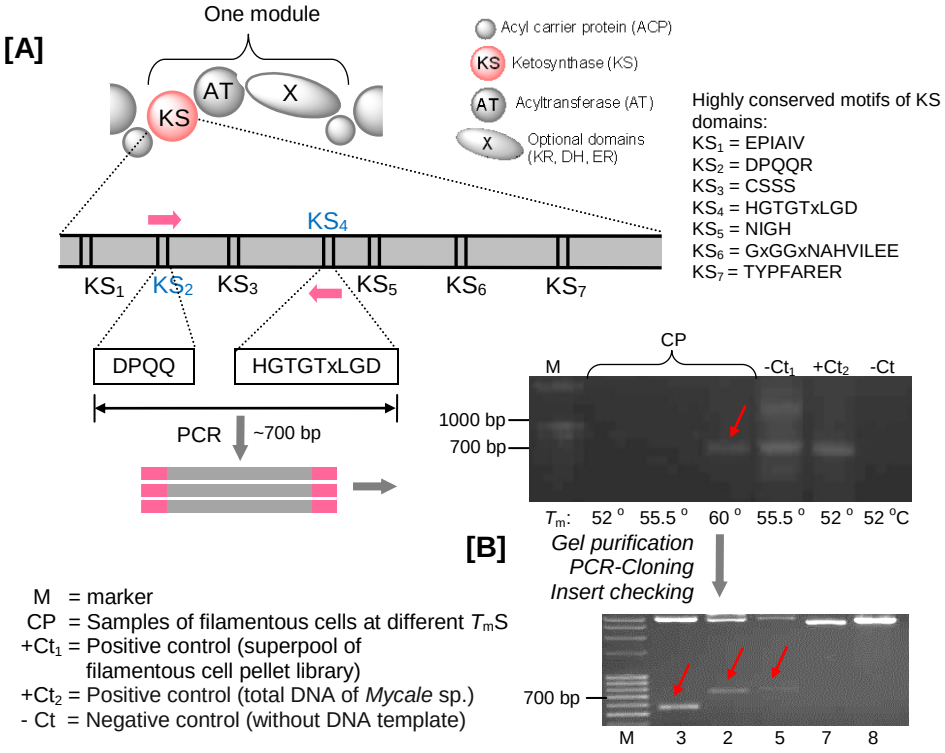


Figure 3.1.5 Cloning of KS sequences from the uncultivated filamentous bacterial fraction of the Japanese marine sponge, *T. swinhoei* chemotype white. **[A]** PCR detection of 700-bp DNA fragments at three different T_ms (52^o, 55.5^o, 60^{o}C). **[B]** The target PCR product amplified at optimal T_m (60 ^oC) was cloned into *E. coli* using pBlueScript II SK(-). Some resulting white clones were subjected to digestion with *BamH*I, *EcoR*V and *Hind*III. The clone harboring inserts (shown with red arrows) were sequenced.

Sequence analysis indicated that all of the five sequenced clones belong to KS-domains of *trans*-AT PKSs. To get insights into the diversity of PKS genes, more KS genes were isolated from the filamentous fraction, leading to the isolation of eight additional PKS gene fragments. Three KS sequences shared high similarity (70-78%) with those known from *cis*-AT PKS modules, McyG in the microcystis (Tillet *et*

al., 2004), JerC in jerangolid (Julien *et al*., 2006), JamM in jamaicamide (Edwards *et al*., 2004) biosynthesis (supplementary S3.1.2). Other KS sequences exhibited high similarity (67-78%) with those known from *trans*-AT PKS modules, DszA in the disorazoles (Carvalho *et al*., 2005), CorL in the corallopyronin A (Erol *et al*., 2010), SorD sorangicin (Irschik *et al*., 2010), Bat3 in the batumin (Mattheus *et al*., 2010), and StiG in the stigmatellin (Gaitatzis *et al*., 2002) biosynthesis (supplementary S3.1.3) (Table 3.1).

As described previously, our work on PCR detection of KS genes in the entire metagenome of *T. swinhoei* led to the isolation of KS sequences involved in the lipid or methyl-branched fatty acid biosynthesis (see above). None of these metagenome-derived KS sequences is involved in the biosynthesis of complex polyketides. However, using the cell fraction separated from the sponge *T. swinhoei*, we isolated ten KS genes predicted to be involved in the biosynthesis of complex polyketides, suggesting that we have mechanically enriched PKS genes in a fraction predominantly consisting of an "Entotheonella sp." symbiont.

Table 3.1 Hypothetical identities of the gene sequences cloned from the filamentous bacterial pellet using degenerate KS primers.

Clone	AA	Protein homolog (BLAST)			I/S (%)	Group
		Homolog	Organism	Accession no.		
EKS14	762	McyG (microcystis)	*Microcystis aeruginosa*	YP_001658877	55/70	*Cis*-AT PKS
EKS30	596	JerC (jerangolid)	*Sorangium cellulosum*	ABK32289	75/85	*Cis*-AT PKS
EKS27	762	DszA (disorazoles)	*Sorangium cellulosum*	AAY32964	53/67	*Trans*-AT PKS
EKS37	387	CorL (corallopyronin A)	*Corallococcus coralloides*	ADI59534	63/78	*Trans*-AT PKS
EKS48	746	SorD (sorangicin)	*Sorangium cellulosum*	ADN68479	59/73	*Trans*-AT PKS
EKS50	739	Bat3 (batumin)	*Pseudomonas fluorescens*	ADD82941	61/74	*Trans*-AT PKS
EKS64	745	DszA (disorazoles)	*Sorangium cellulosum*	AAY32964	58/72	*Trans*-AT PKS
EP2	711	StiG (stigmatellin)	*Stigmatella aurantiaca*	CAD19091	66/79	*Trans*-AT PKS
EP3	699	phenolpthiocerol	*Sorangium cellulosum*	ZP_08782289	52/69	*Trans*-AT PKS
EP5	719	JamM (jamaicamides)	*Lyngbya majuscula*	AAS98784	60/76	*Cis*-AT PKS

Note: AA, amino acids; I/S, identities/similarities

To obtain insight into the NRPS diversity within the filamentous symbiont preparation, NRPS genes were detected by cloning of the A domain sequences responsible for selecting amino acids as building blocks in nonribosomal peptide biosynthesis. To clone the A domain sequences, three degenerate primer pairs, (designated as JP, Öz, and AG) designed on the basis of adenylation catalytic motifs (A2-A8, A2-A9 & A3-A7) were initially used (methodology section Table

4.3.2). The accuracy of these primer pairs were verified with the genomic DNA of *Bacillus amyloliquefaciens* FZB42 as the PCR template. The reason for choosing this bacterial strain as the positive control was due to the presence of the gene clusters in its genome for nonribosomal peptides, such as surfactin, bacillomycin D, fengycin, siderophore bacillibactin, and bacilysin (Koumoutsi *et al*., 2004; Chen *et al*., 2009). All of the three tested primer pairs gave PCR products with the desired sizes, indicating their suitability for being used in targeting the A domain sequences (Fig. 3.1.6).

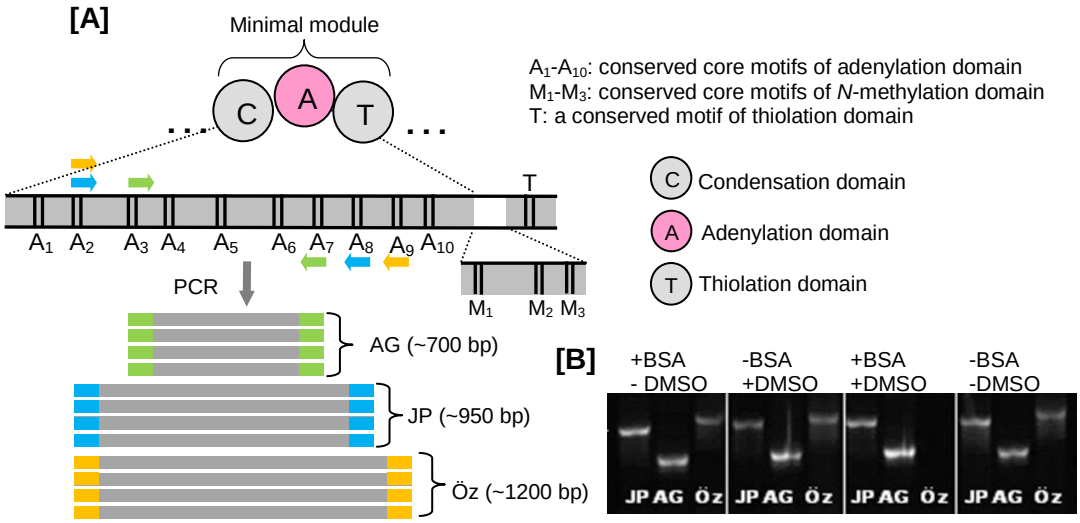


Figure 3.1.6 Detection of bacterial NRPS genes based on A domain sequences. **[A]** Amplification strategy with three different degenerate primers. **[B]** Verification of the primers at various PCR reaction conditions with the genomic DNA of *Bacillus amyloliquefaciens* B42 as the PCR template. A_1-A_{10} = conserved catalytic motifs of A-domain, M1-M3 = conserved catalytic motifs of methylation domain. AG, JP, Öz are three different degenerate primers targeting A-domain sequences (See Methodology section). PCR reaction conditions: +BSA (with bovine serum albumin) and +DMSO (with dimethylsulfoxide).

The three primer pairs (JP, Öz, and AG) were subsequently used to amplify A domain sequences from the filamentous bacterial fraction by whole-cell PCR. The cell suspension was initially rinsed with ddH_2O three times and then used as PCR template. The whole-cell PCR condition was set at three different T_ms (55, 58, 61 ^oC). The results showed that only the primer pair designed based on the A3-A7 motifs gave a target band of the expected size at the T_m of 61 ^oC (Fig. 3.1.7).

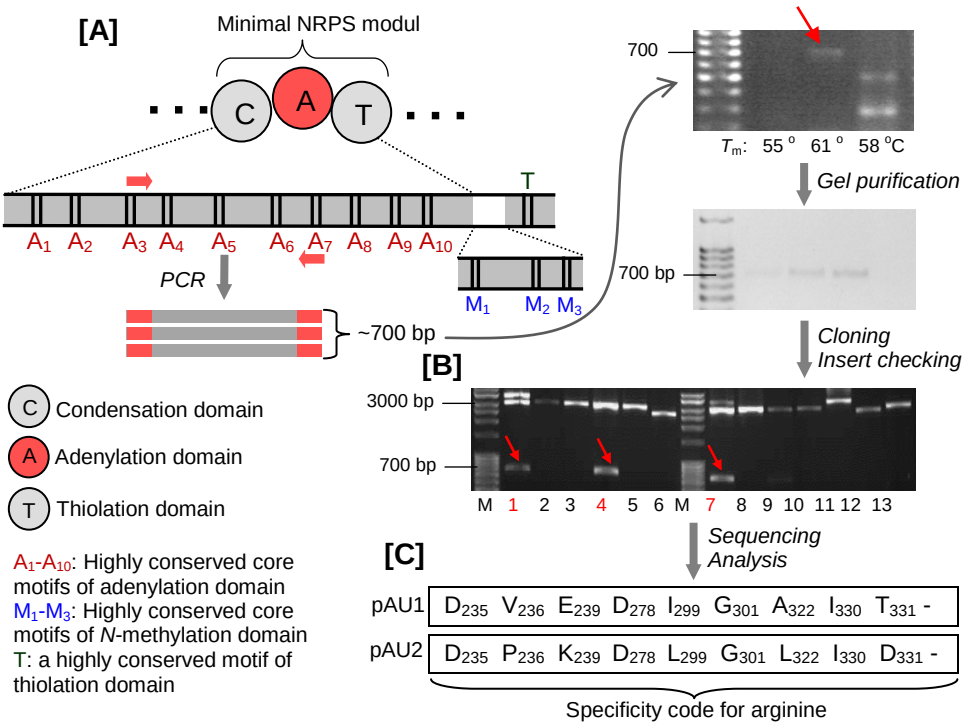
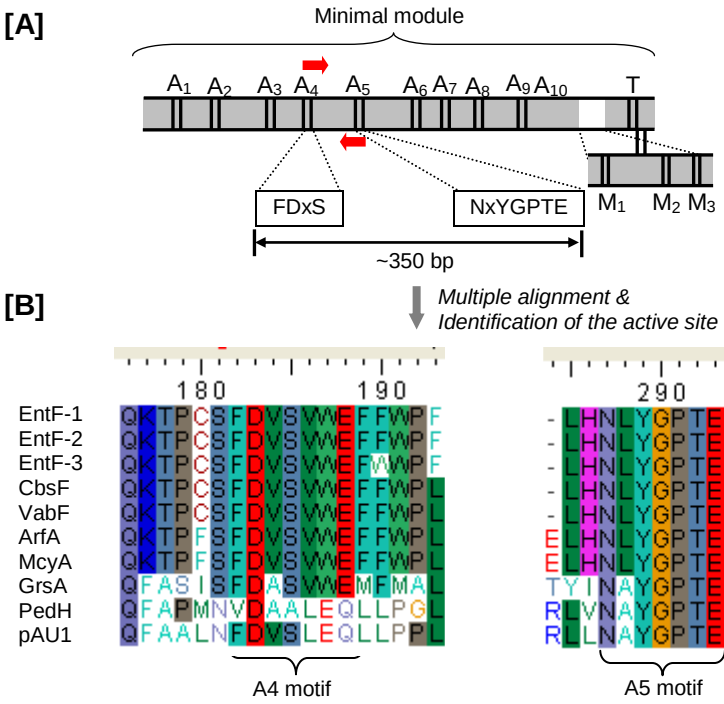


Figure 3.1.7 Molecular detection of bacterial NRPS genes based on A domain sequences in the filamentous bacterial fraction of the Japanese sponge *T. swinhoei*. **[A]** The target amplicon (~700 bp), as indicated by a red arrow, was purified, subjected to nested PCR to get more amplicon, cloned into *E. coli* using pBlueScript II (SK-) vector. **[B]** Some resulting clones were checked for the presence of inserts by *Eco*RI-*Hin*dIII digestion. Three clones containing correct-sized inserts (as indicated by red arrows) were subsequently sequenced. **[C]** Sequence analysis indicated that the isolated genes encoded adenylation domains specific for arginine due to the presence of arginine-conferring specificity codes. The position of the specificity code (ten decisive residues) was based on the residue numbering in PheA of GrsA. The 10th residue position marked with "-"could not be identified, because it was located outside the A3-A7 regions. M = 1-kb DNA marker, S = sample.

Since the target 700 bp fragment was appeared as a single band on the gel, the PCR products were directly purified to remove primers and primer dimmers formed during the PCR reaction. The purified PCR product was subsequently cloned and sequenced. The resulting clone sequences were analyzed using the online software VecScreen (http://www.ncbi.nlm.nih.gov/VecScreen) to remove the vector backbone. The isolated sequences were then compared to each other at the nucleotide level, followed by predicting their putative function using the online software NRPS predictor1 (Rausch *et al*., 2005). We found that three of the four sequenced clones were identical. The two unique clones, designated as pAU1 and pAU2, were each predicted to encode an A domain specific for arginine (ArgA) (Fig. 3.1.7) (supplementary S3.1.4). Because arginine (Arg) or its derivatives are not building blocks forming theonellamide F, the isolated ArgA sequences might not be involved in the biosynthesis of theonellamide F.

To target the biosynthetic pathway of theonellamide F in this work, we focused on cloning of the A-domain sequences specific for an amino residue that is encoded more than once in theonellamide F. Because there are three likely L-serine residues in the theonellamide F structure, the degenerate primers for A domain amplification were designed based on the active site of the serine-activating adenylation domain (SerA). To design the primers, some sequences of A-domains with serine specificity from different biosynthetic gene clusters were initially retrieved from GenBank. These sequences were aligned together in BioEdit (Hall 1999) using Clustal W program (Thompson *et al*., 2007) and compared to those with the specificity for phenylalanine and arginine. The reason of choosing PheA for sequence comparison is that the 3-D structure of gramicidin synthetase A1 (PheA-GrsA1) in a small complex with AMP and phenylalanine has been solved, which provided a basis for the identification of 10 core motifs (A1-A10) generally involved in the coordination of ATP and the substrate (Conti, *et al*., 1997; Schwarzer & Marahiel, 2001; Schwarzer, *et al*., 2003). The reason of choosing the ArgA sequence in the alignment is to avoid selection of amino acid residues conferring arginine specificity. The unique motifs located between A4-A5 on the serine-specific adenylation sequences were used as the basis for designing the degenerate primers (Fig. 3.1.8).

The core motifs A4-A5 are flanked by the A3-A6 motifs that contain all residues involved in forming the substrate-binding pocket, thereby determining the A-domain specificity (Cosmina *et al*., 1993) (Fig. 3.1.8). Subsequently, a similar strategy was used to design primers targeting sequences encoding asparagine-activating domains (AsnA) and tryptophan halogenase (Trp-Hal). AsnA was chosen as target due to the presence of L-Asn in theonellamide F structure, which is presumably activated by AsnA, while Trp-Hal was proposed to be involved in brominating L-Trp of theonellamide F. When the SerA-based primers were tested at different T_ms, one of them, designated as A4F1-A5R1, gave a correct-sized DNA fragment at a T_m of 58 ^{o}C (Fig. 3.1.8).



Protein	Spe	Serine A4 motif (5`-3`)								Serine A4 motif (5`-3`)														
		F	D	V	S	V	W	E	F	H	N	L	Y	G	P	T	E	A	A	V	D	V	S/T	W
EntF	Ser	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CbsF	Ser	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VabF	Ser	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ArfA	Ser	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
McyA	Phe	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I	-	-	-	Y
GrsA	Arg	-	-	A	-	-	-	-	-	I	-	A	-	-	-	-	-	T	T	I	C	A	-	T
PedH	Arg	V	-	A	A	L	E	Q	L	V	-	A	-	-	-	-	-	T	T	I	T	S	-	V
pAU1	Arg	-	-	-	-	L	E	Q	L	L	-	A	-	-	-	-	-	-	-	-	T	A	-	V

[C] *Amino acid to nucleotide conversion*

Primer	Orientation	A4 (5`-3`)	A5 (5`-3`)
A4F1	Forward	TTY GAY GTN AGY GTN TGG GAR TT	
A4F1	Forward	TTY GAY GTN AGY GTN TGG GAR TT	
A4F2	Forward	TTY GAY GTN TCN GTN TGG GAR TT	
A5R1	Reverse		TCN GTN GGN CCR TAN RRT TRT G-
A5R2	Reverse		YTV YTN STN ACR TCN CAN GCN GC

Figure 3.1.8 Strategy for designing degenerate primers specifically to target the A domain for L-serine developed in this work. **[A]** The primers were designed based on the serine active site located in the A-domain core motifs, A4 and A5. **[B]** Several A domain sequences (320 bp) specific for serine, phenylalanine, and arginine were retrieved from GenBank and aligned together with the A-domain for arginine from the filamentous fraction of *T. swinhoei*. **[C]** The catalytic core motifs located in A4 and A5 that are unique for serine specificity were used as the basis for primer design. EntF-1, -2, -3 = Enterobactin synthase component F from *E. coli*, *Shigella flexneri,* and *Enterobacte*r sp., CbsF = chrysobactin synthetase from *Chromobacterium violaceum*, VabF = vanchrobactin sythetase from *Listonella anguillarum*, ArfA = arthrofactin synthetase, McyA = microcystin from *Microcystis aeruginosa*, GrsA = gramacydin S synthetase from *Brevibacillus brevis*, PedH = pederin synthase *symbiont of Paederus fuscipes*, pAU1 = arginine-specific A-domain sequence from the filamentous bacterial fraction of *T. swinhoei*.

This fragment was purified, cloned into *E. coli* XL-blue, and then screened based on ampicillin resistance and blue-white screening. White colonies were picked and cultivated overnight. The recombinant plasmids were checked for the presence of approximately 320-bp inserts by enzymatic restriction. Recombinant plasmids with correct inserts were subsequently subjected to *Hha*I digestion for restriction pattern comparison. All of the selected clones showed similar digestion patterns, and therefore three of them were sequenced (Fig. 3.1.9).

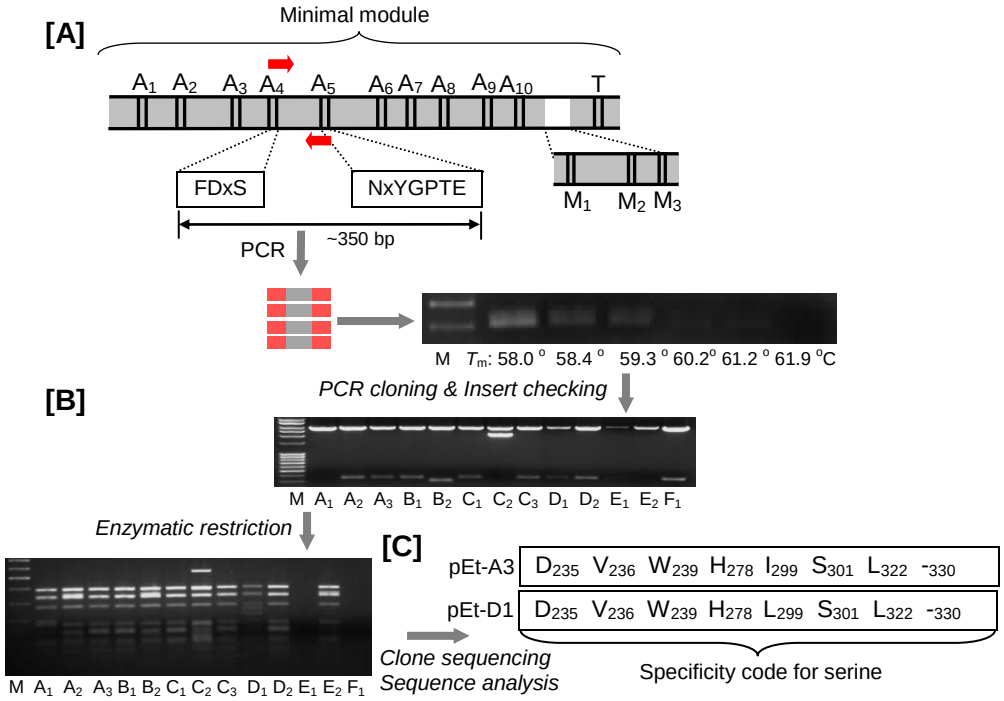


Figure 3.1.9 Cloning of the A domain-encoding fragments specific for L-serine. **[A]** At first, degenerate primers were designed based on the A-domain sequences for serine retrieved from GenBank. The designed primers were applied at different T_ms using the filamentous bacterial fraction as template. **[B]** The PCR products were cloned into *E. coli.* Some positive clones were subjected to HhaI digestion and chosen for sequencing. **[C]** Sequence analysis indicated that two clones harbored different A-domain genes that were predicted to be specific for serine (SerA genes).

The obtained DNA sequences were aligned with functionally verified sequences in public databases using BlastX to predict their putative function. The sequences exhibited highly similarity with L-serine-specific A domains, suggesting that they code for L-serine (supplementary S3.1.5). This was confirmed by using the online software NRPS-predictor1 that showed specificity codes for serine (Rausch *et al.*, 2005). Obtaining two different A domain sequences specific for L-serine (SerA) demonstrated the success in developing an approach to target the NRPS A domain with the substrate specificity of interest. The two isolated SerA homologs might be

responsible for the incorporation of two L-serines present in theonellamide F or alternatively, involved in the biosynthesis of other serine-containing peptides, of which several have been reported from other *T. swinhoei* chemotypes (Fig. 3.1.10).

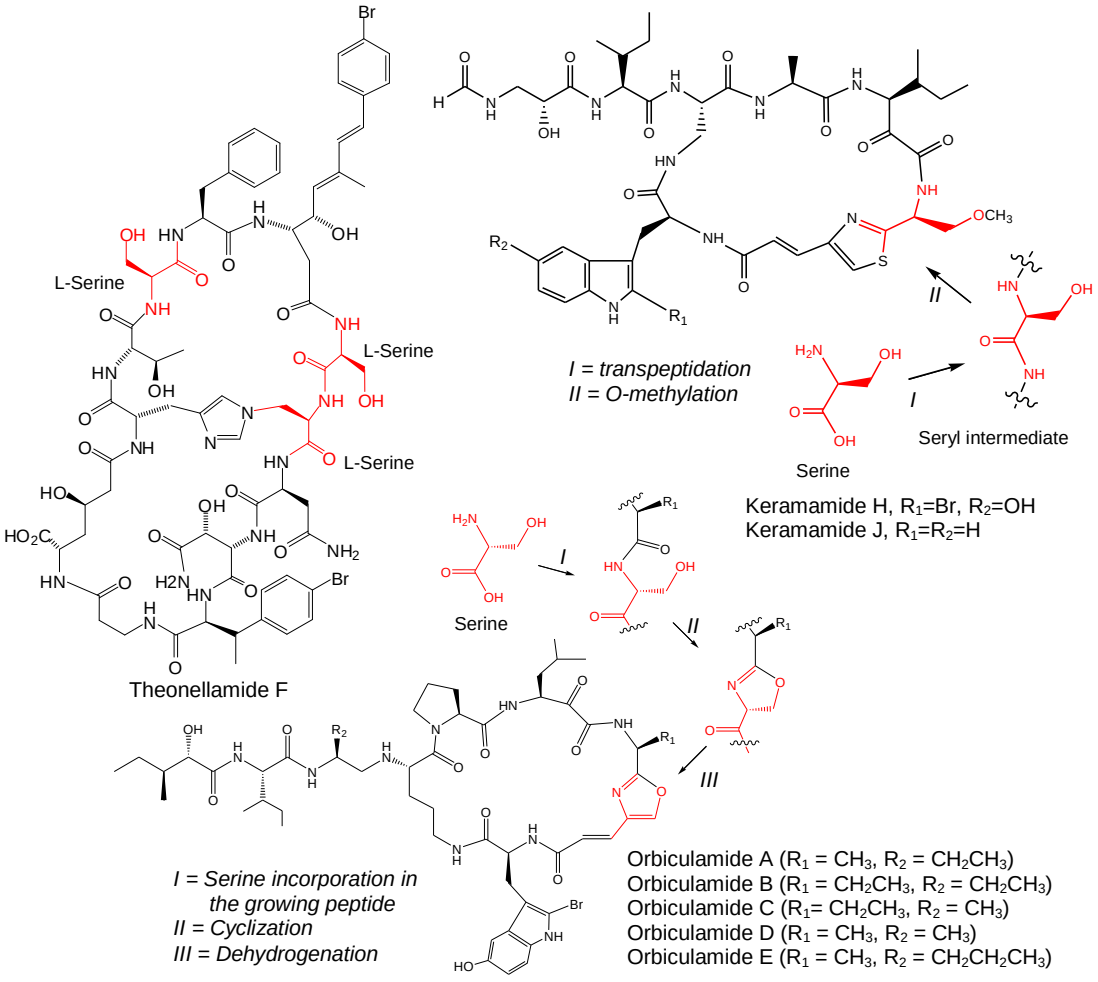


Figure 3.1.10 The predicted functional role of the isolated SerA homologs. They are predicted to be involved in the incorporation of serine in the theonellamide F assembly line, in the formation of O-methylated serine in keramamide biosynthesis, in oxazole formation in orbiculamide biosynthesis, or in the biosynthesis of as-yet undescribed compounds.

For example, in orbiculamide biosynthesis, a serine is expected to be converted into the oxazoline ring through cyclodehydration and subsequently oxidized into the oxazole by an FMN-dependent dehydrogenase, as reported by Li *et al*. (1996) for the oxazole-containing antiobiotic, Microcin B17 synthase. In this case, serine acts as the precursor for the oxazole ring. Another possible role of the isolated *serA* homologue is incorporating serine in keramamide biosynthesis, where such a residue is likely further modified by *O*-methylation, thereby resulting in the methylated serine residue found in this compound class (Fig 3.1.10).

3.2 Metagenomics Discovery of a Polyketide Biosynthetic Pathway

The Japanese sponge *T. swinhoei* used in these studies harbors high density of bacterial symbiotic cells (Fusetani & Matsunaga, 1993). The diversity of the bacteria associated with this sponge derived from at least seven different bacterial divisions, including *Acidobacteria*, *Chloroflexi*, *Actinobacteria*, *Alphaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria*, *Cyanobacteria*, and *Nitrospira* (Hentschel *et al*., 2002). Therefore this sponge can be viewed as highly concentrated reservoirs of so far uncultured marine bacteria, which is termed `a high bacterial abundance sponge` or `bacteriosponge` (Hentschel *et al*., 2002; Hentschel *et al*., 2006). The term `high bacterial abundance sponges` refers to the bacterial symbiotic community in the density range of 6.4×10^8 to 1.5×10^9 bacterial cells per ml of sponge extract, which exceed bacterial isolates from seawater by two to four orders of magnitude (Friedrich *et al*., 2001; Scheuermayer *et al*., 2006) and may consist of hundreds to thousands of microbial species (Taylor *et al*., 2007). Due to the high microbial complexity of this sponge, we expected that the symbiotic assemblage contains numerous homologous genes from diverse pathways (Piel *et al*., 2004a; Piel, 2010). To capture the majority of bacterial pathways present in the sponge, a metagenomic library needs to be generated as large as possible in order to make sure that the biosynthetic pathway of interest is entirely covered in the library. In our work, this involves preparation of metagenome from either the whole sponge or its associated bacteria, cloning of the isolated HMW DNA, amplification of individual clones in the resulting libraries, and subsequent efficient screening strategies for such huge libraries (Fig. 3.2.1).

3.2.1 Metagenomic library construction and screening

As shown previously, mechanical separation not only reduced the microbial complexity, but also enriched the desired biosynthetic genes. In this work, a small library was constructed from the filamentous symbiont fraction of *T. swinhoei*. Our rationale to use the enriched bacterial fraction for library generation was that the library size required to cover a biosynthetic pathway of interest would have to be relatively small, thereby simplifying the screening steps (Fig. 3.2.1). Since the existing filamentous cell pellet was available in a very limited amount, we optimized first the experimental steps for library construction by using the cultivated cells of an engineered strain of *Acinetobacter baylyi* ADP1 harboring a 77-kb gene cluster coding for the antitumor polyketide, pederin. The techniques tested with this small

amount of cultured cells were thus validated and proceeded to generate a library from the filamentous fraction.

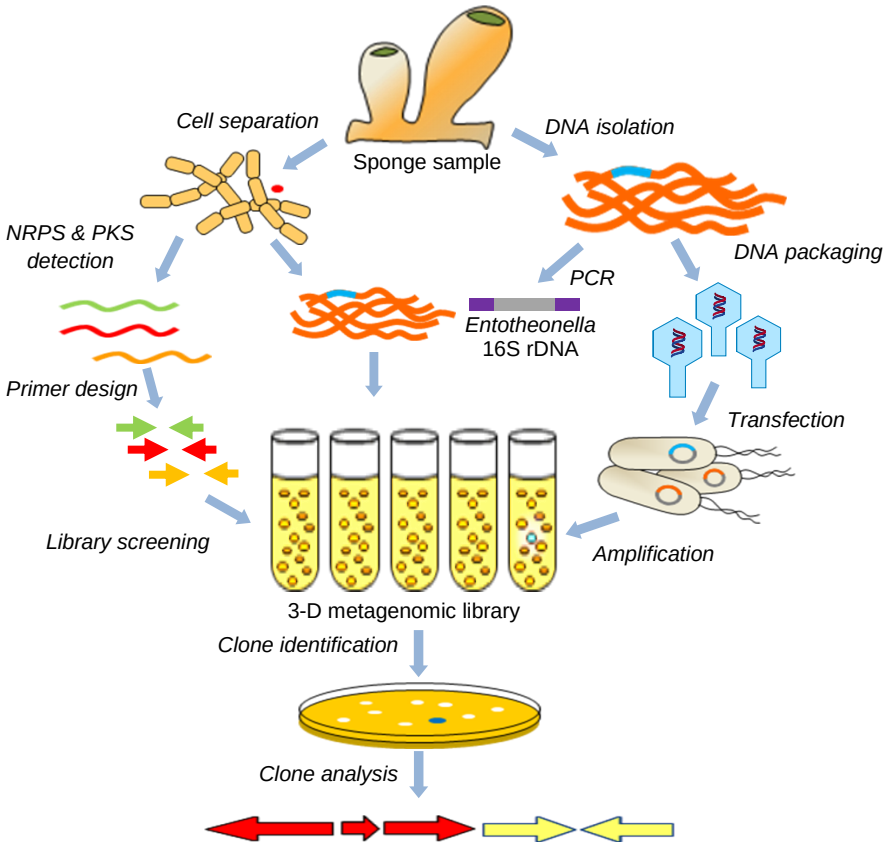


Fig. 3.2.1 General strategies for the construction and screening of a metagenomic library from the Japanese marine sponge *Theonella swinhoei* chemotype W. HMW DNA is initially prepared from either the whole sponge or its associated bacteria and be separated by electrophoresis on a low-melting-point (LMP) agarose gel. DNA fragments with the desired size (~40 Kb) are purified from the gel by Gelase or filtration column, repaired at both termini, and ligated with dephosphorylated fosmids. The ligation products are packed in bacteriophage particles and used to transfect *E. coli.* The resulting library can then be amplified in semi-solid media and screened by clone dilution and whole cell PCR. Before the metagenomic library was constructed, the presence of *Candidatus* `Entotheonella` sp. and SerA-encoding gene were checked first in the isolated total DNA.

We attempted to construct a fosmid library from a very limited amount of the filamentous bacterial fraction. The total DNA was initially isolated from the fraction by a simple heating method (Syn & Swarup, 2000) followed by cloning the DNA into *E. coli* using the fosmid vector pCC1FOS^{TM} (Epicentre), resulting in a ~2000-clone fosmid library distributed into 18 pools in semi-liquid medium (Fig. 3.2.2).

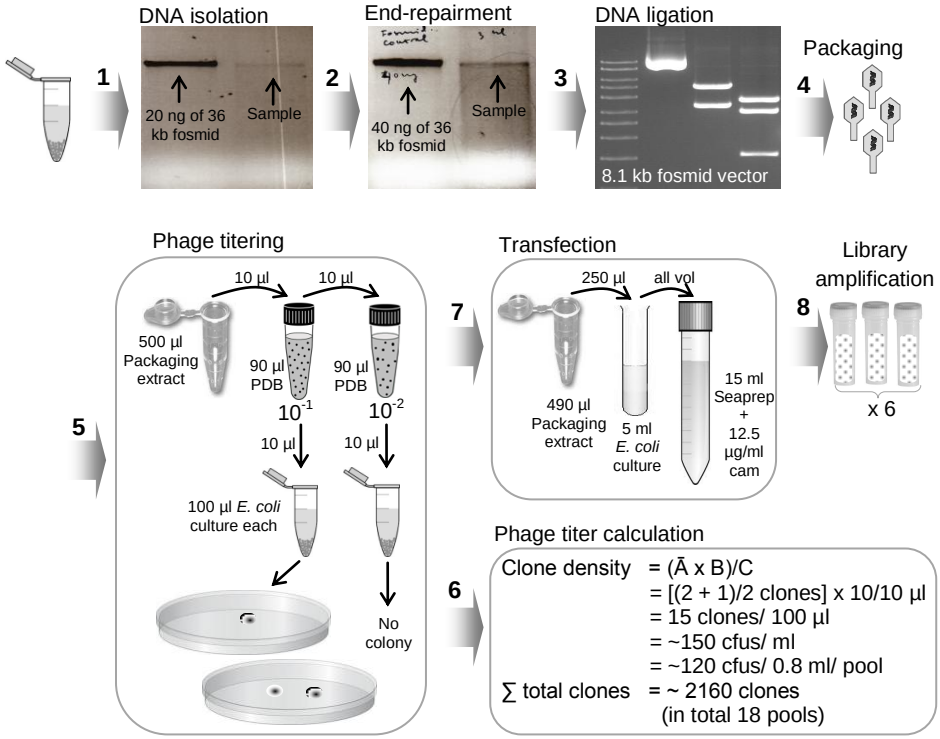


Fig. 3.2.2 Generation of a fosmid library from a filamentous cell pellet prepared from *T. swinhoei* chemotype W. The cell suspension stored in ethanol was rinsed three times with ddH_2O. Then HMW DNA was initially isolated from the rinsed cell pellet (step 1), and subsequently end-repaired (step 2). The end-repaired DNA fragments (~80 ng) were ligated with a fosmid vector (step 3) and packaged into the lambda phage particles (step 4). The titer of the phage particles was determined (step 5 and 6). The phage particles were transfected into *E. coli* cells (step 7), and the recombinant clones were grown in semi-liquid media (step 8) at a density of ~120 cfus per ml. Abbreviations and symbols: PDB, phosphate dilution buffer; vol, volume; HMW, high-molecular-weight; ∑, amount/number; Ā, average number of colonies; B, dilution factor; C, volume taken from the dilution.

We screened the library using primers designed based on three KS-encoding sequences (EP2, EP3, EP5) cloned from the filamentous cell pellet (see Table 3.1). PKS screening of all 18 pools in the library by whole cell PCR (Hrvatin & Piel, 2007) indicated no target PCR product observed on the gel. Subsequently, we tried to optimize the PCR condition to know whether a target PCR product could be observed. The PCR reaction was set up at different T_ms using a combined aliquote from all 18 pools as the PCR template. However, no target DNA PCR product was detected on the gel (Fig. 3.2.3A). Subsequent pool screening with one of the KS primer pairs also did not give positive DNA fragment (Fig. 3.2.3A). Based on these negative results, we assumed that PKS gene cluster regions might not be contained within the library. There are some possible reasons for the low probability or lack of a target gene cluster being present within the library. First, the simple heating method might not be sufficient to lyse the cell membrane of bacteria harboring target

PKS genes, and as a consequence its genome was probably not well represented in the generated library. However, PCR detection of PKS sequences in both cell fraction and DNA isolated from the fraction gave positive results (Fig. 3.2.3B), suggesting the lysis of the cells harboring the PKS genes did occur. Second, the cell fraction used in this work might contain several types of bacteria including filamentous multicellular bacteria. If a sample contains many different bacterial species, then constructing libraries from it would be technically challenging (Riesenfeld *et al*., 2004). Since we do not have 16S rRNA analysis data for this cell fraction as an indicator of bacterial diversity, it is difficult to predict the fosmid library size required to cover the genomes of all different bacterial species present in the fraction. As an example, if there are 9 different bacterial species in the fraction, then the number of clones required to ensure a 99% probability of their genomes (the predicted average size = 4.7 Mb) being contained in a fosmid library of 40 kb inserts is ln (1-0.99/ ln (1-[4 x 10^4 bases/ 4.7 x 10^6 bases]) x 9, which is equal to 4149 clones (Hohn, 1979; Epicentre 2012). The required fosmid library size would be much higher if the fraction sample is contaminated with eukaryotic cells. Third, the target PKS genes might be contained in a very rare member of bacterial species, and not in the dominant filamentous bacteria. If this is the case, then it would be difficult to obtain substantial representation of a rare member in the generated fosmid library.

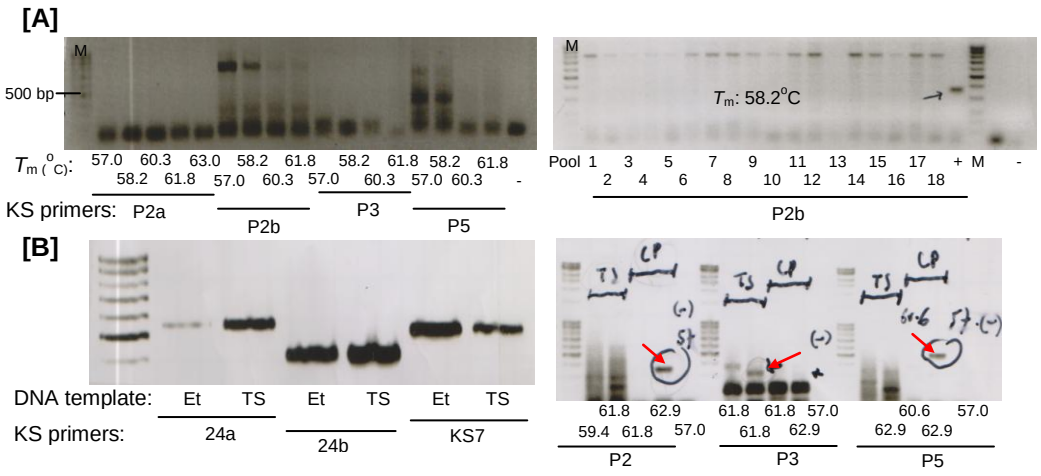


Fig. 3.2.3 PKS screening of the fosmid library generated from the filamentous fraction of the sponge *T. swinhoei* chemotype W. **[A]** Detection of KS-encoding sequences in the superpool (the combined pools) of the library and the 18 pools. **[B]** Detection of KS-encoding sequences in the DNA of *T. swinhoei* chemotype W (TS), DNA of a filamentous cell pellet (Et), and the filamentous cell fraction (CP). The PCR products as indicated by red arrows were sequenced.

Our subsequent screening of the resulting fosmid library was focused on isolating NRPS genes using the specific primers designed based on two isolated SerA domains. This screening led to the identification of 9 positive clones. Restriction analysis by *Pst*I suggested that the nine clones were identical (Fig. 3.2.4). Subsequently, we analyzed one of the clones (pET3.7a) by end-sequencing, PCR amplification, and sub-cloning. End sequencing of pET3.7a clone indicated high homology with phosphokinase coenzyme A and an integral membrane protein, respectively.

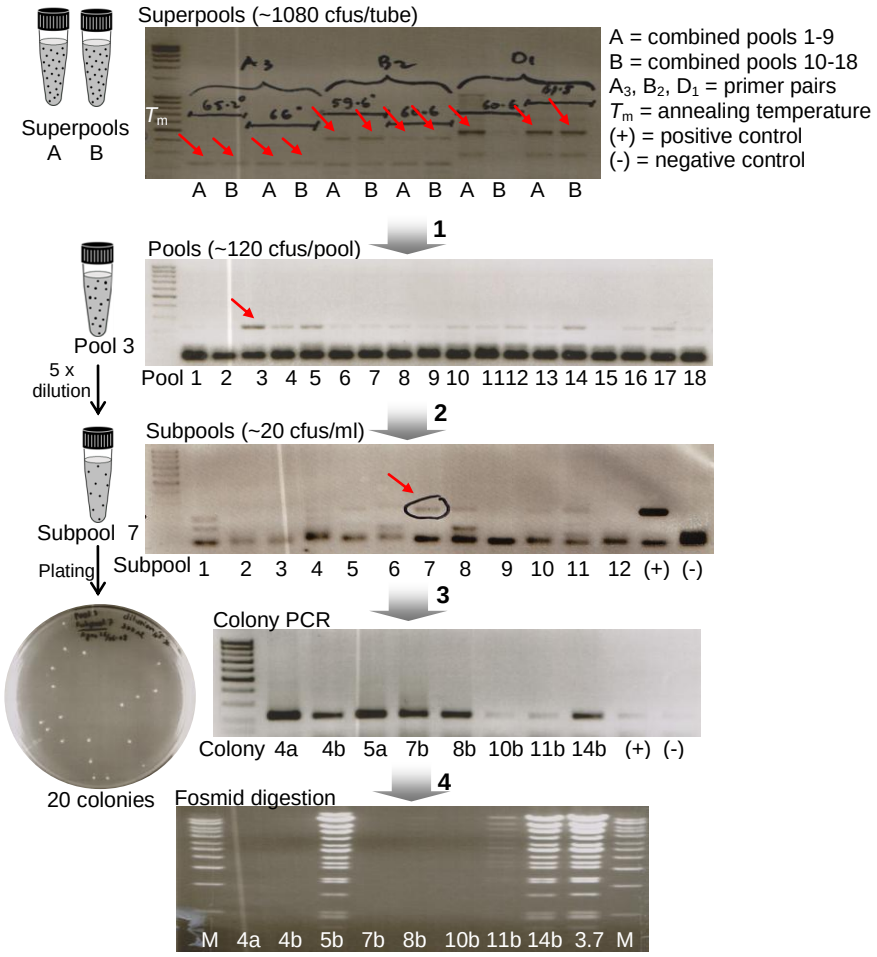


Fig. 3.2.4 NRPS screening of a fosmid library contructed from a filamentous cell pellet prepared from *T. swinhoei* chemotype W. The 18 pools were combined into two superpools. Each superpool was subjected to a whole cell PCR at different T_ms using three different primer pairs specific for adenylation domain-encoding region. All of the pools were screened by A_3 primer pair (step 1). The positive pool (pool 3) was diluted 5 times and the resulting subpools were screened. One of the positive subpools (subpool 7) was plated (step 2). Colony PCR indicated 9 positive colonies (step 3) that were subsequently subjected to enzymatic restriction (step 4). Restriction pattern of the 9 colonies were the same, suggesting that they were identical.

A single amplicon coding for SerA was identified from the clone by using the library screening primers (Fig. 3.2.5A), suggesting the presence of an A-domain sequence

within the fosmid. This SerA-encoding sequence showed highest homology (70% maximal identity) to that present in a peptide synthetase of the cyanobacterium, *Cyanothece* sp. CCY0110 (acc. nu. ZP_01728834). However, when two fragments generated by *Pst*I-digestion of the clone were subcloned and end-sequenced (Fig. 3.2.5B), they only encoded hypothetical proteins instead of typical NRPS components, suggesting that the clone harbored only a small NRPS cluster instead of the expected complex multimodular system.

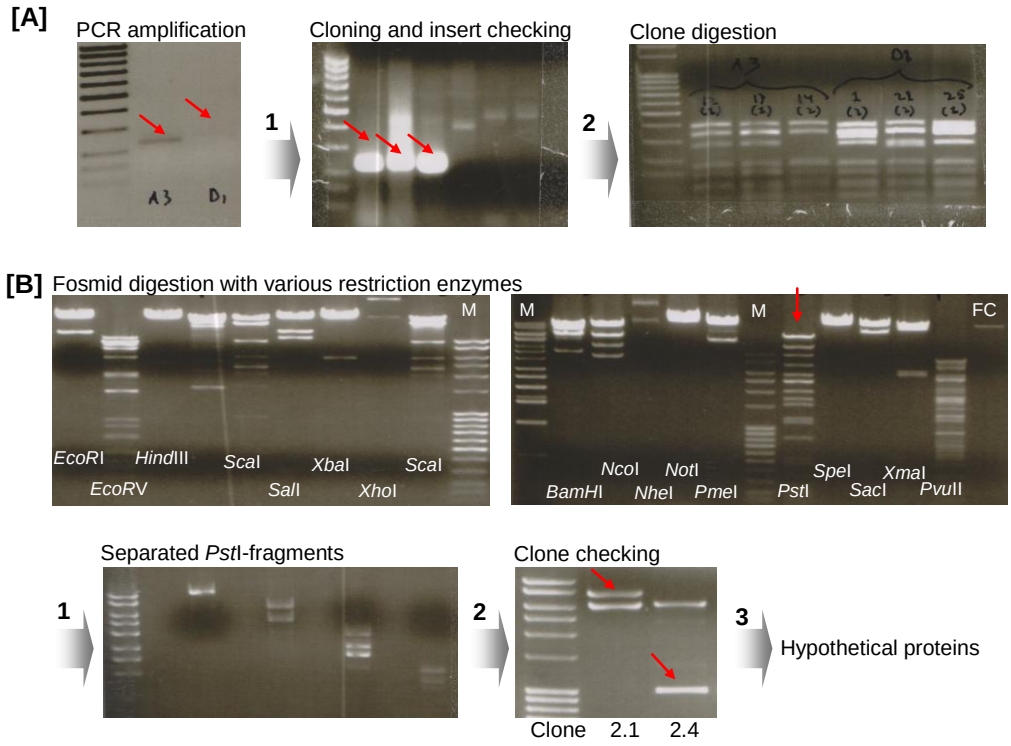


Fig. 3.2.5 Analysis of the fosmid pET3.7a. **[A]** A SerA-encoding regions were amplified from the fosmid using two primer pairs A3 and D1. The PCR products were purified and cloned into *E. coli* (step 1). Some clones were checked by PCR for the presence of A domain encoding inserts (step 2), followed by digestion with the frequent-cutting enzyme *Hha*I. **[B]** Subcloning of pET3.7a. Restriction pattern of pET3.7a digested with different endonucleases, in which the recognition sites are present in the multi-cloning sites of the cloning vector pBlueScript II SK(-). The *Pst*I-restriction pattern was chosen (red arrow). *Pst*I-fragments of pET3.7a were separated by electrophoresis (step 1), individually extracted from the gel, and were individually cloned. Some potential clones representing each *Pst*I-fragment were checked for the presence of inserts, as indicated by red arrows (step 2). Two clones containing inserts (red letters) were subjected to end-sequencing (step 3).

A similar result has recently been reported by Hentschel and coworkers showing the wide distribution of a single small NRPS module in 12 marine sponge species from different geographic locations (Pimentel-Elardo *et al*., 2012). Pimentel-Elardo *et al* (2012) isolated a cosmid clone containing a small NRPS structural gene flanked by efflux protein- and PPTase-encoding genes. This NRPS gene was predicted to

encode two complete NRPS modules (C-A-T-C-A-T-Te) that might be responsible for the biosynthesis of a dipeptide consisting of phenylalanine and tyrosine. The putative bioactivity and function of this dipeptide remain unknown (Pimentel-Elardo *et al*., 2012). A critical issue arises about the reason why a small NRPS gene cluster was detected in the library whereas PKS genes were not. A possible explanation is that the small NRPS and PKS genes were probably contained in different bacterial members in the fraction. Target large PKS genes might be located on the genome of rare bacterial species, and as the consequence, they were underrepresented in the generated fosmid library. In contrast, bacterial species harboring the small NRPS gene is probably abundant, and therefore, its genome was well represented in the library.

Due to the failure in identifying complex PKS and NRPS pathways in the library constructed from the filamentous bacterial fraction, a highly complex metagenomic library was subsequently constructed from the total DNA of *T. swinhoei* chemotype white. Before library construction, we first checked the total sponge DNA for the presence of the SerA domain and "Entotheonella sp." The same specific primers used for previous library screenings were employed to obtain the *serA* amplicon. In addition we detected the presence of *"*Entotheonella sp." in the sponge DNA using a primer pair (238F and 1309R), targeting a unique region indicative of "Entotheonella sp." 16S rRNA genes (Schmidt *et al*., 2000). Cloning, sequencing, and analysis of the target PCR products led to the isolation of a 1-kb 16S gene fragment having a unique region of *"*Entotheonella sp." (Fig. 3.2.6) (supplementary S3.2.1). A similar unique region was previously reported for a Palauan *T. swinhoei* (Schmidt *et al*., 2000) and a Caribbean *Discodermia* sp. (Brück *et al*., 2008).

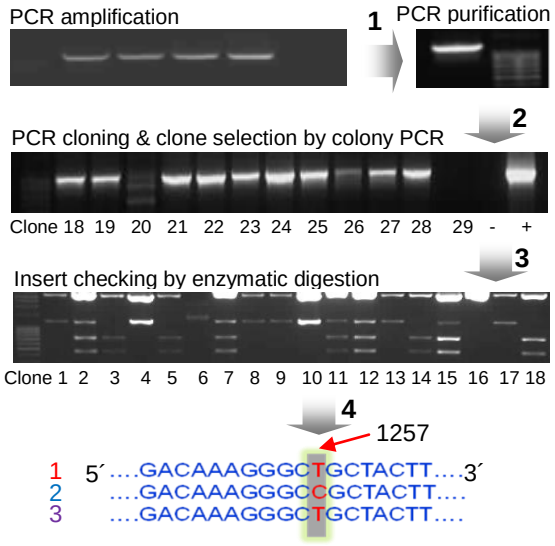


Fig. 3.2.6 Detection of *Candidatus* "Entotheonella sp." in the sponge *T. swinhoei* chemotype W. PCR amplification was carried out using *Entotheonella* 16SrDNA-targetting primer pair (238F-1309R) using metagenomic DNA as PCR template. The PCR product obtained at optimal T_m was purified and cloned into the T-overhang pBlueScript II SK(-) vector. Several clones were selected by colony PCR using the primer pair (238F and 1247R-ad). The positive clones based on colony PCR were subject to a *Pst*I-digestion. Two clones with different *Pst*I-restriction patterns were sequenced. One of the sequenced clones showed a unique region as described by Schmidt *et al*. (2000). Note: 1, Japanese *T. swinhoei* W (in this study); 2, Palauan *T. swinhoei* (Schmidt *et al*., 2000), and 3, *Discodermia* sp. (Brück *et al*., 2008).

For library construction, total DNA was initially isolated from one gram of *T. swinhoei* sample. The isolated DNA was separated on low-melting-point (LMP) agarose gel, and the DNA fragments in the size range of 35-40 kb were purified from the gel. Subsequently the size-selected DNA fragments were end-repaired, blunt-ligated into the fosmid vector pCC1FOS^{TM} (Epicentre), packaged into the lambda phage particles and finally transfected into *E. coli* (Epicentre, Hrvatin and Piel, 2007; Gurgui & Piel, 2010). The recombinant clones were grown in semi-liquid media according to the methods described by Elsaesser & Paysan (2004), Hrvatin & Piel (2007), and Gurgui & Piel (2010) at a clone density of up to 4000 cfus per ml, thereby resulting in a complex 3-D metagenomic library containing ~105,875 clones (Fig. 3.2.7). Subsequently we generated a second metagenomic library containing approximately 36,720 fosmid clones at a density of ~ 510 cfus per ml. Subsequently we expanded the library size to 350,000 clones, thereby increasing the probability of acquiring potential gene cluster.

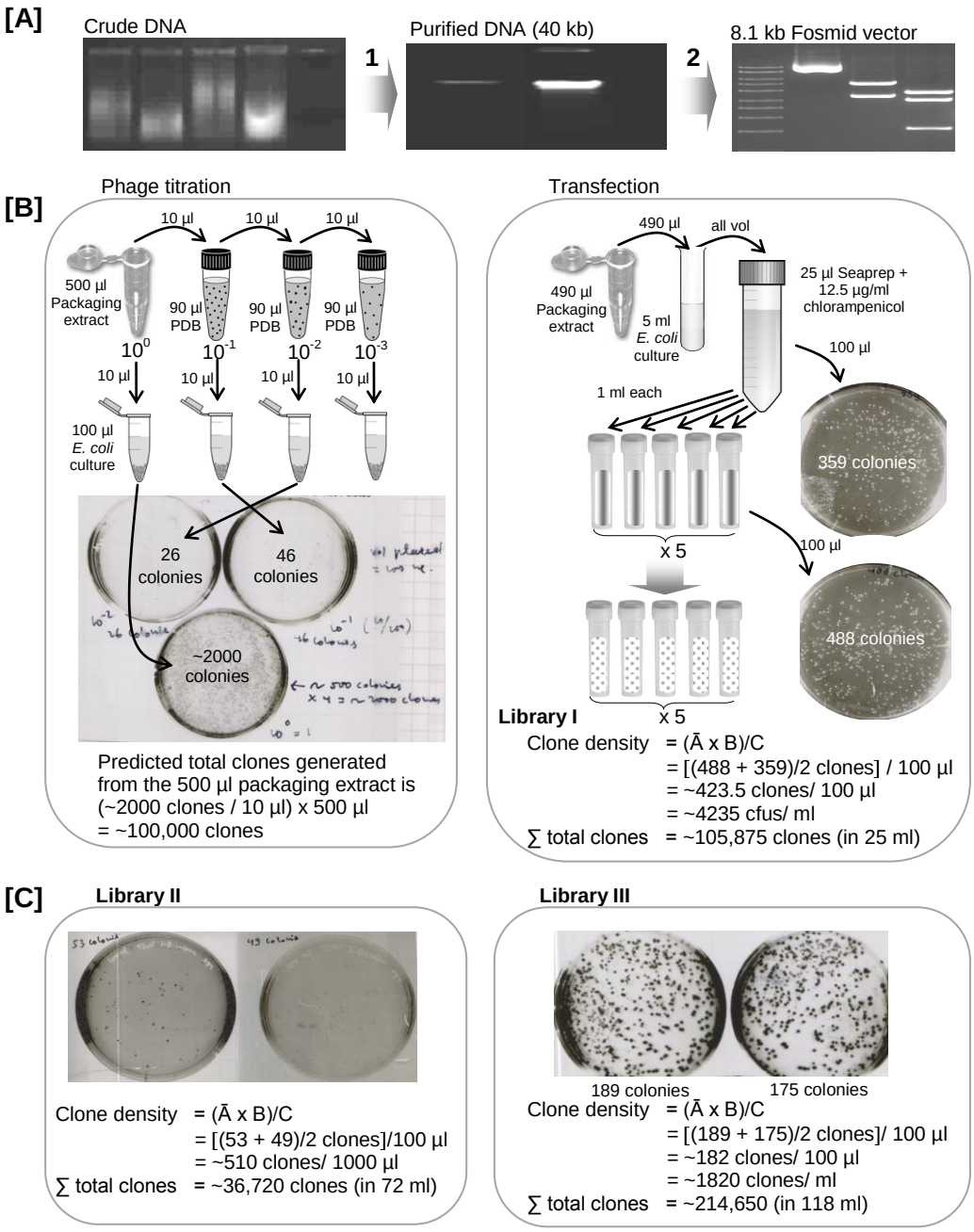


Fig. 3.2.7 Generation of a 3-D metagenomic library from the Japanese marine sponge *T. swinhoei* chemotype W. **[A]** The crude extract obtained from the whole sponge was selected by LMP-gel extraction. The resulting pure ~40 kb DNA fragments was subjected to end-repair and ligated to a fosmid vector. The 8.1 kb fosmid vector used in this work was purchased from Epicentre. It is linearized at the *Eco*72 I (blunt) site and then dephosphorylated. The vector is ready for cloning end-repaired (blunt-end) genomic DNA of approximately 40 kb (Epicentre). **[B]** The number of the phage particles in the packaging extract was estimated. Based on the phage titration, the phage particles were transfected into *E. coli* cells, and the recombinant clones were grown in semi-liquid media at a density of ~4000 cfus per ml. **[C]** The library size was expanded up to ~357,245 clones by generating a second library (~36,720 cfus at a clone density of ~510 cfus per ml) and a third library (~214,650 clones at a density of ~1820 cfus per ml). Abbreviations and symbols: PDB, phosphate dilution buffer; vol, volume; HMW, high-molecular-weight; ∑, amount/number; Ā, average number of colonies; B, dilution factor; C, volume taken from the dilution.

3.2.2 Isolation of a polyketide biosynthesis genes and preliminary analysis

Based on the sequence information of eight KS genes of the *trans*-AT PKS cloned from the uncultured filamentous bacterial fraction as described in section 3.1.2, we designed specific PCR primers to target the *trans*-AT biosynthetic pathway in the 350,000 clone-member metagenomic library. In this work, screening of the highly complex 3-D library was significantly simplified by using whole-cell PCR analysis according to the method described by Hrvatin and Piel (2007) & Gurgui and Piel (2010) (Fig. 3.2.8).

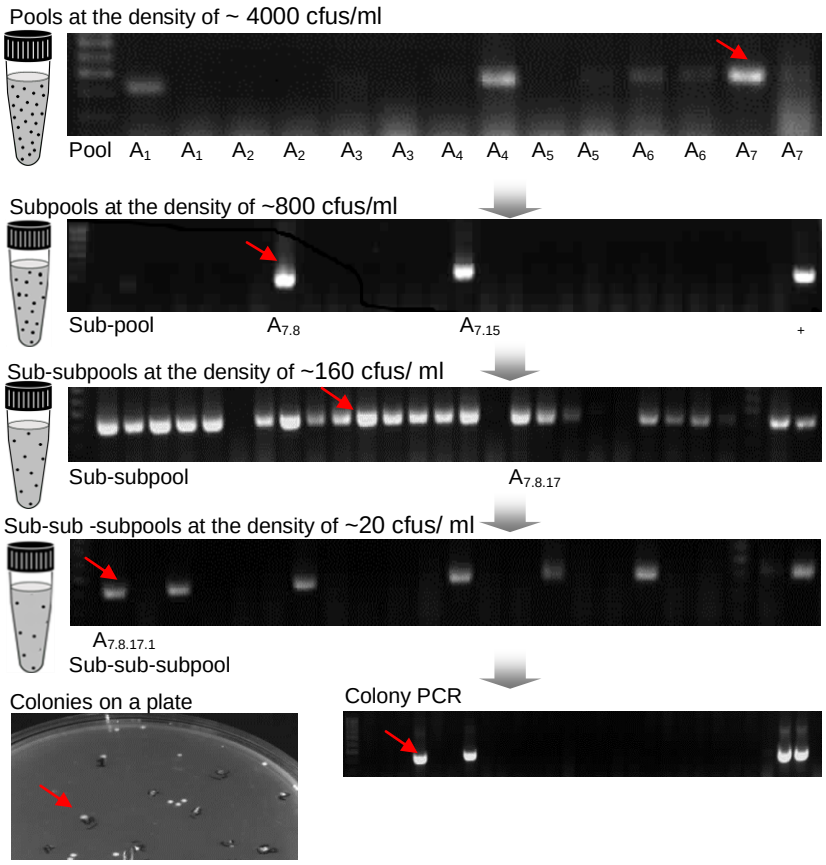


Fig. 3.2.8 Screening of a complex metagenomic library constructed from the Japanese *T. swinhoei* to isolate a PKS gene cluster. The library was initially grouped into pools that are arrayed in three boxes. One of the box (box 1) contains pools with the clone density of ~4000 cfus/ml. Using the specific primers designed based on the isolated *trans*-AT KS domain genes, we detected two positive pools by whole cell PCR. One of the positive pools as indicated by a red arrow was diluted and screened further by PCR. When the clone density was ~20 cfus/ml, a small volume of it was plated at a density of ~200 cfus/ plate. By colony PCR, two positive colonies were detected. One of the colonies, designated as pTSW-AU1, was chosen for further analysis. Abbreviations: cfus, colony forming units.

Clone pools were screened by whole-cell PCR. A positive clone pool (for example pool A7) were re-grown at lower densities and screened further until individual positive clones were detected. The advantages of employing this screening method are that the whole procedure from transfection to the pure clone can be completed in four days by a single person, being highly economic in terms of labor, consumables and freezer space (Hrvatin and Piel, 2007). We screened the library consisting of 4000 cfus/ml, which led to isolation of one candidate 35-kb fosmid clone, designated as pTSW-AU1. The insert had regions with homology to KR and KS domains at its termini, suggesting that it is a part of a large PKS cluster. Then the clone was checked for the presence of the 7 isolated KS sequences of *trans*-AT PKS by PCR amplification. Five corresponding KSs (EKS27, EKS37, EKS50, EKS64, and EP5) were present on the cloned region. Based on this information, the clone was completely sequenced (Fig. 3.2.9).

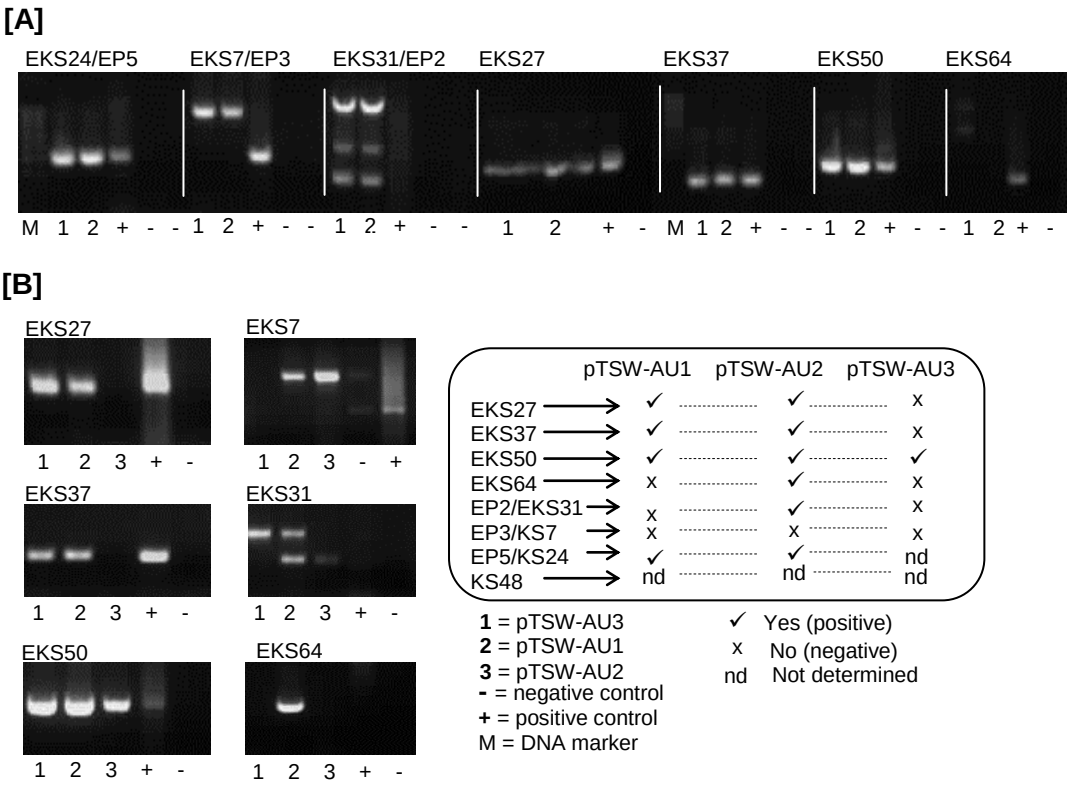


Fig. 3.2.9 PCR detection of the cloned KS sequences in the positive recombinant fosmids isolated from the metagenomic library. **[A]** Checking for pTSW-AU1 fosmid. **[B]** Checking for pTSW-AU2 and pTSW-AU3 fosmid as well as pTSW-AU1.

Screening for additional regions of the PKS gene cluster was conducted by chromosome walking. In this chromosome walking, starting with the known sequence of the isolated pTSW-AU2, we designed specific primers based on the ends of pTSW-AU2 sequence. Then these primers were used to screen the library

by combining clone dilution and whole cell PCR techniques in a similar way as described in Fig. 3.2.3. Employing this screening approach enabled us to isolate two clones, designated as pTSW-AU2 and pTSW-AU3. BLAST analysis revealed that the end sequences of the pTSW-AU2 encoded a KS region and a phosphoenolpyruvate (PEP) synthase. The KS region is overlapped with the previously sequenced clone. To determine whether the PEP synthase gene was located within or outside of the cluster, we digested the pTSW-AU2 clone into five fragments with *Pst*I enzyme and individually sub-cloned them (Fig. 3.2.10A-C). The end sequences of the subclones revealed a 7.5 kb fragment coded for a PEP synthase and an AT, a 3.7 kb fragment encoded ER and DH domains, and a >10-kb fragment for KS and KS-KR domains (Fig. 3.2.10D). The other cloned fragments were parts of the cloning vector. On the basis of this information, we decided to sequence pTSW-AU2 fosmid completely.

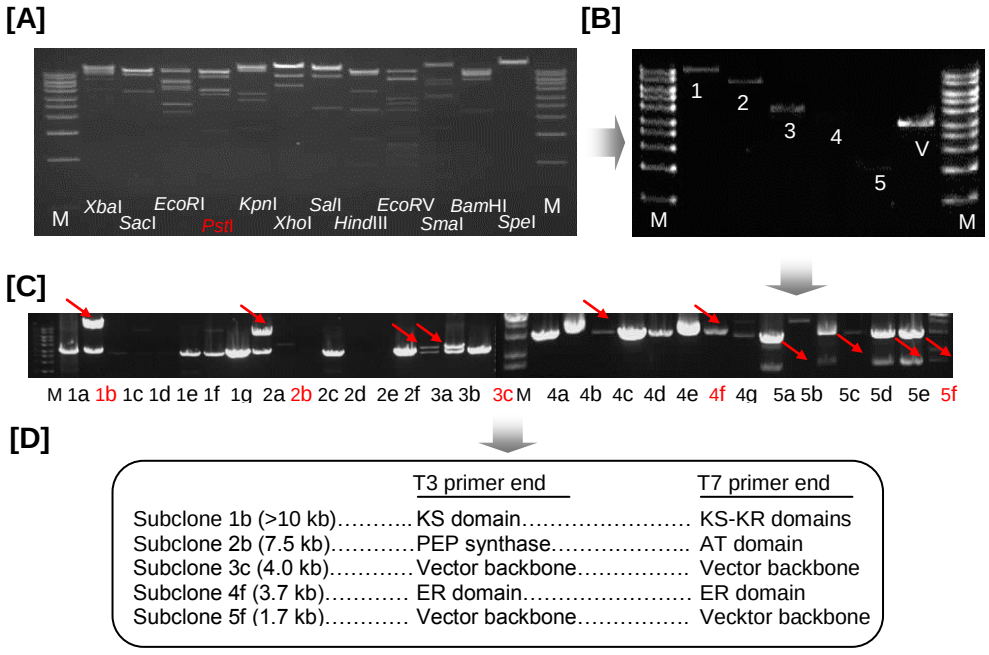


Fig. 3.2.10 Sub-cloning of the fosmid pTSW-AU2. **[A]** Restriction pattern of pTSW-AU2 digested with different endonucleases, in which the recognition sites are present in the multi-cloning sites of the cloning vector pBlueScript II SK(-). The *Pst*I-restriction pattern was chosen (red color). **[B]** *Pst*I-fragments of pTSW-AU2 were separated by electrophoresis and individually extracted from the gel. **[C]** *Pst*I-fragments of pTSW-AU2 were individually cloned. Some potential clones representing each *Pst*I-fragment were checked for the presence of inserts, as indicated by red arrows. Representative clones containing inserts (red letters) were subjected to end-sequencing. **[D]** Subcloning and end-sequencing of pTSW-AU2 fosmid.

We annotated the sequence of PKS gene cluster fragments in pTSW-AU1 and pTSW-AU2 by finding the open reading frames (*orfs*) and searching for the homologous sequences in GenBank. We found four *orfs* on the two overlapping

fragments, which were predicted to code for phosphoenolpyruvate (PEP) synthase, phosphopantetheinyl transferase (PPTase), and two PKS type I proteins (Fig. 3.2.11). The PPTase located upstream of the PKS type I protein is assumed to be responsible for activating ACPs in the downstream modular PKS through the transfer of a phosphopantetheinyl group (Ppant) from coenzyme A to a conserved serine residue of the ACPs. The active *holo*-ACPs facilitate the transfer of intermediates between domain active sites during biosynthesis (Walsh *et al*., 1997; Weissman *et al*. 2004).

We analyzed the deduced amino acid sequences of the two PKS ORFs by identifying conserved residue motifs, allowing prediction of the composition and order of the domains in each module. Our sequence analysis, as described in Figure 3.2.11, revealed that both ORFs harbor 8 modules, in which module 5 is split in two separate proteins. All the PKS modules, except for the first module, do not have integrated AT domains, typical for *trans*-AT PKSs (Piel, 2010). Some examples of compounds encoded by *trans*-AT PKS systems include pederin synthetase (Piel, 2002), leinamycin (Cheng *et al*., 2002; 2003), mupirocin (El-Sayed *et al*., 2003), lankacidin (Mochizuki *et al*., 2003), onnamides (Piel *et al*., 2004b), disorazols (Carvalho *et al*., 2005; Kopp *et al*., 2005), rhizoxins (Partida-Martinez & Hertweck, 2007), and psymberin (Fisch *et al*., 2009).

Although in our case the first module contains an AT fused with a KR, it is assumed that the AT is nonfunctional due to the lack of the conserved active site motif GHSXG. The serine in this signature sequence GHSXG is involved in the formation of the acyl-enzyme intermediate (Cheng *et al*., 2003). We expected that a functional stand-alone AT accepting acyl building blocks was encoded in the cluster as a separate protein (Cheng *et al*., 2003). The module containing the inactive AT displayed 41% homology to MlnB of macrolactin synthase (Schneider *et al*., 2007) and 43% homology to the first module in DszA of disorazole synthase (Carvalho *et al*., 2005). In particular, KS1 bears high homology (74% identity) to the acetyl-accepting KS in the DszA of disorazole biosynthetic pathway, suggesting that module 1 of the PKS identified in this work may play a similar loading mechanism as proposed for macrolactin and disorazole biosynthesis (Schneider *et al*., 2007; Carvalho *et al*., 2005).

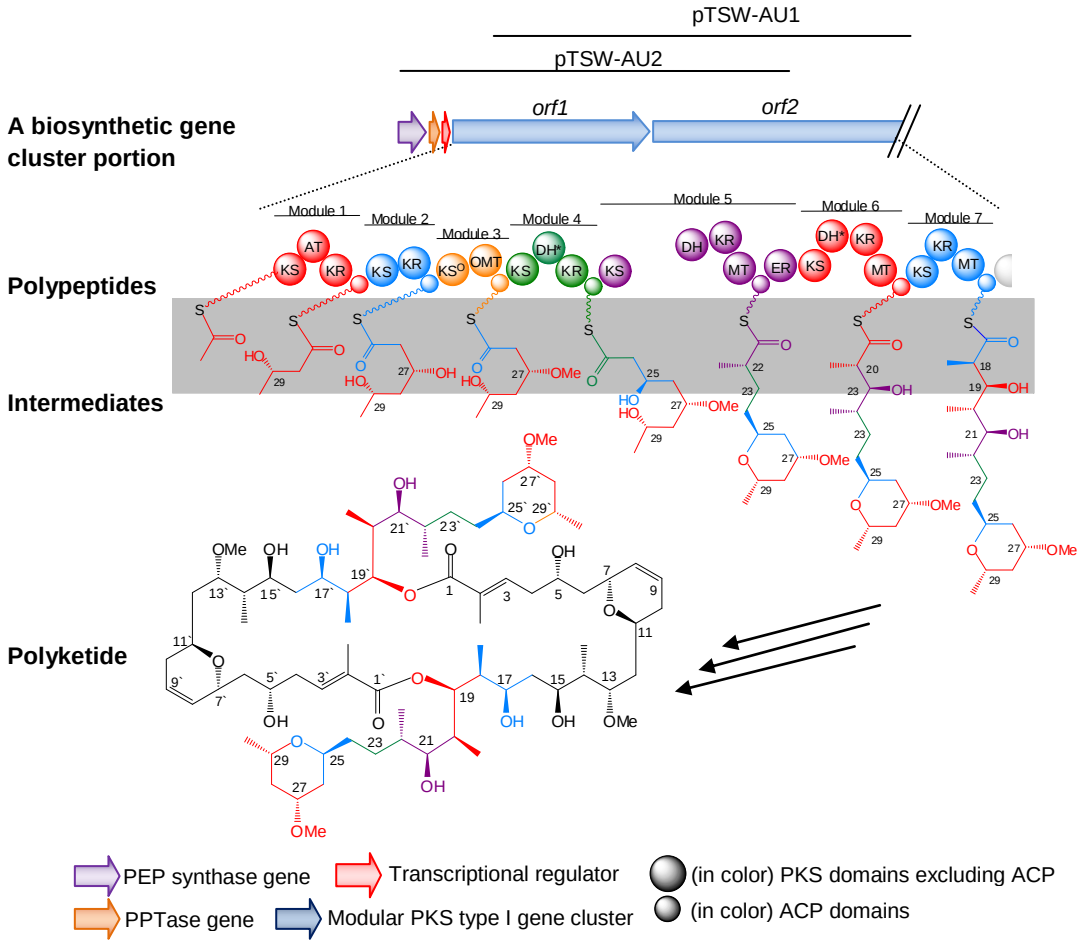


Fig. 3.2.11 Preliminary analysis of two overlapping fosmid clones to predict the biosynthetic pathway encoded on the isolated gene cluster. The segment of misakinolide A covered by the clued PKS section is in the range of C_{15} to C_{29} and $C_{15'}$ to $C_{29'}$. Parts of the gene cluster that are still missing code for the formation of the segment C_1-C_{14} and $C_{1'}$-$C_{14'}$.

We assumed that a malonate would initially be loaded by a separate *trans*-AT onto the ACP1, which is subsequently decarboxylated to generate an acetyl group. The resulting acetyl group would move to the KS1 to create the starter unit that is reloaded by the *trans*-AT onto the same ACP1. This allows the KS1-catalyzed condensation of the acetate starter with the first malonate to generate the first four-carbon intermediate (Carvalho *et al*., 2005). Subsequent β-reduction of the intermediate by KR1 may lead to the formation of β-hydroxymalonyl-*S*-ACP. Sequence analysis showed that KR1 displayed 66% identity to KR19 of rhizopodin synthase that performed β-ketoreduction to form a β-L-OH group (Pistorius & Müller, 2012). The absence of a key D residue in the stereospecific motif of KR1 (AGQM**S**) indicates L configuration of the β-hydroxyl moiety of the substrate loaded onto ACP1 (Reid *et al*., 2003; Caffrey, 2003).

Based on the architecture of module 2 (KS-KR-ACP), it is assumed that KS2 would initially mediate a two-carbon extension. Subsequent β-ketoreduction by KR2 would result in a β-hydroxylated intermediate. KR2 has the stereospecific motif of D configuration (AGVIS**D**), indicating the D configuration of the β-hydroxyl group. BLAST analysis showed the high homology (65% identity) of the ORF1 KS3 to the KS2 of macrolactin synthase that accepts a hydroxylated intermediate (Schneider *et al*., 2007). This suggests that the KS3 would accept the hydroxylated substrate from ACP2. However, the lack of a catalytic HGTGT motif in this KS makes it theoretically incapable of elongating the polyketide assembly line (Davies *et al*., 2000). The OMT domain next to the KS^{o} is predicted to perform α-*O*-methylation of the β-hydroxy group, allowing the formation of an oxy-methyl group. This is consistent with the presence of an OMe group at the C_{27} of misakinolide A structure. Module 4 (KS-DH*-KR-ACP) contains a B-type KR that has the stereospecific motif for D configuration (AGLTR**D**), indicating a β-D-hydroxyl group in the incorporated intermediate. The cognate DH domain in this module, designated as DH*, is predicted to be nonfunctional due to the lack of the essential G residue in the catalytic motif HXXXGXXXXP (Keatinge-Clay, 2008), and therefore it would not be able to perform β-hydroxylation, leaving the hydroxyl group. However, BLAST analysis of KS5 showing its high homology (73% identical) to the KS9 of batumin synthase that was proposed to accept incoming α,β-*E*-double bond (Mattheus *et al*., 2010), suggesting that hydroxyl function might undergo further β-processing to form α,β-double bond by a *trans*-acting `hidden` DH.

Module 5 is split in two separate ORFs. KS5 domain lies on the ORF1, while the other cognate domains (DH-KR-MT-ACP-ER) are located on the ORF2, representing unusual domain organization typical of *trans*-AT PKSs (Piel, 2010). The β-keto functionality generated by the KS5-catalyzed condensation would be reduced by DH5 and KR5 domains to form the α,β-double bond. The presence of an ER5 domain on downstream of the ACP5 subsequently reduces the double bond into a single bond. MT5 would subsequently introduce a methyl group at the α-position. The architecture of module 5 showed that it installs the methylated intermediate bearing α,β-reduced moiety. This is supported by the high homology of KS6 (73% identity) to the KS6 of Rhizopodin biosynthetic pathway that accepts an α-L-methylated substrate with α,β-reduced moiety (Pistorius & Müller, 2012). This is consistent with the C_{22} methyl and the C_{22}-C_{23} saturated bond in misakinolide A structure. Based on module 6 (KS-DH*-KR-MT-ACP) architecture, it is predicted that

it installs an intermediate carrying α-methyl and β-hydroxyl moieties. Highest homologous of KS7 (64%) was the KS4 in pederin biosynthesis that was proposed to accept the α-methylated intermediate with β-L-OH functionality (Piel, 2002). The α-methyl introduced by MT6 is consistent with the presence of the C_{20} methyl in the misakinolide A structure. The β-OH moiety would not be reduced into a double bond because the cognate DH is nonfunctional (Keatinge-Clay, 2008). The lack of the important conserved motif DXXX(Q/H) in this DH domain suggests its inability to perform dehydration (Keatinge-Clay, 2008), leaving a hydroxyl group at C_{21} with L configuration (VMVAQ**S**) based on KR6 analysis. Module 7 is composed of KS-KR-MT domains that were assumed to install the α-methylated and β-hydroxylated intermediate. MT7 is predicted to be responsible for introducing a methyl group at C_{18}. KR-based analysis indicated a key D residue present in the KR7 stereopecific motif (ALILE**D**), suggesting the D configuration of the C_{19}/$C_{19`}$ β-OH moiety. At the end of the PKS assembly line, the $C_{19}$/$C_{19`}$ hydroxyl groups would attack the $C_{1`}$/$C_1$ thioester bonds to form the $C_{19}$-$C_{1`}$ and $C_{19`}$-C_1 ester bonds. Formation of these ester bonds is expected to be catalyzed by a TE to link two resulting monomers and release a dimer from the assembly line. The presence of MT and *O*-MT in modules 6 and 7, respectively, suggest that this *trans*-AT PKS pathway is most likely loaded with malonyl-CoA both as the starter and extender units, instead of methylmalonyl-CoA or other 2-branched malonyl-CoA derivatives.

Taken all these features together, the predicted module architecture and sequence analyses showed that the partial gene cluster isolated in this work seemed highly likely to encode the formation of almost half of misakinolide A or swinholide A, a promising antitumor secondary metabolite. These complex dimeric polyketides are a potent actin inhibitor previously isolated from *T. swinhoei* sponges (Sakai *et al.*, 1986; Kato *et al.*, 1987; Kitagawa *et al*., 1990). However, finding the remaining parts of the gene cluster and subsequent detailed phylogenetic analysis of the entire cluster sequence was required to confirm that we had identified the correct biosynthetic pathway.

3.2.3 Completion of the biosynthetic gene cluster

As described above, our sequence analysis suggested that the isolated partial gene cluster might be responsible for the biosynthesis of misakinolide A. This preliminary result motivated us to isolate the remaining fragments of the gene cluster. Searching for the down-stream parts was performed by chromosome walking, in which the end sequences of the previously isolated fosmids were used as the basis for further screening of the library. Three additional clones, designated 1I9 (pTSW-AU4), 1B9 (pTSW-AU5), 3A1 (pTSW-AU6), were successfully isolated. In the isolation of these three clones, a master student (Ms. Silke Reiter) was involved under my supervision during a 1.5-month practical course in the Piel group. End-sequencing of the clones indicated that pTSW-AU5 seemed promising because it showed a small overlap with the previously isolated misakinolide gene cluster portion. This clone was completely sequenced. Bioinformatic analysis showed that it harbored an extended downstream part of the misakinolide gene cluster as expected. It was predicted that one more additional downstream clone would be required to reach the downstream end of the gene cluster.

Our further screening efforts led to the isolation of five additional clones. Two of them were found to harbor overlapping downstream fragment (pTSW-AU10 or 1E9) along with an additional upstream clone (pTSW-AU11), thereby covering the entire cluster. These two positive clones were entirely sequenced. We analyzed the contigs obtained after the sequencing by removing the vector backbone and assembling the sequences by using the Seqman of the Lasergene program. We filled gaps between discrete contigs by PCR through the following steps: (i) designing PCR-primers flanking the gaps, (ii) PCR-amplification using the primers, (iii) purification and sequencing of the obtained amplicons, and (iv) sequence assembly of the amplified sequences with the existing gene cluster sequences, thereby resulting in a single complete contig. Sequence annotation of the gene cluster was performed by identifying potential ORFs, followed by homology sequence searches in the public databases. ORF identification was carried out using the software FramePlot 2.3.2 (Ishikawa and Hotta, 1999) and Geneious 5.4.6 (Drummond *et al*., 2011). The genetic map of the annotated genes is shown in Fig. 3.2.12.

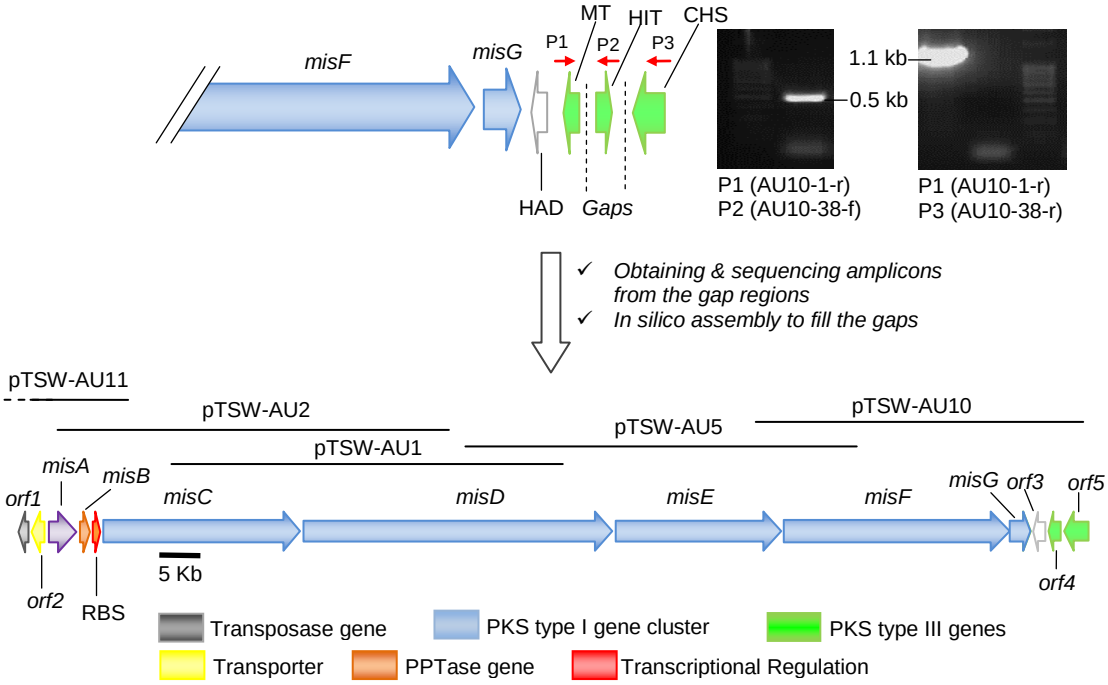


Fig. 3.2.12 Genetic map of the ~103-kb DNA region containing a large *trans*-AT PKS gene cluster. The type I PKS gene cluster is flanked by genes coding for transposase (*orf1*), NhaP-type Na^+/H^+ and K^+/H^+ antiporter (*orf2*), nonspecific acid phosphatase precursor (*orf3*), SAM-dependent methyltransferase (*orf4*), and naringenin-chalcone synthase (CHS) or PKS type III (*orf5*). Abbreviations: RBS, ribosome binding site; P1, P2, and P3 are primers for filling gaps.

The presence of some transposase genes prior to the NhaP-type antiporter (NAH) and PEP synthase genes indicated the upstream end of the misakinolide A gene cluster. The downstream end of the cluster was marked with the presence of a gene coding for a thioesterase (TE) domain. This ~103-kb DNA region harboring biosynthetic genes (*misA-misG*) exhibited typical bacterial gene features, such as the absence of introns, the presence of ribosome binding sites, and small intergenic distances between putative genes, strongly indicating that the producer of misakinolide A is a symbiotic bacterium. However, the taxonomic status of the bacterial producer is still unknown yet and awaits further analysis either by single cell studies or its combination with complete genome sequencing. The screening system used to isolate the gene cluster was based on the sequence information present in KS fragments cloned from the filamentous fraction dominated by “Entotheonella sp.” This preliminary information may provide a clue that the filamentous fraction dominated by “Entotheonella sp.” could be used as the starting material to conduct single cell studies.

3.2.4 Bioinformatic analysis of the entire biosynthetic gene cluster

We performed bioinformatic analysis of the entire assembled sequences derived from complete sequencing of five overlapping fosmids (see supplementary S3.2.2). The resulting continuous sequence covering ~103-kb DNA region was predicted to contain a *trans*-AT PKS gene cluster. The detailed annotation of the individual *orfs*, as summarized in Table 3.2.1, revealed 17 ORFs encoded on the gene cluster. Five of them (*misC*, *misD*, *misE*, *misF*, and *misG*) encode the *trans*-AT type I PKS system consisting of five multifunctional enzymes (MisC, MisD, MisE, MisF, and MisG). Two ORFs, designated as MisA and MisB, are homologous to many PEP synthases and putative phosphopantetheinyl transferase (PPTase), respectively.

We initially predicted the module architecture present in the individual PKS proteins by BLAST analysis (Altschul *et al*., 1990) and confirmed the prediction by finding the highly conserved motifs in each domain (Supplementary 1). Subsequently we checked individual domains for the presence or absence of catalytic important conserved motifs or essential conserved residues. Sequence analysis showed that all the ACPs present in each module contain the highly conserved GXDS motif, in which the serine (S) residue is essential for binding the 4`-phosphopantetheine site. This suggests that all the ACPs are able to undergo posttranslational modification from the inactive *apo* form to the active *holo* form by the transfer of a 4`-phosphopantetheine group to a conserved serine of the individual ACPs (Lambalot *et al*., 1996; Findlow *et al*., 2003).

Sequence analysis of all DH domains in the PKS system showed that two DH domains (DH4 and DH6) are theoretically unable to perform dehydration because of the lack of the essential G residue in the catalytic motif HXXXGXXXXP (Keatinge-Clay, 2008). Sequence analysis of all the 19 KS domains present in this pathway revealed that 15 of them contain the highly conserved TXCSSLV motif (including the active-site cysteine) and HGTGT (including the active site histidine) motif that are essential for chain elongation. Other four KS domains (KS3, KS11, KS14, and KS19) are assumed to be incapable of elongating the polyketide assembly line due to the lack of a catalytic HGTGT motif that is important for decarboxylative condensation (Davies *et al*., 2000).

Table 3.2.1 Deduced functions of the ORFs identified in this work

Protein	Size, aa	Proposed function	Closest protein homolog, (origin)	S/I, %/%	Acc. number
TP1	178	Transposase	NGR_c03400 (Rhizobium sp. NGR234)	80/62	YP_002824890
TP2	227	Transposase	all8065 (*Nostoc* sp. PCC 7120)	70/53	NP_478490
TP3	322	Transposase	PCC8801_4459 (*Cyanothece* sp. PCC 8801)	65/43	YP_002364710
ORF1	803	Esterase/lipase-like protein	Dole_1932 (*Desulfococcus oleovorans*)	49/31	YP_001529813
ORF2	226	NhaP-type Na^+/H^+ and K^+/H^+ antiporters	HCH_05288 (*Hahella chejuensis* KCTC 2396)	64/43	YP_436387
TP4	227	Transposase	all8065 (*Nostoc* sp. PCC 7120)	70/54	NP_478490
ORF3	300	sodium/hydrogen exchanger family protein	PM8797T_03204 (*Planctomyces maris* DSM 8797)	69/49	ZP_01852948
MisA	894	phosphoenolpyruvate synthase	PPSIR1_31263 (*Plesiocystis pacifica* SIR-1)	64/49	ZP_01910935
MisB	250	Phosphopantetheinyl-transferase	ORF16 (uncultured bacterial symbiont of *Discodermia dissoluta*)	61/43	AAY00051
MisC	5870	PKS: KS-AT-KR-ACP-KS-KR-ACP-KS^o-OMT-ACP-KS-DH-KR-ACP-KS	PKS (*Kordia algicida* OT-1)	54/38	KAOT1_04255
MisD	9027	PKS: DH-KR-MT-ACP-ER-KS-DH-KR-MT-ACP-KS-KR-MT-ACP-KS-KR-ACP-KS-KR-MT-ACP-KS	RhiC *(Burkholderia rhizoxinica*)	51/35	YP_003748157
MisE	4474	PKS: KR-ACP-KS^o-OMT-ACP-KS-KR-ACP-KS-KR-ACP-KS^o	Bat2 (*Pseudomonas fluorescens*)	60/44	ADD82940
MisF	8075	PKS: DH-ACP-KS-DH-PS-KR-ACP-KS-KR-ACP-KS-DH-KR-MT-ACP-KS-DH-KR-ACP-KS^o-ACP-TE	PKS module containing protein (*Paenibacillus polymyxa* E681)	66/49	YP_003871371
MisG	354	*Trans*-AT	KAOT1_04365 (*Kordia algicida* OT-1)	73/55	ZP_02163874
ORF5	323	putative nonspecific acid phosphatase precursor	WH5701_02224 (*Synechococcus* sp. WH 5701)	51/34	ZP_01084293
ORF6	145	phage shock protein A	DVU2988 (*Desulfovibrio vulgaris* str. Hildenborough)	50/33	ADP87902
ORF7	236	SAM-dependent methyltransferase	Complement (NZ_AAXU02000001.1) (*Algoriphagus* sp. PR1)	55/37	YP_004054326

Note: TP, transposase; MCAT, malonyl CoA-acyl carrier protein transacylase; Bat2, batumin-related polyketide antibiotics; misA –F, misakinolide PKS genes; aa, amino acid; S/I, similarity/identity; Acc., accession.

At the downstream end of the PKS type I gene cluster, it was found the presence of a discrete AT gene (*misG*) that is seemingly linking the biosynthetic gene cluster of modular PKS type I to that of the PKS type III family called chalcone synthase (CHS). MisG showed homology with many malonyl CoA-acyl carrier protein transacylase (MCAT) genes. It has 55% identity to the *trans*-acting malonytransferase PedD of pederin synthase (Piel, 2002; Jensen *et al*., 2012). MisG contains the conserved active-site GHSXG motif as well as the conserved malonyl-CoA binding motif AFHS (Reeves et al., 2001), suggesting that this discrete AT enzyme is functionally active and presumably specific for malonyl-CoA incorporation. This enzyme showed 54% identity to MlnA involved in macrolactin

biosynthesis in *Bacillus amyloliquefaciens* FZB42 (Schneider *et al*., 2007) and 51% identity to BryP involved in bryostatin biosynthesis (Sudek *et al*., 2007). The MlnA is known as a discrete acyl transferase that acts iteratively in *trans* to load malonyl-CoA onto the PKS modules of macrolactin synthase (Schneider *et al*., 2007). The BryP harbors tandem AT domains upstream of the bryostatin synthase ORFs (Sudek *et al*., 2007), which was shown to load malonyl-CoA in *trans* onto Bry PKS modules based on *in vivo* and *in vitro* functional studies by Lopanik *et al* (2008). Taken together all the information described above, we assume that the putative AT enzyme identified in this work performs in *trans* acylation to provide malonyl-CoA both as extender units in the biosynthesis of misakinolide A as well as in chalcone biosynthesis.

By identifying the domain conserved motifs from the entire DNA sequence data, we predicted the module architecture (Figure 3.2.8). Non-elongation KSs (KS^{o}) were identified in modules 3, 11, 14, and 19, suggesting the inability of these modules to perform chain elongation. The presence of 15 elongation modules correlates with the incorporation of 15 malonyl-CoA extender units. The predicted module architecture resembles a *trans*-AT PKS that exhibits unusual features such as AT-less modules, three modules split between two distinct proteins (modules 5, 10, and 14), presumed nonfunctional domains (KS3, KS11, KS14, and KS19), an ER located behind ACP (module 5) and a duplicated DH domain in module 17. As the consequences of these unusual features, the standard textbook collinearity rules cannot usually be applied to determine the final product structure accurately. However, the Piel group has recently developed a phylogeny-based approach that enabled determination of a polyketide structure from the *trans*-AT gene cluster sequence or KS sequences with high confidence (Nguyen *et al*., 2008). Using this approach, Prof. Piel generated a phylogenetic tree from 521 KS sequences including all of 19 *mis* KS domains in order to obtain insights into the intermediates installed on individual modules (supplementary S3.2.3). In some cases, the stereochemistry of α-branches can be predicted by this phylogeny-based KS analysis. The phylogenetic analysis, as shown in Table 3.2.2 below, revealed that the KSs with similar function were grouped into the same clades, suggesting that the substrate specificity of all *mis* KSs can be predicted according to the clades where they were attributed.

Table 3.2.2 Analysis of KS domains present in the *mis* PKS

Domain	BLAST Search			Substrate specificity predicted phylogenetic analysis[a]	Intermediate moiety predicted based on domain architecture
	Close relative	S/I (%/%)	Predicted substrate specificity		
KS1	disorazole KS1 (AAY32964)	59/74	Acetyl-CoA	Acetyl unit	Acetyl-CoA (acetate starter)
KS2	macrolactin KS6 (CCG49394)	58/76	β-L-OH	β-L-OH	β-OH (macrocylization via an ester bond)
KS3	macrolactin KS2 (YP_001421028)	47/65	β-D-OH	β-D-OH (KS^o)	β-OH
KS4	corallopyronin A KS7 (ADI59534)	62/75	β-Me + α,β-double bond	α-D-OMe	β-OMe (C_{29})
KS5	batumin KS9 (ADD82941)	60/73	α,β-*E*-double bond	β-L-OH or D-pyran	β-OH (C_{27}) converted to pyran ring via ester bond
KS6	rhizopodin KS6 (CCA89326)	58/73	α-L-Me + α,β-reduced moeity	α-L-Me + reduced α,β moeity	α-Me (C_{24}) + reduced α,β (C_{24}-C_{25})
KS7	pederin KS4 (AAS47564)	46/64	α-diMe + β-L-OH	α-L-Me + β-L-OH	α-methyl (C_{22}) + β-OH (C_{23})
KS8	corallopyronin A KS6 (ADI59534)	45/70	β-Me + α,β-*E*-double bond	α-D-Me + β-D-OH	α-Me (C_{20}) + β-OH (C_{21}) formed an ester bond
KS9	rhizoxin KS13 (CAL69893)	31/50	β-L-OH	β-L-OH	β-OH
KS10	psymberin KS6 (ADA82585)	59/72	α-diMe+ β-D-OH	α-L-Me + β-L-OH	α-Me + β-OH
KS11	bryostatin KS12 (ABK51303)	53/71	β-OH	β-D-OH (KS^0)	β-OH converted into β-OMe (C_{15})
KS12	sorangicin KS 19 (ADN68482)	59/78	β-L-OH	β-D-OMe	β-OMe (C_{15})
KS13	onnamide A KS 6 (AAV97877)	71/83	β-L-OH	β-D-OH	β-OH converted into Pyran ring
KS14 (KS^o)	bacillaene KS3 (YP_0014212992)	64/79	KS^o (β-OH)	β-L-OH	KS^o (β-OH) converted into α,β -*Z*-double bond
KS15	bacillaene KS4 (EHM05018)	69/83	α,β-*Z*-double bond	α,β-*Z*-double bond	α,β-double bond
KS16	bryostatin KS9 (ABK51301)	67/79	α,β-*E*-double bond	Pyran ring	α,β-double bond
KS17	bryostatin KS9 (ABM63537)	70/84	β-D-OH	β-D-OH	β-OH
KS18	sorangicin KS12 (ADN68478)	67/84	α-dimethyl + double bond	α-Me + α,β-*E*-double bond	α-Me + α,β-double bond
KS19	rhizoxin KS16 (AEG71853)	60/78	α,β-*E*-double bond	α,β-*E*-double bond (KS^o)	α,β-*E*-double bond

[a] Phylogenetic analysis was conducted by Prof Piel using all the KSs present in the pathway and 521 KS sequences from GenBank (supplementary S3.2.3).

All of the modules, except modules 3, 11, 14, and 19, contain the reductive domain KR. The KR domain is usually responsible for modification of the oxidation state and stereochemistry of the growing ACP-bound intermediates (Valenzano *et al*., 2009). Therefore, the presence or absence of KR domain in each PKS module may lead to different oxidation states (Staunton & Weissman, 2001; Kwan *et al*., 2008). In this work, we predicted the stereochemistry of β-hydroxyl generated by KRs present on individual *mis* modules by performing sequence alignment and identifying the stereo-specific motifs according to Caffrey (Caffrey, 2003; Reid *et al*., 2003; Caffrey, 2005). The presence of a key aspartic acid (D) in the motifs indicates D-configured β-OH groups. In contrast, L configuration is attributed to the absence of a key aspartic acid (D) residue in the streospecific motif (Caffrey, 2003; Reid *et al*., 2003; Caffrey, 2005; Baerga-Ortiz, 2006).

Based on the KS phylogeny (supplementary S3.2.3) and KR analysis (Table 3.2.3), we deduced the functionality and stereochemistry of substrates installed on individual modules (Fig. 3.2.13). In principle, the predicted substrate specificities of *mis* KSs are grouped according to the following functionalities: acetyl unit, β-L-OH group, β-D-OH group, β-L-OMe group, α-L-Me and β-L-OH group, α-L-Me and reduced α,β moiety, α-Me + α,β-*E*-double bond, pyran ring, α,β-*E*-double bond, and α,β-*Z*-double bond.

Table 3.2.3 Analysis of KR domains present in the *mis* PKS

Domain	Sequence alignment analysis		Moiety present in the misakinolide A structure
	Stereo-specificity motif[a]	predicted product	
KR1	AGQM**S**AVSFAEKSLADF	β-L-OH	β-ether bond (C_{29}-C_{25}) in a pyran from β-OH
KR2	AGVIS**D**ALVFNKTHHQT	β-D-OH	β-D-OMe (C_{29}) from β-D-OH
KR4 (+DH*)	AGLTR**D**ASLVNKTVRID	β-D-OH	β-D-OH (C_{25})
KR5 (+DH, ER)	AGLTR**D**ASLVNKTVRD	β-D-OH converted to reduced moiety)	Reduced (C_{22}-C_{23})
KR6 (+DH*)	VMVAQ**S**TPMRSLSEA	β-L-OH	β-L-OH (C_{23})
KR7	ALILE**D**RAIADMDEKTL	β-D-OH	β-L-OH (C_{21}) forms L-ester bond with $C_{1`}$
KR8	AGIAP**S**ESLLNKDIDAF	β-L-OH	β-L-OH (C_{19})
KR9	ALVFK**L**DNSIAKTSEED	β-L-OH	β-L-OH (C_{17})
KR10	AGVIH**D**NFIIKKTPQELE	β-D-OH	β-D-OMe (C_{15}) from β-D-OH
KR12	AGIVR**D**NFVINKTSAEFQ	β-D-OH	β-D-OMe (C_{13})
KR13	SGAVD**S**ETPAFVRKPIAK	β-L-OH	β-D-OH to *Z*-double bond (C_8-C_9)
KS15 (+DH, PS)	AGLTR**D**AYIHNKTREEMI	β-D-OH converted to a pyran ring	Pyran ring (C_7-C_{11})
KR16	AGLIR**D**SFMIRKTVDQLH	β-D-OH	β-D-OH (C_5)
KR17 (+DH)	AIVLA**D**KSLANMDERQFL	β-D-OH converted to a double bond	α,β-*E*-double bond (C_2-C_3)
KR18 (+DH)	AGVIQ**D**NFIIRKAVREFQ	β-D-OH converted to a double bond	α,β-*E*-double bond (C_2-C_3)

[a]) Diagnostic site is bold and underlined

The PKS module architecture suggested that nine *mis* modules (modules 1, 2, 4, 8, 9, 10, 12, 13, and 16) install the β-hydroxyl functions. This is in good agreement with the substrate specificities predicted for the downstream *mis* KSs based on the phylogenetic analysis. In particular the phylogeny-based KS analysis indicated the L configuration of the β-OH functionality accepted by KS2, KS5, and KS9, which is consistent with the hydroxyl configuration predicted based on KR analysis, except for KS5. The hydroxyl groups installed on modules 8, 9, 12, and 16 correlate to those at the C_{17}/$C_{17`}$, $C_{15}$/$C_{15`}$, C_{11}/$C_{11`}$, and $C_5$/$C_{5`}$ of the misakinolide A structure. In the phylogenetic tree, *mis* KS5 was placed in the clade specific for β-L-OH or D-pyran function, suggesting the β-hydroxyl function might undergo further processing towards the formation of a pyran ring.

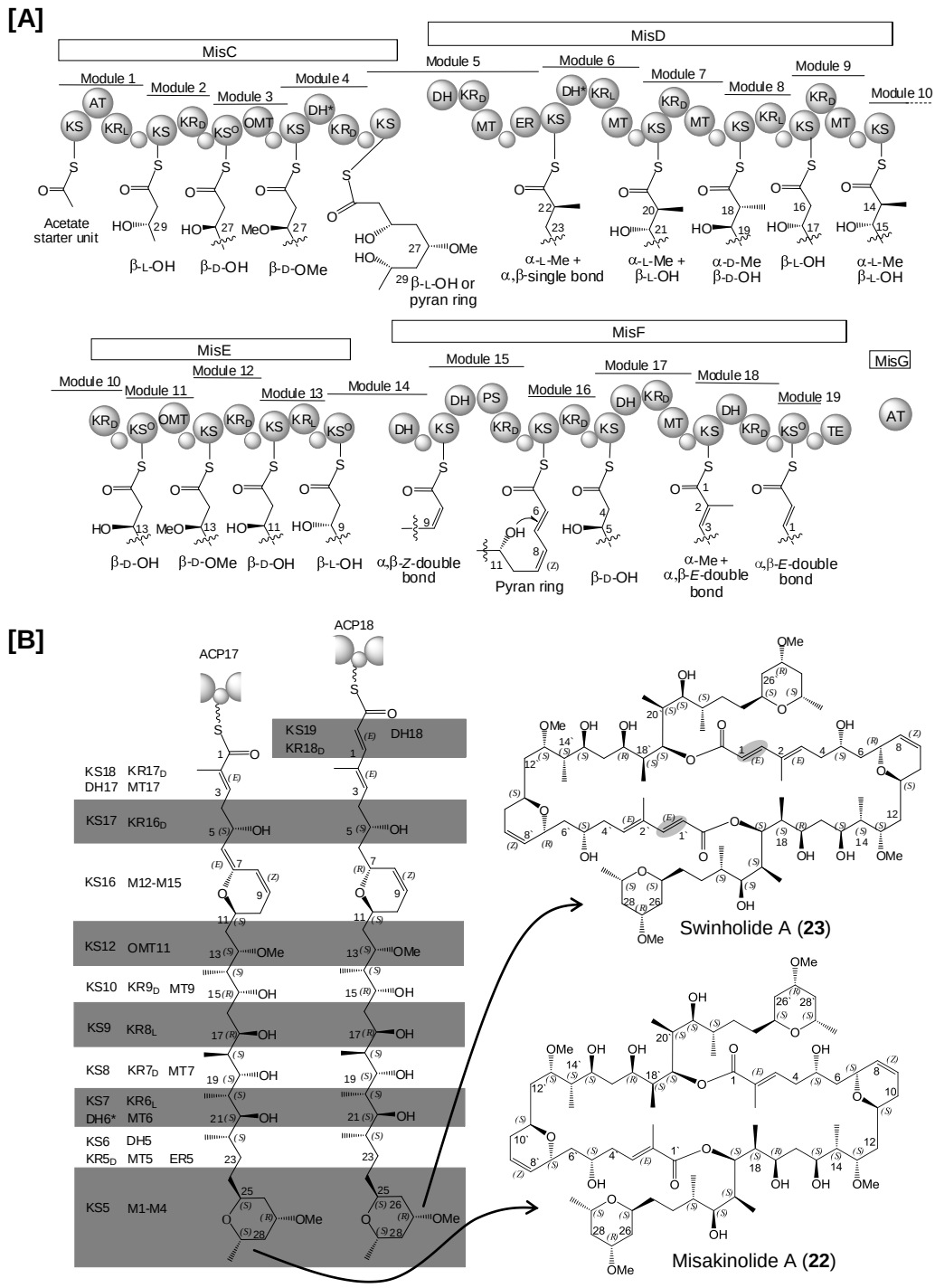


Fig. 3.2.13 Bioinformatic analysis of the isolated entire PKS gene cluster. **[A]** The module architecture and the substrate specificity and configuration predicted based on phylogeny-based KS analysis and KR analysis. **[B]** The correlation between the predicted substrates accepted by KS domains and the structure of misakinolide A and swinholide A.

A possible explanation for the pyran ring formation is that a *trans*-acting functional DH might access the intermediate tethered on module 6, allowing to generate an α,β-*E*-double bond. The highest homology (73% identical) of *mis* KS5 is the KS9 in the batumin pathway that is predicted to give rise to a *trans* double bond (Mattheus

et al., 2010), suggesting the double bond might occur during the assembly line. Subsequent in *trans*-action of a PS domain allows Michael addition of the hydroxyl group to an unsaturated thioester, leading to the formation of a pyran ring, as shown at the C_{25}-C_{29} of misakinolide A structure. This pyran formation mechanism was previously proposed by Piel (2002).

Based on the phylogenetic analysis, two *mis* KSs (KS4 and KS12) were predicted to accept β-D-OMe functions. The predicted specificity of these two KSs is in consistency with the architecture of the upstream modules 3 and 11 (KS^{o}-OMT-ACP) as well as the OMe groups at the C_{27}/$C_{27`}$ and $C_{13}$/$C_{13`}$ of the misakinolide A structure. Module 5 was predicted to install a substrate bearing α-methyl and α,β-reduced moieties. KS and KR analyses indicated the D configuration for the α-methyl group. Three *mis* modules (modules 6, 7, and 9) contain domains required to form the β-hydroxylated intermediate bearing an α-methyl group. KR analysis indicated the L-configuration of the β-OH functionality installed on module 6, whereas the D-configuration of that installed on modules 7 and 9. Phylogenetic analysis of the KSs upstream of the modules provided a clear suggestion that module 6 installs α-D-methyl and β-L-hydroxyl function whereas modules 7 and 9 incorporate α-D-methyl and β-D-hydroxyl function. Three *mis* modules (modules 14, 17 and 18) were assumed to install α,β-double bond functionalities based on the module architecture. The double bond functionality bound on the ACP14 is derived from the DH14-catalyzed conversion of β-L-OH group into the *cis* olefinic moiety. The β-L-OH group is generated through the β-ketoreduction activity performed by KR13 that has the stereospecific motif for L configuration. The double bond functionalities in modules 17 and 18 are predicted to be *trans*, because the cognate KRs contain the D configuration motifs. The presence of a MT corresponds with the α-methyl branch at the substrate installed on module 17. This is supported by the phylogenetic analysis that classified KS15, KS17 and KS19 into the clades of *cis* double bond, α-methylated and *trans* olefinic moiety, and the *trans* double bond, respectively. This is in good agreement with the C_2/$C_{2`}$ α-methyl and the C_2-C_3/$C_{2´}$-$C_{3´}$ olefinic moieties present in the misakinolide A structure. KS16 was predicted to accept a pyran ring functionality. This is in good agreement with the architecture of module 15 that contains all domains required for the formation of a pyran ring. Since KR15 has a key D residue of D configuration (AGLTR**D**), it is expected to produce a β-D-OH group that would subsequently be converted by DH15 to form α,β-*E*-double bond functionality. The presence of PS domain in module 15 would catalyze Michael

addition of the $C_{12}/C_{12`}$ hydroxyl group to the *cis* double bond, forming ether bonds. This is in consistent with the pyran rings generated via the $C_7$-$C_{11}/C_{7`}$-$C_{11`}$ ether bonds. The domain architecture, phylogeny-based KS analysis and KR analysis described above revealed that the number and deduced functions of the five PKS type I genes perfectly correlate with the structure of misakinolide A or a related compound. This provides convincing evidence that we have successfully cloned the correct biosynthetic pathway. A PKS type III located on the downstream of the type I PKS is most likely not involved in misakinolide A biosynthesis.

3.2.5 Proposed model of the discovered biosynthetic pathway

Based on our sequence analysis of the entire *trans*-AT PKS gene cluster as described above, we proposed a model of misakinolide A biosynthetic pathway (Fig. 3.2.14). The pathway is catalyzed by giant multienzymes consisting of 5 separate proteins, designated as MisC, MisD, MisE, MisF, and MisG. Based on the phylogenetic analysis, *mis* KS1 belongs to the acetyl-accepting KS clade, suggesting its function as a loading domain that specifically accepts acetate as the starter unit. *Mis* KS1 showed 74% identity to the acetyl-accepting KS in the DszA of disorazole biosynthetic pathway, which was proposed to generate an acetate starter through the “backloading” process (Carvalho *et al*., 2005). The “backloading” process proposed for *mis* module 1 begins with loading a malonate unit by a separate *trans*-AT (MisG), followed by subsequent KS1-catalyzed decarboxylative condensation of the loaded malonate to generate an acetate starter (**24a**) onto the KS1 of MisC. Subsequent KS1-catalyzed condensation of the activated acetate starter with the first ACP1-loaded malonate would produce the β-keto functionality. Subsequently the β-keto functionality would be reduced by KR1, thereby resulting in a β-L-hydroxymalonyl-*S*-ACP1 (**24b**) (Carvalho *et al*., 2005; Schneider *et al*., 2007). This is supported by the phylogeny-based KS analysis showing that *mis* KS2 belongs to the KS clade of β-L-OH function. The architecture of module 2 (KS-KR-ACP) reveals that it would elongate the resulting β-L-hydroxylated intermediate (**24b**) and then reduce the β-keto functionality of the elongation product. This begins with the transfer of the **24b** intermediate tethered onto ACP1 to the KS in module 2 (KS2). Following loading a malonyl extender unit by the *trans*-AT (MisG) onto ACP2, KS2 would elongate **24b** by catalyzing its decarboxylative condensation with the ACP2-bound malonyl unit to form an intermediate bearing a β-keto functionality. Because *mis* KR2 has the stereospecific motif for D configuration

(AGVIS**D**ALVFNKTHHQT), it would reduce the β-keto functionality to generate a β-D-hydroxylated intermediate (**24c**).This is consistent with the phylogeny-based analysis of *mis* KS3 showing its predicted substrate specificity towards a β-D-OH functionality.

Mis module 3 contains a non-elongating KS, and therefore it would not elongate the intermediate **24c** accepted from the upstream module, leaving D-OH group at the β-position. However, the presence of OMT3 would methylate the β-hydroxyl group of **24c**, resulting in an intermediate with the β-D-oxymethyl moiety (**24d**). This is supported by phylogenetic analysis of *mis* KS4 showing its predicted specificity towards the α-D-OMe functionality. A homologoue of KS4 is present in the KS7 of CorL in corallopyronin biosynthesis, which was proposed to accept a β-methylated intermediate bearing an α,β-double bond (Erol *et al*., 2010). This is in good agreement with the presence of the β-D-OMe moieties found at the C_{27}/$C_{27`}$ of misakinolide A structure. It is assumed that this non-elongating KS would mediate the transfer of **24d** from the ACP3 to the downstream KS4. Subsequent KS4-catalyzed condensation of **24d** with the ACP4-loaded malonyl unit, followed by β-ketoreduction of the elongation product by KR4 (D configuration motif) would lead to the formation of a β-D-hydroxylated intermediate (**24e**). DH4 is predicted to be nonfunctional, and therefore it would not perform β-dehydration, leaving **24e** with the β-D-hydroxyl group (at C_{25}/$C_{25`}$). However, the KS-based phylogenetic anaylsis revealed that the downstream *mis* KS5 belongs to the clade for β-L-OH or D-pyran, suggesting that the hydroxylated intermediate might undergo further β-reductions towards the formation of a pyran ring upon its attachment on the ACP4 or during its transfer to the downstream module. A possible explanation for the pyran formation is that a functional *trans*-acting DH might access the β-L-hydroxylated intermediate tethered on module 4 (**24e**), allowing conversion of the β-L-OH function into an α,β-*E*-double bond. Subsequent in *trans* action by a DH-like domain, termed pyran synthase (PS), either during or post the assembly line would allow Michael addition of the hydroxyl group to an unsaturated thioester, leading to the formation of a pyran ring, as shown at the C_{25}-C_{29} of misakinolide A structure. Pyran formation mechanism was previously proposed by Piel (2002).

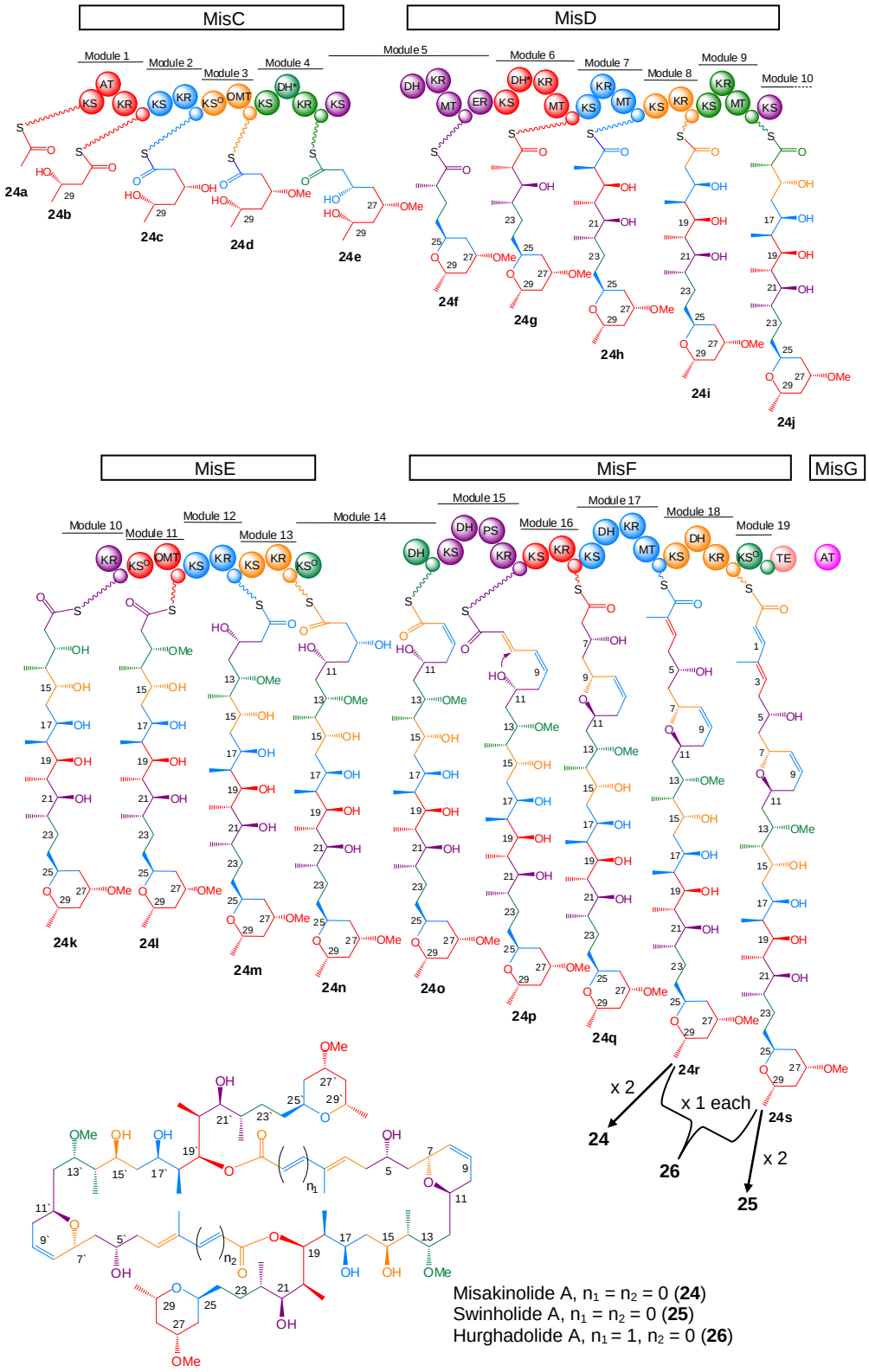


Figure 3.2.14 The proposed biosynthetic pathway for misakinolide A (**24**) isolated from the metagenome of the Japanese sponge *T. swinhoei*. It consists of multifunctional enzymes (MisC-MisG). It contains 15 elongation modules that correspond with the incorporation of 15 malonyl extender units. MisF contains module 18 that presumably facilitates the formation of swinholide A (**25**) and hurghadolide A (**26**).

Mis Module 5 (KS-DH-KR-MT-ACP-ER) is split in two separate proteins (MisC and MisD), in which KS5 is present on MisC whereas the remaining module 5 domains are on MisD. KS5 would accept and elongate the β-hydroxylated intermediate (**24e)** through decarboxylative condensation of it with the ACP5-loaded malonyl extender unit. The elongation product would be passed to the reductive activities for β-processing. At first, KR5 harboring a stereospecific motif for D configuration (AGLTR**D**) would reduce it to produce a β-D-hydroxyl functionality. Then DH5 would convert the β-D-hydroxyl moiety into a *trans* double bond that is further reduced by ER5 to generate an α,β-single bond. Interestingly, the ER domain is positioned after ACP5, which is quite unusual on this *trans*-PKS system. MT5 would add a methyl group at the α position, resulting in an ACP5-tethered intermediate bearing a (2*S*)-methyl branch and an α,β-single bond (**24f**). Phylogenetic analysis predicted that the downstream KS6 accepts the reduced substrate bearing α-L-Me moiety, suggesting the incorporated methyl group has L configuration. This is consistent with the presence of the C_{23}-C_{24} saturated bond as well as the C_{22} L-methyl of misakinolide A structure.

The intermediate **24f** would be transferred to the downstream module 6 (KS-DH*-KR-MT-ACP). KS6 would initially condense it with one decarboxylated malonyl unit loaded onto ACP6. KR6 (VMVAQ**S**) and MT6 would catalyze β-ketoreduction and α-methylation, respectively, resulting in an intermediate bearing α-methyl and β-L-OH groups (**24g**). The β-L-OH function installed on ACP6 would not be reduced further due to the non-functionality of DH6. These intermediate functionalities were supported by phylogentic analysis of KS7 showing that it belongs to the KS clade accepting α-methyl and β-L-OH functions. This is in perfect agreement with the presence of methyl moieties at the C_{19}/$C_{19'}$ as well as L-OH groups at the C_{20}/$C_{20'}$ in the misakinoldie A structure. As in module 6, module 7 would generate an intermediate bearing L-OH and methyl groups (**24h)** through the activities of KR7 and MT7. The intermediate incorporated into module 8 would bear L-OH at the β position (**24i**) due to the presence of KR8 with L stereospecific motif. The KS in module 9 would elongate the intermediate **24i** and its cognate KR and MT domains would subsequently introduce a β-D-hydroxyl group as well as an α-methyl group, thereby generating the intermediate **24j**. These predicted substrate specificities were supported by the phylogenetic anaylsis.

Module 10 contains KS and KR that would extend the intermediate **24j** with the ACP-loaded decarboxylated malonyl unit and subsequently reduce the β-keto moiety into β-D-OH group, thereby resulting in the intermediate **24k**. Module 11 contains a nonfunctional KS, and therefore it would not be able to elongate **24k**. The presence of its cognate OMT would methylate the β-OH, thereby producing the intermediate **24l** that bears an OMe group. This is consistent with the presence of OMe groups at the C_{13}/$C_{13`}$ in the misakinolide A structure, respectively. The presence of two consecutive modules (modules 12 and 13) with the same domain composition (KS-KR-ACP) would elongate the intermediate **24l** with two malonyl units to form the intermediate bearing β-D-OH (C_{11}) and β-L-OH (C_9) (**24n**). Module 14 (KS^o-DH-ACP) is split in two separate proteins (MisE and MisF). This KS^o (KS14 in MisE) forms a biomodule with the upstream module (module 13), which would lead to the formation of a double bond, a similar process reported by Chen *et al*. (2006) in the Difficidin pathway. In the process of generating a double bond, KS14 would initially accept the β-L-hydroxylated intermediate (**24n)**, but would not elongate it due to the absence of the catalytic HGTGT motif. When the intermediate **24n** is tethered at ACP14, the cognate DH in MisF would reduce the β-L-hydroxyl moiety into the α,β-*Z*-double bond to form an intermediate containing α,β-*Z*-olefinic moiety (**24o**). The KS13 is 83% identical with the KS6 of onnamide A synthase (Piel *et al*., 2004b), and therefore it was also proposed that KS13 would accept a β-hydroxylated intermediate from module 12. During the assembly line, the β-OH moiety would form an ether bond catalyzed by a PS domain, resulting in the formation of a pyran ring. This is in consistent with the presence of a pyran ring at the C_7-C_{11} and $C_{7`}$-$C_{11`}$ of the misakinolide A structure. The KS14 shares 79% identity to the KS3 of Bacillene synthetase (Chen *et al*., 2006), and therefore it is assumed to be specific for a β-hydroxylated intermediate. Both *mis* KS14 and *bae* KS3 are non-elongating KSs. Because module 14 contains an additional DH domain, the β-OH moiety would be converted into a double bond that is observed at the C_8-C_9 and the $C_{8`}$-$C_{9`}$ of the midakinolide A structure.

Interestingly, module 15 contains a repeated-like DH domain, in which the second copy, annotated as a pyran synthase (PS) domain, is proposed to be responsible for generating the tetrahydropyran ring (Fig. 3.2.15). KS15 would initially accept the intermediate **24o** that contains an α,β-*Z*-double bond (C_8-C_9) and subsequently catalyze decarboxylative condensation between **24o** and the downstream malonyl-*S*-ACP substrate to generate a β-keto-*S*-ACP intermediate (**24o-1**). Further β-

ketoreduction by KR15 (AGLTR**D**) would generate a β-D-hydroxylated intermediate (**24o-2**). The cognate DH would dehydrate the C_9 β-D-OH group of **24o-2** into an α,β-*E*-olefinic moiety (C_6-C_7), resulting in **24o-3**. Subsequent Michael addition of the hydroxyl group (C_{11}) to the unsaturated thioester would lead to closure of the pyran ring (C_7-C_{11}) (**24p**) (Fig. 3.2.15). A similar mechanism is previously proposed for the biosynthetic pathways of pederin (Piel, 2002) and bryostatin (Sudek *et al*., 2007).

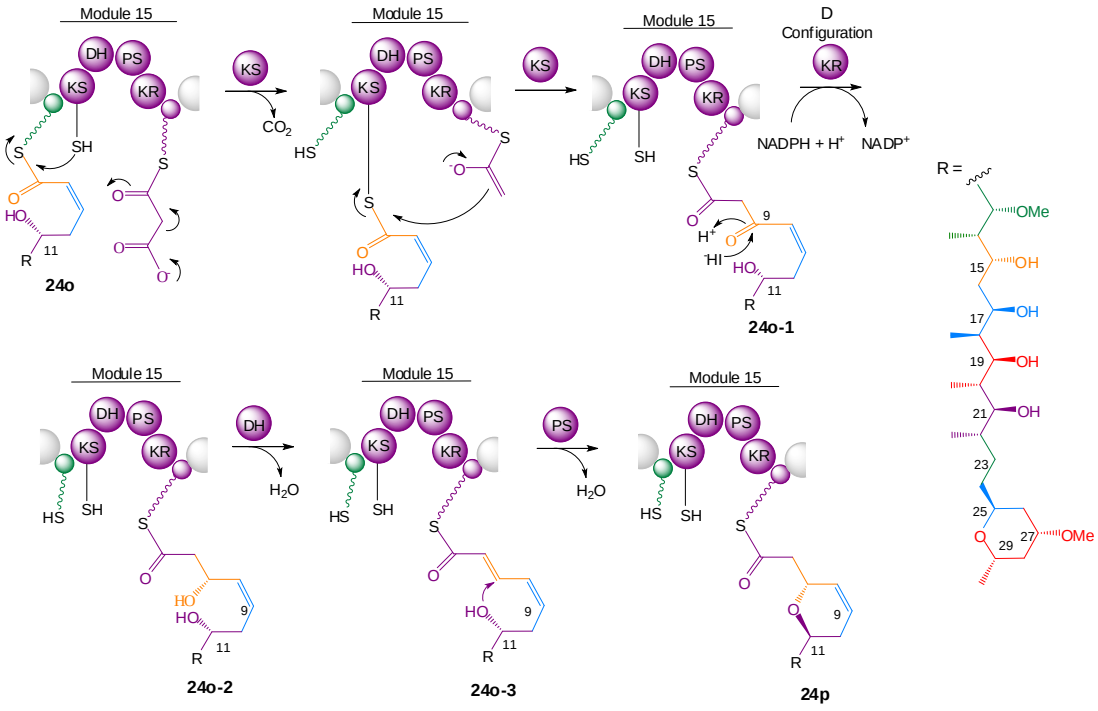


Figure 3.2.15 Formation of the intermediate bearing a pyran ring in module 15. The upstream *cis*-2,3-enoyl-*S*-ACP intermediate (**24o**) would be transferred to KS15. KS5 would catalyze the decarboxylative condensation between **24o** and downstream malonyl-*S*-ACP to generate a β-keto-*S*-ACP (**24o-1**). This elongation product would undergo β-ketoreduction catalyzed by DH15 to form an intermediate bearing β-D-OH moiety (**24o-3**). Subsequent PS-catalyzed Michael addition of the C_{11} hydroxyl group to the C_7 olefinic moiety (**24o-3**) would lead to the formation of an intermediate bearing the C_7-C_{11} pyran ring (**24p**).

The KS in module 16 would accept the intermediate **24p** from the upstream module 15 and catalyze the subsequent decarboxylative condensation between **24p** and the ACP15-bound malonyl extender unit incorporated by a *trans*-acting AT domain (MisG). Subsequent β-ketoreduction by KR16 would generate an intermediate bearing β-hydroxyl in the D configuration (**24q**) (Fig. 3.2.14). Module 17 contain three optional tailoring domains (KR, DH, MT), and therefore the elongation product tethered on the ACP17 would be modified by the additional domains. First, KR17 would reduce the β-ketone moiety to from a β-D-hydroxyl moiety. Subsequent β-dehydration at the C_5 and α-methylation at the C_4 would generate a methylated intermediate with α,β-*Z*-olefinic moiety (**24r**). This intermediate **24r** is an ACP-bound

monomeric subunit of misakinolide A made from one round of biosynthesis. Interestingly, this pathway contains an additional module (module 18) downstream module 17. This module (KS-KR-DH-ACP) is predicted to elongate **24r** with the decarboxylated malonate-*S*-ACP and modify the β-functionality into a *trans* double bond, resulting in an ACP-bound monomer of swinholide A (**24s**) (Fig. 3.2.16).

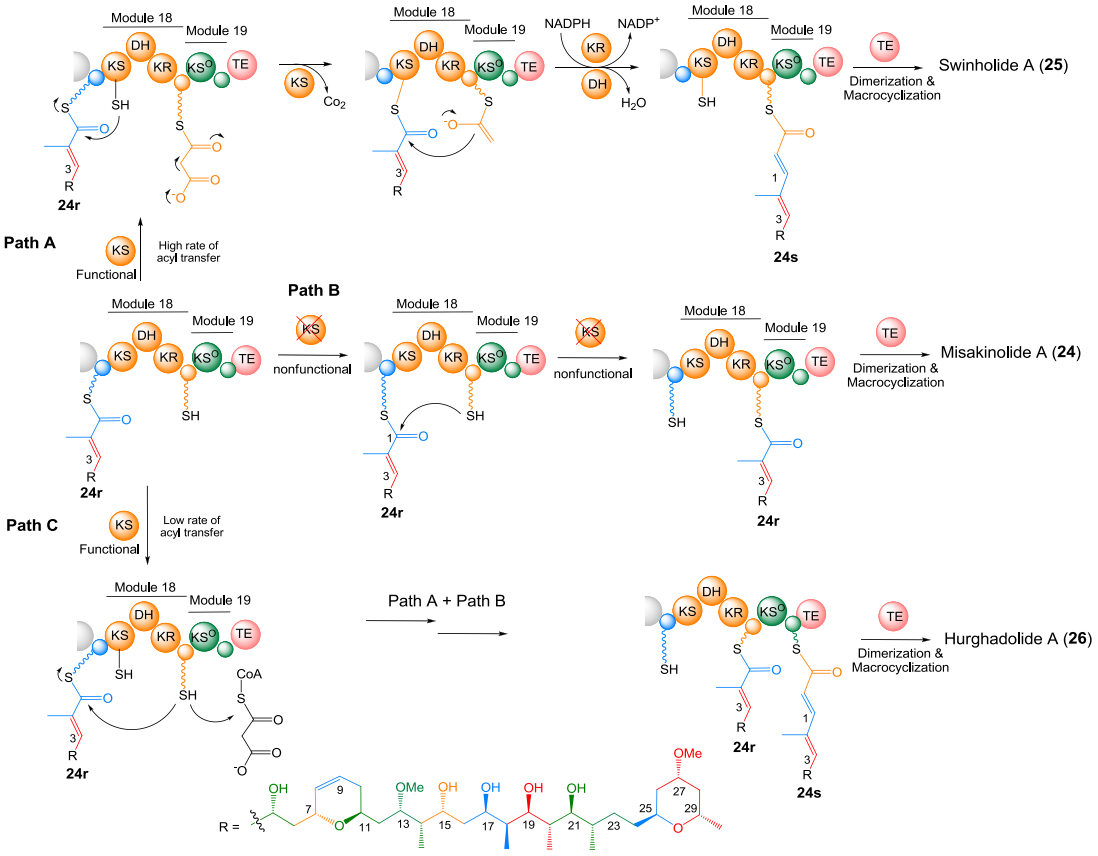


Figure 3.2.16 Proposed role of KS18 in the formation of swinholide A, misakinolide A, and hurghadolide A. [**A**] If KS18 is functional, the misakinolide A monomer bound onto ACP17 (**24r**) would be transferred to KS18 and subsequently undergo elongation with the decarboxylated malonyl-*S*-ACP18. Subsequent β-processing by KR18 and DH18 would lead to the formation of the ACP18-bound swinholide A monomer (**24s**). [**B**] However if KS18 is nonfunctional as proposed in this work, it would not accept the misakinolide A monomer bound on ACP17. Instead KS18 might play in facilitating the chain transfer from ACP17 to ACP18 through a mechanism proposed by Thomas *et al* (2002). The KS18 active site is assumed to recognize and dock the prosthetic groups of both ACPs simultaneously, leading to the transfer of **24r** to ACP18. [**C**] This ACP-to ACP transfer could be enhanced if the priming rate of ACP18 is slowed, thereby increasing the chance for **24r** to be primed onto ACP18. At the point where ACP-to-ACP transfer is controlled so that ACP18 could be primed by malonyl unit in one side and by **24r** in another side might lead to the formation of ACP19-bound swinholide A monomer as well as ACP18-bound misakinolide A monomer. TE-catalyzed dimerization and macrocyclization of these two monomers could form hurghadolide A.

Since this pathway contains an additional module that would match to further elongated structures such as swinholide A or hurghadolide A rather than misakinolide A, an interesting issue is about how misakinolide A would be produced

on this pathway template. A possible explanation is that the KS in this additional module (KS18) has lost its function and therefore no elongation takes place although there is no sequence based evidence for nonfunctionality. We assume that the reduced or lost function of this KS might lead to the transfer of a complete monomeric subunit of misakinolide A (**24r**) from ACP17 to ACP18 through a mechanism resembling a module skipping as described by Leadlay group for a hybrid polyketide synthase (Thomas *et al*., 2002) (Fig. 3.2.16). Therefore KS18 might play a key role in controlling the type of structures produced by this pathway, whether swinholide A (**25**), misakinolide A (**24**), or hurghadolide A (**26**).

Module 19 (KS^{o}-ACP) is presumably unable to perform chain elongation due to the presence of a non-elongating KS^{o} domain, raising a question regarding the role of this non-elongating module in this *trans*-AT PKS system. The importance of non-elongating modules was investigated by altering the active serine residue of the KS^{o} in a mupirocin module to an alanine as reported by Thomas and coworkers (El-Sayed *et al*., 2003), in which they found this residue alteration abolished mupirocin production (El-Sayed *et al*., 2003). Subsequently, the ACPs downstream of the KS^{o} in the misakinolide A synthase are assumed to be catalytic functional due the presence of the catalytic conserved GxDS motif, in which the active site serine (S) is essential for the attachment of the 4`-phosphopantetheine (Lambalot *et al*., 1996; Findlow *et al*., 2003). Other studies by Müller and colleagues revealed that ACPs in non-elongating modules from rhizopodin biosynthesis can be primed with malonyl-CoA (Pistorius & Muller, 2012). This suggests the important role of non-elongating modules (KS^{o}-ACP) in the biosynthetic assembly line. Based on this information, the three non-elongating modules in the misakinolide A biosynthesis (*mis* modules 11, 14, and 19) may play role in facilitating the transfer of the growing polyketide chain to the next module, as previously proposed by Julien and coworkers for disorazole biosynthesis (Carvalho *et al*., 2005). The possible mechanism of chain transfer by such three non-elongating modules are that the KS^{o} initially accepts the polyketide chain from upstream ACPs and subsequently transfers it to the next module, as proposed by Müller and colleagues for the biosynthesis of myxochromide (Wenzel *et al*., 2006) and rhizopodin (Pistorius & Muller, 2012).

At the downstream end of the misakinolide assembly line, a TE domain on the C-terminus of MisF is predicted to act as a chain termination catalyst. The general process of TE-catalyzed cyclic polyketide release usually involves transferring the

acyl chain loaded on the terminal ACP to an active site serine hydroxyl of the TE, thereby generating a TE-bound O-acyl moiety. The acyl chain of this intermediate is then usually cyclized to form a cyclic polyketide macrolide (Kohli & Walsh, 2003). However, for a dimeric polyketide such as misakinolide A, the TE presumably plays a role not only in cyclization but also in dimerization, provoking an interesting question about how macrocyclization occurs to release this dimeric polyketide from the assembly line. As described in Fig. 3.2.17, the possible mechanism of misakinolide A release might resemble the TE-catalyzed dimerisation demonstrated *in vitro* for thiocoraline (Robbel *et al*., 2009) and gramicidin S through peptide-bond formation (Hoyer *et al*., 2007) as well as the TE-mediated cyclotimerisation for enterobactin through ester-bond formation (Shaw-Reid *et al*., 1999).

Some interesting features typical of *trans*-AT PKSs were observed in the domain architectures in the misakinolide A biosynthetic pathway, such as AT-less modules, three modules split between two distinct proteins, presumed nonfunctional domains, the presence of a duplicated DH domain in module 15. Split modules are located between MisC and MisD (module 5), MisD and MisE (module 10), and MisE and MisF (module 14). In module 14, KS^{o} is located on MisE and DH-ACP is on MisF, forming a bimodule that presumably plays a role in creating a double bond, a similar process reported by Chen *et al*. (2006) in the Difficidin pathway. Another intriguing feature of this pathway is the presence of multiple MTs and *O*-MTs within the PKS. Five MT domains, located in MisD and MisF, are proposed to introduce methyl groups at the α-positions of the extending polyketide through *S*-adenosylmethionine (SAM)-dependent α-methylation. Interestingly, the pathway contains a penultimate module (module 18) that is predicted to facilitate the formation of swinholide A as well as hurghadolide A. It is assumed that the KS domain in this module has lost its function and therefore no elongation takes places. All three compounds are potent actin polymerization inhibitors (Camely *et al*., 1985; Kitagawa *et al*., 1990; Youssef & Mooberry, 2006).

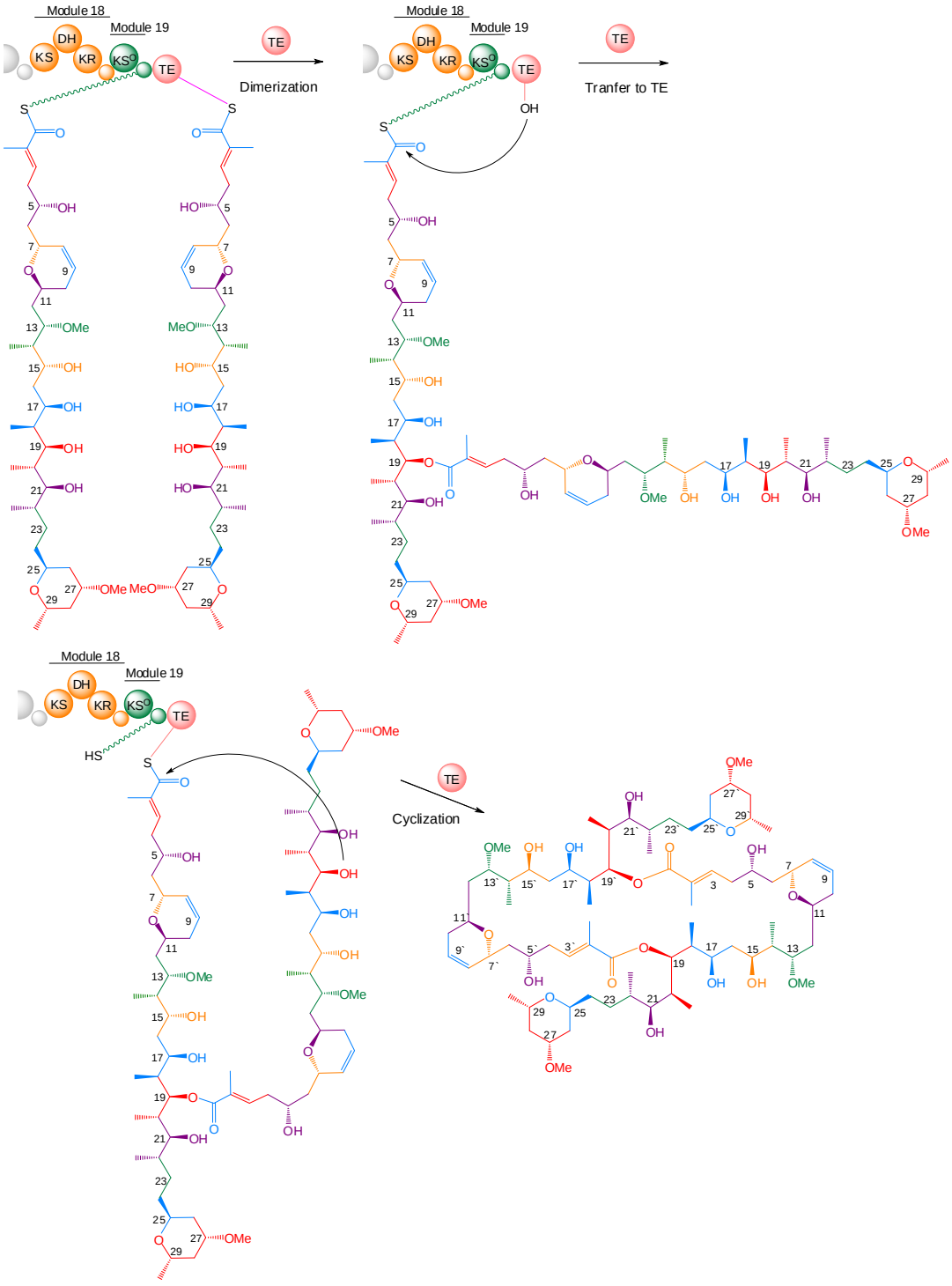


Fig. 3.2.17 Proposed TE-catalyzed release in misakinolide A biosynthesis presumably involves dimerization, in which one polyketide monomer of misakinolide A might bind to the terminal ACP while another monomer is bound to the TE. The two monomers might be ligated by TE, resulting in an acyl-*S*-ACP that is transferred to the TE and cyclized to form the dimeric polyketide.

Misakinolide A is closely related to many compounds that have been reported from remarkably diverse sources, including luminaolide from algae (Kitamura *et al*., 2009) and lobophorolide from seaweed (Kubanek *et al*., 2003). Our data suggest that all of these compounds are made by symbiotic bacteria. Based on a compound localization study, swinholide A was previously found in high concentration in the unicellular bacterial preparation of *T. swinhoei* (Bewley *et al*., 1996; Bewley *et al*., 1998). Gerwick and coworkers found swinholide A in free-living cyanobacteria collected from two different locations (Andrianasolo *et al*., 2005). However, our genetic work on misakinolide A suggests that this seemingly contradictory result might have been caused by export and absorption or sequestering of the polyketide. Some other possible explanations are that (i) misakinolide A is made by a bacterial contaminant present in our Entotheonella fraction, (ii) swinholide A and misakinolide A are made by different bacteria, and (iii) swinholide A moves from the producer to another bacteria (absorption scenario).

3.3 Metagenomics Insight into the Biosynthesis of a Nonribosomal Peptide

As described in the previous section (section 3.1.2), we have successfully cloned two gene regions coding for different serine-specific adenylation domains, which might or might not be involved in the biosynthesis of theonellamide F. Another possibility is their association with other cyclic peptides containing L-serine, orbiculamide and keramamide that were also reported from *T. swinhoei*. This suggests the possible involvement of these genes in the biosynthesis of the two later compounds. To clarify this issue, we needed to isolate the NRPS pathway harboring such serine-specific A domains. This was carried out by screening of the fosmid library generated from *T. swinhoei* metagenome followed by analysis of the cloned NRPS gene cluster.

3.3.1 Isolation of a non-ribosomal biosynthetic pathway

Using the sequence information of one of the genes, we designed two specific primers and employed them to screen the complex metagenomic library (~350,000 clones) constructed from *T. swinhoei* chemotype W. This screening effort finally led to the isolation of two potential 40 kb fosmid clones (2G4 and ADU2a) (Fig 3.3.1).

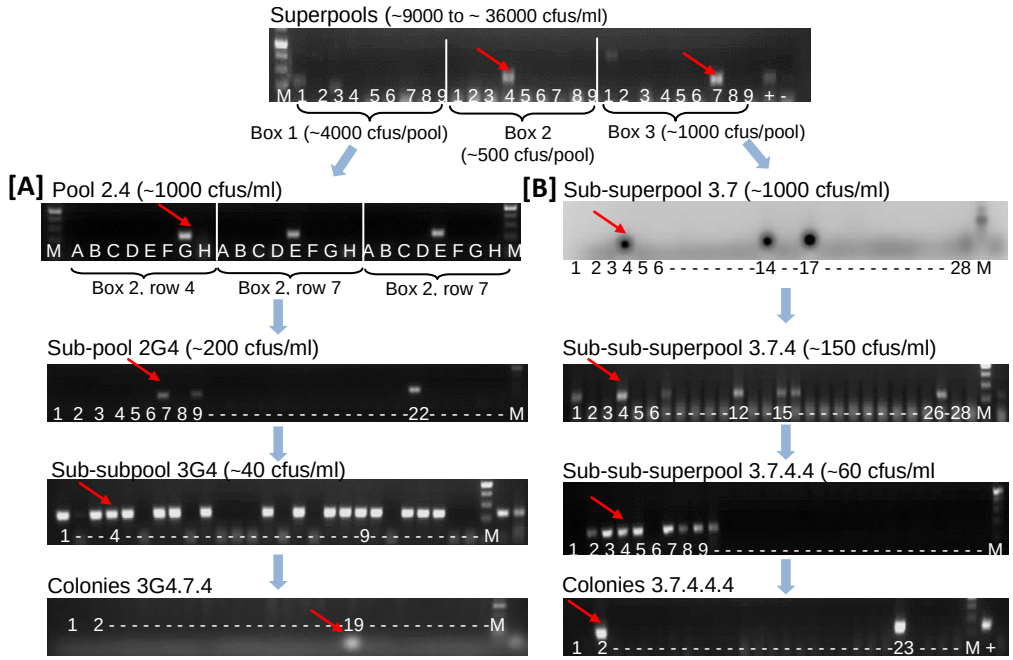


Fig. 3.3.1 Screening of a complex library constructed from *T. swinhoei* to isolate a NRPS gene cluster. The library was disributed into pools arrayed in three boxes. To simplify the screening, pools from the same columns were combined into superpools. Using the primers designed based on the serine-specific adenylation domains, two positive superpools were detected as indicated by red arrows. **[A]** Subsequently, the pools from the corresponding superpool were screened. The positive pool (red arrow) was diluted and screened further. When the clone density was ~40 cfus/ml, colonies were plated and then screened by colony PCR, leading to the isolation of a single positive clone. **[B]** Another superpool was screened in a different way. Instead of checking individual pools, this positive superpool (~9000 cfus/ml) was diluted to ~1000 cfus/ml. Several dilutions were done and whole cell PCR screening of each dilution step was conducted. When the clone density was ~60 cfus/ml, clones were plated and screened, which led to the identification of two positive colonies. One of the colonies (ADU2a) was chosen for further analysis.

We analyzed the two fosmid clones by end-sequencing and subsequent subcloning in order to obtain more sequence information. The end-sequences of 2G4 showed high similarity with condensation domains and hypothetical proteins, respectively. The end sequences of ADU2a (supplementary S3.3.1) were highly similar to many PP-binding domains of PCPs and condensation domains, respectively, suggesting that the entire clone might be a part of a giant NRPS gene cluster. Subcloning was conducted by digestion of both clones with *Sal*I, followed by gel extraction and cloning of the individual *Sal*I-fragments (Fig. 3.3.2).

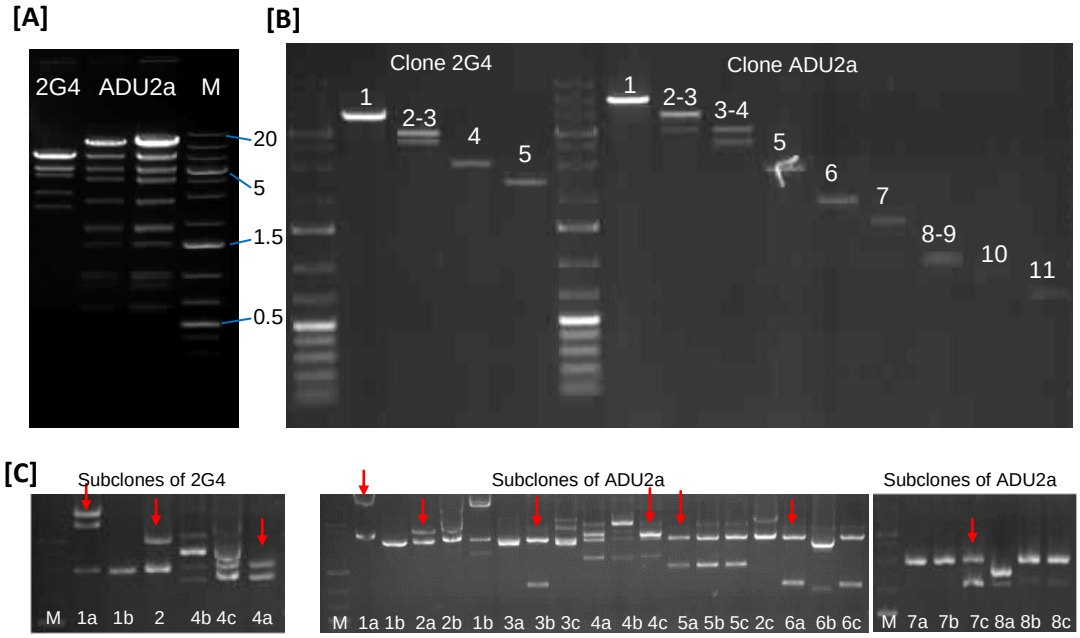


Fig. 3.3.2 Subcloning of two potential fosmid clones. **[A]** Each clone was digested with a variety of restriction enzymes with recognition sites are present in the multicloning site of the vector pBlueScript II SK(-). **[B]** *Sal*I restriction gave well-separated fragments in the agarose gel, and therefore it was used to generate *Sal*I-fragments in larger scale. Individual *Sal*I-fragments were separated on the gel and purified. **[C]** The purified individual *Sal*I-fragments were cloned into the *Sal*I-linearized pBlueScript II SK(-). Some clones were selected and checked for the presence of the correct-sized inserts. Clones harboring individual *Sal*I-fragments (shown with red arrows) were sequenced at both ends.

The clones with correct-sized inserts were subjected to end-sequencing. The protein sequences of the 2G4 subclones containing a 7.5 kb fragment exhibited homology to a kinase-associated protein phosphatase (KAPP)-loop domain protein and methytransferase. The end sequences of the subcloned 5.0 kb fragment encoded proteins that were similar to a hypothetical protein and a MaoC domain protein dehydratase, whereas the other fragment (~3.4 kb) coded for a protein that showed similarity to 1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR) carboxylases (Table 3.3.1).

Table 3.3.1 Subcloning of the fosmid 2G4

Subclone 2G4		BlastX analysis of end-sequences	
Fragment (code)	Size kb	Primer T3	Primer T7
1 *(2G4-1a)*	~7.5	KAP P-loop domain protein & Methytransferase	-
2-3 *(2G4-2)*	~5.0	hypothetical protein MaoC domain protein dehydratase	MOSC domain-containing protein
4 *(2G4-4a)*	~3.4	1-(5-phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR) carboxylase	-

The clone pADU2a consisted of 11 *Sal*I-fragments, of which 10 were subcloned and end-sequenced. End-sequence results (Table 3.3.2) revealed that some fragments

shared homology to sequences coding for typical NRPS components, such as C domains, PCPs, and TEs (supplementary S3.3.2). Interestingly, one of the end sequences of a ~15 kb subclone showed high similarity to McbC-like oxidoreductases. These enzymes are responsible for oxidizing a cysteine-derived thiazoline or serine-derived oxazoline into a thiazole or oxazole moiety, respectively (Schneider *et al*., 2003, Schmidt *et al.*, 2005, Engelhardt *et al*., 2010). Another interesting feature is the end sequence of the ~2.8-kb fragment that is similar to many tryptophan halogenases. Because orbiculamides and keramamide H from the Japanese sponge *T. swinhoei* are known to contain halogenated tryptophan and oxazole/thiazole residues (see structures in Section 3.1.2, Fig. 3.1.10), the sequence data are in agreement with the structures and motivated us to perform a more detailed analysis.

Table 3.3.2 Sub-clones of the fosmid ADU2a

Sub-clone ADU2a		BlastX analysis of end-sequence	
Fragment (code)	Size kb	Primer T3	Primer T7
1 *(ADU2a-1a)*	~15.0	Condensation domain	mcbC-like oxidoreductase, Acyl-protein synthetase, PP-binding
2-3 *(ADU2a-2.1a)*	~8.5	Vector backbone	-
3-4 *(ADU2a-2.2a)*	~5.0	69% GMC oxidoreductase 54% NAD-dependent epimerase/dehydratase	NADH:flavin oxidoreductase/NADH oxidase
5 *(ADU2a-3b)*	~2.8	Tryptophan-5-halogenase	Tryptophan-5-halogenase
6 *(ADU2a-4c)*	~1.8	Galactose oxidase	Galactose oxidase
7 *(ADU2a-5a)*	~1.5	Peptidyl carrier protein, Thioesterase	Peptidyl carrier protein, Thioesterase
8-9 *(ADU2a-6a)*	~1.0	Condensation and reductase	Condensation and reductase
10 *(ADU-2a-7c)*	~0.8	Mostly vector part	-

Based on the sub-cloning results, the clone pADU2a looked very promising and therefore we decided to sequence the isolated fosmid entirely. For the purpose of complete sequencing, the fosmid DNA was initially treated with ATP-dependent DNase provided by Epicentre (http://www.epibio.com/) in order to remove the linear host chromosomal DNA. This enzyme has no activity on nicked or closed-circular dsDNA, thereby leaving circular fosmids intact. The vector backbone was then removed by treating it with the rare-cutting enzyme *Not*I. The resulting NRPS insert was completely sequenced through a *de novo* sequencing approach in collaboration with Dr. Vladimir Benes at EMBL-Heidelberg. Assembling of the sequencing reads resulted in 7 contigs: 7.5 kb, 14 kb, 5.6 kb, 1.7 kb, 1.1 kb, 2.9 kb, and 1 kb, respectively. Subsequently, we ordered the contigs and filled in the gaps between them by PCR (Fig. 3.3.3). We initially obtained amplicons that close the gaps between contigs 1 & 2, contigs 2 & 3, and contig 6 & 7. Filling these three gaps involved Mr. B. Pandey, a master student who was doing a 6-month practical work

under my supervision in the Piel Group. Subsequently, we were able to obtain amplicons from the gaps between contigs 3 and 4 as well as between contigs 5 & 6.

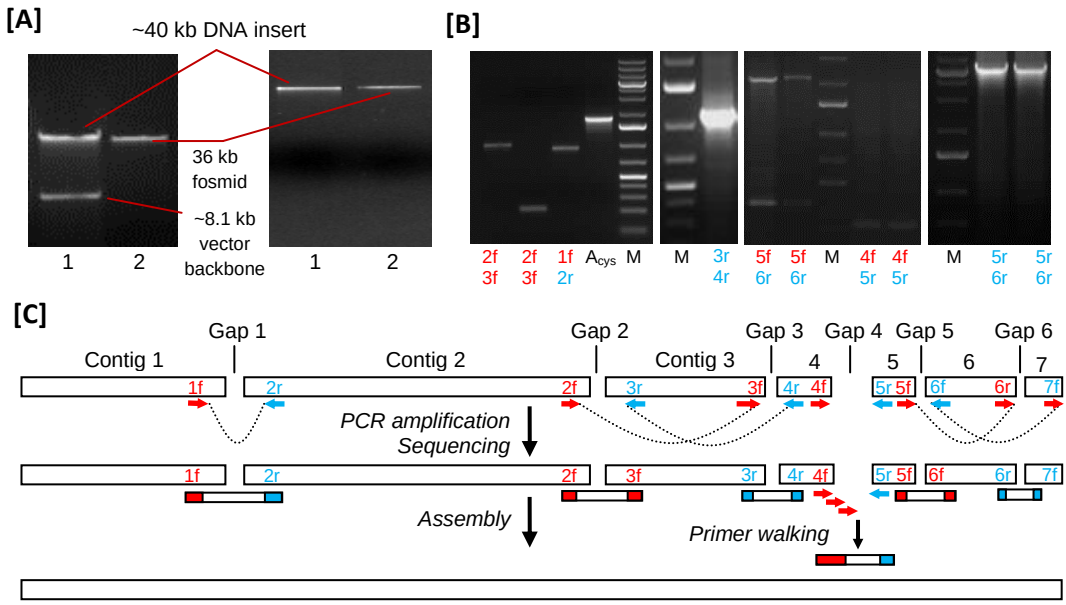


Fig. 3.3.3 Preparing the fosmid pADU2a for complete sequencing and filling the sequence gaps to generate a continuous sequence. **[A]** The circular fosmid extracted from a recombinant *E. coli* clone was treated with ATP-dependent DNase to remove the host chromosomal DNA. Then the vector backbone was removed by digestion with the rare-cutting enzyme *Not*I. On the gel 2% (left side): lane 1 is ADU2a sample (10 ng), and lane 2 is 36-kb linear fosmid control (10 ng). On the gel 1% (right side)): lane 1 is ADU2a sample (5 ng), and lane 2 is 36-kb linear fosmid control (10 ng). **[B]** Complete sequencing of pADU2a and subsequent sequence assembly resulted in 7 contigs. To generate a continuous sequence, we filled in the gaps between the contigs by PCR and sequencing. To fill in the gaps, we initially designed specific primers that bind to the ends of individual contigs. Using these primers with ADU2a DNA as the PCR template, 5 gaps between the contigs were closed, leaving one gap (gap 4 between contigs 4 and 5). Using primer walking, this remaining gap was finally filled in. **[C]** A scheme showing the strategy we used to close the gap between contigs. M is DNA marker. Primer pairs used for gap filling are 1f/2r, 2f/3f, 3r/4r, 4f/5r, 5f/6r, and 6f/7f. A_{cys} is DNA sequence coding for cystein-specific adenylation domain.

However, due to the large gap between contigs 4 and 5, we did not obtain any PCR product. Therefore, we applied primer walking to obtain more sequence data that cover the whole gap between contigs 4 and 5. For primer walking, sequence data obtained using a single primer was used as the basis for designing a new primer to sequence downstream regions. We reassembled all of the sequences derived from both contigs and gaps to generate a 40.9 kb continuous sequence.

3.3.2 Analysis of the non-ribosomal synthase region

The entire nucleotide sequence was translated into amino acid sequence. Open reading frames (ORFs) or potential genes were identified and annotated to predict the functions (supplementary 3.3.3). The annotation results of the sequenced clone

revealed that the 40.9 kb insert of the isolated clone pADU2a harbors 20 complete *orfs*. The genetic map, as shown in Fig. 3.3.4, shows that five *orfs* encode proteins homologous to NRPSs, one *orf* for a PKS, and five *orfs* for tryptophan-5-halogenase, cytochrome C, galactose oxidase, a transcriptional regulator, tryptophan 2,3-dioxygenase, an oxidoreductase, and a monooxygenase. The detailed annotation results were presented in Table 3.3.3 below.

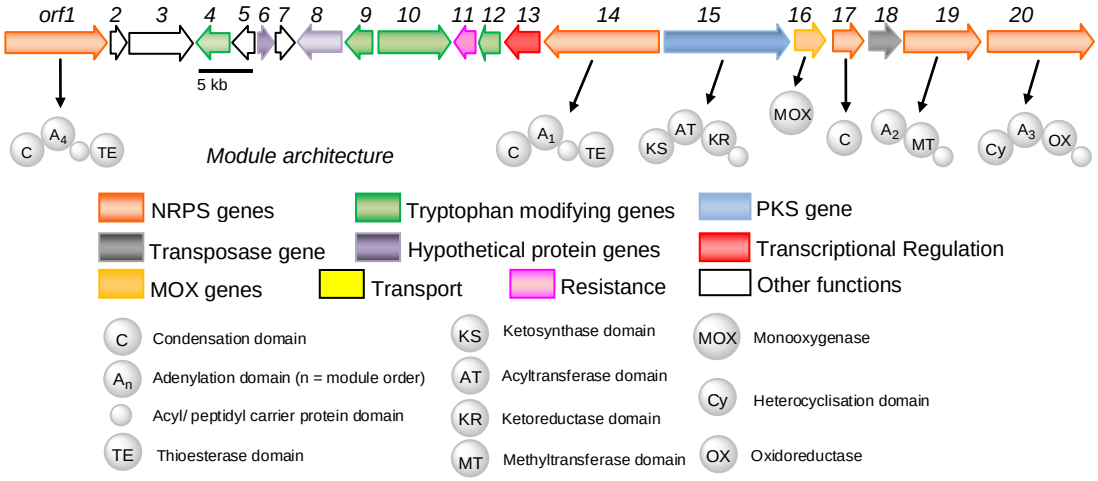


Fig. 3.3.4 Genetic map of the completely sequenced fosmid pADU2a, showing a mixed PKS/NRPS gene cluster and its module architecture.

Interestingly, there are two TE domains encoded on the clone, which are present in different proteins. One TE domain is located in the C-A-T-TE module encoded by ORF1, in which the A domain is predicted to be specific for homopropargylglycine (HPG) (Table 3.3.4). The other TE domain is present in the NRPS module encoded by ORF14, in which the A domain is proposed to have specificity for phenylalanine (Table 3.3.4). Due to the presence of some proteins that are not involved in the NRPS biosynthesis between these two TE-containing modules such as galactose oxidase, it is most likely that both TE-domains are individually involved in two separate NRPS pathways. Between the two TE-containing modules there are a series of genes presumably encoding enzymes involved in the bromination and degradation of tryptophan encoded by ORF9, 10, and 12.

A portion of an NRPS/PKS biosynthetic pathway was identified for the proteins encoded by *orfs14*, *15*, *16*, *17*, *19*, and *20* (Fig. 3.3.4). Four NRPS modules indentified in this pathway contain adenylation domains that play a role in substrate activation. Using NRPSpredictor2 (Rausch *et al*., 2005; Röttig *et al*., 2011), we

predicted the substrate specificity code of all adenylation domains present in the individual NRPS modules (Table 3.3.4).

Table 3.3.3 Deduced function of genes in the completely sequenced fosmid pADU2a.

Gene	aa	Protein homolog (BLAST)			I/S (%)	Proposed function
		Homolog	Organism	Acc. No.		
orf1	1272	ApnD	*Planktothrix agardhii* NIVA-CYA	ABV79988	47/65	NRPS, C-A-TE domains
orf2	205	SCO1/SenC	*Mesorhizobium ciceri biovar biserrulae* WSM1271	YP_004142340	32/52	Electron transport protein SCO1/SenC
orf3	806	Galactose oxidase	*Methylobacterium extorquens* DSM 13060	EHP91138	36/48	Galactose oxidase
orf4	469	NADH:flavin oxidoreductase	*Lyngbya majuscula*	ZP_08430361	47/65	NADH:flavin oxidoreductase
orf5	228	Electron transport complex protein RsxA	*Vibrio nigripulchritudo* ATCC 27043	ZP_08735482	29/45	Electron transport complex protein
orf6	202	Hypothetical protein	*α-proteobacterium* BAL199	ZP_02187391	43/58	Hypothetical protein
orf7	153	cytochrome c class I	*Chloroflexus aurantiacus* J-10-fl	YP_001634601	52/64	cytochrome c class I
orf8	215	Hypothetical protein	*Crassostrea gigas*	EKC28949	27/40	Hypothetical protein
orf9	545	FAD dependent oxidoreductase	*Burkholderia graminis* C4D1M	ZP_02882432	28/43	FAD dependent oxidoreductase
orf10	911	MibH of microbisporicin	*Microbispora corallina*	ADK32563	46/65	tryptophan-5-halogenase
orf11	325	tryptophanyl-tRNA synthetase	*Plesiocystis pacifica* SIR-1	ZP_01909982	28/43	tryptophanyl-tRNA synthetase
orf12	226	tryptophan 2,3-dioxygenase	*Ruegeria* sp. TM1040	YP_614220	26/40	tryptophan 2,3-dioxygenase
orf13	530	IclR family transcriptional regulator	*Rhodococcus erythropolis* PR4	YP_002764363	29/48	Transcriptional regulator
orf14	1371	OciC of Cyanopeptolin	*Planktothrix rubescens* NIVA-CYA 98	CAQ48258	39/57	NRPS, C-A-T-TE domains
orf15	1560	JamP	*Lyngbya majuscula*	AAS98787	50/65	PKS, KS-AT-KR-ACP domains
orf16	406	CtaG	*Paenibacillus mucilaginosus* KNP414	YP_004643249	47/65	Flavin utilizing monoxygenases (MOX)
orf17	418	linear gramicidin synthetase subunit D	*Stigmatella aurantiaca* DW4/3-1	ZP_01467592	45/63	NRPS, C domain
orf18	365	Transposase	*Roseiflexus* sp.	YP_001274820	47/63	Transposase
orf19	992	KtzH	*Kutzneria* sp. 744	ABV56588	44/61	NRPS, A-MT-A-T
orf20	1356	HctF	*Lyngbya majuscula*	AAY42398	53/69	NRPS, C-A-OX-T

Notes: aa, amino acid; Acc. No., Accession number; I/S, identity similarity.

The A domain in ORF15 was predicted to activate phenylalanine (60% match). BLAST analysis revealed that the C domain in ORF14 was most homologous (44% identical, 61% similar) to that in CrpC of criptophycin biosynthetic pathway (Magarvey *et al*., 2006). The CrpC catalyzed the incorporation of *O*-methyl-tyrosine as well as L-phenylanalanine (Magarvey *et al*., 2006), suggesting the flexibility of ORF14 to accept not only L-phenylalanine but also tyrosine and tyrosine analogs. In

addition to that, there might be flexibility of ORF14 towards L-tryptophan, as has been observed in *in vitro* adenylation reactions with domains from the NRPS of tyrocidine (Schaffer & Otten, 2009).

Table 3.3.4 Key amino acid residues for substrate specificity of the adenylation domains of the proposed biosynthetic pathway.

Adenylation domain	Specificity-conferring code of A domain										Predicted Substrate[c] & % match
	Relative residue position	235	236	239	278	299	301	322	330	331	517	
A_4 (ORF1)	Consensus sequence[a]	D	I	F	L	L	G	L	L	C	K	HPG (80%)
	Specificity sequence[b]	D	A	L	H	L	G	L	V	V	K	
A_1 (ORF14)	Consensus sequence[a]	D	A	W	T	I	A	A	V	C	K	Phe (60%)
	Specificity sequence[b]	D	A	L	H	I	G	N	V	A	K	
A_2 (ORF19)	Consensus sequence[a]	D	V	W	H	L	S	L	I	D	K	Ser (90%)
	Specificity sequence[b]	D	V	W	H	I	S	L	I	D	K	
A_3 (ORF20)	Consensus sequence[a]	D	L	Y	N	L	S	L	I	W	K	Cys (80%)
	Specificity sequence[b]	D	L	Y	D	M	S	L	I	W	K	
	Variability (%)[d]	3	16	16	39	52	13	26	23	26	0	

[a] Signature sequences derived from A domains activating the same substrate, as predicted by Stachelhaus *et al* (1999)
[b] The positions of the ten specificity-conferring decisive residues identified in this work, which was predicted using NRPSpredictor2 (Rötting *et al*., 2011; Rausch *et al*., 2005)
[c] The predicted substrates, HPG: 4-hydroxy-phenyl-glycine, Phe: phenylalanine, Ser: serine, Cys:cysteine
[d] Residue variability are shown in percentage and variable constituents within specific-conferring codes are indicated by red color (Stachelhaus *et al*., 1999)

ORF15 is a PKS module that is homologous to JamP in the jamaicamide A pathway (Edwars *et al*., 2004). This protein consists of a set of KS, AT, KR, and ACP domains. The presence of the integrated AT indicates a *cis*-AT type PKS, which is predicted to be specific for loading malonyl-CoA derived from acetate, the same substrate loaded by the AT in JamP (Edwards *et al*., 2004). The malonyl unit would then be transferred to the ACP of the ORF19 (Fig. 3.3.5). The β-keto group of the acyl-*S*-ACP would then be reduced by the KR to generate a β-hydroxy-acyl-*S*-ACP. The stereospecificity of the KR can be predicted by the presence or absence of an LDD motif on the upstream of the conserved motif GVxHxA (Caffrey, 2003, Caffrey, 2005). Due to the absence of this LDD motif, it is assumed that the KR probably generates an L-3-hydroxyacyl moiety.

Downstream of the PKS ORF15, we identified a protein (ORF16) that shared homology to a flavin-utilizing monooxygenase (MOX) in CtaG of cystothiazole A (Feng *et al*., 2005) and in the MelG of melithiazol (Weinig *et al*., 2003) biosynthesis pathways. MOX in CtaG is embedded in the A-domain between the A_4 and A_5 motifs and is responsible for removing the carbon backbone of the ACP-bound amino residue, thereby releasing a terminal amide structure (Feng *et al*., 2005). It is not clear whether ORF16 may serve a similar function as the MOX of CtaG or not, in which the carbon backbone of the ACP-bound serine is removed to release a

terminal amide hydroxy-malonyl structure. Instead, an interesting question araising whether this MOX may be involved in introducing the keto group at the α-position located between the valine and the methylated serine residues. BLAST analysis indicated that ORF16 showed homology (67%) to the MicA of microviridin biosynthetic pathway from *Planktothrix rubescens* NIVA-CYA 98 (Rounge *et al*., 2009) as well as to MOX within NRPS NpnB in Nostophycin biosynthesis (Fewer *et al*., 2011). The NpnB MOX is proposed to oxidize the Ahoa polyketidyl chain following its transfer to the PCP, generating α-hydroxyl moiety (Fewer *et al*., 2011).

There are some possible scenarios for the formation of the α-keto group. ORF16 MOX might perform α-oxidation of the incorporated malonyl unit to form the α-hydroxyl function. Subsequent oxidation of the α-hydroxyl group by another MOX would lead to the formation of the corresponding α-keto group. A similar way of the α-keto group formation was previously reported for rapamycin (Molnár *et al.*, 1996; Aparicio *et al*., 1996) and FK506 (Motamedi *et al*., 1996, 1997, 1998) biosyntheses. Another possibility is the incorporated precursor in the PKS ORF15 is (2*R*)-hydroxymalonyl-CoA instead of malonyl-CoA. This is supported by BLAST analysis of the ORF15 KR showing its homology (53% identity) with KR5 present in the ZmaA of Zwittermicin A biosynthetic pathway that generates β-hydroxyl moiety at the incorporated (2*R*)-hydroxymalonyl-CoA (Kevany *et al*., 2009). The α-hydroxyl moiety in the hydroxymalonyl-*S*-ACP would then be oxidized by the MOX (ORF16) to generate α-keto moiety upon the intermediate transfer to the downstream module. Enzymatic α-oxidation is exemplified by glycolate oxidase (EC 1.1.3.15) that catalyzes the conversion of various 2-hydroxy acids into the corresponding 2-keto acids (Adam *et al*., 1997; Kroutil *et al*., 2004). However, the possible involvement of this MOX domain in α-oxidation awaits functional studies.

Downstream of this PKS module, we identified a NRPS module in two separate proteins (ORF17 and ORF19) that is split by a transposase (ORF18). ORF17 is homologous to a C domain, and ORF 19 contains A, *O*-MT, and PCP domains. The A domain present in ORF19 is predicted to activate a serine (Table 3.3.4). The *O*-MT is embedded in the serine-activating A-domain of ORF19. This domain shares high similarity with the *O*-MT integrated in the module KtzN of the kutzneride biosynthetic pathway, which was proposed to methylate the hydroxyl group of the PCP-bound serine (Fujimori *et al*., 2005). In the assembly line, serine is activated by the A domain in ORF19, which should subsequently be *O*-methylated by the *O*-MT

domain integrated within the A domain. The β-hydroxy-acyl intermediate mentioned above would be transferred from the PKS module (ORF15) to the PCP domain of the downstream module (ORF19) with the help of the C domain (ORF17). Therefore we proposed that this NRPS module is involved in the incorporation and *O*-methylation of serine as well as in the formation of a C-N bond to integrate the β-hydroxy-acyl chain from the upstream PKS module with the amino group of the downstream methylated seryl-*S*-PCP. The methylated seryl intermediate is then presumably transferred to ORF20.

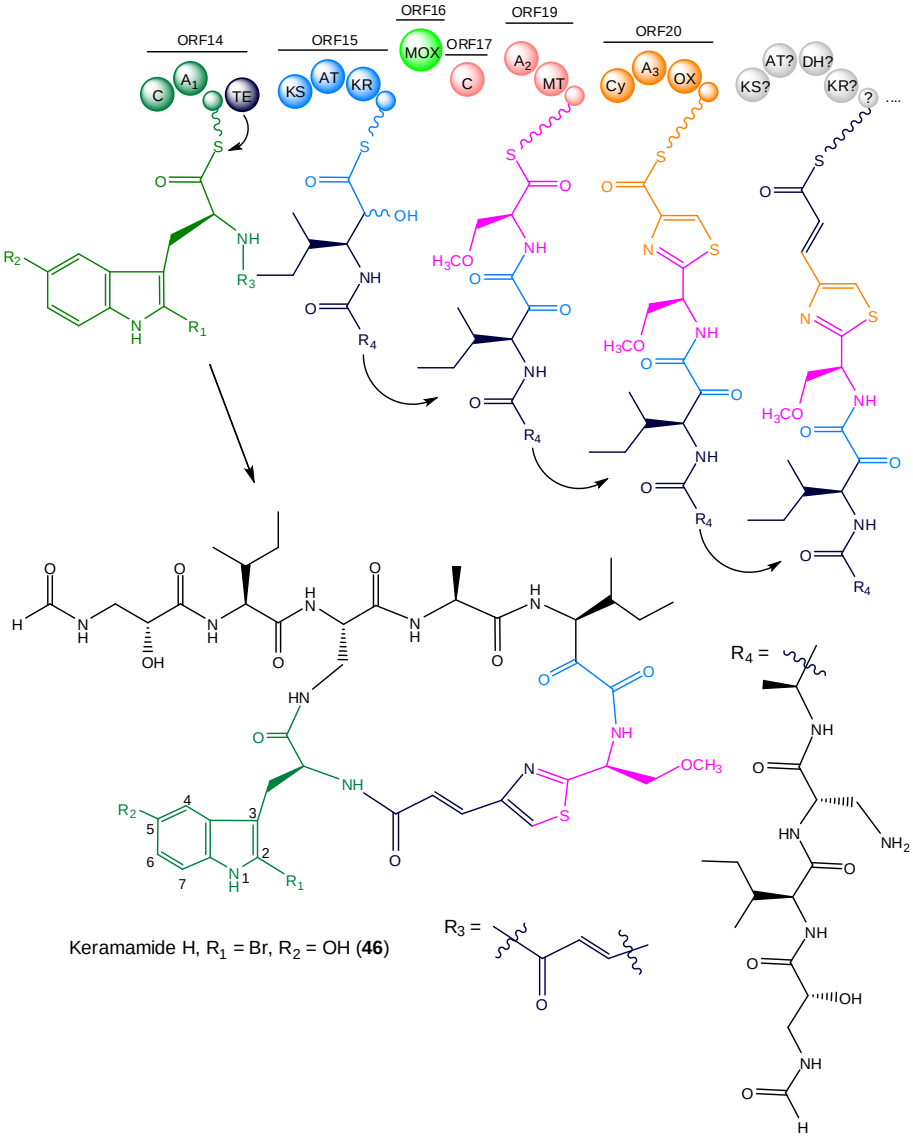


Fig. 3.3.5 A putative PKS/NRPS pathway cloned from the Japanese sponge *T. swinhoei*. It presumably undergoes chain elongation by incorporating malonyl-CoA, L-Ser, and Cys, leading to the formation of half of keramamide series.

ORF20 is highly homologous to HctF of the hectoclorine biosynthetic pathway and contains all the domains usually involved in thiazole ring formation (Ramaswamy *et al*., 2007). In particular, the A-domain of ORF20 is presumably responsible for the activation of cysteine. The incorporated cysteine residue would be cyclized by the Cy domain (Rawling, 2001) (Fig. 3.3.6). The DAxxxDA motif in Cy is assumed to play a role in the heterocyclization of the incorporated cysteine residue (Stachelhaus *et al*., 1998). The FMN-dependent oxidoreductase (OX) in the ORF20 protein is presumably involved in the thiazole formation. A thioesterase domain is located at the C-terminal end of ORF14 and is proposed to catalyze hydrolysis and cyclization of the polyketide from the biosynthetic assembly line (Fig. 3.3.6).

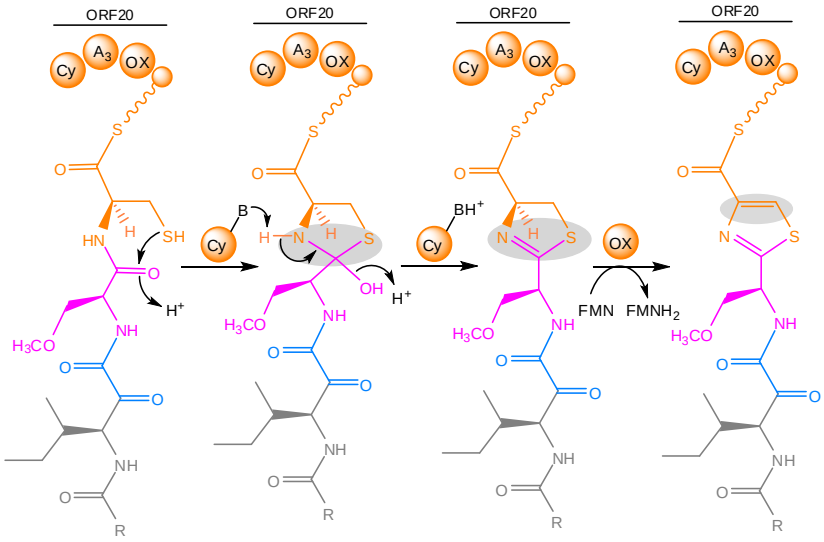


Fig. 3.3.6 The proposed mechanism of thiazole formation in keramamide biosynthesis. This begins with a Cy-catalyzed attack on the serine carbonyl by the cysteine thiol, resulting in the cyclodehydration of the cysteine residue to form a thiazoline ring. The thiazoline would be oxidized by ORF20 oxidoreductase (OX) to generate a thioazole ring.

Upstream of the genes encoding for NRPS and PKS modules, there are a series of genes presumably encoding enzymes involved in the bromination and degradation of tryptophan encoded by *orfs* 9, 10, and 12 (see Table 3.3.3). The protein encoded by ORF10 showed close homology to many tryptophan halogenases, suggesting the role of the protein in catalyzing the halogenation of L-tryptophan. ORF10 is particularly highly similar to the flavin-dependent tryptophan-5-halogenase of microbisporicin (MibH) from the actinomycetes *Microbispora corallina*, which is known to chlorinate tryptophan at C-5 before binding to the NRPS (Foulston and Bibb, 2010). It is difficult to predict whether ORF10 is able to incorporate bromide (Br) or chloride (Cl). The mono-bromination of L-tryptophan at C-5 resulting in 5-bromotryptophan has been shown in the pyrroindomycin B biosynthesis (Zehner *et*

al., 2005). However, since many know cyclic peptides isolated from marine sponges contain 2-bromotryptophan units, such as konbamide (Kobayashi *et al*., 1991), orbiculamide A (Fusetani *et al*., 1991), and various keramamides (Kobayashi *et al*., 1991; Kobayashi *et al*., 1995), it is assumed that ORF10 may prefer to catalyse the bromination of L-Trp at C-2 position.

An interesting question arises about the precise timing of this halogenation in the biosynthetic pathways, whether it occurs before tryptophan-binding to the NRPS, during its binding to the NRPS-ACP, or after NRPS biosynthesis (post-NRPS halogenation). Due to its close similarity to MibH, it is likely that tryptophan is brominated first and then the resulting product is attached to the NRPS assembly line. The $FADH_2$-requirement of the ORF10 product for bromination is suggested by the presence of the N-terminal flavin-binding site motif, GxGxxG, as well as another conserved motif, WxWxIP (Keller *et al*., 2000; Unversucht *et al*., 2005; van Pée and Patallo, 2006). The $FADH_2$ is presumably provided by FAD-dependent oxidoreductase encoded on *orf9* or *orf4* through FAD reduction using NADH. The requirement of $FADH_2$ for halogenating activity was observed by van Pée lab for tryptophan-7-halogenase and monodechloroaminopyrrolnitrin 3-halogenase in pyrrolnitrin biosynthesis (Hohaus *et al*., 1997; Keller *et al*., 2000). Recently Heemstra and Walsh (2008) reported two flavin-dependent halogenase genes and a putative flavin reductase gene encoded together within Kutzneride gene cluster. They observed halogenating activity in the presence of soluble Cl^- ions, FAD, NADH and a recombinant flavin reductase, suggesting that the reduced flavin ($FADH_2$) was required for halogenation (Heemstra & Walsh, 2008).

An interesting question arises about the reaction mechanism for this putative flavin-dependent tryptophan 5-halogenase (ORF10). Some reaction mechanisms have been proposed for flavin-dependent halogenases (Keller *et al*., 2000; Unversucht *et al*., 2005; Dong *et al*., 2005; Yeh *et al*., 2005). All the proposed mechanisms require the reduction of FAD to $FADH_2$ catalyzed by a flavin reductase and subsequent halogenation reaction. The halogenation begins with the reaction of the halogenase-bound $FADH_2$ with O_2, resulting in a flavin hydroperoxide (van Pée and Patallo, 2006). The resulting flavin hydroperoxide would react with a double bond of a tryptophan to form an epoxide ring (Keller *et al*., 2000) or a hypochlorite (HOCl) (Unversucht *et al*., 2005; Dong *et al*., 2005). Subsequent nucleophilic attack of a halide ion and accompanying H_2O elimination would lead to the formation of a

halogenated tryptophan (Keller *et al*., 2000; Unversucht *et al*., 2005; Dong *et al*., 2005).

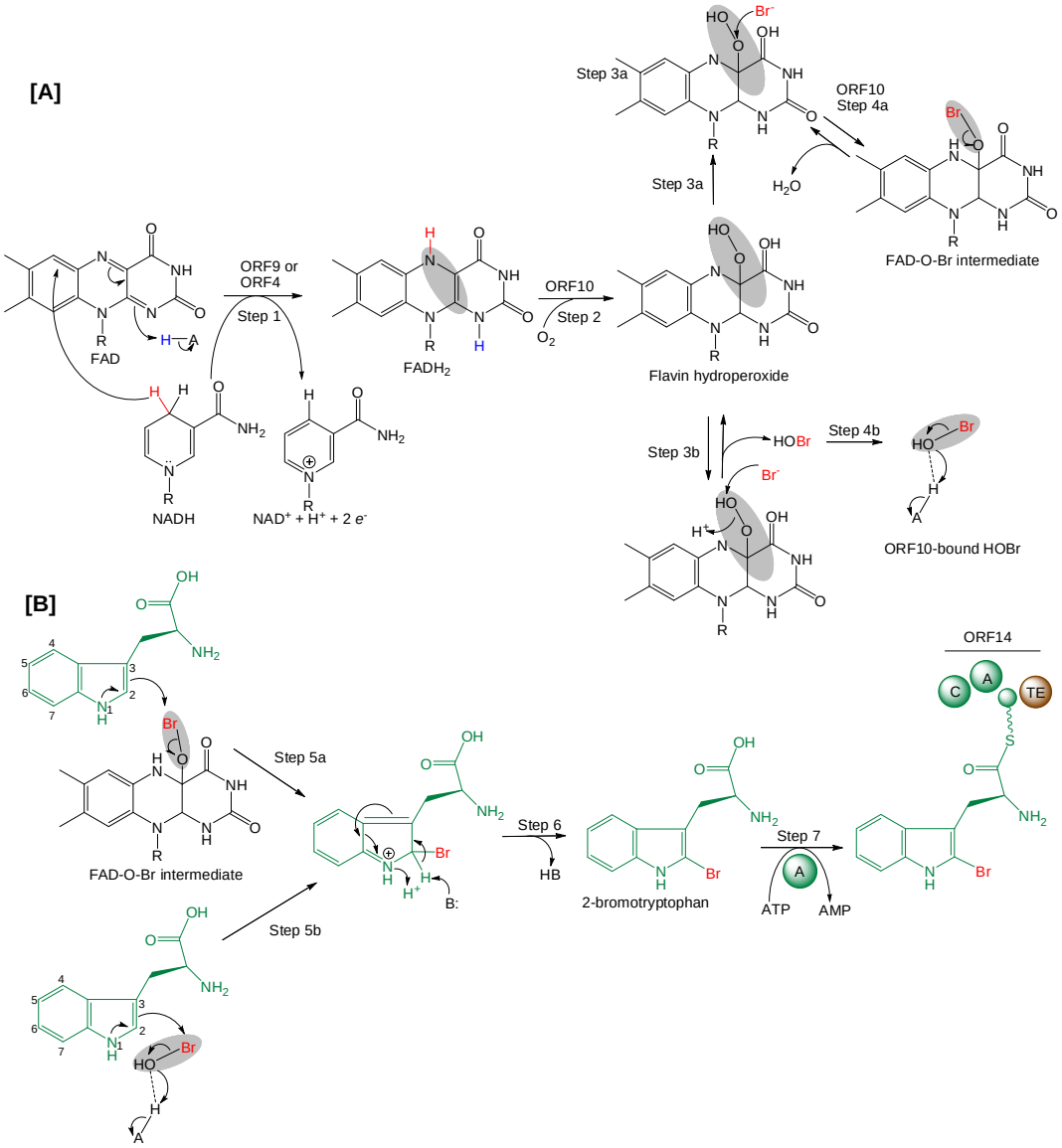


Fig. 3.3.7 Hypothetical role of ORF4/ORF9 and ORF10 in tryptophan bromination in the biosynthesis of keramamides. [**A**] It is assumed that ORF4/ORF9 and ORF10 are required to activate Br by increasing its electrophilicity. A flavin reductase encoded on *orf4*/*orf9* would provide $FADH_2$ for halogenase catalysis. $FADH_2$ would first be generated by ORF4/ORF9 (flavin reductases) using NADH for the reduction of FAD (step 1). The resulting $FADH_2$ would bind to the ORF10 (tryptophan-5-halogenase). The presence of a flavin binding site (GxGxxG) in ORF10 indicates the strict dependence of ORf10 on $FADH_2$ (van Pée and Patallo, 2006). The reaction of the ORF10-bound $FADH_2$ with O_2 would lead to the formation of a flavin hydroperoxide (ORF10-bound FAD-OOH) (step 2). Based on L-tryptophan halogenation suggested by Yeh *et al* (2005) for the formation of 7-chlorotryptophan in rebeccamycin biosynthesis, the flavin hydroperoxide would then react with Br^- ion to form a FAD-O-Br intermediate (steps 3a-4a). An alternative bromination was based on tryptophan chlorination proposed by Dong *et al* (2005), in which the flavin hydroperoxide would first be decomposed by Br^- (step 4b) to form HOBr. An acidic residue would make a hydrogen bond to HOBr, thereby activating Br. [**B**] Attack of the aromatic electrons on the FAD-O-Br (step 5a) or on the ORF10-bound HOBr (step 5b) and subsequent abstraction of a proton at C_2 would lead to the formation of 2-bromotryptophan (step 6) that would further be activated by A domain of ORF14. Pictures A and B were adapted from Yeh *et al* (2005) and Dong *et al* (2005).

An alternative mechanism was described by Yeh *et al* (2005) for the formation of 7-chlorotryptophan in rebeccamycin biosynthesis (Fig. 3.3.7). The resulting flavin hydroperoxide would first be chlorinated. Subsequent attack of the tryptophan`s aromatic electrons on the FAD-O-Cl and proton abstraction would result in 7-chlorotryptophan (Yeh *et al*., 2005). Dong *et al* (2005) elucidated the structure of tryptophan 7-halogenase (PrnA) in the biosynthesis of pyrrolnitrin from *pseudomonas fluorescens*. Based on the three-dimensional structure, Dong *et al* (2005) suggests the flavin hydroperoxide is decomposed by Cl^- followed by chlorination of tryptophan to form a chlorinated product (Dong *et al*., 2005). We postulate that the mechanism for tryptophan bormination in the biosynthesis of keramamides might resemble that suggested by Yeh *et al* (2005) or Dong *et al* (2005) (Fig. 3.3.7). Functional studies of this putative tryptophan 5-halogenase are currently underway in collaboration with van Pée lab at TU Dresden.

ORF12 shows homology to many tryptophan 2,3-dioxygenases. ORF12 is predicted to encode tryptophan-2,3-dioxygenase (TDO) that cleaves the pyrrole ring of L-Trp, to form *N*-formylkynurenine (Colabroy & Begley, 2005). TDO was previously reported in the biosynthesis of anthramycin (Hu *et al*., 2007), which was proposed to convert L-Trp to the *N*-formylkynurenine (NFK). The 4-methyl-3-hydroxyanthranilic acid half of anthramycin is derived from N-formylkynurenine (Hu *et al*., 2007). Some mechanisms for the convertion of L-Trp to NFK have been proposed (Efimov *et al*., 2011). One of earliest hypothetical mechanisms was the base-catalysed abstraction mechanism suggested by Hamilton (1969). In this mechanism, an active site base of TDO initially catalyzes the deprotonation of the indole NH group of L-Trp. The distal histidine was proposed to play such a role (Zang *et al*., 2006). However recent computational studies have revealed that the proton abstraction process is not energetically favorable (Chung *et al*., 2008). Therefore, Efimov *et al* (2011) and Basran *et al* (2011) have recently suggested a radical addition mechanism for the the convertion of L-Trp to *N*-formylkynurenine, as described in Figure 3.3.8. The presence of this L-Trp modifying enzyme (TDO) in the partial gene cluster isolated in this work raises a question whether the tryptophan-incorporating module (ORF14) could also accept *N*-formylkynurenine. An example showing the relaxed substrate specificity between L-tryptophan and *N*-formylkynurenine can be observed from the structures of Brunsvicamides A and B. Brunsvicamide A isolated from the cyanobacterium *Tychonema* sp. is a cyclic peptide known to contain a kynurenine moiety. Interestingly, the analog brunsvicamide B has kynurenine is substituted with

tryptophan (Müller *et al*., 2006), suggesting a relaxed substrate specificity of the A domain in brunsvicamide biosynthesis.

The resulting 5-bromotryptophan or *N*-formylkynurenine would be activated by A domain followed by the covalent binding via a thioester bond to the cognate PCP of ORF14. ORF14 is homologous to the NRPS OciC of cyanopeptolin synthetase in *Planktothrix rubescens* (Rounge *et al*., 2009). The ORF14 C domain would catalyze the condensation reaction between the PCP14-bound bromotryptophan or *N*-formylkynurenine and the peptidyl chain tethered to the other PCP. The full length peptidyl chain would be transferred to the hydroxyl of the active site Ser of TE domain in ORF14 to form a peptidyl ester. The TE-bound peptidyl-O-Ser intermediate would subsequently be attacked by an amino group of the intermediate to release a macrocyclic peptide (Du & Lou, 2010). For the macrocylization of keramamides, it is assumed that the amino group of L-2,3-diaminopropionic acid residue would attack the TE-bound peptidyl intermediate.

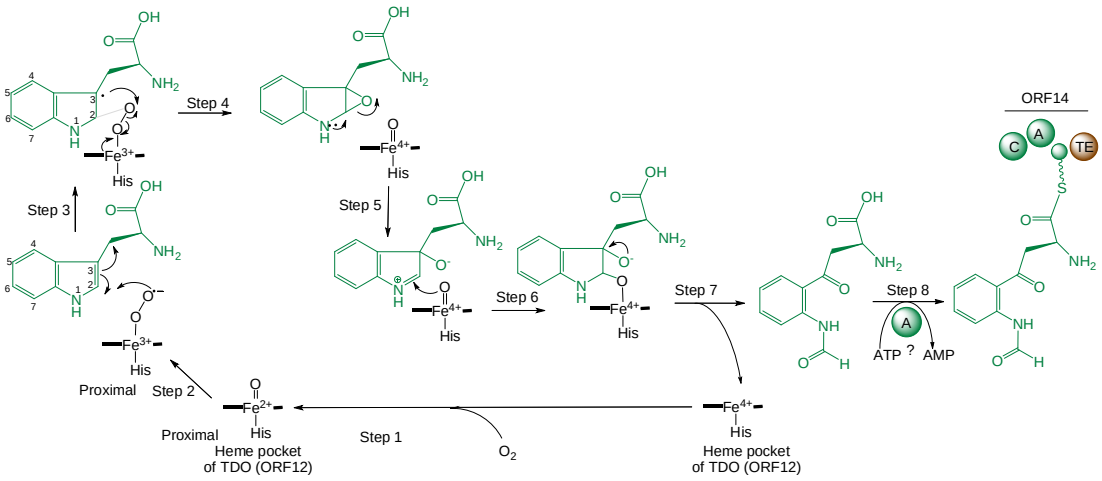


Fig. 3.3.8 Proposed mechanism for the convertion of L-Trp to NFK, as suggested by Basran *et al* (2011). The reductive activation of O_2 by a TDO (ORF12) would generate a highly oxidized iron intermediate (a +3 ferrous oxy intermediate) (step 1). Subsequent stepwise insertion of O_2 to the L-Trp (step 3) would lead to the formation of an epoxide ring and accompanying rapid cleavage of the O-O bond would lead to a +4 ferryl iron intermediate (step 3). Due to the nitrogen lone pair at the N_1, the epoxide ring could undergo C_2-O bond cleavage that would lead to opening of the epoxide ring (step 5). Further C_2-C_3 bond cleavage would generate NFK (Efimov *et al*., 2011; Basran *et al*., 2011). The resulting brominated tryptophan or NFK would be activated by A domain in ORF14 followed by the covalent binding via a thioester bond to the cognate PCP (Marahiel *et al*., 1997). The picture was adapted from Efimov *et al* (2011) and Basran *et al* (2011).

ORF7 is homologous to a transcriptional regulator of the IclR protein family, which probably plays a role in controlling the degradation of aromatic compounds (Molina-Henares *et al*., 2006), including L-Trp, *N*-formylkynurenine, and L-kynurenine. The involvement of an IclR protein was shown in the transcriptional regulation of

tryptophan biosynthetic pathway of *Corynebacterium glutamicum* (Brune *et al*., 2007).

Based on the description above, we have cloned a partial NRPS gene cluster from the metagenome of the marine sponge, *T. swinhoei* chemotype white. Bioinformatic analysis of the partial NRPS pathway revealed the module architecture and substrate specificity loaded by individual modules. The domain order and the predicted function of the A-domain in individual modules are consistent with the amino acid residues incorporated into the proposed structures, suggesting that a portion of this gene cluster presumably encodes the biosynthesis of keramamides.

3.4 Genomics-Guided Identification of a Biosynthetic Pathway

As described above, efforts to target the biosynthetic pathway of theonellamide F has led to the discovery of other interesting clusters likely encoding misakinolide A and keramamide H biosynthesis. Another strategy to target biosynthetic pathway consists of genome sequencing of the enriched filamentous cell fraction, which could potentially lead to the identification of gene clusters for other known metabolites as well as biosynthetic pathways for previously undetected metabolites.

3.4.1 Purity analysis of uncultured bacterial fraction

The major obstacle to sequence the genome of an uncultured bacterial symbiont, exemplified by "*Candidatus* Entotheonella sp.", is obtaining the uncultured cell fraction in the desired purity and quantity. However, the unusually large morphology of "Entotheonella spp." can allow for the enrichment of those bacteria by simple mechanical separation. In collaboration with Dr. T. Mori, Prof. H. Takeyama (Waseda University), Dr. T. Wakimoto, and Prof. S. Matsunaga (University of Tokyo), we obtained a significant amount of the uncultured filamentous fraction not only from the white chemotype but also from the yellow chemotype of the sponge, *T. swinhoei* (Fig. 3.4.1).

Figure 3.4.1 The uncultivated filamentous bacteria isolated from the yellow and white chemotypes of Japanese *T. swinhoei.*

Before the cell fractions were subjected to genome sequencing, we detected first the presence of "Entotheonella sp." by PCR using *Entotheonella*-targeting primers designed by us in this work as well as by Schmidt *et al* (2000). The primer design was based on the unique regions of several *Entotheonella* 16S rRNA genes both originally isolated from enriched cell pellet from *T. swinhoei* and retrieved from the public database, in comparison with 16S rRNA genes from other bacterial species that are commonly found in sponges. Figure 3.4.2 showed three unique regions

within *Entotheonella* 16S rRNA genes (as indicated by blue, red, and yellow lines). The forward primer 238-F designed by Schmidt *et al* (2000) was based on an upstream region at the relative positions of 270 to 240 bp (circled with a blue line). Using this unique region, we designed an alternative forward primer, termed Ento271F.

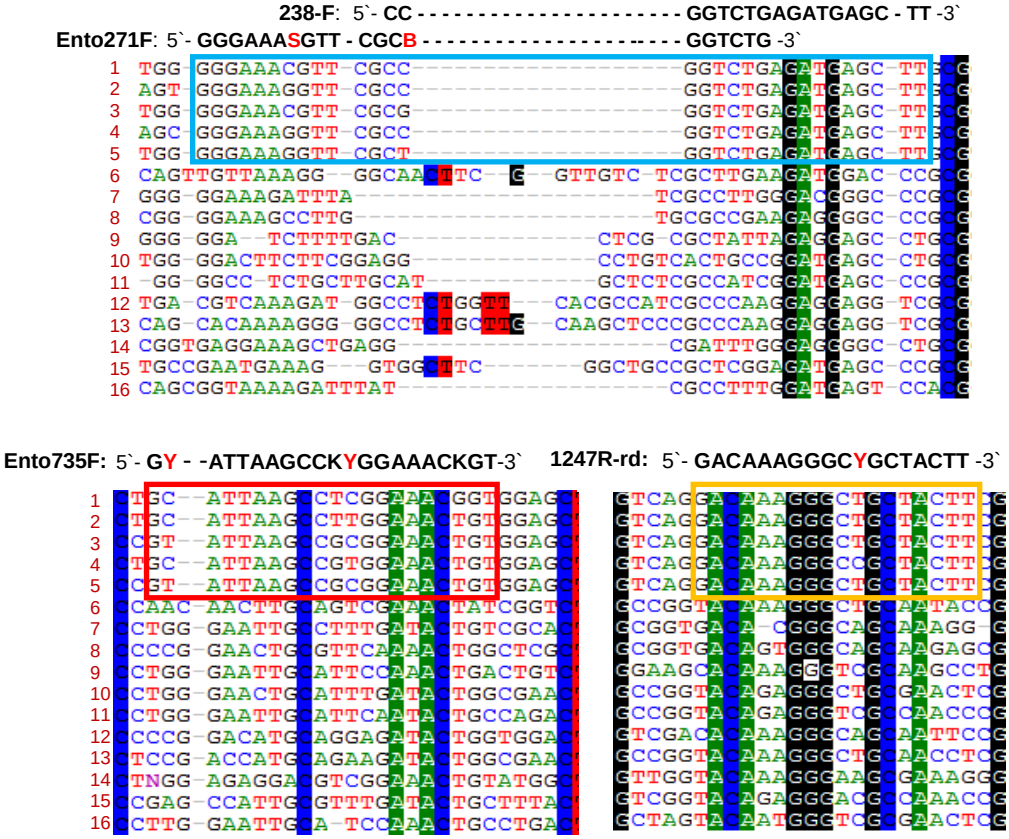


Figure 3.4.2 Sequence alignment of 16S rRNA genes of *Entotheonella* sp. with those from other bacterial species that are commonly found in sponges. The unique regions indicated by blue, red, and yellow circles were used as the basis for designing *Entotheonella*-targeting primers. Highly conserved bases are highlighted, in which black color for G, green for A, blue for C, and red for T. 16S rDNA sequences included in this alignment are from the filamentous cell pellet of Japanese *T. swinhoei* Y (1) isolated by Dr. T. Mori, the metagenome of Japanese *T. swinhoei* W (2) cloned by us in this work, *Entotheonella* sp. clone Dd-Ent-A94A of *Discodermia dissolute* (AY897123.1) (3), *Entotheonella palauensis* of Palauan *T. swinhoei* (AFI130847.1) (4), *Entotheonella* sp. clone 343 of soil (AY493913.1) (5), Acidobacteria clone PAUC34f of *T. swinhoei* (AF186412.1) (6), α-proteobacteria clone TK51 of *T. swinhoei* (AJ347042.1) (7), α-proteobacteria clone TK03 of *Aplysina* aerophoba (AJ347025.1) (8), α-proteobacteria clone CN60 of *C. nucula* (AJ850097.1) (9), γ-Proteobacteria clone E01-9C-26 of (AJ581351.1) (10), γ-Proteobacteria clone TK38 (AJ347044.1) (11), Gemmatimonadaceae clone CN82 of *C. nucula* (AM259917.1) (12), Gemmatimonadaceae clone AD004 (EF076125.1) (13), Chloroflexi clone TK23 of A. aerophoba (AJ347038.1) (14), Poribacteria clone 236 of *A. aerophoba* (AY485289.1) (15), and δ-proteobacteria clone EC214 (DQ889875.1) (16).

Subsequently we indentified another unique region at the nucleotide position of 740-780 bp (circled with a red line) and used it to design a reverse primer. Schmidt *et al*

(2000) identified a downstream unique region and used it to design the reverse primer 1247R. In our work, we de-generated a single nucleotide (T) within the 1247R primer sequence into Y (termed 1247R-rd primer). Two other reverse primers used in this work (1309R and 1290R) were designed based on the downstream region that are conserved well among 16S rRNA genes from different bacterial groups. However, combination of these reverse primers with those primers mentioned earlier has enabled us to target 16S rRNA genes of "Entotheonella sp." At first, we prepared total DNA from the cell pellet of filamentous symbiotic bacteria (Fig. 3.4.3A). The DNA purification was based on the anion-exchange resin commercially provided by QIAGEN, which is designed for direct isolation of chromosomal DNA 20-150 kb in size. Using the total DNA purified from the filamentous cell fractions, we detected the presence of correct-sized PCR products from four combinations of primers (238-F & 1247-rd, 238-F & 1309-R, Ento271-F & Ento1290-R, Ento735-F & Ento1290-R) (Fig. 3.4.3B). All of these four primer combinations also gave correct-sized PCR products for the metagenome from the sponge *T. swinhoei* chemotype W.

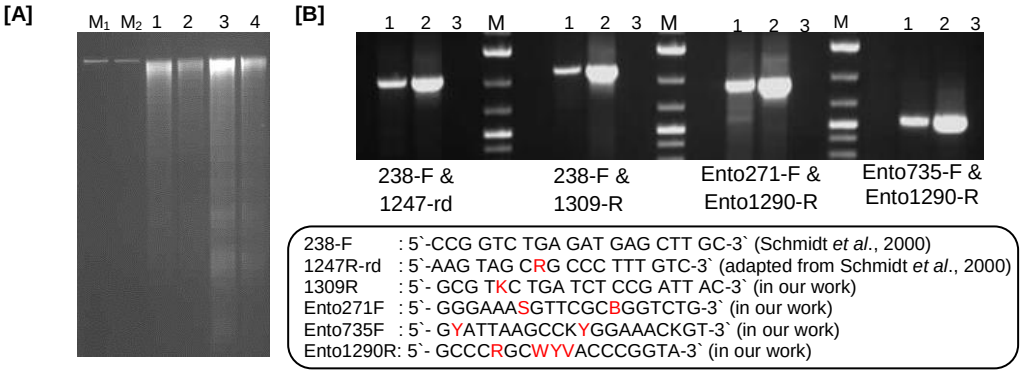


Figure 3.4.3 Detection of "Entotheonella sp." in the filamentous cell fraction of *T. swinhoei*. **[A]** DNA isolation from the cell fraction. The DNA quantity and quality was measured. For the chemotype W cell pellet, the total amount is ~100 µg, A_{260}/A_{280} of 1.8, and A_{260}/A_{230} of 2.36. For the chemotype Y cell pellet, the total amount was ~120 µg A_{260}/A_{280} of 1.9, and A_{260}/A_{230} of 2.17. M_1, 36-kb fosmid control (20 ng); M_2, 36-kb fosmid control (10 ng); lane 1: DNA sample from the chemotype W (507 ng); lane 2: DNA sample from the chemotype W (253.5 ng); lane 3, DNA sample from the chemotype Y (668 ng); and lane 3, DNA sample from the chemotype Y (334 ng). **[B]** Four combinations of *Entotheonella*-targetting primers were tested. Lane 1, DNA of *T. swinhoei* chemotype W; lane 2, DNA from the filamentous cell fraction; and lane 3, negative control (without DNA template); and M, DNA marker.

Subsequently we checked the cell purity using universal 16S rRNA primers in order to obtain a quantitative impression on how dominant are the "Entotheonella sp." cells in the uncultured bacterial cell fraction in comparison with other interfering bacterial cell types. Universal 16S primers, 27F and 1492R (Lane, 1991;

Stackebrandt & Lieasack, 1993), were used for this biodiversity analysis. The PCR products (Fig. 3.4.4A) were cloned into the vector pBlueScript II SK(-). Around 47 clones were selected and checked for the presence of inserts by *Xho*I and *Pst*I digestion (Fig. 3.4.4B). The recombinant clones were subsequently subject to restriction length polymorphism (RFLP) analysis by *Rsa*I and *Hha*I digestion (Fig. 3.4.4C). Fifteen clones showing unique RFLP patterns were sequenced.

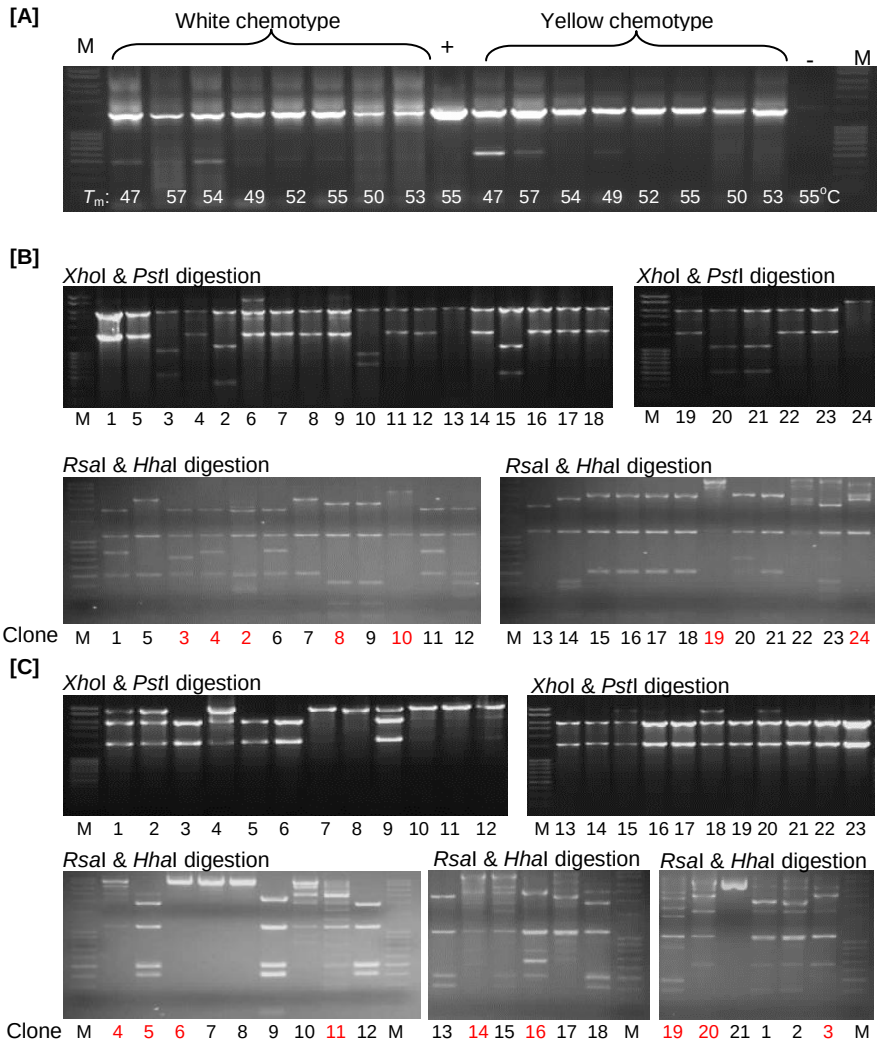


Figure 3.4.4 Genetic diversity of bacteria present in the filamentous cell fractions of *T. swinhoei*. **[A]** PCR amplification of 16S rRNA genes from the filamentous cell fractions of white and yellow chemotypes. **[B]** Insert checking and RFLP analysis of some selected clones from the chemotype W. **[C]** Insert checking and RFLP analysis of some selected clones from the chemotype Y. Insert checking was done by the *Xho*I-*Pst*I digestion of individual clones. Clones containing inserts were subsequently subjected to RFLP analysis using the frequent cutting enzymes *Rsa*I & *Hha*I. Several clones showing unique RFLP pattern (as indicated with red numbers) were sequenced. Numbers on the pictures refer to clone codes.

The results (Table 3.4.1) of the chemotype Y showed that 4 clones (Y3, Y5, Y6, Y19) belong to "Entotheonella spp." 16S-rRNA genes, two (Y14, Y16) were *Spongiobacter* spp. 16S-rDNA fragments, and one (Y4) belong to a member of the

Rhodobacteraceae, a novel marine methyl halide-oxidizing bacteria (Schafer *et al*., 2005). However, the RFLP patterns of the four clones harboring “Entotheonella sp.” 16S rDNA were the same as 10 additional unsequenced clones (Y20, Y5, Y9, Y12, Y13, Y18, Y19, Y7, Y8, Y21). This suggests that the majority of the amplicons belonged to “Entotheonella sp.” 16S rDNA, assuming little cloning bias. The clone Y19 showed 97% identity to *Entotheonella palauensis*, a novel symbiont of the Palauan sponge *Theonella swinhoei* (Haygood *et al*., 1999). This suggested that the filamentous fraction from the chemotype Y is dominated by “Entotheonella sp.” cells, which was only contaminated with other bacteria. In contrast, the 16S rRNA analysis of the filamentous fraction from the white chemotype showed that all amplicons belong to the *Chloroflexi* (W4), *Spongiobacter* (W2, W3, W10) and *Spirochaeta* (W19, W24) genera, while none were “Entotheonella sp.” 16S rDNA fragment (Table 3.4.1).

The negative result from chemotype W was unexpected and there are some possible reasons for this. First, the enzymatic treatment (lysozyme) step in DNA isolation was probably insufficient to disrupt the membrane of “Entotheonella sp.” cells from the chemotype white. Second, lysis of “Entotheonella sp.” cells might occur during mechanical separation from the sponge sample, for instance, due to cellular osmotic pressure. Heldal *et al* (2012) reported the importance of Mg^{2+} ions in cellular osmoregulation in marine bacteria, especially during periods of non-growth and dormancy when carbon and nutrient sources are limited. This suggests that using Mg^{2+}-free artificial water to process the cell fraction could potentially lead to the disruption of certain bacterial cell types. Other possibility is that the isolated cell fraction was not dominated by “Entotheonella sp.” cells, and the abundant filamentous cells observed under the microscope were probably other bacteria. If “Entotheonella sp.” cells were present in very low quantity compared with other non-target bacterial members in the fraction, this could affect the relative ratio of 16S rDNA amplicons. As the consequence, 16S rDNA amplicons amplicons of “Entotheonella sp.” were not well represented in clone libraries. Another possibility is that “Entotheonella sp.” cells from chemotype W might be sensitive to lysis when they were prepared from a frozen sponge sample (personal communications with Dr. T. Mori and Prof. J. Piel). This unique phenomenon was also observed in our recent effort to obtain DNA from the filamentous bacterial cells isolated from a frozen sponge, in which we did not get the DNA.

Table 3.4.1 The genetic diversity of the uncultured filamentous bacterial fraction isolated from the Japanese sponge *Theonella swinhoei* in this study.

Clone	Product Length (bp)	Taxonomic affiliation	Source
Y3	1477	*Entotheonella* sp.	FBC-TSY
Y4	-	*Rhodobacteraceae*	FBC-TSY
Y5	1419	*Entotheonella* sp.	FBC-TSY
Y6	1477	*Entotheonella* sp.	FBC-TSY
Y11	-	*Alpha-proteobacterium*	FBC-TSY
Y14	-	*Spongiobacter* sp.	FBC-TSY
Y16	-	*Spongiobacter* sp	FBC-TSY
Y19	1428	*Entotheonella* sp.	FBC-TSY
Y20	-	*Spongiobacter* sp.	FBC-TSY
pEntoSpo	1253	*Entotheonella* sp.	M-TSW
W2	1563	*Spongiobacter* sp., *Endozoicimonas elysicola*	FBC-TSW
W3	740	*Spongiobacter* sp.	FBC-TSW
W10	1247	*Spongiobacter* sp.	FBC-TSW
W4	850	*Chloroflexus* sp.	FBC-TSW
W19	1300	*Spirochaeta* sp.	FBC-TSW
W24	-	*Spirochaeta* sp.	FBC-TSW

Note: Y6= pEntoSpo; Y3 = Y5 = Y19; FBC, filamentous bacterial cell fraction; M, metagenome; TSY, *Theonella swinhoei* Y; TSW, *Theonella swinhoei* Y

The clone W2 showed 93% identity to *Endozoiciomonas elysicola* (Accession no. AB196667), a bacterium isolated from the marine mollusc *Elysia ornata* (Kurahashi & Yokota, 2006 unpublished). It exhibits 94% identity to *Spongiobacter nickelotolerans* isolated from a marine sponge (Acc nu. AB205011) (Nishijima *et al*., 2005 unpublished). The clone W4 showed high homology (94%) to uncultured *Chloroflexus* sp., sponge associated bacteria (Montalvo & Hill, 2011), and 88% identity to uncultured *Caldilinea* sp., chloroflexi from the intestine of the earthworm *Eisenia fetida* (Rattray *et al*., 2010). The clone W19 has 98% identity to the uncultured sponge associated bacteria *Spirochaetes* clone XD1F11 (Montalvo & Hill, 2011).

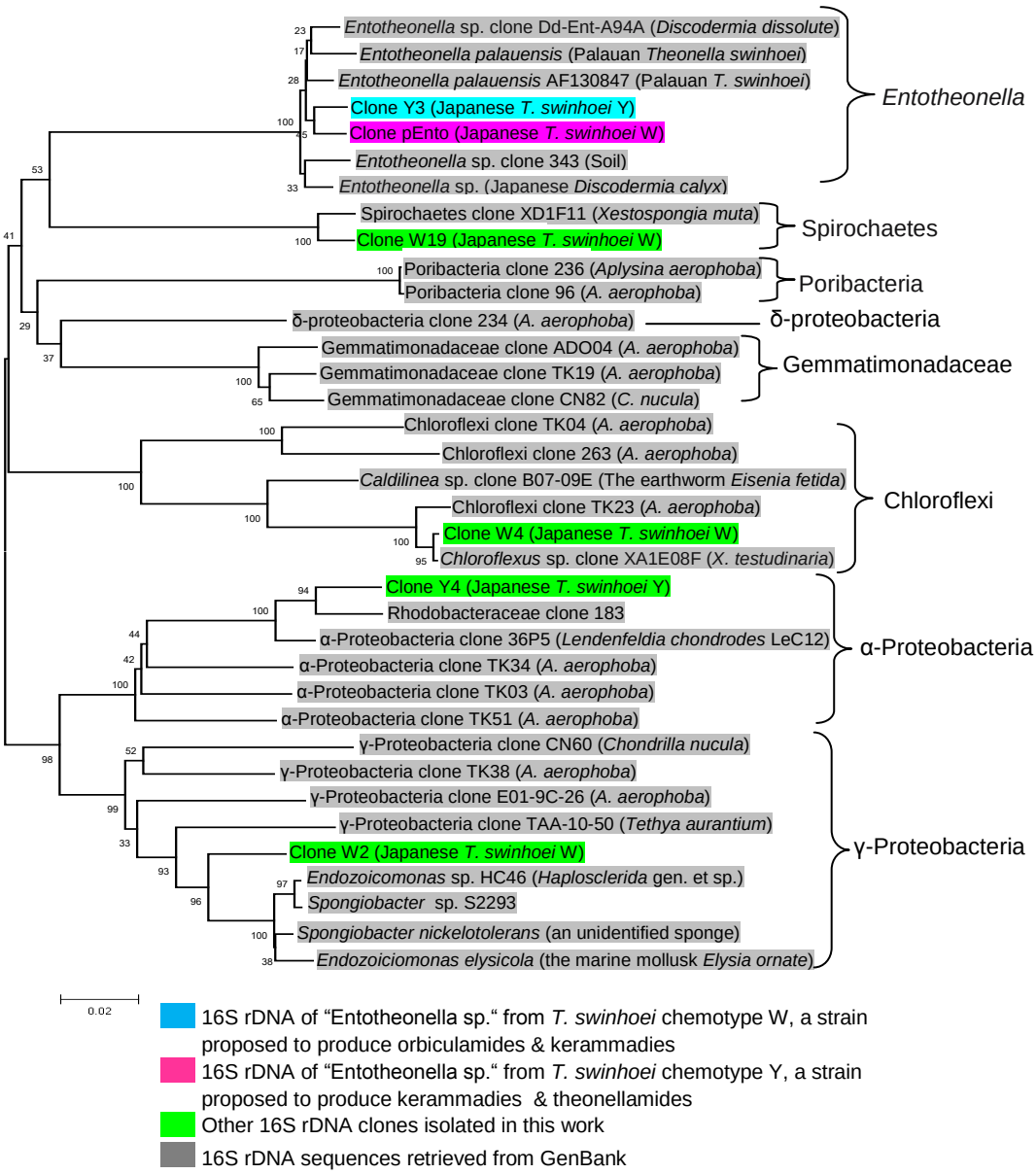


Figure 3.4.5 Phylogenetic analysis of 16S rDNA sequences that belong to "Entotheonella sp." from the Japanese marine sponge, *T. swinhoei*, placing them in the same clade as *Entotheonella palauensis* from the Palauan *T. swinhoei* and *Entotheonella* sp. from *Discodermia dissoluta*. The distance three was based on ~1000 bp of 16S rRNA gene sequences. Clone Y3, identified as "Entotheonella sp.", from the *T. swinhoei* chemotype Y is proposed to produce keraramide and orbiculamide based on genome sequencing. Clone pEntoSpo, identified as "Entotheonella sp.", from the *T. swinhoei* chemotype W is proposed to produce theonellamide F based on the localization study by Schmidt *et al*. (2000), while misakinolide A and keraramide H based on the localization study combined with metagenome mining in this work.

3.4.2 Genome Analysis

The 16S rDNA analysis suggested that the filamentous fraction from the *T. swinhoei* chemotype Y was more pure because it was dominated by *Entotheonella* sequences. Therefore the total DNA from the cell fraction was subjected to whole genome sequencing through collaboration between our group and that of Dr. Jörn Kalinowski (Bielefeld University) within the GenBioCom framework supported by the BMBF. The whole genome sequencing was based on Sanger, 454 and Illumina sequencing (Margulies *et al*. 2005; Droege and Hill 2008; Rothberg and Leamon, 2008; DiGuistini *et al*., 2009). The first round of genome sequencing yielded 18.280 contigs that contain at least 14 additional KS domains of PKSs and 41 PCPs of NRPSs. The presence of only a single 16S rRNA gene in the sequence data confirmed that the "Entotheonella sp." cell fraction was fairly pure. The onnamide and polytheonamide gene clusters were found. However, there are still many gaps in the genome sequences that are necessary to be filled, indicating that many PKS/NRPS contigs have not been annotated.

Some PKS/NRPS contigs were placed in order into scaffolds based on understanding the multimodular domain organization. We identified a promising scaffold containing 11 contigs, which was predicted to direct the biosynthesis of a keramamide or orbiculamide due to the presence of domains presumably involved in thiazole ring formation. We filled in this promising scaffold by PCR, amplicon sequencing, and sequence assembly. Ten of twelve gaps in the scaffold were initially filled (Fig. 3.4.6A). Another additional gap was subsequently filled in through the help of Mr. Bikram Pandey (a master student who did a practical work under my supervision in the Piel Group) (Fig. 3.4.6B). The amplicon covering the remaining gap 3 is quite large (>1.5 kb). End sequencing of it was not enough to obtain an entire amplicon sequence. Although containing a very small gap, the amplicon sequence could be functionally annotated and predicted to code for acyltransferase (AT) domain. We analyzed the resulting continuous sequence of approximately 26 kb by finding the potential ORFs and subsequent functional prediction (supplementary S3.4.1). The potential ORFs were identified using the software FramePlot 2.3.2 (Ishikawa and Hotta, 1999) and Geneious 5.4.6 (Drummond *et al*., 2011). Subsequent functional prediction of the identified ORFs were carried out using BLAST (McGinnis & Madden, 2004), PKS/NRPS analysis Web-site (Bachmann & Ravel, 2009), and NRPSpredictor2 (Rausch *et al*., 2005; Röttig *et al*., 2011).

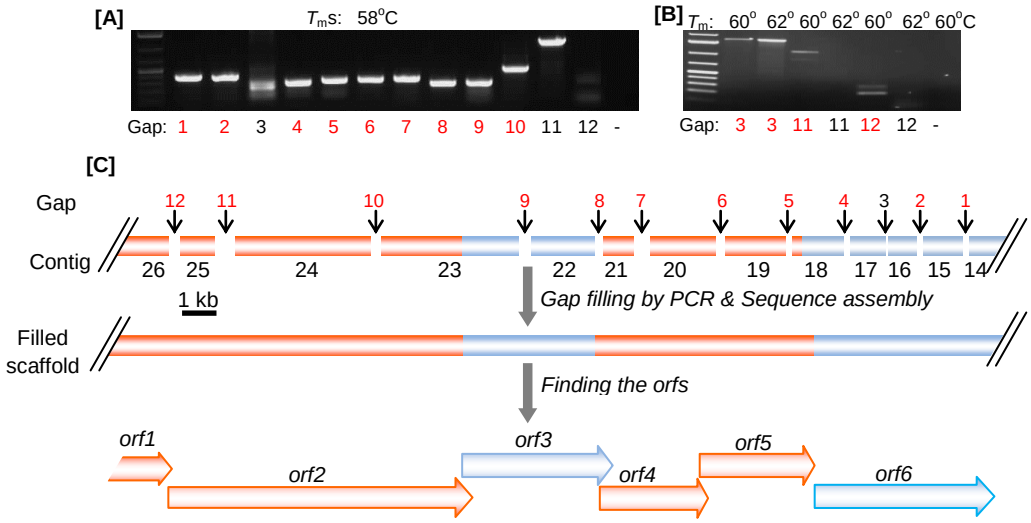


Figure 3.4.6 Filling the gaps in a promising scaffold and identifying potential ORFs. **[A]** Filing the gaps between the ordered contigs was initiated with designing primers that attach each contig at the distance of ~100 bp from the contig ends. The specific primers were then employed to obtain amplions. The individual amplicons were sequenced, and the amplicon sequence data was reassembled with the contig sequences. Ten gaps (shown by red colored numbers) were filled in after reassembly. **[B]** Filling three remaining gaps (gaps 3, 11, 12) was subsequently conducted by obtaining PCR-amplicons at two different elevated annealing temperatures (60° and 62°C). Subsequent amplicon sequencing resulted in filling two additional gaps (gaps 11 and 12) and left a small niche on the gap 3. **[C]** After filling the gaps, the resulting continuous sequence of approximately 26 kb was annotated by finding the potential ORFs and subsequent functional prediction.

Using PKS/NRPS analysis Web-site (Bachmann & Ravel, 2009), we predicted the domain organization on individual ORFS and confirmed the prediction by finding the highly conserved motifs in each domain. The ~26 kb gene DNA harbors six *orfs*: 4 code for NRPS modules and 2 code for PKS modules (Fig. 3.18C). Each NRPS module contains adenylation domain responsible for activating amino acids. Using NRPSpredictor2 (Rausch *et al*., 2005; Röttig *et al*., 2011), we predicted the substrate specificity code of all adenylation domains present in the individual NRPS modules, as shown in Table 5 below. An incomplete ORF (ORF1) at the upstream end contains an A-T module, in which the A domain was predicted to be specific for ornithine (100% identity) or related amino acids. ORF2 harbors two NRPS modules, in which the A-domains were presumably specific for activating proline (90% identity) and isoleucin, respectively. Next to ORF 2, ORF3 encodes a PKS module (KS-AT-KR-ACP) in a separate protein, in which the integrated AT is presumably specific for malonyl-CoA. The KR domain in ORF3 showed 53% identity to KR5 present in the ZmaA of zwittermicin A biosynthetic pathway, which generates hydroxyl moiety at the incorporated (2*R*)-hydroxymalonyl-CoA (Kevany *et al*., 2009). The downstream ORF4 codes for an NRPS module that is responsible for

incorporating alanine (90% identity). The subsequent module, ORF4, contains an A-domain that might be responsible for activating alanine.

Table 3.4.2 Key amino acid residues for substrate specificity of the adenylation domains of the proposed biosynthetic pathway.

Adenylation domain	Specificity-conferring code of A domain										Predicted Substrate[c] & % match
	Relative residue position	235	236	239	278	299	301	322	330	331	517	
A_2 (ORF1)	Consensus sequence[a]	D	V	G	E	I	G	S	I	D	K	Orn (100%)
	Specificity sequence[b]	D	V	G	E	I	G	S	I	D	K	
A_3 (ORF2)	Consensus sequence[a]	D	V	Q	L	I	A	H	V	V	K	Pro (90%)
	Specificity sequence[b]	D	V	Q	F	I	A	H	V	A	K	
A_4 (ORF2)	Signature sequence[a]	D	A	E	S	V	G	A	I	D	K	Arg (60%), Leu, or Ile
	Signature sequence[a]	D	V	A	D	V	G	A	I	D	K	
	Signature sequence[a]	D	A	W	F	L	G	N	V	V	K	
	Signature sequence[a]	D	A	F	F	L	G	V	T	-	K	
	Specificity sequence[b]	D	V	Y	F	V	G	A	V	I	K	
A_4 (ORF4)	Consensus sequence[a]	D	L	L	F	G	I	A	V	L	K	Ala(90%)
	Specificity sequence[b]	D	I	Y	N	N	A	L	T	Y	K	
A_1 (ORF5)	Consensus sequence[a]	D	L	Y	N	L	S	L	I	W	K	Cys (90%)
	Specificity sequence[b]	D	L	Y	N	M	S	L	I	W	K	
	Variability (%)[d]	3	16	16	39	52	13	26	23	26	0	

[a] Signature sequences derived from A domains activating the same substrate, as predicted by Stachelhaus *et al* (1999)
[b] The positions of the ten specificity-conferring decisive residues identified in this work, which were predicted using NRPSpredictor2 (Rötting *et al*., 2011; Rausch *et al*., 2005)
[c] The predicted substrates: Orn, ornithin; Pro, proline; Arg, arginine; Leu, leucine; Ile, isoleucine; Ala, alanine; Cys, cystein
[d] Residue variability are shown in percentage and variable constituents within specific-conferring codes are indicated by red color (Stachelhaus *et al*., 1999)
[e] Specificity code for ArgA based on studies by Demirev (*et al*., 2005) for the oligopeptide moiety of cephabacin
[f] Specificity code for ArgA based on studies by Guenzi (*et al*., 1998) for the syringomycin synthetase gene cluster
[g] According to nostocyclopeptides A1 and A2 primary structures (Hoffmann *et al*., 2003; Becker *et al*., 2004)

ORF5 showed 51% identity to HctE of the hectoclorine biosynthetic pathway that contains domains required in the formation of thiazole ring (Ramaswamy *et al*., 2007). The C terminus of ORF5 is homologous (70% similarity) to the condensation domain, designated as Cy, of the BarG in the barmamide biosynthetic pathway (Chang *et al*., 2002). Cy is known to play a role in the heterocyclization of the incorporated cysteine residue (Stachelhaus *et al*., 1998). The A-domain of ORF20 is presumably responsible for the activation of cysteine. An oxidoreductase (OX) is present in the N terminus of ORF 5. The presence of the cystein-specific A domain, a Cy domain, and an OX domain in ORF5 suggest the role of this module in the formation of a thiazole ring through cysteine cyclodehydration and subsequent oxidation. A PKS module encoded at the downstream end of ORF6 contains the domain components KS-AT-DH-KR-ACP. The integrated AT in this module is responsible for selecting malonyl-CoA. The KR in this module contains the stereospecificity motif AGVLD**D**GILLQQSGARF (diagnostic site in red) that converts the β-keto moiety in the incorporated malonyl-ACP into a β-D-OH. This resulting hydroxyl group is subsequently converted to the α,β-E-double bond by the DH.

3.4.3 Proposed scheme of the identified biosynthetic pathway

Our proposed scheme for this partial biosynthetic pathway, as shown in Figure 3.4.7, begins with incorporation of ornithine onto the T domain of module 1. Proline would be loaded on module 2, resulting in a prolinyl-*S*-T. The C domain in module 2 catalyzes nucleophilic attack of the amino group of the prolinyl-*S*-T donor on the acyl group of an upstream electrophilic ornithinyl-*S*-T chain, resulting in the formation of a peptide bond between these two amino acyl chains. Isoleucine would be loaded in module 3, and the amino group of this T-loaded isoleucine forms a peptide bond with the acyl group of the upstream peptidyl chain. Subsequently the downstream PKS module (module 4) would load a malonyl unit and modifies it by β-ketoreduction. The KS domain accepts an upstream peptide chain and catalyzes the decarboxylative condensation with its cognate malonyl-*S*-ACP, leading to the elongation of the peptide chain with a short carboxylic acid via a C-C bond. Module 5 would incorporate alanine loaded onto the T domain. This thioesterified alanine is integrated with the upstream peptide chain. The A domain in module 6 would activate cysteine. The T-load cysteine residue is subsequently cyclized by the Cy domain, resulting in a thiazoline, which is then oxidized by the oxidoreductase domain to generate a thioazole ring. The C domain in this NRPS module accepts the intermediate from the upstream PKS module, and subsequently mediates a C-N bond formation, allowing the elongation of this intermediate with an alanyl residue. The second PKS module (module 7) at the C terminus of this pathway would load a malonyl unit that is subsequently modified by KR and DH domains, resulting in an α,β-*E*-double bond. The KS domain in module 6 would accept the growing upstream peptide chain from the upstream NRPS module. Then the KS domain would catalyze the formation of a C-C bond to generate a peptide chain with a short carboxylic acid containing α,β-*E*-double bond.

Sequence analysis of the partial biosynthetic pathway revealed that domain composition and substrate specificity loaded by each module is consistent with the formation of amino acid residues incorporated into the proposed structure. The module order and the predicted function of the A-domain in individual modules suggest that a portion of this gene cluster presumably encodes the biosynthesis of an orbiculamide-like structure.

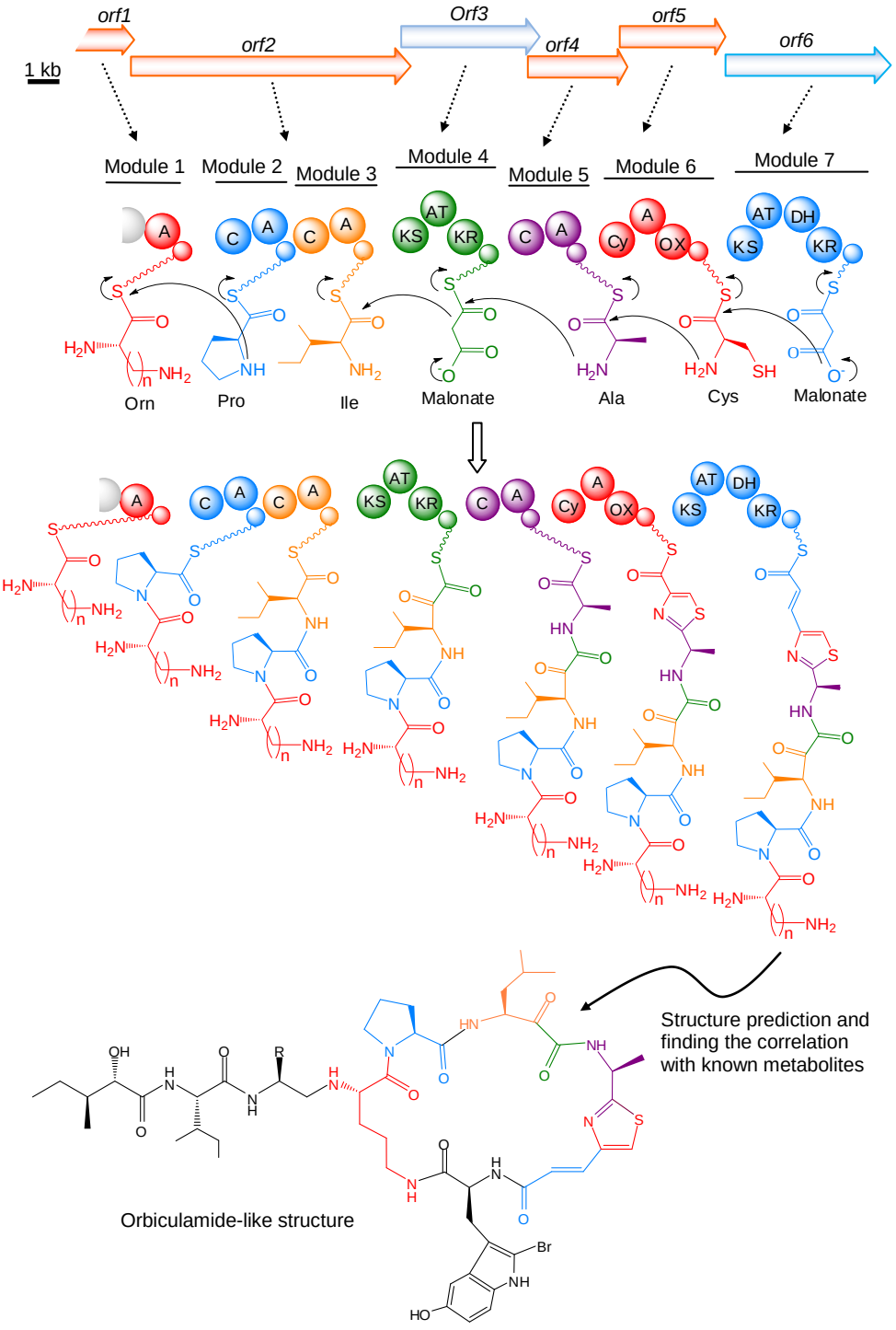


Figure 3.4.7 The proposed biosynthetic pathway of a PKS/NRPS scaffold identified in the sequenced genome of "Entotheonella sp." from the Japanese sponge *T. swinhoei* chemotype Y. Amino acid substrates are activated by A domain in individual NRPS modules through AT hydrolysis. Malonyl units are activated by an AT domain integrated in individual PKS modules. These unstable intermediates are then transferred onto T or ACP domains and bound as a thioester to an enzyme-bound 4`-phosphopantetheinyl (4`-PP). The thiol-activated intermediates can undergo modifications. Thioesterified substrate amino acids are then integrated into the peptide product through a step-by-step elongation by a series of transpeptidation reactions.

3.5 Summary and Outlook

Highly complex consortia of sponge-associated bacterial symbionts have been frequently proposed to be the true source of many sponge-derived metabolites. However, the majority (>99%) of these symbionts are uncultivable (Amann *et al*., 1995; Hugenholtz *et al*., 1998; Friedrich *et al*., 2001; Webster and Hill, 2001), thereby hampering efforts to prove the symbiont hypothesis and to study their biosynthetic potential. However, using cultivation-independent approaches recently developed in the Piel group have enabled us to pinpoint biosynthetic pathways of interest within diverse genetic mixtures from sponge symbiotic systems (Piel *et al*., 2004b; Hrvatin & Piel, 2007; Nguyen *et al*., 2008; Fisch *et al*., 2009; Freeman *et al*., 2012).

The main goals of this project were to obtain further metagenomic insights into the metabolic potential of uncultured symbiotic bacteria by detecting the key genes involved in secondary metabolism, to clone at least one new complete biosynthetic pathway of pharmacological relevance, and to provide taxonomic information of the uncultured bacterial symbionts harboring the biosynthetic pathway(s) of interest. The sponge *T. swinhoei* has been chosen in this work due to the wide variety of pharmaceutically important secondary metabolites isolated from this sponge as well as the high complexity of the associated bacteria, which might play an important role in metabolite biosynthesis (Matsunaga and Fusetani, 2003; Matsunaga *et al*., 1989; Matsunaga *et al*., 1992).

Achieving such goals mentioned above is a very challenging task due to the high complexity of the microbial communities in many sponges. The microbial consortia can account up to 60% of the total biomass in many sponges and can consist of thousands of species with at least 18 bacterial and archaeal phyla (Taylor *et al*., 2007). The genetic variation within a single species as well as possible dominance of different species makes the symbiotic community even more complex. To clone new biosynthetic pathways from this complex symbiosis model, we applied and developed metagenome mining strategies. There are three challenges that we encountered in metagenome mining to obtain a high probability of finding a target biosynthetic pathway: (i) obtaining high molecular weight sponge metagenome, (ii) creating complex metagenomic libraries that represent genomes present in the sponge system, and (ii) rapid library screening, which is also highly economic in

terms of labor, consumables and required space (Hrvatin & Piel, 2007; Uria & Piel, 2009; Gurgui & Piel, 2010).

The first challenge, obtaining high molecular weight sponge metagenome, was overcome in this work using a CTAB-based DNA extraction in combination with gel-based DNA purification developed by Piel group (Piel *et al*., 2004b; Hrvatin & Piel 2007; Gurgui & Piel, 2010). Subsequently, we performed the PCR detection of PKS genes in the metagenome of *T. swinhoei* chemotype W. However, all of the amplicons obtained from the sponge metagenome came from lipid metabolism, with none of them predicted to be responsible for complex polyketide biosynthesis. The difficulty to detect the PKS genes involved in complex polyketide biosynthesis might be due to the huge amounts (86% in average) of fatty acids contained in the sponge, in comparison with the very small percentage of *trans*-AT polyketides (around 2%), based on the previous study by Piel, Hentschel and coworkers on 20 different sponge species (Fieseler *et al.*, 2007).

To tackle such challenge, we used a filamentous cell fraction predominated by "*Candidatus* Entotheonella sp.", a heterothropic delta-proteobacterium associated with *T. swinhoei* chemotype W. The species diversity of this fraction was previously simplified by differential centrifugation. This mechanical separation not only reduced the complexity of the bacterial community associated with the sponge host and enriched the target bacterial population, but also remove contaminants from the sponge tissue such as carbohydrates. Based on the earlier localization studies by Faulkner and coworkers (Bewley *et al*., 1996; Schmidt *et al*., 2000), we hypothesized that theonellamide F (structurally similar to theopalaumide, see Introduction) was located in "Entotheonella sp." Therefore, our initial studies were focused on gaining access to the biosynthetic pathway of theonellamide F from this filamentous fraction. We hypothesized that theonellamide biosynthesis involves a PKS/NRPS system, and theferore, we initially used universal PKS and NRPS primers to target the biosynthetic genes coding for the hybrid system. By PCR cloning and sequencing, we obtained 8 KS amplicons belonging to complex PKS systems as well as two adenylation sequences of NRPS with the predicted specificity for arginine (ArgA). This demonstrated that reducing the microbial complexity by mechanical separation has enriched PKS and NRPS genes.

However, the isolated ArgA sequences mentioned above might not be involved in the biosynthesis of theonellamide F, because arginine (Arg) or its derivatives are not building blocks forming theonellamide F. This prompted us to develop a strategy to clone the A domain sequences specific for amino acid residues encoded more than once in theonellamide structure. One of the targeted residues is L-serine that is encoded three times in theonellamide structure. Using the strategy, we obtained two amplicons coding for serine-specific A domains (SerA) from the filamentous fraction from *T. swinhoei* chemotype W.

To obtain biosynthetic gene clusters that harbor the KS- and SerA-encoding DNA regions, we constructed a fosmid library from the filamentous bacterial fraction. Screening of the library using the specific primers based on the KS amplicons did not give a positive clone. However, when using the primers based on SerA amplicons, we isolated a positive fosmid clone. Analysis of this clone indicated that it harbored only a small NRPS cluster instead of the expected complex multimodular system.

Due to the failure in identifying complex PKS and NRPS pathways in the library constructed from the filamentous bacterial fraction, a highly complex metagenomic library (~350,000 clones) was subsequently constructed from the metagenome of *T. swinhoei* chemotype W. Screening of this library led to the isolation half of a *trans*-AT PKS gene cluster. By chromosome walking, we isolated additional PKS regions, resulting in an entire *trans*-AT PKS gene cluster. Our bioinformatic analysis predicted that the isolated PKS gene cluster encoded the biosynthetic pathway of misakinolide A. This dimeric polyketide is a potent actin inhibitor isolated from *T. swinhoei* sponges (Sakai *et al.*, 1986; Kato *et al.*, 1987; Kitagawa *et al*., 1990). With misakinolide A biosynthetic genes in hand, the stage is now set to perform functional studies and to establish a heterologous expression system to create a renewable misakinolide A source. Furthermore, using the information present in the SerA amplicons, we isolated a NRPS gene cluster from the sponge metagenomic library. Our bioinformatic analysis revealed that the isolated NRPS genes are a good candidate for the biosynthesis of keramamide H. Finding the additional regions of the putative keramamide H gene cluster as well as functional studies will be pursued.

Two different gene clusters isolated in this work as mentioned above exhibited typical bacterial gene features, such as the absence of introns, the presence of ribosome binding sites, and small intergenic distances between putative genes, strongly indicating that the producer of both misakinolide A and keramamide H is a symbiotic bacterium. The taxonomic status of the bacterial producer(s) is still unknown yet. However, due to the screening system used to isolate the gene clusters was based on the PKS and NRPS amplicons from the filamentous fraction dominated by “Entotheonella sp.“, we assumed that both misakinolide A and keramamide H could be produced by “Entotheonella sp.” To confirm this taxonomic status, further analysis either by single cell studies or its combination with complete genome sequencing is currently underway using the filamentous fraction as the starting material.

Preliminary genome sequencing of the uncultured “Entotheonella sp.“ fraction from a different chemotype of *T. swinhoei* (chemotype Y) yielded 18.280 contigs that contain at least 14 additional KS domains of PKSs and 41 PCPs of NRPSs. We identified a PKS/NRPS scaffold containing 11 contigs, which was predicted to direct the biosynthesis of a keramamide or orbiculamide due to the presence of domains presumably involved in thiazole ring formation. We filled in this scaffold by PCR, amplicon sequencing, and sequence assembly. Our bioinformatic analysis confirmed that the identified gene cluster encoded the putative biosynthetic pathway of orbiculamide-like structure. The presence of only a single 16S rRNA gene of *Entotheonella* sp. in the sequence data confirmed that this symbiontic bacterium is the producer of the orbiculamide-like structure. The onnamide and polytheonamide gene clusters were also found. However, there are still many gaps in the genome sequences that are necessary to be filled, indicating that many PKS/NRPS contigs have not been annotated. Next rounds of genome sequencing for obtaining the complete sequence data of uncultured “Entotheonella sp.” are currently underway. Subsequent genome mining could potentially lead to the identification of other biosynthetic pathways.

Chapter 4
Methodology

This methodology section described the methods applied or developed in this work to pinpoint biosynthetic pathway of pharmaceutical relevance among diverse pathways present in a sponge symbiotic system. In principle, the methods include generating a highly complex metagenomic library that presumably contains representative bacterial genomes present in the sponge and subsequent screening the library rapidly. All of the experimental approaches used for a sponge sample in this work can be divided into following main experimental parts: (i) finding key biosynthetic genes, (ii) generating a complex metagenomic library, (iv) rapid screening of a metagenomic library for obtaining biosynthetic gene cluster(s) of interest, and (v) clone analysis. In addition, genetic analysis of an uncultured bacterial type, in which the microbial complexity was reduced by cell separation, was also described. Some basic knowledge behind the experimental steps is briefly described.

4.1. Materials

4.1.1 Organisms and Vectors

This subsection lists all of the organsisms and vectors used in this work. In addition, most of genetically modified bacteria as well as DNA constructed made in this work are presented.

Table 4.1.1 Organisms used in this work

Microorganisms/ bacterial hosts	Source/genotype
E. coli EPI300™-T^{1}R Phage T1-resistant	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻ rpsL nupG trfA dhfr*
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F'*proAB lacI^q ZΔM15 Tn10* (Tet^R)]
An engineered strain of *Acinetobacter baylyi* ADP1	harboring a 77-kb gene cluster coding for pederin
The sponge *Theonella swinhoei* W (white) specimen	collected by scuba diving at Hachijo-jima Island, Japan, at a depth of 15 m. Immediately after collection, it was shock-frozen in liquid nitrogen and stored at -80 ^{o}C. Provided in 2002 through collaboration with Prof. S. Matsunaga (University of Tokyo)
Uncultivated filamentous bacterial fraction 1 (in sub-chapters 3.1 & 3.2)	prepared from *T. swinhoei* chemotype white. Provided in 2002 through collaboration with Prof. S. Matsunaga (University of Tokyo)
Uncultivated filamentous bacterial fraction 2 (described in sub-chapter 3.4)	prepared from *T. swinhoei* chemotype yellow. Provided in 2010 through collaboration with Dr. T. Mori, Prof. M. Takeyama (Waseda University), Dr. T. Wakimoto, and Prof. S. Matsunaga.
Uncultivated filamentous bacterial fraction 3 (described in sub-chapter 3.4)	prepared from *T. swinhoei* chemotype white. Provided in 2010 through collaboration with Dr. T. Mori, Prof. M. Takeyama, Dr. T. Wakimoto, and Prof. S. Matsunaga

Table 4.1.2 Vectors used in this work

Vectors	Source/genotype
pBluescript II SK (-)	Stratagene (Amsterdam, Netherlands)
pCC1FOS	Epicentre (Hess. Oldendorf, Germany)

Table 4.1.3 Genetically modified organisms (GMO) generated in this work

Microorganisms/ bacterial hosts	Source/genotype
Fosmid library of an engineered strain of *Acinetobacter baylyi* ADP1 hosted in *E. coli* EPI300™ using pCC1FOS	An engineered strain of *Acinetobacter baylyi* ADP1 harboring a 77-kb gene cluster coding for the antitumor polyketide, pederin
Fosmid library of filamentous bacteria (~2000 clones, 18 pools) hosted in *E. coli* EPI300™ using pCC1FOS	The filamentous bacteria were prepared from *T. swinhoei* chemotype white
Fosmid library of sponge metagenome I (~200,000 ; 4000 cfus/ml) hosted in *E. coli* EPI300™ using pCC1FOS	The metagenome were prepared from *T. swinhoei* chemotype white
Fosmid library of sponge metagenome II (~200,000 ; 1000 cfus/ml) hosted in *E. coli* EPI300™ using pCC1FOS	The metagenome were prepared from *T. swinhoei* chemotype white
Fosmid library of sponge metagenome III (~200,000 ; 500 cfus/ml) hosted in *E. coli* EPI300™ using pCC1FOS	The metagenome were prepared from *T. swinhoei* chemotype white

Table 4.1.4 Overview of the DNA constructs made in this work

Vectors	Host	Descriprion
pAU1, 3, 4, 5, 6, 7	*E. coli* XL1-Blue	pBluescript II SK (-) with a A domain sequences for arginine
pEt-A3, -B2, -D1	*E. coli* XL1-Blue	pBluescript II SK (-) with a A domain sequences for serine
pAU12	*E. coli* EPI300™	pCC1FOS with a pederin gene cluster fragment
pAU14	*E. coli* EPI300™	pCC1FOS with a pederin gene cluster fragment
pET3.7a	*E. coli* EPI300™	pCC1FOS with a probably small NRPS gene cluster
pTSW4, 8, 9, 12, 30	*E. coli* XL1-Blue	pBluescript II SK (-) with KS fragments of PKS-like FASs
pEKS14, 30, pEP5	*E. coli* XL1-Blue	pBluescript II SK (-) with KS fragments of *cis*-AT PKSs
pEKS27, 37, 48, 50	*E. coli* XL1-Blue	pBluescript II SK (-) with KS fragments of *trans*-AT PKSs
pEKS64, pEP2, 3	*E. coli* XL1-Blue	pBluescript II SK (-) with KS fragments of *trans*-AT PKSs
pTSW-AU1	*E. coli* EPI300™	pCC1FOS with a misakinolide A gene cluster fragment
pTSW-AU2	*E. coli* EPI300™	pCC1FOS with a misakinolide A gene cluster fragment
pTSW-AU5	*E. coli* EPI300™	pCC1FOS with a misakinolide A gene cluster fragment
pTSW-AU10	*E. coli* EPI300™	pCC1FOS with a misakinolide A gene cluster fragment
pTSW-AU11	*E. coli* EPI300™	pCC1FOS with a misakinolide A gene cluster fragment
pADU2a	*E. coli* EPI300™	pCC1FOS with a keramamide gene cluster fragment
pAG	*E. coli* EPI300™	pCC1FOS with a small NRPS gene fragment
pW2, pW3, pW10	*E. coli* XL1-Blue	pBluescript II SK (-) with 16S rDNAs of *Spongiobacter* sp.
pW4	*E. coli* XL1-Blue	pBluescript II SK (-) with 16S rRNA fragment of *Chloroflexus* sp.
pW19, pW24	*E. coli* XL1-Blue	pBluescript II SK (-) with 16S rRNA fragment of *Spirochaeta* sp.
pY3, pY5, pY6	*E. coli* XL1-Blue	pBluescript II SK (-) with 16S rRNA fragment of *Entotheonella* sp.
pY4	*E. coli* XL1-Blue	pBluescript II SK (-) with 16S rDNA of *Rhodobacteraceae*
pY11	*E. coli* XL1-Blue	pBluescript II SK (-) with 16S rRNA fragment of *α-proteobacterium*

4.1.2 Chemicals, solutions, and instruments

Table 4.1.5 Chemicals utilised in this work

Chemical	Source
acetic acid	VWR
Agar	Merck-Eurolab,
Agarose, NEEO ultra	Carl Roth GmbH
Agarose, low melt	Carl Roth GmbH
Ammonia, liquid	Merck-Eurolab
Ammonium chloride	Merck-Eurolab
Ammonium sulfate	Fluka/ Sigma-Aldrich
Ampicillin	Merck-Eurolab
Boric acid	Merck-Eurolab
Bromophenol Blue	Sigma-Aldrich
Calcium chloride dihydrate	Fluka/ Sigma-Aldrich
Calcium carbonate	Merck-Eurolab
Casaminoacids	Life Technologies (Gibco BRL)

Chloramphenicol	Sigma-Aldrich
Chloroform	SDS GmbH
Coomassie® Brilliant Blue G250	Sigma-Aldrich
CTAB (Cetrimonium bromide)	Fluka
Deoxynucleoside triphosphates (dNTP's)	Invitrogen GmbH
Dimethyl sulfoxide (DMSO)	Carl Roth GmbH
Dithiothreitol (DTT)	Fluka/ Sigma-Aldrich
D-Sorbitol	Carl Roth
Ethanol	SDS GmbH
Ethidium bromide	Carl Roth GmbH
Ethidium bromide solution (1%)	Merck-Eurolab
Ethyl acetate	SDS GmbH
Ethylenediamine-tetraacetic acid (EDTA)	Fluka/ Sigma-Aldrich
Glacial acetic acid (99%)	Merck-Eurolab
Glucose	Merck-Eurolab
Glutamic acid	Fluka/ Sigma-Aldrich
Glycerol	Merck-Eurolab
IPTG (Isopropyl-ß-D-thiogalactopyranoside)	Carl Roth GmbH
Isopropanol	Aldrich
Kanamycin	Sigma-Aldrich
LMW-SDS Marker Kit	GE Healthcare Bio-Sciences
L(+)-Arabinose	Acros Organics
Lithium chloride	Alfa Aesar
Lysozyme	Merck-Eurolab
2-Mercaptoethanol	Carl Roth
Magnesium chloride hexahydrate	Merck-Eurolab
Magnesium sulfate heptahydrate	Merck-Eurolab
Manganese chloride tetrahydrate	Merck-Eurolab
Mercaptoethanol	Serva
Methanol	SDS GmbH
Nicotinamide adenine dinucleotide (ß- NAD^+)	Sigma-Aldrich
Nitrogen, liquid	Institut für Strahlen- und Kernphysik
Orange G	Sigma-Aldrich
Peptone	Carl Roth GmbH
Phenol-chloroform-isoamyl alcohol mixture (25:24:1)	Carl Roth GmbH
Potassium acetate (KOAc)	Carl Roth GmbH
Potassium chloride	Carl Roth GmbH
Potassium dihydrogen phosphate	Merck-Eurolab
di-Potassium hydrogen phosphate	Merck-Eurolab
Potassium sulfate	Fluka/ Sigma-Aldrich
2-Propanol	Merck-Eurolab
SeaPrep Agarose	Biozym
Sodium acetate	Carl Roth GmbH
Sodium chloride	Merck-Eurolab
Sodium dodecyl sulfate (SDS)	Carl Roth GmbH
Sodium hydroxide	Merck-Eurolab
Sodium hydroxide solution 50%	Sigma-Aldrich
Sodium tetraborate	Fluka/ Sigma-Aldrich
Sucrose	Merck-Eurolab
TES-buffer (*N*-[Tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid)	Fluka/ Sigma-Aldrich
Tris-Base (Tris-[hydroxymethyl]-aminomethane)	Merck-Eurolab
Tryptone	Carl Roth GmbH
Urea	Roth
X-gal (5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside)	Life Technologies (Gibco BRL)
Xylene Cyanol FF	Sigma-Aldrich
Yeast Extract (Select®)	Life Technologies (Gibco BRL)

Table 4.1.6 Kits used in this work

Name	Manufacturer
Epicentre® CopyControl™ Fosmid Library Production Kit CCFOS110	Epicentre Biotechnologies, Madison, WI 53713, USA
QIAquick Gel Extraction Kit	Qiagen GmbH, 40724 Hilden
RNeasy® Mini Kit	Qiagen GmbH, 40724 Hilden
peqGOLD Gel Extraction Kit	PeqLab Biotechnologie GmbH
QIAquick Gel Extraction Kit	Qiagen
GFX PCR DNA and Gel Band Purification Kit	GE Healthcare
CopyControl™ Induction Solution	Epicentre
RNeasy Mini Isolation from Bacteria	Qiagen
Expand High Fidelity plus PCR system	Roche
SuperScript™ III One-Step RT-PCR System	Invitrogen
QIAGEN Genomic-tip 20/G	Qiagen

Table 4.1.7 Enzymes employed in molecular biological experiments

Enzyme	Manufacturer
Antarctic phosphatase	New England Biolabs, 65926 Frankfurt
Bovine serum albumin (BSA)	New England Biolabs, 65926 Frankfurt
DNase (RNase-free)	Promega, 68199 Mannheim
GoTaq® DNA Polymerase (5u/µl)	Promega, 68199 Mannheim
RNAse A	Invitrogen GmbH, 76131 Karlsruhe
T4 DNA Ligase	New England Biolabs, 65926 Frankfurt
Restriction endonucleases	New England Biolabs, 65926 Frankfurt
Proteinase K	Carl Roth GmbH, 76185 Karlsruhe

Table 4.1.8 Molecular weight markers for gel electrophoresis

Name	Source
1 Kb DNA Ladder	Invitrogen GmbH, 76131 Karlsruhe
1 Kb DNA Extension Ladder	Invitrogen GmbH, 76131 Karlsruhe
GeneRuler™ DNA Ladder Mix	Fermentas, 68789 St-Leon-Rot

Table 4.1.9 Solution, buffers and media

Solution, Buffer and Media	Composition	Store
Alkaline lysis solution I (P1)	50 mM Tris-HCl pH 8, 10 mM EDTA, 100 µg/ml RNAse	4 °C
Alkaline lysis solution II (P2)	200 mM NaOH, 1% SDS	RT
Alkaline lysis solution III (P3)	3 M KOAc pH 5.5	
Gel loading dye	0.05% bromophenol blue, 0.05% xylene cyanol, 0.05% orange G, 40% sucrose	4°C
Sarkosyl lysis buffer pH 7.6	8 M urea, 2% sarkosyl, 350 mM NaCl, 50 mM EDTA, 50 mM Tris-HCl, pH 7.5	
TAE 50X	242 g Tris-base, 7.1 ml Glacial acetic acid. 100 ml EDTA (0.5M), add 1 l H_2O	RT
TE buffer	10 mM Tris-HCl (pH 7.5) 1 mM EDTA (pH 8.0)	RT
Luria broth (LB)	1% tryptone, 1% NaCl, 0.5% yeast extract	RT
LB agar	1% tryptone, 1% NaCl, 0.5% yeast extract, 1.5% agar	RT
Semi-liquid SeaPrep medium	1% tryptone, 1% NaCl, 0.5% yeast extract, 0.6% SeaPrep agarose	RT

All media were autoclaved at 121°C. Put in microwave in 5 min

Table 4.1.10 Antibiotics for the counterselection of recombinant strains

Antibiotic	Concentration Medium (µg/ ml)	Concentration Stock solution (mg/ ml)	solvent
Ampicillin	100	100	water
Chloramphenicol	25	25	ethanol
Kanamycin	25	25	water

Table 4.1.11 Enzymes

Enzyme	Source
Antarctic phosphatase	Roth, Karlsruhe
Hot Start Polymerase	Jena Bioscience, Jena
Lysozyme	Roth, Karlsruhe
Restriction enzymes	NEB, Frankfurt/Main, Jena Bioscience, Roth, Karlsruhe
RNase	Roth, Karlsruhe
RNase-free DNase set	Qiagen, Dusseldörf
T4 DNA ligase	NEB, Frankfurt/Main, Jena Bioscience, Jena
Taq DNA Polymerase	NEB, Frankfurt/Main
Plasmid-Safe™ ATP-Dependent DNase	Epicentre

Table 4.1.12 Equipments

Equipment	Manufactory
Agarose gel electrophoresis	Biometra, Göttingen
Autoclave V65	Systec, Wettenberg
Balance	Sartorius, Germany
Cellophan paper	Roth, Karlsruhe
Concentrator 5301/ speedvac	Eppendorf, Hamburg
Clean bench	Bio-flow Technik, Meckenheim
Electroporation cuvette 2mm	Bio-Rad, München
Filter	Merck, Darmstadt

Equipment	Manufactory
Gel documentation Gene Genius	Syngene, Cambridge (UK)
Glass wool	Roth
IKA-combimag	Janke and Kunkel
Shaking incubator	Sartorius, Göttingen
Incubator	Thermo, Langenselbold
Membranvacuumpumpe	Vacuubrand, Wertheim
Microcentrifuge Micro200	Hettich, Tuttlingen
Microcentrifuge, cool 5417R	Eppendorf, Hamburg
Microcentrifuge, cool Micro200R	Hettich, Tuttlingen
Microprocessor pH meter	Hanna instruments
MicroPulser Electroporator	Bio-Rad, München
Microwave Lifetec	Medion, Essen
Mixer uzusio VTX-3000L	Harmony
Photometer (Biophotometer)	Eppendorf, Hamburg
Photometer BioMate 3	Thermo Electron, Cambridge (UK)
Pipett Pipetman P2 - P10 ml	Gilson, Middleton (USA)
Poly-prep column	Bio-Rad, München
Rotamax 120	Heidolph, Kelheim
Sonipulser	Badelin electronic, Berlin
Thermomixer	Eppendorf, Hamburg
Thermocycler T-Gradient PCR	Biometra, Göttingen
Pipette tip	Sarstedt
Ultraviolet crosslinker	UVP, Cambridge, UK
UV cuvettes	Sarstedt
Pipette tip	Sarstedt
Ultraviolet crosslinker	UVP, Cambridge, UK
UV cuvettes	Sarstedt

4.2. General Molecular Biology and Microbiology Techniques

This section describes general biomolecular and microbiology methods regularly used in this work.

4.2.1 DNA isolation and electrophoresis

The basic steps for isolation of bacterial DNA include bacterial cultivation, cell lysis, protein-nucleic acid complex disruption, protein removal, and DNA precipitation. PCR applications often require less purified DNA or crude DNA extract (Chachaty and Saulnier, 2000). The protein is dissociated from nucleic acids by detergents such as SDS. Protein removal is usually done by extracting cell lysates with phenol:chloroform:isoamyl alcohol (Kirby, 1964). Subsequent extraction with chloroform removes traces of phenol from the DNA preparation. Before extracting

with such organic solvents, proteins are often degraded by treating in the nucleic acid solution with proteolytic enzymes, such as proteinase K (Sambrook & Russel, 2001).

4.2.1.1 Bacterial cultivation

Overnight cultures were performed either for genomic DNA extractions, preparation of circular vectors (*e.g.* plasmids), or preparation of competent cells. At first, a sterile test tube was filled with 3-5 mL of LB medium containing an appropriate antibiotic. Using a sterile toothpick or innoculating needle, bacterial cells were picked from a single colony growing on a plate or from the stock culture frozen at -80 °C frozen, and then innoculated into the LB medium. The tube was placed in a shaking incubator at 37 °C overnight (~16 h) at 200 rpm. For long term storage of the bacterial strains and vectors, 850 µL of an appropriate overnight culture were transferred in 2 ml screw-capped tubes and added with 150 ml of 100% steril glycerol. After a short time vortex, the tubes were stored at -80 °C for a long term. Glycerol has been well known for its utility in preserving viable *E. coli* cells at very low temperatures.

4.2.1.2 Genomic DNA isolation

To prepare genomic DNA from bacteria, we used the procedure described by Jones and Bartlet (1990) with slight modifications. An overnight culture (1 ml) was transferred to a 1.5 ml eppendorf tube and centrifuged at 10,000 rpm for 5 minutes to precipitate the cells. The free-cell supernatant was decanted, and another 1 ml culture was added, followed by centrifugation. The resulting cell pellet were suspended with 438 µl lysis buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.8, 1 mM EDTA, 2.5 mM $MgCl_2$, 1% SDS) and then stored at -80 ^{o}C for 15 minutes. Then 50 µl of 100 mg/ml lysozime was added into the tube. The tube was incubated on dry ice for 20 minutes and then at room temperature until the frozen lysate melts. Subsequently the lysate was treated with 2 µl of 30 mg/ml proteinase K, followed by 1 hour incubation at 37 ^{o}C. The nucleic acid solution was extracted by adding a mixture of phenol:chloroform:isoamyl alcohol (25:24:1), followed by centrifuging it at 14,000 rpm for 2 minutes. The upper layer was transferred into a new eppendorf tube, and then extracted once with chloroform. One volume of isopropanol was added, mixed gently, and centrifuged at 14,000 rpm for 30 minutes (4 ^{o}C). The resulting DNA pellet was washed with 500 µl of 70% ethanol, and then vacuum-

dried for around 15 minutes. The dried DNA pellet was dissolved in 20 µl of EB buffer (Qiagen).

The following procedure used for high molecular weight DNA isolation from uncultivated bacterial cells was reported by Syn and Swarup (2000) with slight modification. Cell suspension was initially centrifuged. The resulting cell pellet was washed once with 750 µl NE buffer (0.15 M NaCl, 50 mM EDTA) to remove extracellular exopolysaccharides (EPS) and then repelleted by centrifugation. The cell pellet was then resuspended in 750 µl of TES (10 mM Tris-HCl, 10 mM EDTA, pH 8.0, 2% SDS), and incubated at 75 °C for 5 min (15 min) to improve cell lysis. The lysate was treated with two rounds of phenol:chloroform:isoamyl alcohol (25:24:1, v/v) extraction, followed by a chloroform extraction. DNA precipitation was done by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 1 volume of isopropanol (room temperature). The sample was mixed by gently inverting the tube several times. The DNA was precipitated by centrifugation at 14,000 rpm for 30 minutes (room temperature). The DNA pellet was washed with 70% ethanol and centrifuged at 14,000 rpm for 10 minutes. It was then air-dried, and resuspended in 15 µl of EB buffer (Qiagen).

The following DNA isolation method was developed in this work. The cell pellet stored in 1 ml ethanol at 4 ^{o}C was suspended by inverting several times. A half volume of the cell suspension (200 µl) was taken out and rinsed three times with dH_2O. The rinsed cell pellet obtained was weighed and suspended in 230 µl of buffer lysis I (10 mM Tris-HCl pH 7.0, 1 mM EDTA, 100 mM NaCl, 0.08 mg/ml RNase). To cell suspension were then added 20 µl of 20 mg/ml lysozyme and incubated at room temperature for 30 minutes. Subsequently, the mixture was treated with 250 µl of buffer lysis II (10 mM Tris-HCl pH 6.0, 50 mM EDTA, 100 mM NaCl, 5 mM $MgCl_2$, 2% SDS, 4 M urea) and 3 µl of 20 mg/ml proteinase K containing 0.02 mM $CaCl_2$, followed by a 10-min incubation at room temperature. An equal volume of phenol/chloroform/isoamyl alcohol (25/24/1) (Roth) was added to the lysed mixture, mixed gently by inverting it several times, and centrifuged at 13,000 rpm for 5 min at 4 ^{o}C. The upper layer was pipetted into a 1 ml centrifuge tube, and re-extracted with equal volume of phenol/chloroform/isoamyl-alcohol as before. The upper layer was transferred into a 1 ml centrifuge tube containing 200 µl chloroform, mixed gently and centrifuged at 13,000 rpm for 5 min. An equal volume of cold isopropanol and 1/10 volume of 3 M sodium acetate was added to the DNA

solution, mixed gently and stored at -20 ^{o}C for 20-30 minutes. The precipitated DNA was recovered by a 30-min centrifugation at the top speed (14,000 rpm), rinsed with 500 µl of 70% ethanol, and air-dried. The DNA pellet was dissolved in 20-30 µl EB buffer (Qiagen) diluted two times with dH_20.

4.2.1.3 Plasmid or fosmid DNA isolation

Preparation of plasmid or fosmid DNA from recombinant *E. coli* cells was based on the alkaline lysis in combination with SDS (modifified from Birnboim and Doly, 1979). The modified alkaline lysis procedure of the cells described below was performed for small-scale preparation of plasmid/fosmid DNA (1.5 ml overnight culture). An overnight *E. coli* culture (0.75 ml) was transferred into a 1.5 ml Eppendorf tube and centrifuged at 10,000 rpm for 3 min. The cell-free supernantant was discarded, and another *E. coli* culture (0.75 ml) was transferred into the same tube. After centrifugation at 10,000 rpm for 3 min, the cell-free supernatant was removed. The resulting cell precipitate was suspended with 200 µl P1 buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 100 µg/ml RNAse), and then mixed well by inverting the tube or by vortexing. The RNAse solution was boiled for 2 min before being added into the P1 buffer to inactivate DNAse activity. The suspended cells were added with 200 µl of P2 buffer (200 mM NaOH, 1% SDS), and mixed by inverting the tube several times. The suspension was added with 200 µl P3 (3M KOAc pH 5.5), mixed again and centrifuged at 13,000 rpms for 5 min. The supernatant was transferred to a new Eppendorf tube containing 500 µl chloroform, mixed well and centrifuged at 13,000 rpm for 5 min, resulting two layers of solution. The upper layer containing plasmid DNA was transferred to a 1.5 ml tube containing 350 µl isopropanol, mixed well, and centrifuged at top speed for 20 min. The supernatant was poured off and the resulting small white pellet was rinsed with 500 µl of 70 % ice-cold ethanol. After spinning the pellet down shortly, the ethanol was carefully poured off and the remaining liquid was removed. The pellet was then vacuum-dried for 20 min and dissolved in 50 µl water. One µl of the plasmid solution was checked by restriction digest and agarose gel analysis. For large-scale preparation of plasmid or fosmid DNA from 150 ml overnight culture, the volume of all the buffers and solutions used were multiplied by 100. The Eppendorf tubes were substituted with 50 ml facon tubes.

4.2.1.4 DNA purification and quantification

Low-molecular weight contaminants in the isolated DNA samples may inhibits

downstream applications such as PCR amplification, enzymatic restriction, DNA ligation or DNA sequencing. Inhibitors can be removed from crude DNA preparations either by drop dialysis (Sambrook and Russel, 2001), gel permeation chromatography (Zadražil *et al*., 1974; Jakson *et al*., 1997), affinity chromatography (Chockalingam *et al*., 2001), or gel purification using agarase digestion (Zhou *et al*., 1996). In our work, additional DNA purification step was usually carried out using gel purification kits that combines spin-column technology and silica-gel matrices or anion-exchange carriers, such as QIAquick Gel Extraction Kit (Qiagen), the GeneJETTM Gel Extraction Kit (Fermentas), and the illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare).

The target DNA fragment was first excised from the agarose matrix after the electrophoretic separation. One of the gel extraction kits were then used according to the manufacturer's instruction manual. To avoid DNA damage by exposure to UV light, ethidium bromide was not added into the agarose gel. A larger well was made by sealing some wells in the comb with a gel-sealing tape. Using this comb with a larger well formed a wide slot on the solidified gel. Then the DNA sample was loaded into a wide slot of the agarose gel, with a DNA marker on the right and left sides. After electrophoresis, both sides of the gel including the marker lines were cut off with a sharp scalpel, and then stained with 1 µg/ml ethidium bromide solution for 20 minutes. The stained parts of the gel were observed under UV light and the position of target DNA band(s) was marked with a scalpel. The gel parts were subsequently reassembled, and the target fragments were marked, excised from the gel, and subjected to gel extraction. The concentration and purity of the purified DNA samples were then measured using a spectrophotometer. This spectrophotometric measurement is based on the amount of UV irradiation absorbed by the DNA bases at wavelengths of 260 nm and 280 nm. Using a cuvette with a seam of 1 cm, the OD_{260} of 1 corresponds to a concentration of 50 µg/ml for double-stranded DNA and 40 µg/ml for single-stranded DNA or RNA. The ratio between OD_{260} and OD_{280} indicates the estimated purity of the DNA preparations. Pure DNA samples have OD_{260}/OD_{280} of 1.8 to 2.0 (Sambrook & Russel, 2001).

4.2.1.5 DNA electrophoresis

The first modern electrophoresis apparatus, reffered to horizontal "submarine" gel system, was developed by Walter Schaffner, who was then a graduate student at

Zurich. This gel electrophoresis model has developed as the standard technique widely applied to separate, identify and purify DNA fragments (Sambrook & Russel, 2001; Rickwood & Hames, 1982). The DNA separation on the gel is based on the negative charge of DNA molecules due to the negatively-charged oxygen of the phosphate group of their backbone. Therefore, when the gel containing DNA fragments are placed in an electrical field, they migrate toward the positive pole of electrophoresis chamber. Their speed of movement in the gel decreases as their length increases, which is proportional to electric field strength (Sambrook & Russel, 2001).

An agarose gel was made by mixing agarose with 1X TAE and dissolving by heating. The concentration of agarose gel depended on the size of analyzed DNA, normally ranged from 0.5-2%. We routinely used 1% agarose gels for the DNA size range of 500 bp to 10 kb, whereas 1.5-2.0% for the DNA fragment size of less than 500 bp. While the agarose solution was cooling to <60 ^{o}C, it was added with ethidium bromide (EtBr) to a final concentration of 0.1 µg/ml followed by stirring to mix it. The warm agarose solution was poured into the plastic mold/tray and an appropriate comb was placed in the molten agarose. When the gel hardens, a small amount of electrophoresis buffer was poured on the top of the gel and then the comb was carefully removed. The gel along with the mold was placed in the electrophoresis tank, and the electrophoresis buffer was poured to cover the gel. Individual DNA samples were mixed with 1/10 volume of the loading dye. This dye contained a mixture of glycerol, xylene cyanol FF (migrates at ~5000 bp), bromophenol blue (migrates at ~300 bp) and orange G (migrates at ~50 bp). Each DNA sample mixture was loaded into the slots of the gel. Size markers were usually loaded on the right side. The lid of the gel tank was closed and a voltage of 80-120 V was applied. Then the DNA samples in the gel were visualized under UV light in the gel documentation system. If EtBr was not directly included in the gel matrix, staining of the gel can be done after electrophoresis by immersing it in 1X TAE containing 1 µg/ml EtBr for 15-20 min. Gel analysis was performed using a UV transilluminator and a digital camera (Sambrook & Russel, 2001)

4.2.2 PCR Amplification

Since polymerase chain reaction (PCR) was first described in 1986 (Mullis *et al*., 1986; Saiki *et al*., 1988), it has rapidly developed as an invaluable tool for molecular biology research (Bej *et al*., 1991; Sambrook and Russel, 2001; Jena Bioscience;

Invitrogen). PCR is intended to amplify a target DNA to yield millions of identical copies. It is routinely used in laboratories around the world for a widespread of applications, including cloning, sequencing, gene expression analysis, and mutagenesis (Sambrook and Russel, 2001; Jena Bioscience; Invitrogen).

4.2.2.1 Standard PCR amplification

A PCR procedure commonly includes three major steps: denaturation, annealing, and elogation (Sambrook & Russel, 2001). At denaturation step, the double-stranded DNA is denatured to single-stranded DNA by heating to 94 ^{o}C. At the annealing step, the temperature is lower to 40-60 ^{o}C that allows primers to anneal to the single-stranded DNA. At the elongation step, polymerase binds to the primer-DNA template hybrid for DNA synthesis in direction 5'-3'. Sometimes, the first cycle is preceded by a predenaturation step at 94 to 95 ^{o}C for 3-5 min. The elongation step is often followed by final elongation step by heating to 70-72 ^{o}C for up to 10 min to ensure that the single-stranded DNA is completely extended. A PCR typically consists of about 20-35 cycles, relying on the concentration of PCR compositions (Bej *et al*., 1991; Sambrook & Russel, 2001, Jena Bioscience). The following PCR composition and program are generally used in the Piel group.

PCR component		Volume
10X ThermoPol buffer (NEB)	:	2.5 µl
10 mM of dNTPs 10 mM	:	0.5 µl
50 µM of forward primer	:	0.25 µl
50 µM of reverse primer	:	0.25 µl
Taq polymerase (NEB)	:	0.125 µl
DNA template	:	0.5 µl
dH_2O	:	up to 25 µl
PCR program		**Temperature & duration**
Lid preheating	:	105 ^{o}C
Predenaturation	:	95 ^{o}C 2 min
Denaturation	:	95 ^{o}C 30 sec
Annealing	:	T_m ^{o}C 1 min (T_m is annealing temperature)
Elongation	:	72 ^{o}C 1 min
Final elongation	:	72 ^{o}C 5 min
Storing	:	4 ^{o}C pause
Thermal cycles	:	35

4.2.2.2 Hot-start PCR amplification

To reduce nonspecific amplification and obtain the higher yield of the desired PCR product, we used Hot-start polymerase (Jena Biosciences). The activity of this enzyme is inhibited at ambient temperatures, thereby reducing nonspecific priming and primer oligomerization during PCR set-up (Birch *et al*., 1996; Chou *et al.*, 1992; Kellogg *et al*, 1994). During the initial denaturation step of the PCR program, the enzyme becomes fully active (Jena Bioscience). The Hot start polymerase has proofreading activity but also has the template-independent activity that creates a single additional adenosine at 3`-OH termini (Jena Bioscience), thereby facilitating direct ligation with a vector carrying T-overhang (Clark, 1988). For DNA targets with a high G + C content, primers are often difficult to reanneal at the secondary structure regions. To solve this problem, certain organic chemicals such as DMSO (dimethyl sulfoxide) often need to be added into the PCR reaction (Bej *et al*., 1991). The addition of DMSO in low concentration can often greatly improve yield and specificity without affecting Taq DNA polymerase activity (Bej *et al*., 1991; Varadaraj and Skinner, 1994). Furhermore, the addition of either bovine serum albumin (BSA), dithiothreitol, or glycerol in the PCR reaction mixture allowed the specific amplification, which were previously failed under standard conditions (Nagai *et al*., 1997). In this work, we used hot start polymerase (Jena Bioscience), sometimes in combination with DMSO and BSA, to obtain target amplicons from the sponge metagenome or uncultered symbiotic bacteria. The PCR composition and program are as follow.

For crude sponge DNA as the template:

The PCR composition for a 25-µl PCR mix: 10x buffer, 0.5 µl of 10 mM dNTPs, 0.25 µl of 100 mM forward primer, 0.25 µl of 100 mM reverse primer, 1.5 µl of 25 mM $MgCl_2$, 1.25 µl of 100x BSA, 2.5 µl of 100% DMSO, 2.5 µl of DNA sample, 0.125 µl Hot-start DNA polymerase, 1 µl DNA template (diluted 10x, 100x, and 1000x) and ddH_2O up to 25 µl. The PCR reaction program was composed of lid heating at 105 ^oC, predenaturation at 96 ^oC for 5 minutes, denaturation at 96 ^oC for 45 seconds, annealing at 55, 58, 61 ^oC, elongation at 72 ^oC for 1 minute, and final elongation at 72 ^oC for 7 minutes.

For pure sponge DNA as the template:

The PCR components for a 25-µl PCR mix are composed of 21.25 µl ddH_2O, 2.5 µl of 10x thermophol buffer (NEB), 0.5 µl of 10 mM dNTPs, 0.25 µl of 50 mM forward

primer, 0.25 µl of 50 mM reverse primer, 0.125 µl of Hot-start DNA polymerase or high-fidelity polymerase, 0.4 µl pure sponge DNA or bacterial cell suspension.

For suspension of bacterial cells as the template:

The PCR composition for the total volume of 25 µl consisted of 17 µl ddH_2O, 2.5 µl of 10x thermophol buffer (NEB), 0.5 µl of 10 mM dNTPs, 1 µl of 25 mM $MgCl_2$, 2 µl of 50 mM KSDPQQF primer, 2 µl of 50 mM KSHGTGR primer, and 0.125 µl Hot-start DNA polymerase or high fidelity polymerase. The thermal cycle program was composed of lid heating at 105 ^oC, predenaturation at 95 ^oC for 2 minutes, denaturation at 95 ^oC for 1 minute, annealing at 55, 58, 61 ^oC, elongation at 74 ^oC for 1 minute, and final elongation at 74 ^oC for 10 minutes.

4.2.3 DNA Modifications

A wide range of commercially available enzymes can be used to modify DNA, such as restriction/ digestion, phosphorylation/ dephosphorylation, ligation, and polymerization. In this work, we used a variety of endonucleases and exonucleases for DNA digestion, phosphatases for removing 5` phosphate groups to prevent self-ligation, a DNA polymerase for adding additional nucleotides on the DNA termini, and DNA ligase to combine DNA molecules. In addition, combination of T4 DNA polymerase and terminal phosphotransferase is usually used to repair the DNA termini. T4 DNA polymerase fills the recessed 3` termini and remove the protruding 3` termini, thereby converting the frayed ends into blunt ends. Terminal phosphotransferase converts the hydroxyl group at the 3` termini to a phosphate group using that is also present in the end-repairing enzyme mix (Brown, 2006).

4.2.3.1 DNA restriction

Two µl of the total 50 µl miniprep was used to check the presence of an inserted amplicon by treating it with suitable restriction enzymes that flank the insert. The following composition was used for double digestion: 5.5 µl of ddH_2O, 2 µl of DNA, 1 µl of NEB buffer, 0.2 µl of enzyme 1 and 0.2 µl of enzyme 2. The reaction mixture was incubated at 37 ^oC for 3-5 hours. Then the digestion products were run on the electrophoresis to check for the correct insert size. The clones harbouring correct-size inserts were subjected to restriction fragment length polymorphism (RFLP) in order to obtain insight into the difference in restriction pattern. The restriction enzymes used for RFLP in this work were *Rsa*I, *Hha*I, or combination of both. The

clones showing unique restriction patterns were sequenced. The sequence data was analyzed by BLAST search to find the homology in the public database. For KS and adenylation domain sequences, the highly conserved motifs were identified.

4.2.3.2 Preparation of T-vectors for PCR cloning

PCR cloning was based on the ability of *Taq* polymerase to deoxyribonucleotide to the 3`-OH terminus of a PCR amplicon through its template-independent terminal transferase activity (Clark, 1988). As the result, the PCR products tend to harbor A-overhang at their ends, usually adenosine, which can readily be ligated with a phagemid vector harboring T-overhang at both ends (Marchuk *et al*., 1990). This template-independent activity of Taq polymerase can be exploited to create a cloning scheme which has the efficiency of sticky end cloning, but requires no additional enzymatic modification of the PCR product. To make T-cloning vector, the phagemid pBluescript II (SK^-) was liniearized first by using *EcoR*V (Jena Bioscience) to create blunt-end sites. At first, pBluescript (Stratagene, La Jolla, CA) plasmid (10 µl from 50 µl standard miniprep) was added with 32 µl ddH_2O, 5 µl NEB buffer 3, 0.5 µl BSA, and 2.5 µl of *Eco*RV. After 2-hours incubation at 37^oC, the digestion reaction was inactivated by heating at 80^oC for 20 min. The digestion product was treated with 0.5 µl of *Taq* polymerase (1 unit/ µg plasmid/20 µl volume) and 10 µl of 10 mM dTTP and incubated at 70 ^oC for 2 hours. The linear plasmid was recovered by chloroform extraction and isopropanol precipitation. The resulting T- vector is ready to ligate with PCR product containing A-overhang (Marchuk *et al*., 1990).

4.2.3.3 Preparation of vectors for subcloning

For the subcloning of cosmid DNA fragments into pBluescript II SK (+) vector, both insert and vector DNA were digested with the same restriction enzyme. The following standard protocols were regularly used in the Piel group:

Reaction for insert DNA:

Cosmid DNA	:	50.0	µl
Endonuclease	:	10.0	µl
Buffer	:	15.0	µl
BSA (100x)	:	1.25	µl
Purified water	:	23.75	µl

Incubation time at 37 °C was extended to 3 hours to ensure complete digestion. The restriction reaction was applied to a gel without ethidium bromide. The DNA fragments were excised separately, as far as possible, and were then recovered from the gel for usage in ligation reactions.

Reaction for vector DNA:

Vector DNA	:	20.0 µl
Endonuclease	:	2.5 µl
Buffer (10x)	:	5.0 µl
BSA (10x)	:	5.0 µl
Purified water	:	17.5 µl

After 2 hours of incubation at 37 °C, inactivation of the enzyme was achieved by heating at 65 °C for 20 minutes. The DNA was then dephosphorylated (section 4.2.3.4).

4.2.3.4 Dephosphorylation of vectors

Antarctic phosphatase was used to catalyses the removal of 5´ phosphate groups from a vector DNA. The following treatment was used for the dephosphorylation of a linearized vector according to NEB Inc:

Dephosphorylation reaction for a linearized vector DNA:

Linear pBluescript DNA	:	40 µl
10x Antarctic phosphatase reaction buffer	:	6 µl
Antarctic phosphatase	:	2 µl
dH_2O	:	12 µl
Total	:	60 µl

The dephosphorylation mixture was incubated for 15 min at 37 ^{o}C for 5` extensions or blunt ends, and then 60 min for 3` extensions. This was followed by heat inactivation at 65 ^{o}C for 5 min and subsequent purification using gel extraction kit. The vector was ready to use for ligation (section 4.2.3.5).

4.2.3.5 DNA ligation

One µl of each DNA sample is checked on a gel. The relative DNA concentrations are estimated. For a ligation, approximately three times more insert than vector is used. The following composition is typically used: 6 µl insert, 1 µl vector, 2 µl of T4

ligase buffer, and 1 µl DNA ligase. The ligation mixture is mixed briefly and incubated for three hours at 16 °C or for best results overnight at 4 °C. The enzyme activity is deactivated at 65 °C for 10 min and it proceeds with DNA transformation (Jena Bioscience) (section 4.2.4).

4.2.4 DNA Cloning

The routine cloning of PCR-generated or enzymatically generated DNA fragments used in our work include preparation of electrocompetent cells, ligation of vector and DNA fragments (section 4.2.3.5), electroporation/ transformation of ligated products into the competent cells, clone slection, and clone analysis by restriction (section 4.2.3.1) and clone sequencing.

4.2.4.1 Preparation of electrocompetent *E. coli* cells

A 5 ml overnight *E. coli* culture was transferred to 200 ml of LB medium in a 500 ml Erlenmeyer flask, followed by incubation at 37 °C and 200 rpm. When the cell density reached the OD_{600} between 0.4 and 0.6, the culture was put immediately on ice. All subsequent steps were conducted on ice. The 200 ml culture broth was divided into 50 ml falcon tubes followed by centrifugation at 5000 rpm for 5 minutes. The cell-free supernatant was discarded and the resulting cell pellet was resuspended in 25 ml of 10% glycerol and sedimentation. The cell pellets were washed twice with 10 ml of 10% glycerol. The two tubes were combined and added with 5 ml of 10% glycerol to a volume of 25 ml. After centrifugation at 5000 rpm for 5 min, the cell pellet was resuspended in 5 ml of 10% glycerol followed by centrifugation. The cell pellet was resuspended in 1 ml of 10% glycerol. The cell suspension was distributed to eppendorf tubes (each eppendorf tube contained 70 µl) and immediately shock-frozen in liquid nitrogen followed by storage at -80 °C.

4.2.4.2 Electroporation

This electric transformation termed “electroporation” was initially used for *E. coli* by Dower *et al*. (1988) and Calvin & Hanawalt (1988) by subjecting concentrated cell suspensions and vector DNA to electrical fields at a very high amplitude. Because electroporation is simple and fast, it becomes a regular method of choice for bacterial transformation in many laboratories, not only using small DNA molecules but also large molecules such as DNA from bacterial artificial chromosomes (BAC), ranging from 7 to 240 Kb as reported by Sheng *et al*. (1995). The electroporation

procedure we used was based on that described by Sambrook and Russel (2001). At first, one vial (70 µl) of electrocompetent *E. coli* cells per transformation was taken from –80 °C freezer and thawed on ice. The pulser was set to "bacteria" (i.e. 1.8 kV). The ligation mixture (1-10 µl) was added to a vial of cells, and mixed gently by pipetting up-down three times. The suspension was transferred to an ice-cold 2 mm cuvette and placed into the cuvette holder of the pulser Bio-Rad electroporator set at 2.5 kV. The "flash" button was pushed to create electric current. The cuvette containing cell suspension was immediately added 1 ml LB, mixed and transferred into an empty 1.5-ml eppendorf tube, incubated at 37 °C with shaking for 1 hour.

4.2.4.3 Lac-based clone selection

Lac selection is based on the use of an antibiotic and x-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (isoprophylgalactoside). Clones harboring no insert appear blue on the selective media because *lac*Z´ gene on the vector is not disrupted, and as a consequence such gene is expressed to produce *β*-galactosidase that breaks the x-Gal down to a deep blue colored product. IPTG here functions as an inducer of the enzyme. The colonies with a disrupted *lac*Z´ gene are unable to produce β-galactosidase, and therefore they look white on the selective media (Brown, 2006). For Lac selection, at first a LB plate containing an appropriate antibiotic was spread with 40 µl of 20 mg/ml X-Gal and 40 µl of 0.1 M IPTG. Then 100-300 µl of 1 ml transformed cell suspension (section 4.2.12) was spread on an LB plate containing antibiotic as well as IPTG and X-Gal. The plate was incubated at 37 °C overnight. Some white colonies were picked up and individual colonies were cultivated in 5 µl LB containing appropriate antibiotic. The plasmid from each culture was prepared (Section 4.2.1.3). The presence of inserts was then checked by enzymatic digestion (section 4.2.3.1).

4.3 Molecular Diagnostics of Key Biosynthetic Genes (methods for chapter 3, section 3.1)

Detection of polyketide and non-ribosomal peptide biosynthetic genes in this work was performed by PCR using primers based on KS and A domains, respectively. This consists of (i) isolation of metagenomic DNA, (ii) PCR amplification of biosynthetic key genes, (iii) cloning of PCR products, and (iv) clone selection and analysis. Primers were specifically designed and employed to target certain A

domain encoding sequences. Subsequent cloning of the PCR products was streamlined using TA-cloning (Clark, 1988).

4.3.1 Isolation of metagenomic DNA from a sponge sample

Most sponges contain polysaccharides in their tissue (Bucior and Burger, 2004; Parish *et al*., 1991; Esteves *et al.*, 2011). The polysaccharides can co-precipitate with the isolated DNA and may inhibit enzymatic reactions in DNA modifications (Kaufman *et al.,* 1999). Therefore removing polysaccharides is very important step when dealing with obtaining pure DNA from a sponge sample. The cetyltrimethylammonium bromide (CTAB)-based DNA isolation procedure modified by Piel *et al* (2004b) was found very useful to remove polysaccharides from the sponge cell extract (Schirmer *et al*., 2005; Hrvatin & Piel, 2007; Gurgui & Piel, 2010). In solutions of high ionic strength, CTAB is a cationic detergent that can bind nucleic acids to form an insoluble complex, leaving polysaccharides and proteins (Jones and Walker, 1963). The presence of NaCl can break the complex down, thereby leaving nucleic acids that can be concentrated by ethanol precipitation (Rogers & Bendich, 1985; Wilson, 2001). This CTAB procedure was subsequently optimized by adding lauryl sarcosyl and urea to improve cell wall lysis. CTAB and NaCl solution are added at concentrations of more than 0.5 M to remove the polysaccharides (Hrvatin & Piel 2007; Gurgui & Piel, 2010).

The following CTAB-based DNA isolation protocol was applied in our work to obtain total DNA from one gram of a sponge sample (Hrvatin & Piel 2007; Gurgui & Piel, 2010). At first, the sponge sample (1 gram) was cut into small pieces. The sample pieces were pulverized in the mortar under the liquid nitrogen and the nitrogen was evaporated. The homogenized sample was immediately transferred into a 50 ml Facon tube and suspended with 10 ml of lysis buffer (50 mM Tris-C pH 7.5, 50 mM EDTA, 700 mM NaCl, 2% Sarkosyl, 8 M EDTA), followed by a 10-min incubation at 60 ^{o}C. The cell suspension was extracted two times with an equal volume of phenol/chloroform/isoamylalcohol (in ratio 25:24:1) and centrifuged at 11.000 rpm at 5 min for each extraction. The upper layer was transferred into an empty tube and the remaining phenol was removed by adding an equal volume of chloroform. CTAB (previously warmed at 55 ^{o}C for 15 min) was added to the final concentration of 0.5%, mixed carefully and incubated at 60 ^{o}C for 10 min. The CTAB-treated extract was then centrifuged at 11,000 rpm to precipitate the polysaccharides. The

polysaccharide-free supernatant were subsequently added with 2 volumes of 100% cold ethanol, mixed gently by inverting the tube several times, stored at -20 ^{o}C for 30 min to improve the DNA precipitation, and centrifuged at the maximal speed (4 ^{o}C) for 30 min to precipitate the DNA. The resulting DNA pellet was washed two times with 70% ethanol to remove the remaining salts and CTAB traces, dried through air flow, and dissolved in 200 µl of ddH_2O or EB buffer (EDTA 10 mM, Tris-Cl 50 mM, pH 7.0).

4.3.2 PCR amplification of biosynthetic genes

The total DNA obtained by the CTAB-based DNA isolation protocol described above (Section 4.3.1) can be used as the template for PCR amplification. In our experiences, total DNA isolated from sponge tissue often gives variable results in PCR amplification due to the impurities in DNA preparation that can inhibits DNA polymerase activity. To solve this problem, a series of PCR reactions containing different dilutions (1/10, 1/100, and 1/1000) of the extracted DNA was set up. Each sample dilution was then used as the template for PCR optimization. The PCR composition used for the amplification of KS or A domain genes were based on the protocol described in the Section 4.2.2.2 for crude sponge DNA. As an alternative, the extracted DNA was purified by gel purification (PegLab, Epicentre). Subsequently the purified DNA was used for PCR amplification (section 4.2.2.2 for purified sponge DNA).

To simplify PCR detection of target biosynthetic genes, we used the uncultured bacterial fraction as the PCR template. The filamentous bacterial fraction obtained was mainly dominated by large filamentous cells "*Candidatus* Entotheonella sp." based on microscopic observation. The bacterial fraction was prepared from the sponge *T. swinhoei* chemotype W through differential centrifugation (Piel *et al*., 2004b). This mechanical separation was not only intended to reduce the complexity of the sponge symbiotic assemblage but also remove polysaccharides. The symbiotic cell pellet was stored in 1 ml ethanol at 4 ^{o}C. To detect the PKS/NRPS genes in the fraction, whole cell PCR amplification was performed. At first a small amount of cell suspension previously stored in ethanol were rinsed 3 times with sterile ddH_2O. The rinsed cell pellet was then used as template for the PCR-amplification (Section 4.2.2.2 for bacterial cell suspension). The oligonucleotides used for the PCR amplification of KS genes were based on two highly conserved motifs in the KS domain: DPQQ and HGTGT (Piel, 2002). The degenerate KS

primer pair, designated as KSDPQQF and KSHGTGR (Table 4.3.1), was applied for the PCR amplification of KS fragments using total sponge DNA.

Table 4.3.1 Existing primers used in identifying KS sequences

Primer	Ort	Sequence (5`-3`)	Target motif	*T*m (°C)	Reference
KSDPQQF KSHGTGR	For Rev	MGNGARGCNNWNSMNATGGAYCCNCARCANMG GGRTCNCCNARNSWNGTNCCNGTNCCRTG	DPQQ HGTGT	58	Piel (2002)

To clone the A domain sequences, three degenerate primer pairs based on the adenylation catalytic motifs (A2-A8, A2-A9 and A3-A7) (Marahiel *et al*., 1997) were used (Table 4.3.2). Those three primer pairs were checked and optimized first using the genomic DNA of *Bacillus amyloliquefaciens* B42 as the PCR-template. All of the three tested primer pairs gave PCR products of the expected sizes (~700 bp, ~950 bp, ~1200 bp). The genomic DNA was prepared according to the protocol described in Section 4.2.1.2.

Table 4.3.2 Existing primers used in identifying adenylation domain-encoding genes

Primer	Ort	Sequence (5`-3`)	Target	*T*m (°C)	Reference
JP-for JP-rev	For Rev	GCN GGN GGN GCN TAY GTN CC CCN CKD ATY TTN ACY TG	A_2 NRPS A_8 NRPS	58	Rolf Müller
Öz-for Öz-rev	For Rev	- -	A_2 NRPS A_9 NRPS	55	Provided by Özlem
AG-for AG-rev	For Rev	GCS TAC SYS ATS TAC ACS TCS GG SAS GTC VCC SGT SCG GTA S	A_3 NRPS A_7 NRPS	61	Ayuso-Sacido & Genilloud (2005)

To target sequences coding for the A domains activating serine, we designed a series of degenerate primers based on the unique motifs located between A4-A5 on the serine-specific adenylation (SerA) sequences (Marahiel et *al*., 1997). At first, sequences of some A domain sequences with serine specificity as well as those with other amino acid specificity from different biosynthetic gene clusters were initially retrieved from the GenBank and subsequently aligned together in BioEdit (Hall 1999) using Clustal W program (Thompson *et al*., 2007). The unique conserved motifs responsible for serine specificity located on the A4-A5 region were identified and used as the basis for designing degenerate primers (see Figure 3.1.8, Chapter 3). All of the designed primers are listed in Table 4.3.3.

Table 4.3.3 NRPS primers targeting certain A-domain sequences

Primer	Ort	Sequence (5`-3`)	Target	Reference
A4F1 A4F2	For For	TTY GAY GTN AGY GTN TGG GAR TT TTY GAY GTN TCN GTN TGG GAR TT	SerA (with A5R1 or A5R2) SerA (with A5R1 or A5R2)	In this work
A5R1 A5R2	Rev Rev	TCN GTN GGN CCR TAN RRT TRT G- YTV YTN STN ACR TCN CAN GCN GC	SerA (with A4F1 or A4F2) SerA (with A4F1 or A4F2)	In this work
A4AsnF A5AsnR	For Rev	TTY GAY YTN CAN GTN CAN WSN ATH CAT RCA NCC NAC NAC NGY RCT NGT	AsnA (with A5AsnR) AsnA (with A4AsnR)	In this work
TH-for TH-rev	For Rev	ATHTGYGGNGGNGGNYTNGCNGGN YT AR NGG DAT NRD CCA NAC CCA RTA	Tryp halogenase gene Tryp halogenase gene	In this work

Notes: ort, orientation; For, forward; Rev, reverse

The designed degenerate KS-specific primers were used using total DNA from a sponge as the PCR template. Since none of the amplicons likely falls to into the group of complex polyketides, uncultured filamentous bacterial fraction prepared from the same sponge was used as the PCR template. The designed degenerate NRPS primers were also applied using filamentous bacterial fraction as the PCR template (section 4.2.2.2 for bacterial cell suspension).

4.3.3 Cloning of PCR products into T vectors

Subsequently the amplified products with the desired size were checked on a 1 % agarose gel. If a single target fragment appears on the gel, the PCR product can directly be purified by using *Illustra*^{TM} *GFX PCR DNA and Gel Band Purification Kit* (GE Healthcare). The purified PCR products with around 700-bp were ligated into a T-vector (Section 4.2.3.5). The T-cloning vector was prepared (Section 4.2.3.2). The ligation products were transformed into competent XL-Blue MRF` cells via electroporation (Section 4.2.4.2). The competent cells of *E. coli* XL-Blue MRF` used in this work were prepared by washing the cells several times with glycerine to remove the salts from the growth media. The prepared competent cells were suspended in 10% glycerol and stored at -80 ^{o}C. The transformed cells were grown overnight on the LB plates containing ampicilin, X-Gal and IPTG, thereby allowing Lac selection (Section 4.2.4.3). Some white colonies were picked up and cultivated in liquid media. The plasmid DNA was then extracted from the individual recombinant cultures (Section 4.2.1.3) and checked for the presence of the right insert (Section 4.2.3.1)

4.3.4 Clone analysis and sequencing

Each clone with the right insert was subjected to restriction fragment length polymorphism (RFLP) analysis using *Rsa*I or another frequently cutting enzyme (Section 4.2.3.1). Inserts showing a unique restriction pattern were sequenced. Sequencing of some DNA samples was conducted at GATC Biotech AG service or 4baseLab on an ABI 3730*xl* DNA Analyzer (Applied Biosystems) using the dideoxy mediated chain termination method (Sanger *et al*., 1977). The employed dideoxy nucleotides were linked with four fluorescent dyes, each base type carrying a specific colour. Therefore the nucleotide sequences were determined by means of a fluorescence detector subsequent to sample separation via capillary electrophoresis in a polyacrylamide matrix (Sambrook & Russel, 2001). The sequence data obtained

were analyzed using BlastX to predict their functional properties in relation to the correspondent target compounds. This advanced BLAST X program was used to search a protein database provided on-line by NCBI using a translated nucleotide query that provides information about putative gene products and their functions (Altschul *et al*., 1997). Based on the data, specific PCR primer pairs were subsequently designed, which were further employed for screening of highly complex metagenomic libraries.

4.4 Construction and Screening of Complex Metagenomic Libraries (methods for chapter 3, sections 3.2 and 3.3)

To capture the majority of bacterial pathways present in the sponge, a metagenomic library needs to be generated as large as possible in order to make sure that the biosynthetic pathway of interest is entirely covered in the library. The technical challenges in generating a complex metagenomic library containing clones with relative large inserts include obtaining pure HMW metagenomic DNA (in the size range of 35 to 50), the stability of the cloned large fragments, amplification of clones in the library in a high density, and rapid screening of a highly complex library.

Construction of genomic library was initially tested using the cultured cells *Acinetobacter baylyi* ADP1. Genomic DNA was initially prepared from this bacterium (Section 4.2.2) and subsequently cloned into *E. coli*. Clone amplification was done in the microtitter plates instead of in the semiliquid media since this library was generated from a single bacterial genome. This was initiated by transferring 400 clones from the petri dishes into microtiter plates. After overnight incubation, individual culture in the same rows were pooled and be then screened by PCR. The clones which correspond to the positive row pool(s) were individually screened to detect single positive clones. A library construction steps optimized for the cultured cells *Acinetobacter baylyi* ADP1 were applied for the uncultured bacterial fraction. For this fraction, the resulting recombinant clones were generated and amplified in 3-D fomat termed “semi-liquid library” as described by Hrvatin and Piel (2007). Since no complex biosynthetic gene cluster was identified in this small library and enlarging the library size was not possible due to the extremely limited amount of the existing uncultured bacterial sample, the library construction was subsequently carried out using total DNA of the *T. swinhoei* sponge specimens from which the

uncultured bacteria were derived (Sections 4.4.1-4.4.5). With this sponge sample, the library size could be expanded up to around 350,000 clones.

This section describes methods used in this work for the construction of complex metagenomic libraries from the sponge *T. swinhoei*. Generating a fosmid library was initiated by purification and size selection of the HMW DNA (Sections 4.4.1 & 4.4.2) followed by repair of the DNA fragment ends (Section 4.4.2). Subsequently the repaired DNA was joined with a dephosphorylated fosmid (Section 4.2.3.5) and packed into phage particles (purchased from Epicentre). The packaged viral particles were used to transfect *E. coli* (Section 4.4.3). The resulting infected cells were amplified in the semi-liquid media, producing complex 3-D cultures (Section 4.4.4) (Fig. 4.4.1).

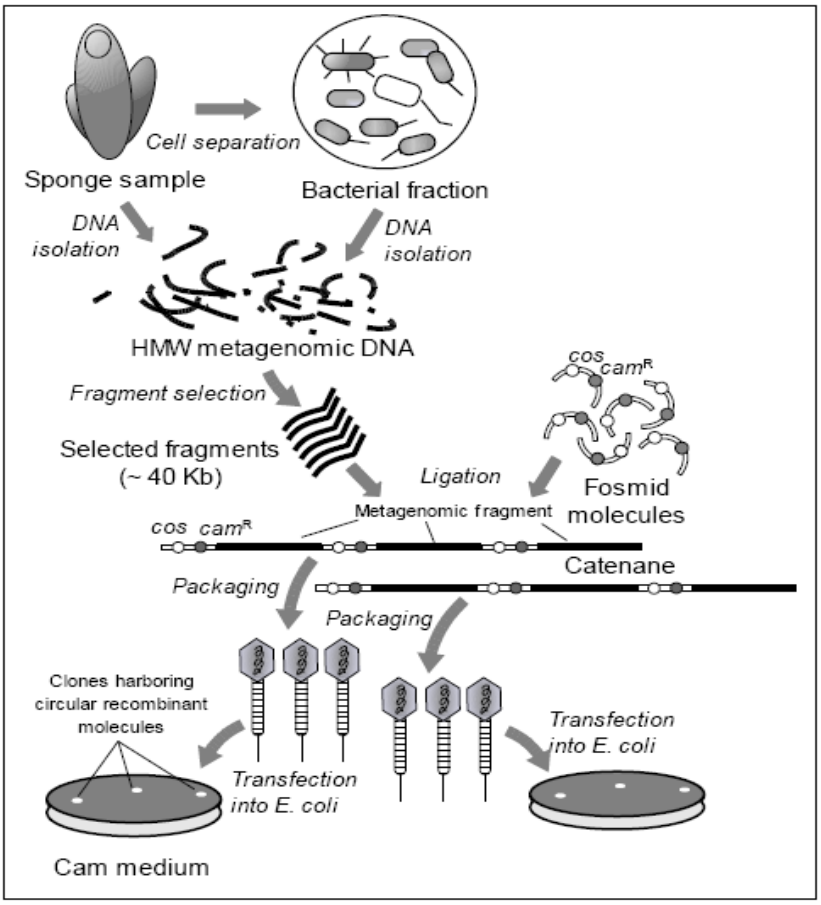


Fig. 4.4.1 Construction of metagenomic library. HMW DNA is prepared from either sponge or its associated bacteria and be separated by electrophoresis on low-melting-point (LMP) agarose gel. DNA fragments with desired size (~40 Kb) are purified from the gel, repaired at both termini, and ligated with dephosphorylated fosmid molecules. The ligation products are packed in bacteriophage particles and used to infect *E. coli.* The resulting library can be amplified in semi-solid media.

4.4.1 Isolation of high-molecular-weight (HMW) metagenomic DNA

There are three main important things that we considered in isolating metagenomic DNA for library construction: the DNA quality, size, and yield. The methods described below were used in this work to meet these three criteria as a prerequisite for generating a complex metagenomic library. Obtaining pure HMW DNA from a sponge sample was based on the CTAB-based DNA isolation as described previously (Section 4.2.1.2). DNA isolation from this filamentous fraction was performed using a simple heat treatment described by Syn and Swarup (2000) (Section 4.2.1.2).

4.4.2 Size-selection and end-repairment of HMW DNA fragments

The DNA fragments obtained either from a sponge sample or a bacterial cell pellet were separated b low-melting point (LMP) agarose gel electrophoresis in order to remove the remaining contaminants including RNA. However, handling a LMP gel before and after electrophoresis is difficult because the LMP gel is easily broken. By making two gel layers consisting of standard agarose gel (lower layer) and LMP gel (upper layer) (Sambrook & Russel, 2001) as shown in this work, the LMP gel break can easily be prevented. The desired large fragments with the average size of 40 kb are subsequently recovered from the gel either by enzymatic gel extraction using GELase or by column-based gel extraction. GELase commercially provided by Epicentre contains β-agarose-digesting enzyme that is able to break down the carbohydrate backbone of LMP gel, thereby releasing small, soluble oligosaccharides (http://www.epibio.com/item.asp?id=297). The standard agarose gel here serves as a support to strengthen the LMP gel. In addition, obtaining large DNA fragments (at least 30 kb) is necessary for subsequent construction of a metagenomic library based on fosmids and cosmids.

The following procedure was used in this work to select DNA fragments of around 40 kb (Epicentre, Gurgui & Piel, 2010). At first, a large well on an electrophoresis gel comb was made by covering the small wells with a clear sticky tape. The large well was flanked by two small ones, which are used for loading the DNA size marker. An agarose gel without ethidium bromide was prepared in two layers. The first layer was prepared by pouring 1% standard agarose exactly below the comb`s wells on the electrophoresis` template. When completely solidified, the comb was place on the surface of the first layer and 1% LMP agarose was poured on it. The functions of

the first layer are (i) to support the LMP layer, thereby avoiding it from breaking, (ii) to minimize the amount of the expensive LMP agarose required, and (iii) to reduce the thickness of the gel slice taken out. The DNA control of 36 kb was loaded into the two small wells flanking the large well. Then the DNA sample was uniformly loaded into the large well. The electrophoresis was run first at 70 V for 10 minutes to allow the DNA entered the gel. ~100 ng of 40 kb (or 36 kb) control DNA was loaded into each of the outside lanes of the gel. The isolated DNA was loaded in the lane between the 40 kb control DNA marker lanes. The gel electrophoresis was first run for 15 minutes at 60-70V and then overnight at 40-50V. Using a sharp razor blade, a 2 cm slice of gel was removed from both the left and the right sides of the gel. The edge pieces were stained with 1 µg/ml ethidium bromide solution. Two water baths were set up, one at 45 °C and the other at 70 °C. The band of HMW DNA in the stained edge slices of gel was located using a UV light. The position of the top and the bottom of the 40 kb marker was marked on each gel slice. Using the marks, the stained side pieces was put as guides to locate the band of HMW DNA in the unstained portion of the reassembled gel (Epicentre, Gurgui & Piel, 2010).

The DNA repairing step is very important because the ends of the DNA fragments obtained by extraction and purification are fray or not blunt without phosphate residues. Therefore the DNA fragment termini were repaired prior being used for a blunt ligation. The end-repairing enzyme mix contains T4 DNA polymerase to fill the recessed 3` termini and remove the protruding 3` termini, thereby converting the frayed ends into blunt ends. The hydroxyl group at the 3` termini is converted to a phosphate group using terminal phosphotransferase that is also present in the end-repairing enzyme mix (Brown, 2006). The following protocol was used in this work to repair the ends of the isolated HMW DNA fragments (Epicentre, Gurgui & Piel, 2010). The following components were put into a 1.5 ml eppendorf tube: sterile dH_2O (up to the final volume of 40 µl), 4 µl of 10x end-repair buffer, 4 µl of 2.5 mM dNTP mix, 4 µl of 10 mM ATPs, HMW genomic DNA (2.5-5.0 µg), and 2 µl of end-repairing enzyme mix. All resulting bubbles were removed and the reaction was incubated at room temperature at exactly 45 minutes. The reaction was then inactivated by adding 1.6 µl of 500 mM EDTA (up to final concentration of 10 mM), followed by incubating at 70 ^{o}C for 10 min. The DNA was recovered by isopropanol precipitation, and then dissolved in 20 µl of EB buffer (Qiagen). The concentration of the blunt-ended DNA was determined by running 0.5–1 µl on an analytical 1%

agarose gel with known amounts of Fosmid Control Insert DNA (Epicentre, Gurgui & Piel, 2010).

4.4.3 DNA packaging and transfection

To deal with the instability problem of the cloned large fragments, a high-capacity vector that is present in low copy number was used in this work to clone large DNA fragments in the range of 30-40 kb. Using a low copy number vector not only maintains the stability of large inserts but also reduces significantly undesired modifications such as rearrangements and deletions as has been shown by Kim *et al* (1992) for complex mammalian genomic DNA inserts. The fosmid vector termed pCC1FOS was purchased from Epicentre. The low copy number (1-2 copies per cell) in the fosmid-based clones is controlled by the F factor system present in pCC1FOS. The presence of an *ori*V replication origin allows the copy number to increase to about 10-50 copies per cell under the condition when the vector is introduced into an appropriate host harboring a *trfA* gene, such as EPI300^{TM} T1^{R}. The *trfA* transcription is controlled by the *ara*C-P_{BAD} promoter/regulator system. Since this system is inducible by L-arabinose, the DNA yields can be increased by adding this sugar into the growth media (Westernberg *et al*., 2010, Epicentre, Wild *et al*., 2002). High yields of a cloned DNA are required especially for applications such as subcloning, sequencing, and recombination. The library construction steps using pCC1FOS followed the method described by Epicentre, which was subsequently optimized by Gurgui and Piel (2010).

The following procedure was used in this work DNA packaging (Epicentre, Gurgui & Piel, 2010). Five mL of LB broth containing 10 mM $MgSO_4$ was inoculated with a colony of EPI300-T^{1}R *E. coli* cells (Epicentre) a fresh plate, and incubated overnight at 37 °C in shaking incubator (at 200 rpm). On the next day, the 5 mL overnight culture was inoculated into 50 mL of LB broth containing 10 mM $MgSO_4$ and incubated at 37 °C with shaking (200 rpm). When the OD_{600} reached 0.8–1.0, the culture was stored at 4 °C until it is needed. One tube of the MaxPlax lambda packaging extracts (Epicentre) was thawed, and half of its volume (25 µl) was immediately transferred to a 1.5-mL microcentrifuge tube placing on ice. The remaining 25 mL of the MaxPlax Packaging extract was placed back to the −80 °C freezer for later use. The ligation reaction (10 µl) was added to the thawed extracts, mixed gently by pipetting the solutions up and down several times. The packaging

reaction was incubated at 30 °C for 1.5 hours in the water bath, and then added with the remaining 25 mL of MaxPlax lambda packaging extract. After incubating the reaction for an additional 90 min at 30 °C, PDB buffer was added to the final volume of 1 ml and mixed gently. The packaging reaction can be stored at 4 °C until needed. Before plating the library, the titer of the phage particles was determined to predict the number of clones inside the library. Serial dilutions of the 1-ml packaged phage particles was made using phosphate dilution buffer (PDB) buffer, ranging from 10 to 1000 times with the interval of 10. Then 10 ml of each dilution was added to 100 mL of EPI300-T^{1}R *E. coli* host cells prepared before, followed by incubation for 20 min at 37 °C in a water bath. The remaining cell culture was stored at 4 °C until the 2nd day, as they will be used again when plating the entire library. The infected EPI300-T^{1}R *E. coli* culture was spread on LB plates containing 12.5 µg/µl of chloramphenicol. The number of colonies was accounted and the titer of the packaged phage particles with the formula was calculated by multiplying the colony number with dilution factor and subsequent multiplying the calculation result with 100 (Epicentre, Gurgui & Piel, 2010).

4.4.4 Amplification and screening of 3-D libraries

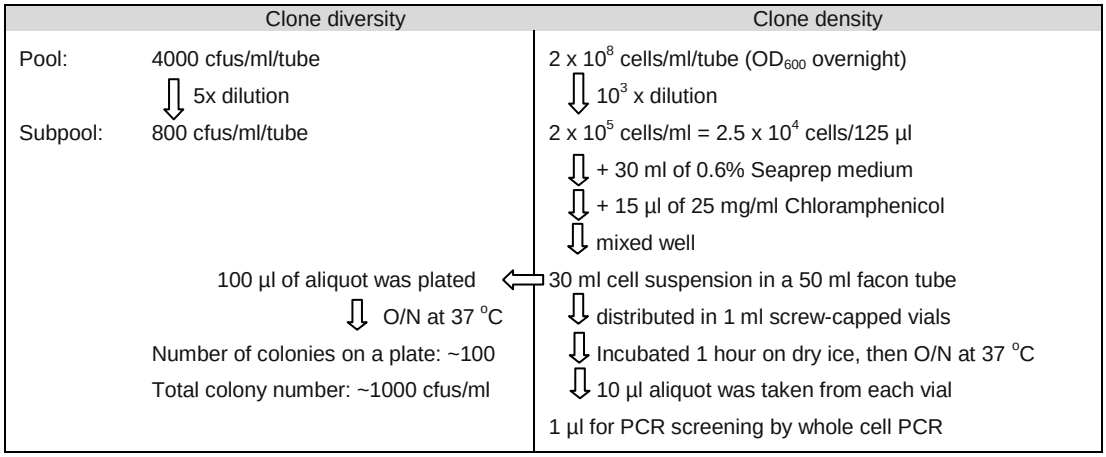
Screening of a complex metagenomic library containing a large number of clones requires effective methods that can be used rapidly at a high-throughput format. The resulting library can be generated in different formats, usually on plates or microtitters. However, libraries on agar plates cannot be stored for a longer time. As an alternative, clones growing on a plate can be combined into pools. These pools can be mixed with glycerol and stored at -80 ^{o}C for a long time. All clone pools can be screened by PCR. The positive pools are plated out at low colony number and rescreened until a single clone is detected (Piel, 2002). However, this approach is only effective for medium-sized metagenomic library (less than 100,000) (Hvartin & Piel, 2007). Another way is transferring individual clones from agar plates into microtiter wells, and subsequent screening of clone copies either by PCR or hybridization. Using this approach, glycerol can be added to the microtiter plates and then stored in the freezer for later screening. However for a highly complex library with a large number of clones, this is inefficient because it needs a lot of microtiters, lots of picking and larger storing space. In addition, performing hybridization by directly blotting individual colonies from agar plates onto Nylon membranes is not efficient because very large numbers of expensive membranes

are necessary for metagenomic screening (Uria & Piel, 2009). Recently Hrvatin and Piel (2007) have developed an elegant method for screening of a complex metagenomic library, which is highly economic in terms of labor, consumables and required space. The method termed a liquid gel pool analysis involves the construction of a metagenomic library in a 3-D semiliquid format, followed by rapid screening by whole-cell PCR.

This library amplification described below allows growing well-separated clones at high densities (approximately 1000 clones/ ml). All the clones are organized into pools, in which each pool contains approximately 1000 clones in one ml semi-liquid medium (Hvartin and Piel, 2007). The following library amplification was based on the procedure described by Hvartin and Piel (2007) and Gurgui and Piel (2010). Sterile 2 ml screw cap vials as well as LB media containing 0.5% SeaPrep Agarose (Lonza, Switzeland) were at first prepared (Table 4.5). It is required one tube for each 100-1000 clones. One liter of media can be used to grow approximately 10^6 colonies. 500 µl of the packaged DNA was added into 5 ml of the EPI300T^{1}R culture (OD_{600} of 0.8-1.0) in a 50 ml Facon tube, and then mixed gently by using a 10 ml Pasteur pipette. The mixture was incubated at 37 ^oC for 20 minutes. 35 ml of the LB medium containing SeaPrep Agarose was added into the transfected EPI300T^{1}R culture and mixed slowly by a Pasteur pipette. Chloramphenicol solution was added to a final concentration of 12.5 µg/ ml and then mixed gently by using a Pasteur pipette. 100 µl of the transfected cell culture mixed with SeaPrep Agarose was plated on the LB media containing 12 µg/µl of chloramphenicol, and incubated at 37 ^oC overnight. The number of colonies on each plate was counted. The remaining 40 ml SeaPrep medium was distributed into screw-capped vials (1 ml in each vial), incubated on dry ice for 1 hour and subsequently at 37^oC overnight. Colonies were clearly visible (even after 10-12 hours) as spatially growing in the complete volume of the media. Each screw-capped vial was mixed by vortexing until the suspension becomes homogenous (5 seconds) (Hvartin & Piel, 2007; Gurgui and Piel, 2010).

To obtain the complete biosynthetic pathway of a complex polyketide or nonribosomal peptide, it usually needs more than one clone to cover the entire gene cluster involved in the biosynthesis. This can be performed by the homology-based screening relying on the ability of the sequence or probe to specifically recognize the target biosynthetic pathway. The following screening procedure was based on the

method described by Hvartin & Piel (2007). The library constructed in this work was grouped into pools that are arrayed in three boxes. In box 1, the clone density of each pool is ~4000 cfus/ ml. In boxes 2 and 3, the clone density per pool is ~500 cfus/ ml and ~1000 cfus/ ml, respectively. For screening of the library to find individual positive clones, a small aliquot of all individual pools were combined together to super-pools that were then screened by whole-cell PCR. Pools corresponding with a positive super-pool were subsequently screened by PCR. Below is an example about how to calculate the volume of pool aliquot required for cell dilution.



The resulting positive clone pools were then diluted in new tubes at lower cell density and re-screened by PCR. Pool dilution was performed in the same way as described above. Combination of cell dilution and PCR screening until the clone density was ~20-40 cfus/ml. Then colonies were plated at the density of around 200 cfus/ plate. Colony PCR was then carried out to isolate a single positive clone. The specific primers used for library screening was designed based on the sequence data of PKS-KS and NRPS-A genes cloned from the filamentous fraction of *T. swinhoei* (Tables 4.4.1 to 4.4.3).

Table 4.4.1 Specific primers based on sequences for serine-specific adenylation domains of NRPS

Spesific Primer	Sequence (5`-3`)	Lenght (bp)	GC (%)	T_m (°C)
A3-F	CCCGAAGGCCATCAGGATCCGAGTTAT	27	56.0	67.0
A3-R	TGTGGGGCCGTAGAGGTTGTGTAGTTC	27	56.0	64.0
B2-F	AGTGCTTACCTGGTGCGTCTGCT	23	59.0	59.0
B2-R	GTGGGGCCGTAAAGGTTGTGTA	22	57.5	57.5
D1-F	TTTTCTGGCCGTTGATGAGCG	21	59.7	59.7
D1-R	GTGTGGGGAAGTAGATGGGCTGATT	25	58.4	58.4

Table 4.4.2 Specific primers based on sequences for *cis*-AT KS domains of PKS gene cluster

Spesific Primer	Sequence (5`-3`)	Lenght (bp)	GC (%)	T_m (°C)
EtP5-F	TCAGGGCTTGATCGGGGTGTATG	23	57.0	58.8
EtP5-R	CGTCTAAACTCGGGGCGGTGAA	24	59.0	58.6

Table 4.4.3 Specific primers designed based *trans*-AT KS domains to screen the library

Spesific Primer	Sequence (5`-3`)	Lenght (bp)	GC (%)	T_m (°C)
EKS27-f	CTTGGGCCGCGTTGGAAGAT	20	55.7	60.0
EKS27-r	CCGCTTCGCCAAGCACAAAG	20	55.3	60.0
EKS31-f	TTGAAGATGCGGGGCATGCC	20	60.0	57.4
EKS31-r	TGGGACAAAGCCATCCGCAC	20	60.0	55.3
EKS37-f	TGATTGGTTCCGCGTTGCACA	21	52.4	56.7
EKS37-r	GGTGCCGTGGGTTTCGATATAGCTA	25	52.0	56.1
EKS50-f	GACCCGCAGCAGCGTATGTT	20	57.4	60.0
EKS50-r	TTTTGCAGCGCCCATCCTG	19	59.2	57.9
EKS64-f	ACCCGCAGCAGCGCTTACTGTT	22	59.1	61.8
EKS64-r	ATCCAAGAGCGTATCCGGGGAC	22	59.1	60.0
Et-KS2-F	GGCCCGAGTTTTGCCCTTGATACC	24	58.0	61.0
Et-KS2-R	TCTCCTAGCGATGTGCCGGTGCC	23	65.0	62.0
Et-KS3-F	ATCGACTACAATCTTCTTGCGCGC	24	50.0	57.0
Et-KS3-R	GATCGCCCAGTGAGGTTCCGGT	22	64.0	60.0
Et-KS5-F	TAGCGAGGCGGTGCCGATGGAC	22	68.0	62.0
Et-KS5-R	GGTCGCCAAGGGAGGTGCCCGT	22	73.0	64.0

4.4.5 Chromosome walking, fosmid analysis and sequencing

Once a part of a target gene cluster is isolated, the remaining parts can be searched by chromosome walking. Particularly, the biosynthetic gene clusters investigated in this work were large up to hundreds of kilobase pairs. To find the remaining up- and down-stream parts of the gene cluster, specific primers (Table 4.4.4) were designed based on the end sequences of the previously isolated fosmid and employed to screen the additional portions of the gene cluster. This process was repeated until all target genes were isolated.

Table 4.4.4 Primers for chromosome walking to find additional regions of misakinolide A gene cluster

Spesific Primer	Sequence (5`-3`)	Lenght (bp)	GC (%)	T_m (°C)
EKS27FP-f	TGAATCAGCATATGGGCATTGGTGC	25	48.0	58.4
EKS27FP-r	TGGTATTAAAGCGCCTTAGCCGAGC	25	52.0	57.4
EKS27RP-f	TAGAGCATCCGGGATTTTCCTG	22	50.0	52.0
EKS27RP-r	AGTTGTTCACACGCGGGTGTTC	22	54.5	53.5
EKS27FP2-f	GGACCGCACGTGGTTCTTTTCTGTGAA	27	51.9	61.5
EKS27FP2-r	TGGAGGTCAACGCCCTTAACGAAGTCT	27	51.9	59.6
PEP-f	GATGAGTGGTCAAGCCTGTGTCGTGAA	27	51.9	59.1
PEP-r	CGCTATGAATTTGCTGCTGTACAACCAC	28	46.4	57.6
AT-f	CAACGTCTTTAAATGTCAAAACATCATCGC	30	36.7	57.6
AT-r	GATGACCTCTTCATCAGAGACTTGCAAAA	29	41.4	56.0

A positive clone isolated from a metagenomic library need to be analyzed further in order to know whether it harbors a biosynthetic pathway of interest or not. The clone to be analyzed were initially prepared by a scaled-up standard miniprep and subsequently treated with ATP-dependent DNase provided by Epicentre (http://www.epibio.com/) to digest the linear host chromosomal DNA. The activity of this enzyme has no effect on nicked or closed-circular dsDNA, thereby leaving circular fosmid molecules that are relatively free of host chromosomal DNA. The pCC1Fos vector backbone of the purified clone can be removed by treating it with rare-cutting enzyme(s) such as *Not*I. The clone analysis in this work was initiated by end-sequencing and sometimes sub-cloning (Sections 4.2.3 & 4.2.4), followed by complete sequencing and bioinformatic analysis. The sequencing reads were assembled into contigs. The gaps between contigs were filled by primer walking, resulting in a continuous sequence. The primers used for filling in the gaps were designed to flank the gaps, as listed in Table 4.4.5 for PKS gene cluster and Table 4.4.6 for NRPS gene cluster, which are complementary with the regions located approximately 100-200 bp from the ends of each contig library.

Table 4.4.5 Specific primers for filling the gaps in the misakinolide A gene cluster

Spesific Primer	Sequence (5`-3`)	Lenght (bp)
1AU5.3-r	GGCAAAGACTGTATTAGCGACGTGC	25
1AU2.2-f	CGACACAGGCTTGACCACTCATC	23
2AU2.1-r	GAATGAAGCCACCCAAGTGCGC	22
2AU2.1-f	CCTCCCGCTGAGGACAATAACTG	23
3AU5.2-r	GTGCAACGATAGGTTGACGCCTC	23
3AU5.2-f	CAATATTGGCCATACGGAGTCGG	23
4AU5.1-r	CGCGTTCATCTGGTGTTTGAGGTG	24
4AU5.1-f	GAATCCAGCAATACCGGCAGCC	22

The entire nucleotide sequence was translated into amino acid sequence. Open reading frames (ORFs) or potential genes were identified and searched by BlastX for homology in GenBank. Finding the potential ORFs was carried out by the softwares, FramePlot 2.3.2 (Ishikawa and Hotta, 1999) and Geneious 5.4.6 (Drummond, *et al*., 2011). The composition and order of the PKS domains were predicted by finding the typical residue motifs. Based on the predicted domain organization and function, the encoded biosynthetic pathway was predicted.

Table 4.4.6 Specific primers for filling the gaps in the keramamide H gene cluster

Primer pair	Sequence (5`-3`)	Lenght (bp)
NewContig1r NewContig1f	CTCCTGTTTCGCTATCGTTGG CAAGAGCTTCGGACGCTTC	21 19
NewContig2f NewContig2r	CGTTATCAGACGAAGCTTAC GGAAGTTGCCAATAAACGCC	20 20
NewContig3f NewContig3f	GGCGTTTATTGGCAACTTCC CTATCTGGACCACGCCATC	20 21
NewContig4f NewContig4r	CAAACGGTCGACGGACCTC CAATCGTAGGTGTTTCAAATAG	19 22
NewContig5f NewContig5r	CTTCTCAATATCCTTTGTCAG GCTTCAGCATCCTGTTCAGC	21 20
NewContig6r NewContig6f	GATAAACAGGGTGTGCAGCAG CAACAGTTGCGCAGCTGAATG	21 21
NewContig7f NewContig7r	GATGCGCGAAGACGACCAC CGACACAGATCGATTTGATC	19 20

The resulting NRPS insert was completely sequenced through a *de novo* sequencing approach (Margulies *et al*. 2005; Droege and Hill 2008; Rothberg and Leamon, 2008; DiGuistini *et al*., 2009) in collaboration with Dr. Vladimir Benes at EMBL-Heidelberg. Assembling of the sequencing reads was conducted using using the Seqman of the Lasergene program and Geneious 5.4.6 (Drummond *et al*., 2011). The potential ORFs were identified using the software FramePlot 2.3.2 (Ishikawa and Hotta, 1999) and Geneious 5.4.6 (Drummond *et al*., 2011). Subsequent functional prediction of the identified ORFs were carried out using BLAST (McGinnis & Madden, 2004), PKS/NRPS analysis Web-site (Bachmann & Ravel, 2009), and NRPSpredictor2 (Rausch *et al*., 2005; Röttig *et al*., 2011). Advanced BLAST X program searches a protein database using a translated nucleotide query that provides information about putative gene products and their functions (Altschul et al., 1997). The computation was performed using the BLAST network service that utilises the NCBI (National Centre for Biotechnology Information).

4.5 Genome Analysis (Methods for chapter 3, section 3.4)

Before genome sequencing of an uncultured bacterial fraction, it is necessary to check the purity of the prepared fraction by genetic diversity analysis. Bacterial diversity analysis can be carried out using 16S rRNA subunit as a taxonomic chronometer.

4.5.1 Genetic diversity analysis

To obtain information about how dominant is “Entotheonella sp.” cells in the uncultured bacterial cell fraction in comparison with other interfered bacterial groups, genetic diversity analysis was carried out. Several criteria making 16S rRNA subunits and their genes the most useful chronometer used today for determining the taxonomic indentity include (i) the ribosome function has not changed for very long time (about 3.8 billion years) (ii) the universal presence of16S rRNA genes among all cellular living organisms, (iii) the 16S rRNA gene size of 1540 nucleotides makes them easy to analyse, and (iv) the primary structure is invariable, more or less conserved to highly variable regions, and (iv) lateral transfer of these genes is a rare event and has been noticed mainly among closely related strains (Stackebrandt, 2009). The huge numbers of 16S rRNA gene sequences (from more than 7000 types of prokaryotes) in public databases enables to identify a 16S rDNA fragment cloned from any source (Stackebrandt, 2009).

In this work, universal 16S primers, designated as 27F and 1492R (Weisburg *et al*., 1991; Stackebrandt & Lieasack, 1993), were used for the biodiversity analysis of a sponge sample as well as an uncultured bacterial fraction. In addition to that, primers were designed in this work to detect the presence of “Entotheonella sp.“ in the fraction as well as in the total sponge DNA (Table 4.4.7). The primer design was based on the unique regions on several *Entotheonella* 16S rDNAs both isolated from the *T. swinhoei* cell pellet and retrieved from the public database, in comparison with 16s rRNA genes from other bacterial groups that are commonly found in sponges.

Table 4.4.7 Primers for 16S rDNA amplification

Primer	Sequence (5`-3`)	T*m* (^{o}C)	Target	Reference
27F	AGAGTTTGATCMTGGCTCAG	52	Universal	Lane (1991)
1492R	GGTTACCTTGTTACGACTT	47	Universal	Stackebrandt & Lieasack (1993)
238F	CCG GTC TGA GAT GAG CTT GC	55-59	*Entotheonella* sp.	Schmidt *et al*. (2000)
1309R	GCG TKC TGA TCT CCG ATT AC	57	*Entotheonella* sp.	In this work
1247R-ad	AAG TAG CRG CCC TTT GTC	57	*Entotheonella* sp.	Adapted from Schmidt *et al* (2000)
Ento271F	GGGAAASGTTCGCBGGTCTG	57	*Entotheonella* sp.	In this work
Ento735F	GYATTAAGCCKYGGAAACKGT	57	*Entotheonella* sp.	In this work
Ento1290R	GCCCRGCWYVACCCGGTA	57	*Entotheonella* sp.	In this work

The PCR products were purified by gel extraction using *QIAquick Gel Extraction Kit* (Qiagen). The purified PCR products were subsequently cloned into *E. coli* by TA cloning (Sections 4.2.3-4.2.4). Several recombinant clones from the filamentous

fraction were subjected to restriction length polymorphism (RFLP) (Section 4.2.3.1) by the *Rsa*I and *Hha*I digestion. Clones showing unique RFLP patterns were sequenced. The 16S rDNA sequences obtained were analyzed and a phylogenetic tree was constructed to distinguish closely related microbial species based on sequence variation within the 16S rRNA genes.

4.5.2 Gene cluster identification

Based on genetic diversity, it was found that the filamentous bacterial fraction from *T. swinhoei* was fairly pure. Subsequently total DNA was prepared from the cell fraction for whole genome sequencing (around 60 ng for the first round of sequencing). One of simple ways to prepare genomic DNA for genome sequencing is using ion exchange chromatography based on resin. In principle, the cell extract is passed into a column containing resin. Because this extract contains very little salt, all the negatively charged molecules including DNA, RNA and protein bind to the resin. If salt solution at gradually increasing concentration is loaded onto the column, the molecules become unbound, starting with protein, then RNA, and finally DNA. The first salt solution is usually sufficient to elute the protein and RNA, thereby leaving DNA on the column. By passing a second salt solution at a higher concentration, the DNA is eluted, which can be concentrated and deslated using isopropanol/ethanol precipitation (Brown, 2006). In this work, purification of genomic DNA from a small amount of bacterial cell fraction was based on the anion-exchange resin commercially provided by QIAGEN, which is designed for direct isolation of chromosomal DNA 20-150 kb in size. The procedure was done according to the manufacturer's instruction manual (QIAGEN), which involved careful cell lysis and binding of genomic DNA to the resin. A medium-salt buffer is used to remove traces of RNA and protein as well as to disrupt nonspecific interaction that causes removal of nucleic acid-binding proteins. Using a higher-salt buffer, the DNA is efficiently eluted from the column (QIAGEN). The purified DNA from the cell fraction was subjected to whole genome sequencing in collaboration with Dr. Jörn Kalinowski (Bielefeld University). The whole genome sequencing was based on a novel sequencing-by-synthesis technology, called 454-pyrosequencing. The most significant advantage of this new sequencing technology, when compared with the whole genome shotgun approach, is that there is no need for cloning of the target DNA fragments, thereby avoiding cloning biases resulting from toxic sequences killing their cloning hosts (Huson *et al*., 2007). Although the read lengths

are significantly shorter, in the range of 100 to 500 bp (Hall, 2007; Frasser-Liggett, 2008; Wommack *et al.,* 2008), newer algorithms have allowed for the assembly the short reads effectively (Huson *et al*., 2007).

The genome sequence data was analysed using Geneious 5.4.6 (Drummond, *et al*., 2011) to annotate each contig. Some contigs containing PKS/NRPS domains were tried to be ordered into scaffolds based on module architecture. Primers flanking the gaps between contigs were designed. Using the primers (Table 4.4.8), gaps in some scaffolds were filled in by PCR amplification and subsequent sequencing of the PCR products. The resulting sequences were then aligned together with the contigs in a scaffold to produce a continuous sequence. The nucleotide sequence was translated into amino acid sequence and the protein function was predicted. Based on this, module architecture was predicted, and the biosynthetic pathway was predicted.

Table 4.4.8 Primers for filling the gaps in NRPS contigs in sequenced “Entotheonella sp.” genome

Primers	Sequence	Contig	Lenght (bp)	Position	Distance from the contig end
14C-f	CATCATCCCAAAGTTTGATCCC	14	1646	1546-1568	100 bp
15DHKR-r	CCGCAGATACGGTCACACGTCAG	15	807	99-122	99 bp
15DHKR-f	GACGACATCATTTTGCACGGCAAC	15	807	707-731	100 bp
16DH-r	GAGGATAACGCTCACACACTATTTG	16	1264	98-121	98 bp
16DH-f	GAGCGTGCCGAAAGCATGCCGCGT	16	1264	1164-1187	100 bp
17KS-r	ACTATTTGGCGAGCCAGACCGGCG	17	1047	82-105	105 bp
17KS-f	CGCCAGGTGAACCGCGACCAGACCC	17	1047	947-971	100 bp
18AOPK-r	TGGAGCACGCCGGCTTGGATGTG	18	3242	101-123	123 bp
18AOPK-f	GGCGGCCCGTTGCTTCAGTTGCT	18	3242	3078-3100	100 bp
19PCy-r	GGCATGAAGAAGAGGGGACACCG	19	1873	100-122	100 bp
19PCy-f	CATTGGGTGTCAGGGGCAAGCTGT	19	1873	1772-1795	101 bp
20A-r	TGGATTACCAGGTCAAACTGCG	20	1465	101-122	122 bp
20A-f	CGGCAACTGAACAATAGGCTGA	20	1465	1365-1386	100 bp
21C-r	GAGCGGGATTGGAGCCGCAATC	21	1372	100-121	100 bp
21C-f	GTCCCTCTTTGGAAACGTCTTACG	21	1372	1271-1294	101 bp
22AKP-r	GTCTGCGGGACGCCTTGCACCTC	22	2091	79-101	101 bp
22AKP-f	CGGACGCGAATCACGGGCGTGGG	22	2091	1990-2012	101 bp
23PKA-r	CCAATCTAACTGGCACCTGGATG	23	3825	86-107	107 bp
23PKA-f	GGTCAGCATCGGTCATGTCATAG	23	3825	3725-3748	100 bp
24APC-r	CGGGCCAATCAGCTGGCCCAC	24	2823	100-121	121 bp
24APC-f	GTTGTGCATTGAGATAGGCTAACA	24	2823	2723-2746	100 bp
25C-r	GCCTATGCCCATCAGGATATTCCG	25	739	80-104	104 bp
25C-f	GACACCCATCGACCAGACGTCTGA	25	739	638-661	100 bp
26AP-r	GCAACAGATTCACCCGGAGCTG	26	1653	101-123	123 bp

Supplementary

Color highlights used to indicate conserved motif residues in PKS or NRPS domains:
KS/ C, AT/A, KR, OMT/ MT, DH, ACP/ T, TE

Supplementary for Sub-chapter 3.1

S3.1.1 The protein sequences encoded on the *sup* (sponge ubiquitous PKS) amplicons cloned from *T. swinhoei* chemotype W:

>TSW4
READEMDPQHRMMLETSWQALEDAGIDPDRLKESLTGVYTGISNDEYRMLVVDSTKPSGAAECLYALSGTNLNGTSGRVSFVLGLMGPAKAVDAACA
SSLVSVHDAVADLQQGKADLAIAGGVQAILNSRIYELRADAMMLSPDGQSKAFDASANGYVRGEGCGVVVLKRLSEAEADGDRIWAVIRGAAVNHGG
ASSGLTVPHTPALVQVIETALSQAGIPPSEVDYLEAHGTGTTLGDP

>TSW8
READAMDPQQRMMLETSWLALEDAGIDPAGLKGSRTGVYTGISNDEYRMLVVDSSRPTEAAGSLYALSGTNLNGTSGRVSFVLGLMGPAKAVDAACA
SAMVAVNDAVADLRQGKADLAIADGGVQAILNGRIYELRAEAMMLSPDGQCKTFDASANGYVRGEGCGVVVLKRLSEAEADGDPIWAVIRGTAVNNG
GTSVGLTVPHTPALEQVMETALSDAGVHPSEVEYLEAHGTGTRLGD

>TSW9
EAHPMDPQQRMMLETTWEALEDAGMDPDRLRGSRTGVYTGISNDEYRMLVVESRKPADAASCLYALSGTNLNGTSGRVSFVLGLMGPAKAVDAACAS
SMVSVHDAVADLQQGRADLAIAGGVQAILNSRIFELRADAMMLSPDGQCKTFDASANGYVRGEGCGVVILKRLSEAEEDGDRIWGVIRGSAVNNGGA
STGLTVPHIPALEQVMEAALSDAGVLASDVDYLEAHGTGTSLGDP

>TSW12
EALAMDPQQRMMLETSWQALEDAGIDPGQLKGGQTGVYTGISNDEYRMLVLDTTKPTEAAGCLYALSGTNLNGASGRVSFVLGLMGPAKAVDAACAS
SLVSVHDAVVDLQQGKADLAIAGGVQALLNGRVYELRADAMMLSPDGQCKTFDASANGYVRGEGCGVVVLKRLSEAEADGDRIWAVIRGAAVNHGGA
SVGLTVPNTPALEEVMETALSDAGVSPLEMDYLEAQAAPAPHWVTQ

>TSW30
REAEEMDPQQRMMLETSWLALEDAGIDPVRLKGTRTGVYTGISNDEYRMLVVDSSRPSEAAGSLYALSGTNLNGTSGRVAFILGLMGPAKAVDAACA
SAMVAVNDAVTDLQHGKADLALVGGVQAILNGRIYELRAESMMLSPDGQCKTFDASANGYVRGEGCGVVVLKRLSEAEADGDRIWAVIRGSAVNHGG
ASVGLTVPHTPALEQVMETALSEAGVLPSEVDYLEAHGTGTTLGD

S3.1.2 The protein sequences of *cis*-AT KS amplicons cloned from the filamentous bacterial fraction of *T. swinhoei* chemotype W:

>EKS14 (identity/ similarity 55%/70% McyG *Microcystis aeruginosa*)
REAKAMDPQQRLLLECCWHALEHAGYPPSQHGARTGVFAGSFLPSYLLHCLHGGGFMEPNNPGLGHLTEIGNDKDYLTTRVSHLLNLRGPSLSVQTS
CSTGLVAIATACQALLAGQCDVALAGASSLTLPQGGYQYLEGFVNSRDGQCRAFDADASGTVLGDGVGVVALKRLEDAEAAGDHILAVIKGFAVNND
GNLKADYSAPSVQGQVEVVAAAQAMAGVLAETISYVEAHGTGTCLLGDPIKLSNLD

>EKS30 (identity/ similarity 78%/85% JerC *Sorangium cellulosum*)
WRPLGSRFDSLLGEARLESLDGYQGTGQASSVLSGRLAYVLGLQGPAMTVDTACSSSLVALHLACAGLRQGECDLALAGGVQVMSTPATFVEFSRLR
GLAPDGRCKAFSAEADGTGWAEGCGVVVLKRLRDAERDGDRVLAVVRGSAVNQDGRSQGLTAPNGPSQQRVIDRALSVSGLSPDDTDAVEAHGTGTC
LGDP

>EP5 (identity/ similarity 60%/76% JamM *Lyngbya majuscule*)
SEAVPMDPQQRLFLEITWQALEHAGLDVTSYQGLIGVYAGSSMSSYYLNKVSKMPASSLASDLSAWLGNDKDYLTTQTAYQLNLQGPCVNVQTACST
GLAAVHLACQSLLNQECDVALAGGVTVKVPQESGYIYQENAILSPDGHCRTFDASARGTVFGNGVGMVVLKRLSEAREDGDNILAVIKGSAINNDGS
QKIGFTAPSLDGQAKVIAEAITVARVAPETITYVEAHGTGTSLGD

S3.1.3 The protein sequences of *trans*-AT KS genes from the filamentous bacterial fraction of *T. swinhoei* chemotype W:

>EKS37 (identity/ similarity 63%/78%, CorL *Corallococcus coralloides*)
SALIGSALHIGLNGAGMLSPEGACKAFDQEANGFVLGEAVGVVVLKPLSKAVADGDSIHSLIIGSAINQDGKTNGIMAPNGPSQTALIESVYRKFEI
DPKTISYIETHGTGTSLGDPIKLSNLDPRGRH

>EKS48 (identity/ similarity 59%/73%,SorD *Sorangium cellulosum*)
XGGPSSRGIDSLISEAHAMDPQQRVFLEVAWAALEDAGHANRAISNARCGVFVGAGPSDYVTLMRERGVPGEPQSFWGNSSSLIASRIAYFLNLKGP
SITIDTACSSSLVALHLGCESIRRGECDMVIAGGVYIGVTPHFHIIASNAGMLSPDGQCKTFDQGADGFVPGEGVGAVVLKPLSAALEAGNFIYGVI
RGSGSNQDGRTYGITAPSAASQMELEQRVYRNAKLDPASISYVEAHGTGTTFGDP

>EKs50 (identity/ similarity 61%/74%, Bat3 *Pseudomonas fluorescens*)
XPRVEVSISLIREALAMDPQQRMLLQTVWKTIEHAGYRASELSGGNTGVFVGVANSDYRELVQAAAPGVEAQSSTGTFSSILANRVSYIFNFTGPSE
PIDTACSSSLVALHRALNALSRGECASAIVGGVNALLTPTLHITGGKTGMLSQDGRCKTFDQTADGYVRGEGVAALLLKPLSLAKAHGDEIYALIRG
SSVNHGGHATSLTAPNPRAQAELLIRAYEDAGVDPTTVTYIETHGTGTTFGDP

>EKS64 (identity/ similarity 58%/72%, DszA *Sorangium cellulosum*)
VAAPRVEVR LDREADAMDPQQRLLLQSVWATIEDAGYPAAALASENIGVFVGVQFNDYQQLMKQAGENGAQILTGNAHSVLANRISYLLDVRGPSL
AIDTACSSSLVAVHQGVQAIRNGDCTAAIVGGVSLMLSPDTLLDISQLGVLSLDGRCKTFDRGANGYVRGEGVGSFLLKPLSRALQDRDVIHGIIRG
SAVQHGGKANSLTAPNSEMQAALLLNAYRDAGVGLDTIGYIEAHGTAPARASGI

>EP2 (identity/similarity 66%/79%, StiG *Stigmatella aurantiaca*)
VRPWRWIRSKRLLLEVAWEALEHAGQAPDKLVGSATGVFVGLSTADYFYLMSRGLDAEDIDAYVGVGSTASVAAGRLSYILGLQGPSFALDTTCSSS
VVATHLAMQSLRRAECKMALVGGVNLMLSPMSTIYFSKLRALAANGRCKTFDASADGYARSEGCGIVVLKRLSDAVADGDRVLAVLRGAAVNHDGRS
GGLTVPNGPSQQAVIRKALANAGVEPNEVSYVEAHGTGTSLGD
>EP3 (identity/similarity 52%/69%, phenolpthiocerol *Sorangium cellulosum)*
LRQWTRSSGCLLMTSVEAIQDAGLRPSELRQNETGVFVGASNIDYNLLARQRTGHGDIQAGTGTAFSILANRVSHQLDLKGPSLTVDTACSSSMVAL
DIACEKLLSGACDSALAGGVNILLDPRMFITFSRARMLSLRGEISAFDARADGFVRGEGAATILLKRESDALRDGNRIYAVIRATAINQDGATSTIT
APNEEAQTRLFEKAVALAGIDPKDVTFVEAHGTGTSLGD

S3.1.4 The protein sequences of genes coding for adenylation domains from the filamentous bacterial fraction of *T. swinhoei* chemotype W:

>pAU1 (60% Arginine code: DVEDIGAIT-)
GCRNSIAYAIYTSGSTGQPKGVLIEHGALANHCRTMQDTYGLTASDHVLQFAALNFDVSLEQLLPPllagatvvlggLWPPSLFHKKVQECGLTVVD
LPPAYCQQVLEIWVADPALMRDTPLRLIIVGGEAASPELAAWWQRADLPSVRLLNAYGPTEATITATVFDVPRDIEPARHRLCLPIGRPLANCQVYV
LDAARQLVPIGVPGElylggaglargyLNRPELTEAAFIANPFGSGYLYRTGDLIKLIDTVDLE

>pAU2 (60% Arginine code: DPKDLGLID-)
AYEIYTSGSTVQPKGVVMSGKATHKLVTLSQELQLGSHDGIGSLTSIGFDPIFKQLVCSWSCGATTVLIPDGAMTDGSLFWSYVLTPHDRVLDTTPA
FLMSLLDAPPYDIRLRVLLIGGDVFPAAMLEALRCQLVSMAIYNLCRPTDTCIDATLFDATNHRVGGNVPIGKTQANYSISILDSCLEPLPIGVVGE
lyiaglglargylgrgglTTDRFLSNPYGSPLSRMYRTDDVNEFLQPGEIHFYSGR

>pAU3 (substrate specificity was unidentified because the sequence is too short)
ATITATVFDVPRHIEPARHRLCLPIGRPLANRQVYVLDAARQLVPIGVPgelylggaglargYVNRPELTEAAFIANPFGSGYLYRTGDVNQAYRYR

S3.1.5 The protein sequences of genes coding for serine-specific adenylation domain from the filamentous bacterial fraction of *T. swinhoei* chemotype W:

>pEt-A3 (70% serine code: DVWHISL-)
FDVSVWEFFWPLFTGARLVVAQPEGHQDPSYLVSAIRNHQITTLHFVPPMLSIFLEAPEVETCDTLRRVICSGQALPYELQKRFFSKLSAELHNLYG
PT

>pET-B2 (70% serine code DVWHISL-)
FDVSVWEFFWPLITGACLVVADPGGHRDSAYLVRLLREQHVTVLHFVPSMLEVFLEEPALDDTPLPNLRKVICSGEALTGRHRDRFFEQLPGAELHN
LYGPT

>pEt-D1 (70% serine code: DVWHLSL-)
FDVSVWEFFWPLMSGASLLIAQPEGHQDPTYLCDLIQRERVTFLHFAPSMLRQFISDPNSQNCTSLRVILTSGEALPWDLAQDTAHLLPHTELHNLY
GPT

Supplementary for Chapter 3.2

S3.2.1 16S sRNA gene cloned from the metagenome of *T. swinhoei* chemotype W (unique regions (highlighted) was used as the basis for designing *Entotheonella*-targeting primers:

> 16S sRNA gene cloned from the metagenome of *T. swinhoei* chemotype W
TCCGGTCTGAGATGAGCTTGGGGGAAACGTTCGCCGGTCTGAGATGAGCTTGCGGCCCATTAGCTAGTTGGTGGGGTAATAGCCCACCAAGGCGACG
ATGGGTAGCTGGTCTGAGAGGACGATCAGCCACACTGGCACTGAGACACGGGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGCGCAATGGG
CGAAAGCCTGACGCAGCAACGCCGCGTGAAGGATGAAGGCTTTCGAGTCGTAAACTTCTGTCTGGGGGAAAGATGATGACGGTACCCCCAAAGCAAG
CCCCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGGGGCGAGCGTTGTTCGGAATTATTGGGCGTAAAGGGCGTGTAGGCGGCTTCATAGG
TCTGAGGTGAAAGGATGCGGCTCAACTGCATTAAGCCTCGGAAACGGTGGAGCTTGAGGCTGGGAGGGGCTGGTGGAATCCCCTGTGTAGCGGTGAA
ATGCGTAGAGATGGGGGAGAACACTCGTGGCGAAGGCGGCCAGCTGGACCAGTTCTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGA
TACCCTGGTAGTCCACGCCGTAAACGATGGGCACTAGGTGTCGGCGGTTCTTAATCCGTCGGTGCCGGCGCTAACGCAGTAAGTGCCCCGCCTGGGG
AGTACGGTCGCAAGGCTGAAACTCAAACGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGAAGAACCTTACC
TAGGTTTGACATGGACGGGACCGCCGGTGAAAGCCGGTTTCCCTTCGGGGCTCGTTCACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGAT
GTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGCCCTCTGTTGCTACCGGGTGATGCCGGGCACTCTGAGGGGACTGCCTCGGTTAACGGGGAGAA
AGGTGGGGATGACGTCAAGTCCTCATGGCCTTTATGCCTAGGGCTACACACGTGCTACAATGGTCAGGACAAAGGGCTGCTACTTCGCAAGAAGATG
CCAATCCCAAAAACCTGGCCCCAGTTCGGATGGTCGGCTGCAACTCGCCGGCTTGAAGCTGGAATCGCTAGTAATCGGAGATCAGCACGCA

Unique regions:
"Entotheonella sp." : GACAAAGGGCTGCTACTT (in this work)
E. palauensis : GACAAAGGGCCGCTACTT (Schmidt *et al*., 2000)

S3.2.2 The protein sequences of the proposed misakinolide A biosynthetic pathway:

>**TP1**, position 489-1022, length 178 aa, Transposase
MISLQLCWRTHSVLQMRKISPNVSTARSITHRGPSMTTVTNNLNQSTWDCKYHGVFTPKYQKKVLYGAIRRDLREVFHRLAKQKECTIESGHLMPDH
VHMLIAIPPKHSVATIVGFLKGKSSIWVAENVDNKQRNFVGHKFWARGYFVSTVGANEQVIRAYSENQEKADQRFQNLFDR

>**TP2**, position 2225-1545, length 227 aa, Transposase
MCMRVPSSFQDHFASLSDPRSFRTPNLRHDLIDILVMAVCAVIWGAEGSEDLEDYGYAQAAWFSDVLDLPHGLPSHDTFRRVLSQLDPDELTKCFVS
WTTALSDLAGGDIVAIDGKTLRQSFDRATSQSAIPMVSAWATANRLGLGQVKVDDKSNEITAIPKLLKILDLSGATVTLDAMGCQKEIAQVITEQGA
DYVLALKKNHSTLYGDVKLFLDDAQASGFADLD

>**TP3**, position 3664-4659, lenght 332 aa, Transposase
MRFIHPLSHETLQLLNRVHKQSRHHRVRQRAHCIVLSSQGYTTTHRADIFQVDRITIYHWFNAWEHNGFPGLYSRKASGRLPLFTPDQKEQIRQWIK
LFPRNLNKIRSLIHDEFGLKVSKQTIKRVLKSYRLIWKRMRKRVKGQPAPELYQQRKAALEILQTEDCDGIIDLRYFDESGFCLVPYVPYAWQQSGH
TISLESTQSKRLKVLGFMNKRNELEAYTFEGTVDGAVVIRCIDEFCKTLQGPTIVVMDNASLPTRPTFQAQLPRWEQQGLEIFYLPEYSPHLHLIEI
LWRFMKYEWIESWAYTSWSHLVRYIDEVIQNVGDKYKIHFG

>**ORF1** (position 7211-4803, length 803 aa, Esterase/ lipase-like protein)
MSQPGRSTRTYHLDVDCRFERYNYERSRGLHHRISNRATLTMQPPELDLPPINTRNHNLTLFLLRRLESNLDVRIRLHGLTDQLDMGGGIIVANHFT
RLETFVVPFVVHRATNIILRILASHVLFGNKTFGDYLMSIGALPTNHPNKYELIARDILNGGWWLIFPEGLIKDRRVIEKGRLNVTTDAGQVRRRPR
SGSAILAMTVQRYKAALRQALEHNGDIQSICESLGLDDRRPSELKALAFRPTPIIPLNITYYPLNPQENAFKSLATHILPNLPHSTFGQELLEELTV
EGSMLLKGVEIDMRFGNPLLVGDDDLQFVDDWRIVPWSPSPWRRYLNALRSWRPMRPYTHLVDRWAVSSSWRQRQRAWQITRTTMQALYRLANIHIG
HLISAIVLQTLRTTKQRRFEVLELQKRLYLATRRLQEGGNLALHPNLTDPDLQYLLATETPHPFLEDFARRASELELLTVKDQMWEPNTEILLEPWS
FSSIRLKNFIQVYFNEIEPVTEVMQAVRYAMRAELGDYRELFSDAMHNYEDQLYEADHAEFVTPATAAKLTPLVPEHGRPALLRGHGSNRAIGILLL
HGFSASPGEVMPLAKTLNGRGYTVYVVRLRGHGTSPYDLQQRSWQDWYDSVQRGYAYLRTITDVQFVGGMSTGGGLALYLSVHEPTVCGVFAVAAPI
RLTNRAIHLAPIVKTVRGFVNSNPGNPHTNYLAFPVQALQELLQFISAYSKTLQDVTVPTLLIQAQGDTTVRPDSAQYIYDELGTETKELIWKDANQ
HVIVSDAFPDVHDDVLTFLQQHQPSS

>**ORF2**, position 8065-7388, length 226 aa, NhaP-type Na^+/H^+ and K^+/H^+ antiporters
MPWPERAFLAWVAPRGIVAAAMASIFTLRLQAEGYAQADQLVPITFLVIIGCVAVYGLTATPVARWLGVAQDNPQGGLFISASPLLQALADALRAED
IATCMADINWSHIATARLNGFKTHYGNALTENALDNMDLDGIGRVLALTPNDEANALAALHFAELFGHAEAYKLVPSTANRETVPHLRGHFLFDESA
NYNVLSLHLASGATIKTTGLSEQFDYEAFQDY

>**TP4**, position 8253-8933, lenght 227 aa, Transposase
MCMRVPSSFQDHFASLSDPRSFRTPNLRHDLIDILVMAVCAVIWGAEGWEDLEDYGYAQAAWFSDVLDLPHGLPSHDTFRRVLSQLDPDELTKCFVS
WTTALNDLAGGDIVAIDGKTLRQSFDRATSQSAIPMVSAWATANRLGLGQVKVDDKSNEITAIPKLKILDLSGATVTLDAMGCQKEIAQVITEQGAD
YVLALKKNHSTLYGDVKLFLDDAQASGFADLD

>**ORF3**, position 10299-9400, length 300 aa, NhaP-type Na+/H+ and K+/H+ antiporters
MTEHVLLGLASIVALGIAAQWLAWCIRIPSIFLLLLLGLVAGPFTGWLQPQHLFGELLFPAVSLSVAVILFEGGLSLRTVVLRDIGRVLRDLVTLGV
LVTWLITSAAAYVFLGLDGSMAILIGAMLVVSGPTVVGPLLRHVRPVERLRSTLQWEGILIDPIGVLLAVLVFEVILVGELQAATTLVITGMLKMVV
LGSALGWVSAKTVVWLIQRGFIPDYLHNAVTLMIVVVSFTFANVLHHEAGLFTVTVMGIVMANQKDILVRHIIEFKETLTVLLISGLFILLSARLNV
NDLAQLGQG

>***misA***, 10649-13330, 894 aa, phosphoenolpyruvate synthase (PEP)
MSMDSSPEPTRDHQILCVAFPGVTPAPLAEVGGKAHSLMLLAQQGFQVPPGIVLTTAFFQPWFEQIQRTACWAKLLAAEPDEWSSLCRELKALSAGL
TFGEAQQQALEGLREDLHALGDARFAVRSSSLEEDLPQASFAGGYETCLGVRSEDLEIAVRRCFASTLDERVLMYKRERGFDIRSPRLAVVVQQQIH
SEVAGVGFSLNPLNNDYDEAVINANWGLGESVVAGRVSPDHFVLDKVGRQVVERKQGAKQLSIWVSEAGGTTEKQGYRSQEFSLNDAQLSELIDAIC
EIETFYQRPIDIEWSYANDQLYLLQARPITTYIPLAPRMMTQPGERRCLYADMALSSGLTINAHISPLGLDWMREGLVLLADMTFGKLDFDLRFEEA
FCFCAGGRMYQNVSNVLWLTTPEQLAKMSAASDVLMAETQAYIDADRYRSQVKPPWAKVAMLKYLPRVMWRGRYFVGTLLWAILAPEHAWRSYQVKA
DGYKTELTGDLDYSLSLDAFRRQYTSTDNARLMLTFTMSVVWASVGAIAMVDVLVGKSAEGKALGDKLKRGFTGNIVVEMGITLFRLAKMLPPSDFD
DLDQLVERINTRQMPAQFCHAWDAFMAKFGCRGPLEMDLASPRYADDPVLALKQMSLMPLDDERFDPELTHQRQVEEREQAFAALMARSGWLRRRLL
RRVYHVIDLFAGIRDHPKYYMLLFYQGLRKRALSEGERLTREGRLDASEDIFDLTFDDIDSAVRDPALDLRAIRQQRTAFNQILSNQVHTFPAVIDS
RGRILRPPARESKPGEYRGMAVSAGLAKGPVKVLQRPDEKPVDKGDILVVYTTDPGWTPLFVSAAAIVIEVGGALQHGAVVAREYGKPCVVGIDQLI
SQFEDGQQVEVDGTAGMVRLL

>***misB***, 13444-14193, 250 aa, phosphopantetheinyltransferase (PPTase)
MICAMSCLRITLNRQDLTTDAGFAWVHAADYPALHDTCSRVLHVNELAYYQRVPAIRRKTSFLIGRYAAKQALSSLVHESDYTRIEIASGIFQQPIV
KYPSPEPLGVSISHTDKYACAIAFPLIHPLALDVERIDRVSVKAMKSQILPHELQTEALARLPEPTRCAIIWTAKEALSKVLKCGMMSPFCIFETMD
LSQEPDWIVGHFKNFGQYKFHAWVLDDHVVSMAVPKRTDIDVDLSTLLHDLLPIGD

>***misC***, 15451-33060, 5870 aa, KS-AT-KR-ACP-KS-KR-ACP-KS-OMT-ACP-KS-DH-KR-ACP-KS-
LEAKDIAVIGVGVLFPGASHHDLFWENLIAGIDSIQEIPVSRWDIDAYYSPDFNEPGKSISKWGGLIQGIDQFDHQFFNISPREARNMDPQQRLLLQ
EAWHCVEDSGIALKTLQEKVTSVFVGVMAIDYHQHMAQSRQPVDSYACLGNYGSILANRLSFCLGLSGESKAIDAACASSLVAVHDARRSLQAGECD
YAIAAGVSVICHPWKYISFSKSRMLSPSGRCKTFDKSADGYVPGEGVGVLLMQSLERAQEQKCHIYGVLKGSALNNNGLHGQRHADDRGRKKASITA
PSIDAQRRVITDALRSADIDPASVSYVEAHGTGTSLGDPIEIEALKQVFAVPGRTTPSCAIGSVKTNIGHLEAAAGVAGVIKVLLMMKHKMIPQVLH
LNTVNPMIDLTDTSLYLSDQLHSWESKEGIQRRAGVSSFGFGGSNAHAIFEEYVDTRHTPQGTLQRTSGEHLFVLSAKSHSSLDGLLNDWVAFLNHP
QRSAISLEDMCATLMTGRESFPFRFALSVTDKKALRSALISPPESAFGHAEMPSVLLRIGSVNAEQAAVIASWYRKHPGFREPFDACRRLAKMLCSE
EAVRDWTAQFPKVLAEHEALYGFAAAYATARLLLDSGVQPRLICGEGIGHGVALCLSGLWTLSQGLRHAMRLIDQLDLSPQRLDYPIYDQIQGTTLS
PYQPSRGYLQALLKNIQVRDDLFIRDLQKALLLHEHQYTFKNYLESWNSALKEIGLEIYQLIGLHAELQEDISKKFLVYIIVAYSLKQLYEKWDLVE
YDASSDPFLNEFLDLIGDDVLTFKDVVTLIGNGWDTGKEDLFHGITARVIPVSSVKPYEYLQQYQFERFDHIRSWLEKIGDVPLETPIHEDDLLLTV
GDEGDVSRLKIVSNSGHSLSFLTRGQKASHLTEDVRQLWLHGLNIDWQSFVDAGSFTKASLPGYSFDGSPFWIENSRSDSPFIQGEGIKAEPVQNSY
YHFVWQASKHPTVLDRKPDVKNNEAILLFDSDASLAQLLREQWHAEIILIRPARAFDWKSWHDVEIDPDNENHYKLVCTALQQKNLRPVCAIFTWTL
QESTSTGESLLSHPEPWLSKGIYAVYHLLKSLSQSAHPPLSQLLFFFHGNEGQGNPLMEAVAGYGRSLKLTYPQMRFRSIQFETGSPDTIEVVTAVI
REFSLLKGEERQEVDTEVLYRHGQRLIKTPVRLHLDAVAQAPWKQSGVYLITGGMGALGSIWARYLAERYQARLILTGRSVCQGDIGKSIRVFIKQL
QQLGGDAVYLQADITEPHAVRQLIEQAYQAFGPVRGLIHAAGQMSAVSFAEKSLADFQQTLSAKIQGTLVLDRVLHESIEKQALDIVIYFSSAASTL
GDFGQCDYAIGNRFLDGYGRFREGLRTQGQRQGRTVVVNWPLWREGGMHLDSEAEALYLSASGLDYLEAEQGLDALDRILVSPHTQVMVFAGDPANI
DRNLGVNKVVDMQVDETFVASRIEADLQHLAAVLLQMQADEIALDDNLSAYGFDSIAIKEFADQLSQYYQLEISPIVFFAHASLRSLGAYFWEHFAA
DMKAFYAAEPSNQMPEARHAEAPARQVSQPARAEADQDIAIIGMSGVFPGSPDLATFWKHLSRGRDLVGEVPAGRWDWRVYDGDLMREHEAKKTAIK
WGGFIENVDRFDADFFSISRREAEFMDPQHRLFLQCVWATIEDAGYATAALSGENIGVFVGVQFSDYQQLLKQAGERGAQILTGNAHSVLANRISYL
LDVRGPSLAIDTACASSLVAVHQGVQAIRNGDCTAAIVGGVSLMLSPDTFLDINHLGVLSPDGRCKTFDRNANGYVRGEGVGSFLLKPLSRALQDRD
PIHGIIRGSAVQHGGKANSLTAPNSEMQAALLLKAYEDAGIGLDTVSYIETHGTATELGDPVEIEGLKQAYRTYRQQTGAAASRTGACGLGSVKTNI
GHLEPAAGIAGLAKVLLAMRHQILPATLHFDSLNPYIDLSESPFYIVDEKTPWQRRLDEQGEEIPRRAGVSSFGFGGAIAHLVLEEYRVPQDTLENG
DDDTSPYLIVLSAKNEARLKEVATNLYGHLIETASVAPIKLRDLAYTLQVGRDALPQRLALVVDETAHLLDGLAAYVQGEPDRGRWFAGTARRDRCP
LTSLVEAGEAMIAMRIEKEALREFAQSWVQGASINWQQMYPVNKPRRQSLPTYPFAQQRYWFDTSANTKPQERHAAVAPIDNKSNIERIYYRREWAQ
TALLSPVSIETGDQAQFVLLFETNPQIGEVWQTQLEQHGLRCIRVCIGERFRLLEPHYYQLNPRQYDDYASLAQALRQNASWPRRVVHLWQSNEPDV
NHDQHLSLQLERGFYSLCHFTRALMQHTFGDACDILYSYRAEAWQPEPACAAVSGFAHTLALEHPGFSCKTIEIPAFLSYRDARTIDLLLSELQGHA
LSDTTVRYQEQGTLQRWVQHFCEWQASPIEVVSKYPRPHGVYLITGASGSLGAQFARHIAAQVKATFVLCARSERTPACEQLINTLRTMGSEAVYIR
SDLSQRDHVFKLIEEIQARFKALHGIVHSAGVISDALVFNKTHHQIAQVLAPKVYGTVYLDEATKDIPLDFWVLFSSTTAVTGNVGQSDYACANRFM
DEFAVWREQLRGMKQRSGKTLAINWPYWRDGGMSMSEPGYQGWMERTGLQALPTSEGVAAWDDLLNAPSGHAYVVAYGFRELLSRMFARPQSAAAQP
SADERATSSETLYEDTMSWLQHEIASVVKASPQDIDVDTAFHQFGLDSVVMRQTNVVLEKKLGALPKTLLLEYTTIRDLTDHLSTQYIRELNGVLGS
AFLEQSPNGERQVRKRVGSENSMRNGIFSAEPIDSQALEQEQTNNLEEIAIVGLSGRYPQAEDLDEFWENLKAGKSAIADIPDSRWDYRAYYDADPD
QASRGKMYCSQGGFLNDIDQFDPLFFNITPLEAKSMPPEERLMLETVWSTLEDAGYAGEHMRSQKVGVFVGVTSNTYPLECDHSDGVRYQSPIDTSY
FNIPNRISYFFNFTGPSIAIDTACSSSLVAIHMACESLRHGECEMAIAGGVNLYLHPSKYILMCQKRLLSTKISTGMFEEGGDGFIPGEGVGAVVLK
PLTQAVRDRDNIYGVIKASSVFHKGRSNGYLLPSPESQTELLRQTLQQAKVAPETINYIEVQALGSEMVDQVEWRSLIQAYGAKGKSDEYKPYCGLG
SLKPNIGHLEAASGIAQLSKVLLQMKYCQFLPSNVSDRRDPILDASINHPDSPFYFQGEWRVWERVSRQRDDTVGLTVQPRRAGINSFGAGGVNAHV
IVEEYISPVESSHAPRLEEESAALIVVSSKTQQQLRQALVRLRAYLLREANRADGPSDRVSLADIAFTLQQGREAFAERWAMVAKSKAQLLESLTRF

LDDKTVHAPLFVGRVPSTNHENHVFMTHDRIEALWHSHQLEKLAQLWRQGQNISWEHLSYRYSPRRVSLPTYPFEKRHCWWGTSDDKDPVEVTAQSM
PDLISQKTHETRVNMSDIGQVVADYYDQVTESLKQKIGSQEVYLLFAPFQERVEGFSWLKMFFEPEKHREHFELTLQKQHELKSMLYRLVEFANVHR
IFDIGCGYATDLIQLAKRHPQVRGWGFSISLNQVEFGQQRIQDEGLGERVQLSCNDSTKDPLPGLFDLIIGFEVVVHIADKQGVFGNMARHLSNNGT
IVLADCVANTVTSINTSYIGQYTSTAEEYSDILAQNGLAIVDCVDVGPEISNFLYDPNYLENAAYLSSIYPEMVNIEKEHQGWSNFGKALEVNLVRY
VLLTIRKATASQTYDHLLQINLERLGSGALSYAEALQAYPHIEQIVIEQIAETKSPDSNATTTSVADIEQTIVELAAQVLEMMPDEVNTQARFIDFG
VDSLRGLILLLDAVNRRLGLTLQIPVLFDHSSIHELSIYIASQLPSSPQAVVSAPAASVPVVSPAVAVSSVRSVASSHTAHSDIAVIGMSGRFPGAPD
LHTFWQNLEQGIDSISEVPPTRWELTQYYDADRRLPNHSYSKWGGFLDDVDQFDPKFFNISPAEAELMDPQQRLFLQTSWNALEDTGYAGQTLDGMK
CGVYVGVLNHAYDILIDRANISQHRAYALLGNSSSILAARIAYLLNLKGPAVTLDTACSSSLVALHWACKSLRDDEVDMMLAGGVTLYLGEREYIEA
SQAGMLSPTGKCRAFSQEADGITLAEGVGVVVLKRLDKALQDRDAIYGVIKGTGINQDGKTNGITAPSAKSQYELQTEVYTRYGIRPERISYVEAHG
TGTKLGDPIEVEALTQSFRQYTQKQNYCAIGSVKSNIGHTSAAAGIAGLIKILLSLQHHSIPPTLHVEHKNEHINFRDSPFYVNTQLRPWTHEDDLP
RQAAVSSFGFSGTNCHVVVEEPPARIVHQHIASTLDAPVLMVLSAANQDRLRGYVALMIRYLEARKTRQHTRKAWPSDSLTDIAYTLQTGRTPLEQR
LAMVVASEDEWLDKLRRYERGENVIDACYEGQANAEPAKLDLLIEGEEGRAFIESIVRRKNLSKLAQLWVGGTDIDWTLLYHRPSSDAAVLPQRISL
PGYPFARESYWLAAPTLAASSLVVERLPHQDDICFKTVLSGDAFFLRDHCIDRQKVLPGVVYLEMARKAATSFRPDAPQPVLKNVIWLRPLMVNATD
VVAYTRLKQQAQRIDFEISTLDEQNNRVLHAQGQLHDPEPVTPQHPDWPSFDLRLIQERCPRTLEGEAFFQAIQAQGFRTGPSFQVVRKAWYNQTEA
LCELGLPALREADFQDYLLHPSLVNAALESVILLIAKDPSASDAPHLPYTLGTWAYDDTVWPTSLYAYARLTAPQPETEALTAATTVQRYEINLLDR
SGRLIARGKDFCVRALRSSGETRPASSAAQYPVKYYYPRWEACENERVAANIVVERSWLVFETENSLWHALDEKLRALGVTCIQVRPGYRFQVLEKR
RYEIHPERAGDYQKLVAELKRNHAIPERIIHAWSEEQWCDNAELDPRSLAWGFYSVFYLSQALISCWQDTPLHLLYLYRQDQDSPQAHYAAVRGFAQ
SLTHERLDWHWKTLALPTTLPQAQVANMLWDELRYANQLEREIRYDANRQRWVRRWHAGEGTPGQGPAAPDTSLLRQQGVYLITGGGGGLSLLLGEY
LAREYQAKVILVGRAEGTEAQQAKIDSWKRWGGEALYLQGDISRREDVDHVMAQLRSRFPVLHGVFHAAGLTRDASLVNKTVRDIDEVLAAKVYGTI
YLDQATRSEELDFFVLFSSLSGVVGNIGQSDYAYANSFLDHFAERREAQRSMQKRFGLTVAIDWPLWRHGTMQVDATTEKFLARSIGMQALETDDGW
MALQHALAGSKSQLLVLCGDQGRFSQALGVQQHHTVAERAESPSENLSETSSFDLLTHLQIDLRSIMAAVLKLSPQEIEWEREMSSFGFDSISFVAL
ANELNDTFELDLLPAVFFECLNAGDLLKHLHHHHCDALFAYYRRRVPEPAAQEVLGTPGQANVEGPVHVRQLEKPSGLNHHAGTEPIAIVGMSAVMP
QSPDLETFWQHLAAGQDLITEIPKDRWDWQNAGGDTTIRWGGFMPDIDQFDALFFGISPREADLMDPQQRLLLQTVWKTIEHAGYKASDLSGSDTGV
FVGVANSDYRELIHAAGPGIDAQSSTGTFSSILANRVSYIFNFTGPSEPIDTACSSSLVALHHALNALSRGECSSAIVGGVNALLTPTLHITSNKTG
MLSHDGRCKTFDQAANGYVRGEGVAALLLKPLSLAKAQGDEIYALIQGSNVNHGGHATSLTAPNPRAQAELLIHAYEDAGIDPTTVTYIETHGTGTS
LGDPIEISGLKQAFATLYERFGHVADVSAYHRCGLGSVKTNVGHLEAAAGMAGLLKVVLAMHHKTLPAHLHFHNLNPYIDLSGSPFYITRQTCSWDA
LQDAAGKALPRRAGVSSFGFGGVNAHVIVEEYVEDRHEDR

>***misD***, 32948-60028, 9027 aa, DH-KR-MT-ACP-ER-KS-DH-KR-MT-ACP-KS-KR-MT-ACP-KS-KR-ACP-KS-KR-MT-ACP-KS-
MKTDSPNHPENQAINVNHYLMVLSAKTKERLQTVVENLLDYAKANRRLIESRLADVAFTLQTGREEMEARLAFVVSDYHDLMGQLSGYLEDAPNANT
IFRGQVDLFDQSAPVRQHRAAQERDCERMIRTHDLPGLAELFVQGVPIDWRRLYVQGRPRRIALPSYPFAPERHWVARREPDHLAADMPAVRQPHSL
LDGRDDRMSLESGVVFRKLLRINDGVLQDHQVNETPVFPGVAYIEMACEALAQLSFSDPEVLRKTVLNQMVWLQALAVTEAEHEILLCLEKDTHGLA
FEVKDHQTLYAKGRFDQSDDLSQGHQYLELETIKGDCPHEKHRSVIYHDFRLLGLNYGPWFQGLEHTWFSIDSAEGEGEALGYFNMPLICESEFSAY
TFHPGLLDAALQVVASVPGYADSGLLLPYAVEKLEWFHPFPACGYSHVRKLAKHRYDVTICDEQGKVCAKFSDLSVRPSRVAGSEPSLNQLTYRARW
SHQELGERRSPAGRPEETGVVFIIHTPEQQALVQELSSVLADMHGDHRVRRMCLNANDWDFKSLGNWNHRIRAVQDLRDIYFLAAMEHTVSSGQAHV
DPLSFTNRQEHGLLSLFRLLKALDQNQLLLQPLSLKVLTTGNYRVLPSDTICPLSAGLPGFCRSIAQEYPLLALSCIDLGCPDGDAMQPSSLKKLAQ
FIVAEPAHQPHETVGIRELRRYTRKLVPVRFPEPGGQRAFKSRGVYMIVGGAGGIGMVLAQYLAERVQARLVLVGRSALDEEKKQNLSTLTAAGGEA
LYIQADISRLYSMKDAVAQARQRFGNINGVVHSAMVLNDQSLARMSEEGLRASLAPKVEGCIVLQQAMAGEPLDFMLFFSSAISFSGNAGQSNYAAA
STFKDAFAHYLTQVEAYPVFNINWGYWGCVGVVANAAYRERWAARGVHSIDPEQGLQVIESILTNHIDQVAAVNVEPSFLERLGVELSQTQTVLPPR
RSSLIAKLNALPLPETGGFQDNVNRLEIVSAFRWAEVFKHMLDMSSTQGMYARSDLKRQLQISSEHDQLFDVLISMLIQEGYLHQQGKNLTPTPYLN
GEFFQQQREQVALLESELSQTDNAYLNLLSRCLDHYPAMLRGEVAATDVIFPQGSMEWVEGIYRSHASADYFNQQLAQRLVAYVEARLPSLGANESI
HILEIGAGTGGTSAIVLQALQPYAARLRYTYSDISPGFLNYGRKTFGDAYPFVEFRHLNIEHNGEAQGYQAGEYDLVLAANVLHATRHIGNTLARVK
FFMKRHGWLLLNEATQVRHFGTLTFGLLKGWWLFQDAENRLDGSPLLSTTAWQRLLSEEGFKHTLSLCDAVEEESIFQDVIVAESDGVIQMNQSDSP
VQPREPVDVSVDVHHVDDMMEPGQVADVVESRHDFIVNSVSQCMAQSLGVAFTDIHINRPFSLYGVDSITGVDFIHQINSRLGIVLRTTALFDHATI
FDLSRFIDESHGAAIDEQRQRLRSTPLTSSDTKPAERVAPPDEALPHIERGCHQAVVLEKPSQIEHLRFHPVKTREPAAHEVQIAVRAFSLNFGDLL
CVTGLYPTMPAYPFTPGFEVSGHVANVGAAVSHVAVGDEVIALAGQQLGGHSTLLTTAGHLVVHKPRQVSHEEACAFPGVFLTVYHAFQSANIQAGE
KVLIQGAAGGTGLIAVQMAQAVGAEIYATAGSPQKLDYLRERGVPHLINYQTDDFAEVILADTNGYGVDVIFNTAAGDAMQKGLDILAPNGRYLEIA
MVALKRSRSVDLSHLVDNQVFYSIDMRKLIAGNPQLISRYLNRMVEHLAEGHVQPTIGARFALDDIQSAYRLLHDRQNIGKVVVSIPDGMMDNQVGQ
RMVPESSTVEDYRVGNATVEQVAVKQPADMDEPIAVVGMSGRFPGANNLEQFWLNLIGGNNAITEVPADRWDVHQWYDPDPAVLDKTHCRWGGFLRD
IDKFDALFFKLSGREAELTDPQQRLFLEVAWAALEDAGHANRAISNARCGVFVGAGPSDYVTLMRERSVPGEPQSFWGNSSSLIASRIAYFLNLKGP
SITIDTACSSSLVALHLGCESIRRGECDMVIAGGVYIGVTPHLHVIASNAGMLSPDGQCKAFDQGADGFVPGEGVGAVVLKPLSVARQAGNFIYGII
RGSGSNQDGRTYGITAPSAASQTELEQSVYRQAELDPASIGYVEAHGTGTKLGDPIEIDALNRAFQSDAKKQHSCAIGSVKTNIGHAGAAAGIAGFI
KVLLCLQHQQLVPSLNYHTCNEHIDLDNSPFYVNTESKDWFVAQGEPRRAAVSSFGFSGTNAHILLEEYQHHEDEAQALGDDKDVLIVLSAKNQERL
KKAAENLHHFLESRVSYKLSDIAYTLQVGREAMEARVAFMVVSVLDLQQKLVSLLHDPTDEQDMYWGEVRDGEGPAGVAAFEDVALALANKQLDQLA
QWWVSGVEIDWTQLARVVKVRLVPLPTYPFSPDRYWFDSVGPATVVHAEHSASALYPLLNRIAPQASTQIANGCVFETCLKSSDPLLAQHLIDGQAV
FPAVGFLEMVIAGLREIQPDKAARLENVFWLRPLDVMDEQRVHLVFEVKQDRFQYRIVPAESHGDGAYFRGQGYWSNDMNPAIPMDIDRLKSRMRQQ
PLLNESSPRVRLGRLYQGATEGWTDGEQALIAFELPTAQPYRDGSPDQKSYPFHPAAFVAALQAYHWIHEAPKRLGMPFAVQKVELWRHRDQPAYAY
VALDEHGNADVHILDAGGQVCVCYRGMALQTLNNAHHDGRQESTQKDELGKLLFVPVWKRLDTKHGRAHGMHALARGTGVLLVYTHENGVEAEQLMQ
RYREEQCELLSVCLSAADRALDTLAARLKHLQSPLRHVCFLDRHQAALDPADKQIDERLREPQTLTGFLALVRTLSELGYDEQPLQWLVLTSSACAV
NQADPINPYAAAIVGLASCLAQEHRYWQVSVCDRGKGTSFTTIANHGETLRVIAERSGYSYVRHIKELELPPPRQSRLRHKGVYVLVGGAGYVGTLL
SRYLVQTYQAKVIWLGRREQDEQIEQNIALIATLGSRPQYLSLKTGSDDELHRAFAQIKAENGLIHGVFNLVMVAQSTPMRSLSEAQFRAGSFNSKV
NSAQALYRILKKEALDFMVFFSSMQSFAQQLNFMSANMSSYVAGCMFQDAFVHQINCELSYPVKTINWGYWSQESSESEKAKWYETYIADKGIYSLN
VGEAMQALEYILAHEPEQVIVTKVSHRVLSDMGFVQDDSENVDLQPVIDSNDDNTQLEDLIVNGILKTFETLDVLQVFDSACNLDVSRLQAHVPDKY
RRLWRELLLILVQAGKMAPQGAGFALRESYTSLASQLRLMDLDREAGLLRERHPTLASKLALLVPCLKDMAEVLTARKPATEVLFEHASLSRVEGIY
RGNPTTDYFNQRVADIVKAAVQQALPSLSVGKKVRILEIGAGTGGTSAYLFQALNTYADQLHYIYTDISKRFLIHADETFCEQYSFIETKLLNIEQD
VAGQGFDPASIDIVVAANVLHATRNISKTLQHVKDLLKNDGLLVLNELADVSPFLTVTFGLLDGWWLYDDEAIRQPGSPGLSAKNWQLVLKNQRFHN
ISLNDEKALTSEQQIIIAQNESKESVPVSSTVLLGTELIASVEASLIAQLSQCIKIEQNRLNTYTTFSDYGVDSILMLAYVDQINGDLGIDLNPSEL
FNYTTVERLTKFIVDYHGKEIISAKESVERSLLQNQESKFDLNDIDVITKAGNEAPEMARSLLFDNVLQVPSGNVHHHQSDKARPDIAVIGMSGQFP
GATNVQQFWQNLSAGVDSVSELPESRLDRQLYSPQPEAGKCYCQWGGILEEQASFDPLFFEISPKEAAEMNPSQRVFMQEAYRALEDAGLVNAGLSQ
LHCGIYVGCEPNGYTKQGFTGASEAIVASRISYFLNLQGPALVINSGCSSSLVAIHLACESLRHGENDIALAGGVFTNLNNELLVCLSSIEMLSFSG
KCHAFDEAANGTVLCEGVSVLVLKRLQDALDDGDVIYGVIKGSGMNQDGRSNGITAPNGLAQENLIKRIYQDYNIDASKISYIEAHGTGTKLGDPVE
ANAXKNVFSSRNTGLDFRCGIGSVKSNIGHAAAAAGVIGLVKVLLCMQHRQLPGLVHFNKLNPLIDFNDTGLYIVEQLQNWEPASGEPLMAGINSFG
HSGTNAHLVIQEPPVRPVRDSENPFHLICLSAKTEEALKQKISDLSAYLRRKEEPPTLGAISYTLNTGRAHFRKRCAFVVDSLEALQNALMGALDGL
VGNAERLSPQEQAFCKKAFDLIIKDLHDVENLESRADQTIYKDNLLALTGLYLKGFDLDWQRVYKGQSYHRVSLPTYPFSRDTYWIDEPKQISFTEP
GTDKPGNTSRNHQQDQKDEDVLGGQFYIPQWKRLDQTQSYASLSDHGKLTIASDPVVIFYPPTMTSLKDDIQTCYEQLPASQVASFELGCTPSSEVD
VRDPRAFERIFVSPRQALEELLEQGRTLTIYFILPEPDQNPDSGCNLECLEESLSLGLISLFRLIQALLAHHGYASELTLKIVTQQTQVCKASERIV
PYGADVVGFVQSLAKEYELWQISCVDINRQTWWTPAQSKNFAEALIALKPPADGRALCWRDGYVYHRVLTPANLVGVQGSAFKTRGVYLIIGGAGGL
GLVLSQYLLEHVQAKLVWVGRRSLEALAPDALRLFNTPGAHVLYEQADLCDPEALQRAQLQQAKGRFGYIDGVIHSALILEDRAIADMDEKTLRKVLA
PKVQGNVVLQQVLADEPLDFLLYFSAGQSFMSHAGQSNYAAACTFEDAWALCQDQRQNYPIKIINWGYWGEVGIVATADYRKQMARQGVLSIHPQEG
MEAVERCLASCATQAAYFKVEPRALSRLGLDGTGILHASRNRVPPQLIHVVIDATQAWSISMPYAQRYQQSLKKLDRLGCYLLLRVFQAMGVFSQPH
EHYEKQELASRLAISKGYHHLYQECLNLLDTSGFIAINGSHIEVQDVLPEEHDEHHSWLQPYPELAAHEKLLRICLQAFPDIIRGSVAATDIIFPDA
SVELVEGIYQGNPATDALNTKAKDAVVAYVKARLADCQEQIIFLEVGAGTGGMSHEVLKGLAPYGQYIEYYYTDVSEHFIQYARRRYGQDYPFVRFQ
RLDIEQEHSALSEGLLSACDVVLAANVLHATQSVRNTIQHVKALLKRHGLLILNEATTLESFATLTFGLLEGWWRYQDAGDRLPGGPLLSASMWHRL
LEEEGFDQVVSLGEPGEGDMLGQHLIVGRSDGWSLHQSARDDETRSAPVSPVAASDVIQGNRNDKPAELIKHVEAQIVQTLSAALDIAAARVERDVS
FSEYGVDSILSVAIINQLNQRLGIRLKRPDLFNYTTTESMAAHIVATFADHLTAALVEEAEPPTPVTVASPQPIKAAGQQDVLQQNLQTQPQVSQGV
WQTKPEPIAVIGMSACVPGADHVGAFWDNLVAGTDSVTRIPHERWQREAVNELPGGPDTSQRKAGFMDNIAEFDPLFFAISPQEAQWMDPTQRLFLQ

QAWAALEDAGYNPRALAGASCAVMVGCAATGYGVKEGERTGYVSYAMTNSEAILPARISYFLDLKGPSIPINTACSSSLVAIHLACEQLRSGQSDMA
LAGGASALIGSALHIGLNGVGMLSPEGACKAFDKEANGFVLGEAVGVVVLKPLSKAVADGDSIHGLIIGSAINQDGKTNGIMAPNGPSQTALIESVY
RKFEIDPKTISYIETHGTGTQLGDPIEVNALAAAYENFTTSQTQYCAIGSVKTNIGHTTTAAGVCGLIKTLLCLKHQKLVPSLHYKEPNENIEFEKT
PFYVNTQVKDWPGHQNGVRRAAVSSFGYSGTNAYMVVEAFVETSPRAVGQQGPALIVLSARHEVRLKELASNLMAYLKRYGLNPDRPIQLEDMAYTL
QVGRQAMSERLAFVADSVADVVKHLATFVDGEANAQELYRGQVKPFGDKSDLLLEGEAGQAFLNTILRQQDYRKLAALWVAGTDLDWNLLYDGENRA
RRISLPTYPFARETYWLDEGIGSNEKELLRDQTQKELRDRDREQTTELMTFEEIWQEASLNNHSDSIKYKSIVCFLSGSEHQQEMRTAMQAIDPAIH
LTMIAQGHGYTKHSSDRYTVDPADIQSFREAFLDISQTHGEVDALLYLWPTEDWRLLYDPSAVVYMLQALVKSELKPSRCLLVAAFSTALERCYLES
WIGFERSLKLIMPGFQVAGLFQEHAGPDQTISMTHWSQHIVSELKAPKAQTVLIQNHRRYVNQIQKTELAAQGASVASAVKQGGTYLITGGCGGLGF
LVAQHLAEHLGQPAASASGKTVHLILTGRSSPDDAIQSKITQLEALGAKVGYVQADVADAVAMRRGLAQMKNRFGDIHGIFHIAGIAPSESLLNKDI
DAFQNVLGPKIAGTLM**LDD**LCRDEPLDVICYFSSSAAILGDFGACDYAIGNRFQMAYAAYRNHLLQQGMRQGKALVINWPLWQEGGMGFHQDENTKM
YLSSSGQRVLETQEALTVLDHVLNQDKTQHLVLVGQPGRVYGFLGLTEAPMTQMPVTEAPVTQTSTRPELRGLSTEQSVLWDLRNLIVQQLKVPWEQ
LGRDHNLADFGFDSISLLSFANRLSEHYQLDITPALFFGYSTLEQLTQYFLSEHCDFINAFYKEPEKSEIPVKKEAVEERAPIRIVSEETRPKEIDN
AKDEPIAIIGMSGRFPKAYDVEQMWQLLASGECAVEEISPHYFDWRRYFGEPSQEANKTNGKWCANIPGVAEFDPLFFQISPREAPNIDPRQRHLLQ
ESWNALEHAGYGPHQLNSQKIGTFVGAEEGDYYQRVKEGNITSNHSGILAARLAYFLNLKGPTMAINTACSSSLVAAHQACLSLRQGECDTAIAAGV
NLMLTPYGYISMGRAGMLSEDGHCYAFDERANGLVPGEAVVVVVLKRLSRALADGDPIHALILASGVNYDGKTNGITAPNGVSQAELVREVHQRAGV
HPQDIDYIVTHGTGTRLGDPVEVNALNEVFRDSQHLCAITSSKTNFGHTFSASGLLSLVNLVQALRHETIPKSLHFAQPNKFMTWDQGPLYVNTANK
PWLEKPGEKRIGGVSAFGMSGTNAHMLIQDYQLHRKEPRAVPPYALLVLSAKTAEALHLKVKAMVAFLQEGLADTATLADISYTLFEGRHHLQYRLA
IVVQDMTEAMCCWTQAGQPDKKPNIFQGKVPTEFSGQKVTQNYINQLVEQLDRSTGDDDVYRERLLGLADFYCQGYDVPWSRLYGDTRPRRISLPGY
PFAKQQFWLDFSPKDHGEQENVELPIGKMQRQVLNEKNTLETLSYIHKWEQQDKAEVKIEISQVCLLVYFESASQFEKTLASFFLQQGASQVIEIQL
GLETKPVSEHKWRCNVNDVDAFQRCLQAYAQTINSVYFLCVDTLQQGMEESEMQLLRLVKALKSRMQSHAFIDFTLLTLDNHRVDNSATSPNHGGLS
GLGYSVAQSDHRFRVRNIDLSREDLVSAEQQCTLPGMIVSEAPSDRGEVVKLQAGLRYKQRFFKLDWQTLQAAQPEAAGLKTGGVYVLLGGAGTVGA
IITRYLIQKYQANVVWLGRRALTSPAVQEQLHAFRSVDEGCVPLYIQADVTDQASMTQAVKEIKQHHPCIHGAIFSALVFKLDNSIAKTSEEDFNEV
LDTKKKGAQYFYHAFKDENLDFMCYFSSVQSFFFISSRESAGYATGITSADTFVHAIDHDSTFPVGIINWGYWEASLAGTEAEKRLPGHYDLISDHD
GCQFFDGFVPMLRSGLIKQAVCLGVSKAIEDFMGCDATQMITVNKLEADSLIFSLFNENECGQATCEETEIERLLDKDFNNELKQWLPRLLFSQLQQ
LGIFLNRGEQRDCTTWQQQAGVIKKYERWWRECCLGLLETSGYVQCEGERVEVTHHSPADEADQLWQGWESFQNLFLDKPERKPVLDLLESCLRSLT
QILCGDILATDILFPKSSMEKVENIYQRNALSDYCNTFVAKGIEAYIHQRIQAEPQARIRILEVGAGTGGTTALVLPQLRQVQHAIDEYCYTDISKA
FLIHAEKTYGPDYNYLDYKLLNLEQPLAPQGIEPGSYDMVIATNVLHATENIRQTLRHTKTALKRHGLLVINEVVEKSIVGILTFGLLDGWWLYQDE
SLRIPGSPLITLDTWRKILKEEGFARVLLPMRAASKLGQQVMVAESDGLVRREREDLPDEDESAVTDKTSVVQGMTLTPEQRIAQTPSVRELSSETL
REQVAELILTGLAETLGMAQENIDPLVSFSDYGIDSILGVGFIGQLTKSFGITINTAILYDYTNVVRLSDYLVEIYAEDIQNKIQNNRMSPVEKDAV
LYTETQLETTVDRNPEKSQLENSVPNMTSGENKSAKCQEIAVIGMSGQFPRAKDLDAFWQNLIQGRDGVVELPANYLNPALFSPHKQAGTTDCKWGG
ILEQRDCFDPLFFNLTPREAESMNPHQRLILSESWKALEDAGYNPKELDNTRVSVFIGAEPTDYYHETFTGSSDAIVASRLSYYLNLKGPALVVNTG
CSSSALAIHLACESLRRGESSMALTGGIFAAMHEEALVKLAQIDMLSATGRCRTFDQFCDGTVLSEGVGVVVLKRLDDAVRDNDPIYGVIQASGANQ
DGASNGITAPSGLSQEELLTSVYRQYDINPETITYIEAHGTGTSLGDTVEANALKRAFKQFTDKKQFCALGSAKSHIGHTGAAAGVIGLIKVLLSMR
YHQLPKLLHFEELNPSIELEDSAFYINQQQMEWTGQGVPLTAGLNSFGHSGTNVHLVVREYLPLSETKAEMEAVLIPLSAKTLESLRTYAESLSTFL
ENGHSGHGVSLTDLAFTLQLGREAMRERCIFLAGNVPDLVAQLKAFANDKTLNNVGWYGRVGRSKNSHSFNDHESRQMITQWLQQGKLDKIAAAWVQ
GLSIDWAALYRELTARKIHAPTYPFAQEHYWKPEQQSMVQKPVAPETQPETFGTLMLQPCWIDQPLEHS

>***misE***, position 60205-73626, length 4474 aa, function KR-ACP-KS-OMT-ACP-KS-KR-ACP-KS-KR-ACP-KS^{o}-
MAAVDCLVYQSRSENIAERFQDYALRLFTEIQQLLTARAEAKILLQFVIRDLLEQQLFAGLAGLLKTAHLENPNLMGQLLCVAPGDDAEHLVATLKV
NAQTPLDQRVAYRNGKRRIAAWKEVPIANDEDVSIPWKENGTYLITGGLGGLGLLFTKEIAQQIRSGTVMLTGRSPLDESKRQMLQTLQQFCGQSRV
HLEYRQVDVCEPQGVASLIQEINHHVGPLDGVIHAAGVIHDNFIIKKTPQELENVLAPKVKGLVNLDEATKDQPLDFFFFCSSLAGGIGNVGQADYA
CANAFMDCYARYRNQLLQSHQRQGRTLSINWPLWQEGGMRVDAASENMMAQSIGLLAMPTSAGMRALYKAIASQQDQVMVVEGDLAKLRQTFLSLPE
KTAQPEIGDVEAALDTGQLQHKTLEKLKSLFAEMTKLPLERVNEEEPLDSYGIDSVMIAQLNQKLTAELGPVSKTLLYEFHTLEELAHHLVKAYFDR
CVNWTQLNSTLNFSEVSVETRNHKSEKGRYSNEYDMQMPDDTAQQPIAIIGLSGRYPKAKTIAEFWQNLKHGEACISEIPRQRWDWKRYYHADPNEA
VKSGKSYSKWGGFLEDFDQFDPMFFNMTPREAENIDPQERIFLQESWKALEDAGYSPSNLSADLRQRIGVFAGITKQGFNLYRMESGPDRQLACPAT
SFASLVNRVSYVLNLQGPSFPVDTMCSSSFVAIHEACAYIRSGQGDMAVVGGVNLYAHPSSYLELSLGRFVSNGSTCAAFEKDADGFVPGEGVGVVI
LKAYDQALKDRDSIYALIRGSAVNHDGKTNGYKTPNPGPKAAVIRQALQQSGIDARSISYIEAAAQGSDVGDAIEASALTDVFSGRNHVEGAYRIGS
VKPNIGHSESASGMSQLTKVILSLNHQTLPPTLRSRALSPSIDFDQLPFELQQELSEWQPVRVDGIEVPRRAGMMNIGAGGVNAHIIVEEPPPHEEP
VHDGHGPVLFVLSAKNEERLERYVQDWLIYITNGPILDLPSIAYTLQVGRESMACRLAFVVNQSEELATYLTRWLKTRELGDVGFYGDVTQERVKNS
EAVSTLVKAGDASHARELAALWVQGNAIAWDKWHRGTTLGRRSGLPTYPFANRTCWIARDDHDCSEDDRGEEKSDVEPRIDHTLDENKAVEFYTRMT
QDCDEAFQEEYLTFGAFPRRIPGFSMTRVAMNAERYPEELEMIKESQRELRQVLFCREDFDRVHTFLDIGCGHGTDVIQVAALYPHIQTHGFTITES
QAILGNRRIEHMNLGSRARIYHKDSSKEPFPDNCDMILGIEVTFHIRDKDALLQKMADALSDDGKILLMDYIANLRGSIVDPNVEISIPTQSQWIDL
LSQHDLIIDEIIDVSTEVSNYLYDPEHAEHTKHLPKVSQDSYRNYANQAIALEKGWITYCLLKLQRDTQRSEQERREYNAWKIANQTPYAQALEDML
RVGRMPYPLSGGLKMAVQPMDGQPPSTLDESVPDLGQIQTALAGIFVAILGFETRELDFIDQLPLTELGISSVNAVELLEAINTEFDLHLPTSLVFE
CHRLEDLAGHIQVELGRKPIAPAMSNASVAVQAEAESRQKTTVPVEEKAPIANEETAIIGLSCRCAGAKDQEQFWQVVSQGKDCIEDLKEPQWLDYI
AEHGSETFPIRYGVMPERDDFDPLFFHISPREAEAMDATQRVLLEECYRALEDAGYNPSSLKQQAVGTYIGAMGSASLRKGFSHLDMLGSETSILSA
RTAYFLDLTGPALTINTACSSSLVAIEVASQALKSGEINLAIAGGVTLYQHPGTFIAMHNASMLSPTGQCRPFDDDADGIVVGDGVGVVILKRLREA
ERDGDSIYGIIRGCGMNQDGQTASMTVPSFLAQSELESAVYQKANIDIEDLQYIEAHGTATKLGDPIEIHALNKSFQALTSKQKFCAIGSLKANVGH
TTAAAGVLGVIKVLLGMKYQQLPPSIHFNEGNRHIDFDNSPVYVNTTLKPWPKNAKDSRLAAISSFGFSGTNAHLVLEQYMPLPDTDSGVFIPLVLL
LSAESDDALQHYAHNVKCYLEDHPAIQYGHFLYTFQVGRDAFAYRLAIIAENQQELIQKLEQWLTTGGPVTDIHNKPDIGNREEEKASLQSLISGRK
VEPLAQLWEQGASVDWHAFYPDGRCRRLAGLPTYPFAREYFGAPQDEECQYGEAKQAPSEDCEGTIARSSTETLIACPYWKEKSVAGATESFAYSRH
LVLLCGMEHISAAKLQAQIPGVECVHLSSVNRFHRSTHVETIYQDEFTQVFEMIQDVLASKTRGNILIQCVVPQQPNQPCLSAISGMLKTAQLENPK
VSGQLIEVLADETEVSLLSKIHENGRCPEDKQIRYDQGQRVVQAWQELVVSDRACKTPWREGGVYLITGGAGGLGLMVAEEIASQTQRSTLILSGRS
SLSPEEVASFSYGLQALGTRVEYQPADISNPYDVDTLIQRISKAYGSLHGILHCAGIVRDNFVINKTSAEFQAVLAPKVAGVVYVDAASKDLNLDFF
VLFSSIAGAMGNAGQADYATANAFMDAFATYRNDLVRIGQRHGQTLSINWPLWRAGGMQVTAHQEAAMRQSVGLEMLDTARGMQALVKGLASGQSQI
LVAAGHLPTLRGWLFQEPLAPSTSDQPLANVHRVGRQPSLASVDKQRLYEETLVQLKGLLGTVFRLSPNRIKMHDAFERYGIDSVMIVQLNQEMEAV
FPELSKTVFFEYQTSHELVAYLLDEYPQECMVWTGLGAESDIDHGNGESSWPERAWRPKHSLATSVDRDLAQAQNPQIQEAIAIIGVSARFAQAETL
QAYWQNLQAGKDCIGEIPTNRWSLDGFFVDDIDEAIAQRKSYSKWGSFLESFADFDPLFFGIPPRDVLSIDPHERLFLQSSWEALEDAGYTRDTLRE
RYQQQVGVFAGVTKTGFDLYGPELWKRGEQAFPHTSFSSVANRTSYFLNLQGPSMPIDTMCSSALTAVHEACEHLRRRDCRLALAGGVNLYLHPSNY
VQMCALRMLSKDGQCKAFGQGGNGFVPGEGAGVVLLKPLSEAIREGDNIYAVIRASHVNHGGKTNSYTVPNPGAQADLIRQTLDKAGIEARSISYIE
AHGTGTALGDPIEITGLTQAFKTDTDDTGFCAIGSAKSNLGHLEAASGMAGLFKVLLQLKHRQLVPSLHAGSLNPNIAFEKTPFVVQQALSEWKRPI
VESSGEVKEYPRRAGISSFGAGGSNAHLIVEEYEANDEPYAEKAVDESVMIVLSAKNTDQLIQLARNLRRFIAETETKPPRLEDIAYTLQTGREAME
ERLALIVGSLEELEEKLHAFVEGQEPIASLYRGDLKQDLNLLSDFATEEELSNTIDDWMMSKQINKLLSLWVKGLTINWQKFYIYIYHKPRSRRISL
PTYPFAKEFYWFEPQAFNHGISTHSQKPAENQTSSRSIQYLQKQWLLDESEPDESYAKQVVARKIAILSGPQTRLLASLLEEEFPNSQIIEHDNIKE
HLEQSHADWQTHGGCIDLAGCSRDEEAGDALDWMLWLQRLIENGSKTGLTLLGVTRGLEFFQNSEVNLSGASRVGLYRMLQSEYAHIQSRHIDTDPS
LKTSALAEQIAREFVLNDENCEICYRDHKRYRAYLAASSLKAAKISDTDLQFAQEDVLLITGGTGGLGLLCAQHFVEHYGVKRLVLTGREALPSREQ
WADPQSDKVKQKIDVIQSLERQGVKVQVLPMVLSDHDAVRETLGQVTKNFGPIRGLLHCSGAVDSETPAFVRKPIAKVSQVLEPKVSGLQTLLAQCR
PQSLKFAVLFSSVAATIPRLASGLSDYAMANAYMDYVAQATGGKLVSIQWPSWQETGMGAANSRAYHETGLLFHNNQEGLRLLDDILSQQTEPVILP
AVVDSALWAPQQLMKHQSRRESISPVLHHHPQTPMHIEPATPNMASSQLTLNDLVALFSQELGIPTQNLALDKSFADYGLDSIWLAQLASKINQQLV
EPLDPSILYEYSSLNALASWLASEQAPALKSQAHDRDVEQKIDPKNAADSVVPRSSKKQALDIAVVGMSCRFPDANNLDEYWSLLSEGKQAIRPVPA
DRWSRPTPFYAGLIENSSHFDPAFFLIPPEDAKVLDPQALLLEECLNLFYHAGYTLDEMKGRAIGVYLGGRSQHRIDEADLGQARNPIVAAQNYLA
THLSQFFDLNGPSLVIDTACSSALVGMNFAVQSLLSGEIDSAVVGGVSLLADEDAHRIFQQRDLLNRDASFHIFDQRASGITLGEGAGLVFLKTLPQ
AQRDGDQIYAVIKGLVINNDGRTAGPATPNFEALKSVMQRALKQSCQQPQDIVHIEANGSGSEVTDLLELKAMEAVYRSDNKAPCSLGSIKPNMGHP
LCAEGIASFIKLVCMLKYQQQVPFLSGQVPLTHYDLDASPFYFCRQSTKWTEHSPLLVALNCFADGGTNAHVIMEAWTQAESHQQRQPIPPPALNRV
NIRGAVSQHPWW

>***misF***, 73446-97670, 8075 aa, DH-ACP-KS-DH-PS-KR-ACP-KS-KR-ACP-KS-DH-KR-MT-ACP-KS-DH-KR-ACP-KS^{0}-ACP-TE

MDRTLPPSGGAQLFCGWWDQCTCDYGSLDAGRISPATAAHTTTRAKPSEYPRCSQPASLVVRWHILKRNLDNHAVPSYNRINNVKKEFHIHASHPIV
KHHQAYGQALLPGLAYIDLLYQFFREQGYEFAALELRNLSIYRPLLVGEDDGIQLLLQADDMAEGLWSVQLLDGEQRGAPESKAQQLYIKAEMYHHR
EAPVFDDVLDFQALQAVSHTQIPLNDIYAQCRSQGLVHSGMMKAEGTAYTLDDAQIIEVALPDEAASSAAGFMFHPALIDGSGVGSSNLFAALVEGE
QRLFLPLFYERFRASARFERTCLVRVKTASVWRKNDMMYMDIEFFNESGEKIAELKNFVNKLVREAALINPNLKPASQAQSSKLKPYEEASSESQAN
PPRLDKSEALEGMEDLLKEIIAPYVDGSPEQLDPEAGYYELGLDSAILLQVVTALEKRFDAKLMPTLLFEYTTIEELAAYFVNENFATKHVAKEPAT
PLAPVLPGPIQTGHEDDIAIIGISGRYPGANNIEFWQNLVVGKDCISDVPEERWDWRSVAKHKSPSGKSVSKWGGFIQDPDCFDHQFFRISPREAE
VLDPQERLFLQTCWECIEDAGYTPETLVTRQQAPHRRQSVGVFVGVMHKDYVFLGMEAIARGEQFHLSINNAPIANRVSYFCDFHGPSMVIDTVCSS
SLTAIHLALESIRHGESDIALAGGVNLSLHPNKYLTYGLLDMYSSDGRCRAFGRGGDGYVSGEGVGAVVLKPLAQAIQDRDHIYAVVKGSMINHVGK
VTGISVPNPVAQSDMILQCLEKTGIHPRTLSYVEAHGTGTSLGDPIEIQGLCKAFQTHTNDTQFCAIGSVKSNIGHTESAAGISGLSKVALQLHYQT
LVPSLHSEAINPHLDLEASPFFVQRQTQHWERTSDHPRRAGLSSFGATGSNAHVILEEYIPEQSETDLAQPAEGVEQRLPVLIPLSAKNEERLKEAA
RNFYHYLISKPEDVAGHQVARLSDIAYTLQVGRRAMETRAIFLVDHQQQLVESLQAFLDDDHTKGNSWLGKQVKQQRHARRGEYSEAMLQQWMVQGR
LDQLAQAWSEGVPLNWDALYSPESGVAKPSRVSLPTYPFAKARHWVPMEVKAVTPFAALHPLLHRNTSTLGEQRFSSTFSGEEFFLADHRVHGQKML
PGVAYLEMARAAVALATSAGEAQAIRFKNIIWQQPLVIDGTEKTIHIGLYAKANSQVDFTIYTGDEHGSDAPFPDEQNRRVHCRGIAMLTLSDRERE
TYLDLDHLQSQTGQNDFTPEQCYQAFAAMGIDYRDGHQGIAAVHAGADQILAELSLPASVLAHREPFVLHPSMMDSALQAAIGFALSAQVAGAQQPA
SLPYALESLDVRAACTPSMWVWIRYADNNSAGSPVQKFDLDLCDDRGKICVLMRGFSSRIFHAQASPCFAETLLPSQLHPLLHQHVPSENPVHDQFI
AQFTGEEFFLRDHQQVMLGMAYLEMAHRAGQHRASQKILGLKNVVWSQLLFIAGQAREVSVHVQNEVLQHRFRITTQNEDSLQNIHCQGMLVVDQAL
YPERPDQLNIQDIQSRCSASIETTECDQLLQYTHGPSLLTIKQLWHNEHEALALLQLPDGMRSGREDYTLHPSMMNGAILTSVLFSLLNGPQEHLPM
PFALDELWMYQEMPERAYAYVSMSSANNHDIELVDITGKCIVLFKGFTAAIQGAWQSQAPQNNVIYASTQWRESVLSGDEHANTAASAPIVMLAQDN
PELAGLLRQRWPEASIEMLAPIGEDIAKTVQANFLQVFDRVKTCILEQPQRIQPIVVLLPENENDYLHDGLAALFKTARFEYRKIVGKLIRYTIDID
IEELAKSILMEVSHDDQHVDVYYSQTGTRKVKTLNEIYDFEKSHAHLLVEADLLHEGDVVWITGGLGGLGQIFTRHLSHTKGLKLILSGRSPLNAAS
EKVLRELQQSDVEMSYLSCDISDQQAVATLVQSILNKYGKLNGVIHSAGLTRDAYIHNKTREEMIAVLRPKMAGILAIEAACQDIQLNFMVLFSSIA
GVFGSIGQADYAMANAFLDSFAEARNRRVQQGKCFGRTLSINWPLWRNGGMGVDAQNETLMQRNTGMIALETRAGLNAFELALSSEYSHILVAQGET
EKIRSRLLSPRFVEENLSSVEADTTEPLEHQNQSEPSKAEQDQRAWLIRLLLKDLAAFVCRLQKVQPEEIDLEAEFSVYGFDSIGFMSFVNHLNDMY
DLELMPTLFFEHSSLRALADALIEKHQEALLKKYESPKTAAQPSQDRMPIQARAVEMDKRQHPVSATMTPVESETTQGNPTPESLSTTDPIAIIGMS
GKFPGSENLTAFWQHLEANRDLVSEIPEDRWDWRDYYGDHLSESGKTKIKWGGFIEGIDHFDPLFFGISPVEAEIMDPQFRLFLQTVWATIEDAGYR
ASALSGSKTGVYVGVTTTDYKDLMQQARARENSQEYHGMFPFMLANRVSYLFNFRGPSEAIDTACSSSLVAIHRATESIRQGSCEMAVAGGVNIIAN
PQLVLSATQGGMLSEDGRCKTFDQSANGYVRGEGVGAILLKPLARAIEDHDPICGVIRGSAENHGGKATSPTAPNPVAQQELLVSAYAQAGITPETV
GYIEAHGTGTPLGDPIEINGLKSAFAQLYEQIGRPPANEPHCALGSVKTNIGHAEAAAGISGVLKVLLMMKHQKIPGNPHLKEPNPYLQLTGSPFHL
AQDTQDWRALHDEVHNPIPRRAGVSSFGIGGSNAHLVLEEYTAEPAAEPVVCGEASLIILSAKNRACLVEAAGNLYRFLIDNRESSRGELLRLADIA
YTLQVGREEMAHRLGFVASSLTELEERLQTFINVQGENQDLYEGHTKGNRSTLEAFESDDDLQEMLTKWVEGRRFEQILKLWVKGLAVNWESLYRGI
QPRRISLPSYPFARERYWIPASAESLPVPAASPVSPIQVDNAAGVLMCHPVWEEKAAPVQVGSFDPASLHDCQHLILFCHMDVPLGLQDRFGTSSVC
RVFQAQSDDVALGYSYLCEQVLTCLKEILENKPRDRVFVQIVIPGPETSEAEGASGLLHGLSALLKTAHLENPKVLGQLIEVGPRHTGDALIRCVQE
NCQSPDDVQVRYRHGRREVMAWEELPVSEPNAAPMPWKERGVYLITGGAGGLGFIVAREIARRVKAPILILAGRSHLSQAKRDELQQEAQRLGVQIE
YRQADVSQPQEVNDLIRSILRDFTKLNGILHAAGLIRDSFMIRKTVDQLHQVLAPKVFGTVYLDEASKDVDLDFFLLFSSGAAALGNAGQADYAAAN
AFMDAYAEYRHALFASKQRFGRSLSINWPLWQEGGMQVDEPTLTALEQYAGIRALQTGSGLDALYRALSVTPNQIMVLEGDLAKIRQQLFSASTEPY
RNESSRVEVDEERLLARTLHQCKTLFGGIIKLAADRVEAEEALEHYGIDSVIIAQMNLRLEEVFGEVSKTLLYEYQTLGELASYLATEYRDACMTWT
GLTAIPTSDMTTVPDDAIPAKPQITSEPRRSSGVDVHALTPKAPAPSQDEPIAIIGLAGRYAQADTPRAYWENLQSGTDCVTEIPPDRWSLSGFYHE
DLEEAIAQGKSYSKWGSFLNGFADFDPLFFSMSPREAIGIDPQERLFLQSSWEAMEDAGYTRETLRDQFQQRVGVFVGITKTGFSLYGPELWQRGDI
VYPTTSFSSVANRISYLFNLRGPSLPIDTMCSSSLTAIHEACEHIRRGECELAIAGGVNLYLHPSTYVQMCAAHLLSQDGKCKSFGEGGNGFVPGEG
VGTIVLKPLTQAIDDHDHIYAVIRASSVNHGGKTNGYTVPNPNAQAALIEDALDKAGVDARTLTYIEAHGTGTELGDPIEIRGLTQAFQKHTIDTAF
CAVGSAKSNLGHLEAAAGIAGLTKIILQLQHAKIAPSLHASTLNPNINFAKTPFVVPQALSEWKRPIVDMHGEVKEYPRRTGISSFGAGGSNAHLIV
EEYVKGTSVGETRASHHHDVIRNEAQVILLSAKNEDRLHCVAKDLHDYLNAEQGANLSLRDIAYTLQLGREAMDERLGLVVYSLDELRQKLNGFIAG
ESGIEDVYRGQVQGHKKALALFTEDDELQEAIEKWMQRGKFARLVELWSQGLAVDWNKLYCDREVIPYRVSLPTYPFEQHRYWIEDTLHASPKPVAE
SVGQSAVKLHPLIHENTSNFLAQRFRSCFRGDEFFLADHLVQGRRVLPGVAYLEMVRMAVSLAMDTQIEGYDLRFENVVWARPIIVDGQPLNMHTVL
QVEDSTTISFEIYSQVDTEENDADVLHCQGSVEICVLEKEQELEVSHLLGECQKHHLSAAQCYEVFESAGIQYGPGHRGLDTVYVGDQQVLAKLLLP
PSVASTAEQFILHPSLMDAALQASLGLAVGGVGSEPTPSLPYALETLDLCDAGARPSWAWIRESNQSSGEIQKLDIDLCDDQGRICVRMRGFSSRVL
ERAWGSQEDRLTLLSPVWNVISDPPTEGERDDVASEQVIIAGGSDLQRDILGQVHPGAVWLPADDFEAIDQQVKVLPSQMDVVHIIWVAPDQPMTSV
RQESLIREQVQGVLYLFRLMKVLLAHGYGEKELHWTVITTQAQAVRKKDIVNPTHASVHGFIGSLAKEYPHWKIRLLDMDADESWPVHDMLALPGDQ
NGDAVAYRAKEWFQQALVPVLETDQGLPPCYRSQGVYVVIGGAGGIGEAWSRWMIERYQAQIIWIGRRKKDALIQEKLDALDKLGTPPIYIEADAGD
LMALQNAYQQIKRRYGRIHGVIHSAIVLADKSLANMDERQFLAGLRAKVDVSVRLAQVFAKETLDFVLFFSSTQSVAKMPGQSNYAAGCTFKDAFAS
QLAQEWPCQVKVINWGYWGSVGVVAISSYRERMARAGVRSIEPAEGMAALQSLLQSSLDQLAVVNTLPTASIAQPSRNQALEEWITTYPDGLPVCVD
SLPQYLPNQPLPVVDSGTGVPDQTVTIEDLLPQLLLATLQSLGLIEGRAEQRAAALPAATFYRQWLEESLTFLWAKEYLEYDGQGYTNKLPSVDLAR
VWQEWEREKVLWLQDRNQKAIVSLVEACMRALPDILRGAQLATDVIFPDSSVALVEHIYKGNVVADYYNQILQNSVVAYIQARLAQAPTAQIRILEI
GAGTGGTTSGLLAELRPYQRHIVEYCYTDLSKAFLIHAERQYATEHPYLVTQIFDVEKPLAEQGIQANHYDVVIAANVLHATKNIRNTLRNAKAGLR
QHGLLFLNEISTKSLFTHLTFGLLEGWWLYEDAALRITGSPVLSSAIWADVLAEEGFNSITFPALLAAESGHDFGQQIIIAQSDGIVRQKARRQSSP
VLPSTPKKQSVKPKEITKLSPNTETLREKATVYIKKLVASTLQLPSDQIDSAQALEAYGLDSILVVQLTSTLRKAFHNISNTLFFEVQTIDGVVDYL
LERQPEVFSDLLGVDKQKIDERIVETSTPIESLESTRQANGRNHSLQKLNFRTTQKLNALAASKKHETVHDRDIAIIGLSGRYPQAENVNAFWECLK
AGQNCIREIPQERWDWRRYYDAEKGKEGTLYTKWGGFIDDVDKFDPLFFKLSARDAERMDPQGRLFLECAYSCMEDAGYTPSTLCDSRKVGVFVGVV
NSTYRRQPSFWSMANRVSYLFNFQGPSMSVDTACSSSLTALHLAIESLHSGTSECAIAGGVNLILDPIQYLTLSAMTMLSSSDACKSFGNQADGFVD
GEGVGAVILKPLAKAEQDGDAIYGIIKASMINAGGKTNGYTVPNPIAQSHLIREALARAEVDPRTISYVEAHGTATALGDSIEITGLSKAFGLSEDK
QFCPIGSVKSNIGHAESASGIASLTKVLLQLKHRQLVPSLHSQQLNPSIDFAETPFVVQQTLEPWHRPVLTIDGETKTYPRIAGISSFGAGGANAHV
IIQEYMNEKGDAAADAEGQGWPETSLIVVSARDEDRLRELVNHLCLYARACSLADQGKAQTAKLQAMAYTLQVGREAMEVRLGVVVASFAELSEKLQ
GFLDGQRDVDNLYRGQVRPHREAMAVLAGDDVFAETIDAWIHRHKYDKILELWVKGFVFDWNKLYGEQKPPRMHLPTYPFARERYWAKIIDLSLDGF
QVGAPQQVVEWIHPLVHKNTSNLEEQRFSSTFSGQEFFLANNRVDGEKVLPGVVYLEMARVAAEQARPHPSTSLAKSHIQLRNIGWATPFVCDHSSQ
SVHIRIVPGTYKGSDVSFEEMSYEIYSHIDTEREVVVHSQGVIALWQAREVAGLDLPALQASINQKCLSAQQCYDAFAAMGIEYGQGHQGIEDVYMT
PSQAGRCQVLAKLRLPAVVAETSHDFVLHPSMMDAALQACIGLTVEHAVPESGQMSLPFAIDAIEVIEKCTDTMWAWIRYADPHADPIQPLSSRVTK
VDIDLCDESGNLCVRIKGFSTRVVAIPESKAIGTWMSFPVWKEKSWSEHEPSEYAQRTVMLIEMEWVSDRLIESQIEGTTCVHLTTQANSLAERFQD
ISVQVFERVKALMEAKLRGSILIQLLVPHSGETLLFSALAGILRTAHLENPALLGQVIILEPSETQEGVIDKLQRDSSCPEDALIRYQQERRQTMAW
QELPMSEREQAVQVPWKEGGVYLITGGLGGVGFMFAKHLAQHSRSVSLVLTGRSELDEQRREQLKELEALGARVDYLSVDVSLKEAVERLIDYIESE
RGSLTGILHSAGVIQDNFIIRKAVREFQSVLVPKVDGVIHLDQATQALNLDFFVLFSSLSGINGNAGQVDYACANAFLDAYAHYRNRLVAINKRSGQ
TLSVNWPLWKEGGMQIDPQRAEAMFQHLGLVAMETSRGIQALIHGLSSGQSQLMVVAGDLERLKVAHPSRPSSPPSQMDREHNVIGENELDANALAS
RSIHYFKNLIASILKIPAHRIQNDELLELYGIDSIIIMQLINQLERDFGPLSKTLFFEYPTIEEVNQYFVDSYRQKLHVVLGLDKSKAPETDHSAVV
DPARPKVHRRSRLLAPTLDESPLPVAQTQDVAIIGVAGRYPQSPDLESFWENLSQGKNCITEVPESRWDPSLYVDFDPSNSGQVYSKWGGFLDDVDK
FDALFFNISPREAELMNPNERLFLETVWELVERSGYTPERIQEQHENNVGVYIGAMYQQYQALASEPLIESVLSLSSHSAIANRVSYFFDFQGPSLA
IDTMCSSSLIAIQLACDSLLKGECQMAIAGGVNLTIHPNKYIGLSLSQMISRQNTSKSFGEGDGYLPAEGVGAVLLKPLAKAIEDQDSILAVIKSIS
TNHSGHTYEFRVPNPNAQAQLFEANFRKSGIDPRTITYVESAANGSPLGDPIEVSALRKAFQKFTADQGFCAIGSVKSNIGHAEAASGISQLTKVIL
QLQHQQLVPSINADPLNPNLSFDGSPFYLQQQLQPWARPVFNIDGQEREFPRRATVSSFGAGGSNAHLIIEEYIASAGKNAVSPDIELAADRELMVF
SAKSCDRLQAMLQQMRHYVESNQALSLKKMAYTLQVGRMAMTHRIAMLVHNREELIQGIQDALHSLEQPAGIINTTIPIYMGDLEADHSEVKHLLSG
QSGEAMLKVLFAEENLEKIALYWVKGGHIPWLALHQKGQMSMLSLPTYPFSKRYCWVESKALSQVVTTAEAELLSKSTDDHISVDRNEESVLSEILC
RTLGLTAEEFNQKTPLVQYGVDSVMFVQIFQQIKDKIDNRVNLSQLLECQTMQDMIFYLTAKKEDAQTEPELVNLSGAGLSSSHDTRFPELVHMNSS
TRGRPVFWFHGIAGVTVYEPVAEKSQRPFYGVQPCSWINQTAGPSHIQAMVQGYLDAIRSVQPEGPYDFGGYSLGGMLAYEATRQLQDMGESVSTIV
MVDTLQSNPDRKDKYSRKTDYLVVLNRSFASSAWQSSENTVQDMLIRADELDASLSDDEYLTELIVLAKQRGLVKTETEIRASLEYATTLDEFYQTD
PFTVRPLSDPDAVNCYYFRNKNGAMLGDQAPYYFATPQDREIFVTTDQSADSRQWEKNLTNLHIIDIDSSNHMTMFSEEKPRQTILEFCEILYSEAG
LPNAFLDFFMAKAREIHGNIELDR

> *misG*, position 97718-98779, 354 aa, function trans-AT
MIAYMFPGQGSQVREMGKELFSQFPDLIQKADAILGYSIEDLCLHDPHKELNQTQFTQPALYVVNALSYRSVIQATGKKPDFVAGHSLGEYNALEAS
GGVSFEDGLRLVQQRGALMSQAPIGGMAAVIGVTGEQVEALLKKFGFDAIDVANFNSPEQVVISGLKADIDKASDDFQEAGTRYIILNTSGAFHSRY
MQSSKVQFEVFLKGFSFSELQIPVISNVTARPYIQSDIVENLAKQITHSIQWTESMRYLLAQGVVEFQEMGVGTVLTKLITDIKAKEVPREQKEPSG
NWTAQEQVDRWNATHSIGLKVHVDGYENELETRTEAIVLFGHRAAIYMQGYNGYFALDEVTPA

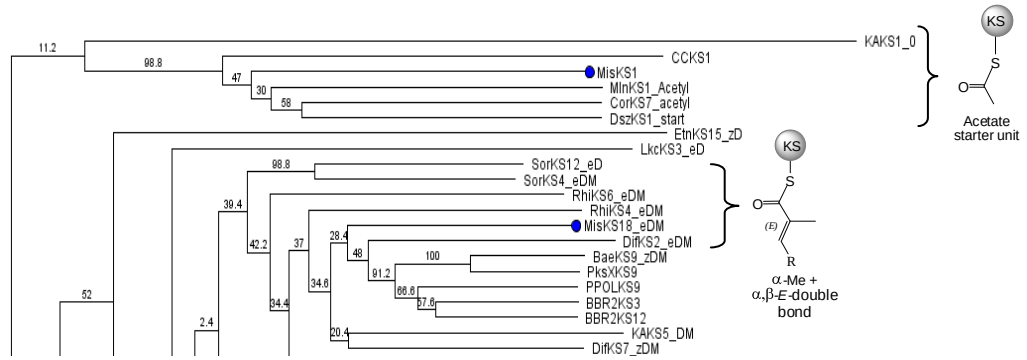
>**ORF5**, position 99817-98849, length 323 aa, function putative nonspecific acid phosphatase precursor
MPMLKQMGMLLLMTCIYLLPILTAAHAQPQAPLPSWQDTANKKAIISFVQSVSTKGSPDFVPREQRVATFDQDGTILIEKPLPVQYEHIFRFFEEPK
PDPKLALVGRKPGEWANIGSGSLLAYQYMTLPEYSENAIKFVKENDHPTCRVPYFDLFYAPMLELIQYLRANKFRVYIVSGSWQIFLRSVAKAKLGL
RKSHFIGTQTGLVDFPANTFPLRIAGTIDNNLLNLNMGKPKNIQIQIGEKPILAFGNSSSDVAMLQYASTNSHKSLSLWLEHDDAEREYVYPSEITG
EKDWLKISMKNDFAKVFDKVNGEQVTCGLLPQ

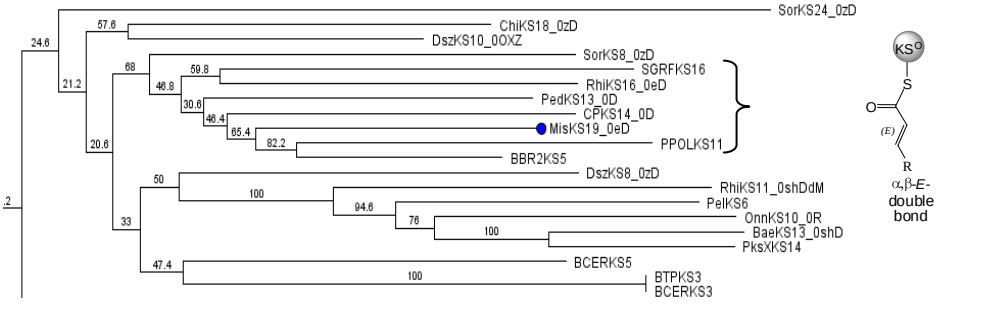
>**ORF6**, position 100789-101223, length 145 aa, function phage shock protein
AMMPCRQRRYMIRKMSAACDCGKWWRNMAQCTTSNVVESVRANTSDFWTVISGYVVVRCRATSQMAGLKSTAVMLMAICWRRAQCTRAMGISAEPVP
ISRTLRDGQFPLDSKRESASREALMPPRKRLTYSISLSAATTSWPSIVA

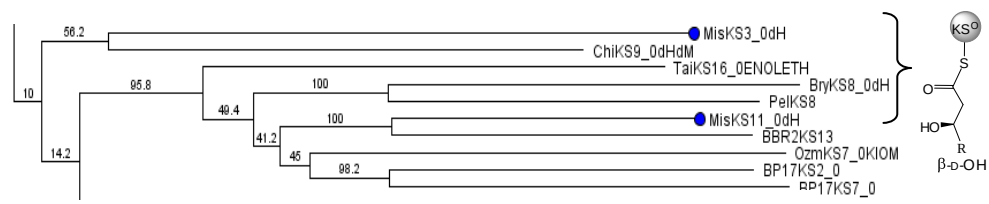
>**ORF7**, position 101270-100563, length 236 aa, function SAM-dependent Methyltransferase
MVDFRQRSTAPEHMDHATIDGHEVVAALNEIEYVNRFLGGINASREALSRLLSSGNWPSLSVLDIGTGSADIPMALVHWARRQQIAINITAVDFNPA
ICEVARQRTTTYPEITVQKSDVFALTDSTTFDVVHCAMFLHHFPQSQAADILRIMYRLCRHGIIVNDLHRHPFAYYSITYLSRLLSLSPMFQHDGPI
SVLRGFQRPDLEALIKMGGLANAEITWRWPFRYVMTAWKLLA

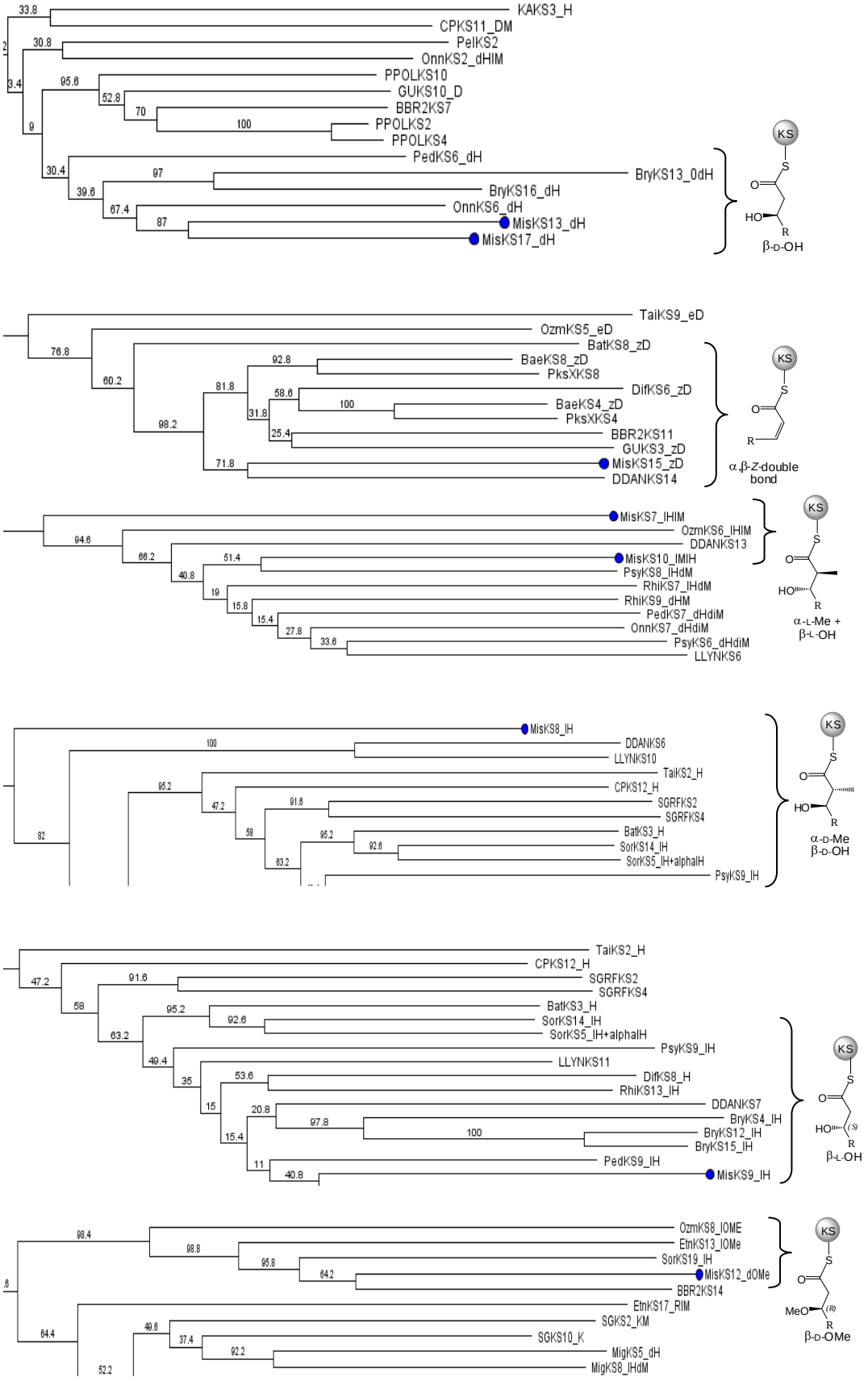
>**ORF8**, position 102567-101275, length 431 aa, function naringenin-chalcone synthase (CHS) or PKS type III
MAISVHDDLRRARYLRLGGCYLRPNMERRLRRQVEGQAQSATSFRFATYCPHVPPATPDKELGVHHEIESTALWSIGTAVPTFQFSQQKALTYALSR
IPPAIARHVKFLYRRSQIDSRHTCSQALYNAIELDTGEAAGDEASTAERMQAYESQAADLAARAAKRALAQAPHFSAADITHLVMVTCTGFVSPGPD
VLLFEQLGLRSDVRRVQVGFMGCHGAMNGLHTAAAICRSEPQAVVVVASLELCTLHFQSTLTEENLVITSLFSDGAAAALLSGKAFTDGQPRCYLDA
FGSRVYPEARDAIAWRVGNQGFDMGLALSNAHDLRPLIPDFVGDWLSSYELHQDDIDLWAIHPGGRAVLDTCEWALGFAPDALASSRWVLRHYGNMS
SPTILFVMDHLLAKTKANRVRRGLSLGFGPGVTLEGILWHEGD

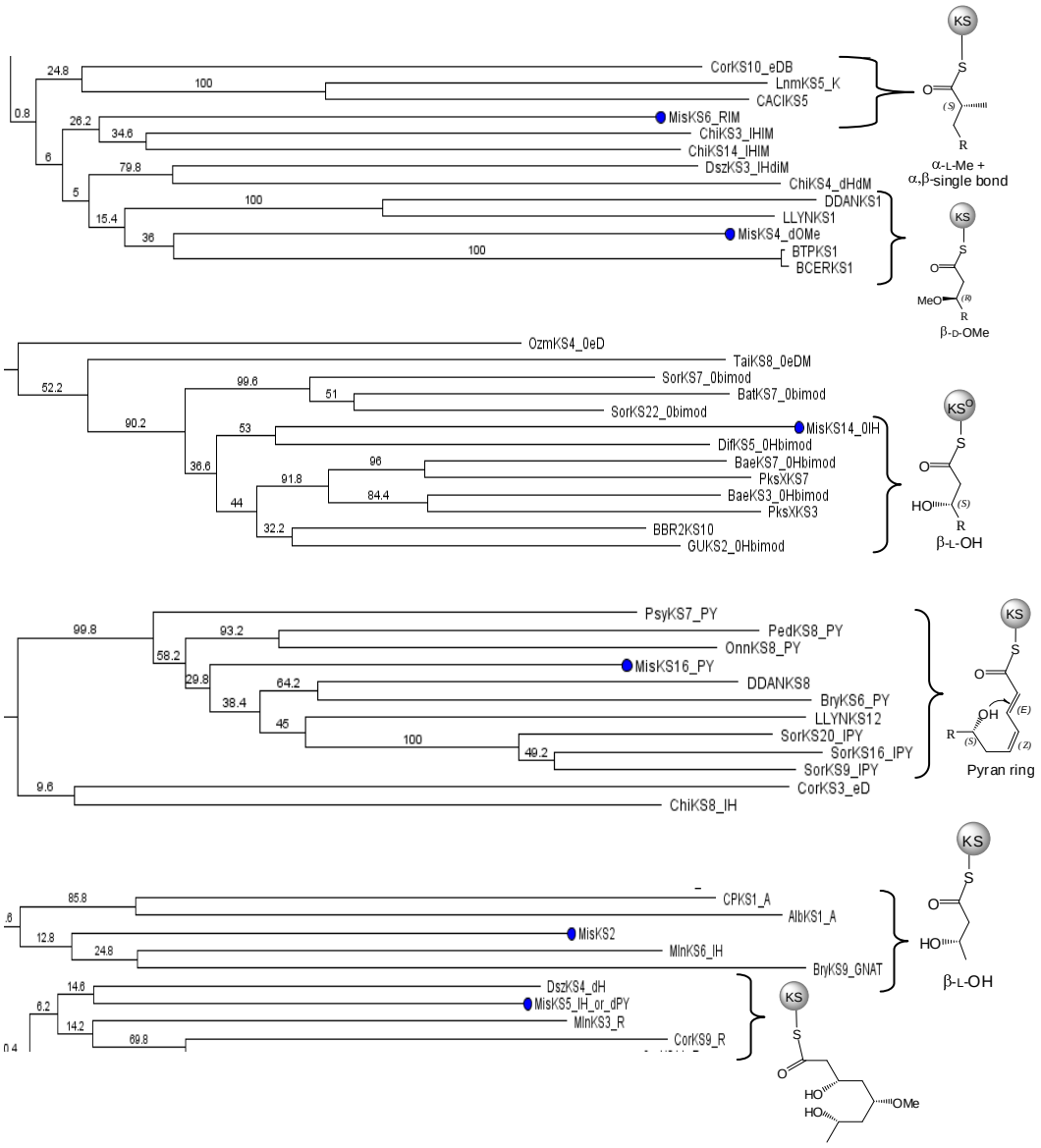
S3.2.3 Phylogenetic analysis of all KS domains (MisKS1-MisKS19) in misakinolide A biosynthetic pathway. Note: Bae, bacillaene; Chi, chivosazol; Dif, difficidin; Dsz, disorazol; Cor, corallopyronin A; Sor, sorangicin; Lkc, lankacidin; Lnm, leinamycin; Mln, macrolactin; Mmp, mupirocin; Onn, onnamide; Ped, pederin; Ta, myxovirescin. Due to the complexity of this phylogenetic tree, it is not present here as the whole tree. Instead only clades containing KSs from misakinolide A pathway are presented.








KAKS3_H
CPKS11_DM
PelKS2
OnnKS2_dHIM
PPOLKS10
GUKS10_D
BBR2KS7
PPOLKS2
PPOLKS4
PedKS6_dH
BryKS13_0dH
BryKS16_dH
OnnKS6_dH
MisKS13_dH
MisKS17_dH
KS
β-D-OH
TaiKS9_eD
OzmKS5_eD
BatKS8_zD
BaeKS8_zD
PksXKS8
DifKS6_zD
BaeKS4_zD
PksXKS4
BBR2KS11
GUKS3_zD
MisKS15_zD
DDANKS14
α,β-Z-double bond
MisKS7_IHIM
OzmKS6_IHIM
DDANKS13
MisKS10_IMIH
PsyKS8_IHdM
RhiKS7_IHdM
RhiKS9_dHM
PedKS7_dHdiM
OnnKS7_dHdiM
PsyKS6_dHdiM
LLYNKS6
α-L-Me + β-L-OH
MisKS8_IH
DDANKS6
LLYNKS10
TaiKS2_H
CPKS12_H
SGRFKS2
SGRFKS4
BatKS3_H
SorKS14_IH
SorKS5_IH+alphaIH
PsyKS9_IH
α-D-Me β-D-OH
LLYNKS11
DifKS8_H
RhiKS13_IH
DDANKS7
BryKS4_IH
BryKS12_IH
BryKS15_IH
PedKS9_IH
MisKS9_IH
β-L-OH
OzmKS8_IOME
EtnKS13_IOMe
SorKS19_IH
MisKS12_dOMe
BBR2KS14
EtnKS17_RIM
SGKS2_KM
SGKS10_K
MigKS5_dH
MigKS8_IHdM
β-D-OMe



Supplementary for Chapter 3.3

S3.3.1 End sequences of the pADU2a fosmid:

>**pSTW-ADU2a**, T3 primer: PP-binding of PCP domain (clone end sequence)
CCPSASRSWLYICTMRTELAPRSKKLVVTLICGSASSWRAVDVNYLIPDSGPPIRLR

>**pTSW-ADU2a**, T7 primer: Condensation domain (clone end sequence)
GAPRQIIHAAAPLHLPITDLQPHAPASQMAEAERLARDELQRPFDLASDLMLRAALLKRNADTHVLLLTLHHIATDGWSFGNLCQELAVLYDAHIQG
KPSPLTTLPIQYADFAYWQRQWEQSDAFAQQLAYWKEQLAGVPERLALPTDRAHPSRITFDGGVVPF

S3.3.2 Protein sequences of the end of individual pADU2a subclones:

>**ADU2a-1a**, T3 primer: Condensation domain
PLPLSLDQERLWFINQFAPENAAYNIHTATRLIGNLDMAVLEKSFNAVIQRHDVLRTSFVAENGQPKQVIAPSLQLSIPIIDLQHLSQAEQDAEAMR
HASEEAEAPFDLSQAPLMRVWLLKMSLTEHVLYFNQHHIITDWWSAKLLFQEMATFYEALLSGQLPVLPEPYQYKDFVLWERERVENQDMARSLAYW
RKHLAEGTFQLDLPIAHRRPVEQTFGGRRQWVTIPPGLAHDLKAFSRRENVTLFITLTAAFKMLLFRYTGQTDITLGTPLANRSQVELERVLGFLIT
MLVLHTDLSGALSFKELVHKVRQVILEGVCASKMYPFAKTLDILQLARGLEPETQSSSFPSPYYDQIILT

>**ADU2a-1a**, T7 primer: mcbC-like oxidoreductase, Acyl-protein synthetase, PP-binding
VTGAKLLSSNRSRGPVSRAARTFTTPADYQIIQLSTTAAIGRDAHGGPNAEIFEHSAFSLFLVGELAAIEPMYGAAARDFCLLEAVYMSQLLMEEAP
GYEIGLCPIGGLNFDPLRPAFEIGDSQVLVHSLVGGGITPEQMRSLPQPEGALQSLEEDIKAYMGQKMPSSMVPKFFVSREAPPVSTNGKVNIQAIP
VPDLTAFEEDVVAPENETEDQIAAHVRDMFSLTQFSVTTKKATSATTPSSRCRCATISGGRGSCKPCGGPSY

>**ADU2a-2.2a**, T3 primer: GMC oxidoreductase, NAD-dependent epimerase/dehydratase
STGPMCIRTIGPRLFCWATGNERLGDFLKIFYLSGLSLCPGAALSPQLHSVAMRLRRPLAPEQIADAILVNLFAESALVAICNSLVLALTCIDQMLR
EPGRQSAHGVRGILITSCREHAAGADVEVRCVVYP

>**ADU2a-2.2a**, T7 primer: NADH:flavin oxidoreductase/NADH oxidase
DDIIFTPLQIGSLEVKNRLFRSSISGRIDNYDGSGTPWRMNFEKRFARGGVGAIISSHVPIAVDARILPDYAMIDRDERIPFWRQVGEAVRQYDHCR
FILQLSHSGRQQDIAGVENLGQRPRSSTNRRDAFHGLRGRRMSQDEIQIVVQQFVdaarraadAGMDAIELHAGNGYLFSQFL
SSAINDRTDPYGGSLENRYRIYREVISGIR

>**ADU2a-3b**, T3 primer: Tryptophan halogenase
PQVKPTFKMGIRFEWGGTTPEHYFNYPFGAHGQLLEAMTYQGHIRDYSLGSMLMSADKMAIFRTGDGAYESRLHYGAISYAYHIDNQRFVSWLQREA
KRAGLHHIDANIQDVVTDDSGQNVVALLTDDGARHEYDFYVDCSGFRSLLLEKALDSEWVPYDQTLFTDRAVTANVPHGGHIKPYTVAETMDHGWCW
NIPMREDDHRGYVFSSAFCDEETAIKEMQAKNPGMSDTKLVTFRSGRHRHFWKGNVVAIGNSYAFVEPLESTGVHMIAEEILAFIGNFPQSTHDDAN
RAALNRYMAAYWDELRWFLGLHFRFNTKLDTPFWKACRA

>**ADU2a-3b**, T7 primer: Tryptophan halogenase
VRPLESLFPRSTDVQMGIRFECRGTTPEHYFNYPFGAHGQLLEAMTYQGHIRDYSLGSMLMSADKMAIFRTGDGAYESRLHYGAISYAYHIDNQRFV
SWLQREAKRAGLHHIDANIQDVVTDDSGQNVVALLTDDGARHEYDFYVDCSGFRSLLLEKALDSEWVPYDQTLFTDRAVTANVPHGGHIKPYTVAET
MDHGWCWNIPMREDDHRGYVFSSAFCDEETAIKEMQAKNPGMSDTKLVTFRSGRHRHFWKGNVVAIGNSYAFVEPLESTGVHMIAEEILAFIGNFPQ
STHDDANRAALNRYMAAYWDELRWFLGLHFRFNTKLDTPFLEG

>**ADU2a-4c**, T3 primer: Galactose oxidase
FQAEIWNPETRTFRLVDSAKEVRNYHAVALLLPNGTVFTGGGGVPPFGKRDAHSSVNDGPGKKIKPNFTAEIYYPPYLFTDDGNGPRLADRPIMTLT
TADADISYGAVITADVSAGSRIDLVTFISLGSMTHSQNMDQRLIKPYFEQVGNELRITLPNNKNLLPPGDYMLHAVDDAGVPSIAEIVQLDQDNVAT
SSDATSVIELARTISLESVNVPGHLVRHATFQAFIAPIDADSTQIERDDAAFRIVNSLAGEGCVSFESRNFPGYYLRGTLPDF

>**ADU2a-4c**, T7 primer: Galactose oxidase
WRPLESTHDLGPSGVELWRESDDDLKFSGQDGRYIDLWTPAEGLVKEAHARPDGAQLSFATDSFCNAAKLLPQGLLLVAGNDFTGRPNGENNTDLFD
YTNNQLQLELADMRYPRWYGSLVRLADDRMVALGGAQSYEHNSFKVWRNGDYDMAGPNYDPALPNGGDDPSYTLGERRHSSIPEIFTYQSGNPAGLW
QEMTGAQNIKTFGRQDNAWWYPRAYLAPNGLIFGVSWNNLWTMDPDGGDATNPGSTAEHGTVAHNVGASGASVMFEGGHLLFAGGGQHSNEEIAPGL
NQATIIDITTPTDPIVTEVNPMTHARNWGIATALPTGEVLVTGGTTDSNRWNDNLAVFQAEIWNPETRTFRLVDSAKEVRNYHAVALLLPNGTVFTG
RGRLAPVWKKG

>**ADU2a-5a**, T3 primer: Peptidyl carrier protein and Thioesterase (PCP-TE)
VLKRQQIGIHDDFFDTGGHSLLAIELLSHIQHVFDQPLPLDFLFQNPTIAQLAAQLGAEEKTSVTHVLPIQPTGTQTPLFCVAGANGYAFYFRDLAT
CLGADQPLYGLEAPGRDGFSPLPDSVEDHASQLVAALRQVQPQGPHLLAGYSAGCAVAFEMASQLEQQGETVRQLIILDTGLIAHPEHFTERSELDM
IWYMVRRVETLKAVSLGLDYQQLVQQPDDQARWQLAAEHLHRPGVLPEYESLSLLQTGLRVGQGLMDNYTIYHPARPIAAPIVLYRAKEIDALLLEE
HRASSHYDLPDWGWQAHTQGSVHVNWVPGNHGSILYEPQVQVLASDLQRRMDFGASW

>**ADU2a-6a**, T3 primer: Condensation and Reductase
GAPRQIIHAAAPLHLPITDLQPHAPASQMAEAERLARDELQRPFDLASDLMLRAALLKRNADTHVLLLTLHHIATDGWSFGNLCQELAVLYDAHIQG
KPSPLTTLPIQYADFAYWQRQWEQSDAFAQQLAYWKEQLAGVPERLALPTDRAHPSRITFDGGVVPFRLDAELTARLRQLSQQADTTLFMTLLAAFQ
ALLARYSGQNDIVVGTAIANRNRRELEPLIGFFVNMLALRADVSGNPTFRELLGQVRQVTQAAYENQDLPFERLVEVLQPARSMEHQPLVQVGFALQ
NMPASHLELSGLHAEP

S3.3.3 Protein sequences of the partial gene cluster of pADU2a fosmid screened from the metagenomic library of *T. swinhoei* chemotype W

>**ORF1**, position 1-3816, lenght 1272 aa, proposed function: C-A-T-TE, specificity code of A = DALHLGLVVK (hpg)
VAAPRVDLEDPPDGAPRQIIHAAAPLHLPITDLQPHAPASQMAEAERLARDELQRPFDLASDLMLRAALLKRNADTHVLLLTLHHIATDGWSFGNLC
QELAVLYDAHIQGKPSPLTTLPIQYADFAYWQRQWEQSDAFAQQLAYWKEQLAGVPERLALPTDRAHPSRITFDGGVVPFRLDAELTARLRQLSQQA
DTTLFMTLLAAFQALLARYSGQNDIVVGTAIANRNRRELEPLIGFFVNMLALRADVSGNPTFRELLGQVRQVTQAAYENQDLPFERLVEVLQPARSM
EHQPLVQVGFALQNMPASHLELSGLHAEPVDLALNTSRLDLEMFLTETEGELDGMCLYKTVLFEASTIERMWGHYQTLLQQVVNDPEQRLSELPLLT
EAEQHQLLVAWNDTHVPFEPGACVYHGVEQQAQRTPHAMAIEYEGQQLTYDAFNQRANQLAHFLRQRGVGPETRVGICLERSVEMIVALLXICKSGG
AYVPLDPAYPRDRLAFMMQDSQMRLLVTQTSLLEQLPIPDGLPQMVCLDRDWPHIATMPSDNPVVQVAAEQLAYIIYTSGSSGAPKAVMVPHRGLRN
LVNWYLREFDVNASDRCIQFVNLSFDAYCLEIWPTLSRGAKLHLVKPELMSVALRLQDWLIKHDITLAHIPTPIAEQLIRLTWPDAMALRLLLTGGE
QLHLSGSASPLPFQFVNLYGPTENSVVSTWYVVSPDSPAIPPIGRALDNQCMYVVDGHLQPVPVGVAGELYVGGIGVGRGYLNRPGLTAEKFVPNPF
AQGRLYRTGDLVRYLPDGNLEFLGRIDNQVKLRGFRIELGEIEAVLSQHEWVKEAVVTLYEADGNRRLAAYVTGDITPEQLAVVKHGIKARLPNYMV
PSQIVVLNALPLNVNGKIDRNALPAPGNWGEDAYEAPRNEIEQRLAEVWSTVLKRQQIGIHDDFFDTGGHSLLAIELLSHIQHVFDQPLPLDFLFQN
PTIAQLAAQLGAEEKTSVTHVLPIQPTGTQTPLFCVAGANGYAFYFRDLATCLGADQPLYGLEAPGRDGFSPLPDSVEDHASQLVAALRQVQPQGPH
LLAGYSAGCAVAFEMASQLEQQGETVRQLIILDTGLIAHPEHFTERSELDMIWYMVRRVETLKAVSLGLDYQQLVQQPDDQARWQLAAEHLHRPGVL
PEYESLSLLQTGLRVGQGLMDNYTIYHPARPIAAPIVLYRAKEIDALLLEEHRASSHYDLPDWGWQAHTQGSVHVSWVPGNHGSILYEPHVQVLASD
LQRRMDFVASL

>**ORF2**, position 3816-4532, length 239 aa, proposed function: SCO1/SenC superfamily
VRVSWLLTAGSPKLAFVYPIHLFVLLVVRIAGGFMATRLARMSCGALLLCLLILTNSGATGSSGELLCAAPGGAQRGKIVDYGWSLMDQDGREFAIA
STQGQVTVLFFGFSRCRGVCPNLLAKMARAEAELGDKAEGVGFVMVSVDGDNDRPKHMKALMRRYSKRFVGLTAAPDIVAPIAKKYGVAFKTSDRPG
DDGLLHSGHIVLLNQQACPVQALPQESSTKALVKAIRSHNKVEKQ

>**ORF3**, position 4532-6949, length 806 aa, proposed function: Galactose oxidase
MTSITVQHISKTLFLIVLSLVALGAYGTSGAPASSPAPTAVVSPSATAAVTLSQAVADRLIDDLVNLEPNAPTHGMWTAPARWPVIPIHAALLPDGR
VLTYGSPEGVLLPEGVDTNDLGPSGVELWRESDDDLKFSGQDGRYIDLWTPAEGLVKEAHARPDGAQLSFATDSFCNAAKLLPQGLLLVAGNDFTGR
PNGENNTDLFDYTNNQLQLELADMRYPRWYGSLVRLADDRMVALGGAQSYEHNSFKVWRNGDYDMAGPNYDPALPNGGDDPSYTLGERRHSSIPEIF
TYQSGNPAGLWQEMTGAQNIKTFGRQDNAWWYPRAYLAPNGLIFGVSWNNLWTMDPDGGDATNPGSTAEHGTVAHNVGASGASVMFEGGHLLFAGGG
QHSNEEIAPGLNQATIIDITTPTDPIVTEVNPMTHARNWGIATALPTGEVLVTGGTTDSNRWNDNLAVFQAEIWNPETRTFRLVDSAKEVRNYHAVA
LLLPNGTVFTGGGGVPPFGKRDAHSSVNDGPGKKIKPNFTAEIYYPPYLFTDDGNGPRLADRPIMTLTTADADISYGAVITADVSAGSRIDLVTFIS
LGSMTHSQNMDQRLIKPYFEQVGNELRITLPNNKNLLPPGDYMLHAVDDAGVPSIAEIVQLDQDNVATSSDATSVIELARTISLESVNVPGHLVRHA
TFQAFIAPIDADSTQIERDDAAFRIVNSLAGEGCVSFESRNFPGYYLRGTLPDFRVDLVQDDGSAAFAAEASFCVREGLAGQGLSFEWAADTSRMMR
HRNTAVWVEQNDGTQLFGQDSSFLIKDAYR

>**ORF4**, position 8508-7102, lenght 469 aa, proposed function: NADH:flavin oxidoreductase
LGKEAVLDDIIFTPLQIGSLEVKNRLFRSSISGRIDNYDGSGTPWRMNFEKRFARGGVGAIISSHVPIAVDARILPDYAMIDRDERIPFWRQVGEAV
RQYDHCRFILQLSHSGRQQDIAGVENLGQRPRSSTNRRDAFHGLRGRRMSQDEIQIVVQQFVDAARRAADAGMDAIELHAGNGYLFSQFLSSAINDR
TDPYGGSLENRYRIYREVISGIRHILPNMPMIVKFSPVDRHNAVTFPLEWSQGNTIEDGVQVAKWIEADGADAIHVTTGSMFPHPWNPAGPLALEQT
AIIYQSMIASGELTFRNYLLFRYRWLRWLPRWAWQRTLGDRVSGEDRWCKLEGLNLEDSRAVKSVVSIPVICTGGFQTARTIRNAIEAGACDAVSMA
RPLLRDPDLPNKMRNASETGHANFEPPSGQECTLCNKCCVTVLEYPFGCYEENRYATRNDMMAHVLSFFAEQPEPETIGEG

>**ORF5**, position 10029-9346, length 228 aa, proposed function: electron transport complex protein RsxA
MALDEGIEAGDVVDVDIADIASDQRHAVYFRCSRKQIVNRMVGIHRTHATPFFRYLFSHGKKATPILLPQGVQRLFQTSGVLLVARAVTSANLLNAF
TNFADRHDTEKQIGILNLFKPVENTNIGGALVQLRDNIRIEQVTHQSLSRLGRSGPGSRVRSSPSSSRERKASLNPFLGGRIRVCSSSSILSRSAST
GLPVSLERAARTPCCSGVSSMCPVSIANLLFPPV

>**ORF6**, position 10182-10787, length 202 aa, proposed function: hypothetical protein
LTPFLIFITAELKGARMIGRLTVVIFALIITIWLISPMVSVAQHTPYAGQQHREIKALSSQEIQSYLSGHGMAFAKAAELNHYPGPMHVLEHAQALQ
LTKRQRTQTQAIFDAMRQEAIELGHQLVQAERHLDHLFSTQAVTEDTLQEATRTIATYRGRLRHVHLRAHLAQRRVLNAKQIQRYDDWRGYPASGFN
HPPHRHHH

>**ORF7**, position 10826-11284, lenght 153 aa, proposed function: cytochrome C class I
VNNRLMGALIYGCAWLMATIPVGCGAAPSGTADATNAEQVAAGQKGFQQHCATCHGVNLEGQPNWKQPLPTGGLPAPPHDDTGHTWHHPDQVLFRIT
KYGGQNGGGAGLKSNMPAFEGTLSDTDIWAVLAFIKSNWSPRARAFQEQANTRGQR

>**ORF8**, position 11172-11816, 215 aa, proposed function: hypothetical protein
LRVRSVIPTFGRCSPLSKVTGRLEPVRFRSKRIRVGNDKRVAAYDLDCHVPCRRTTNPAASPYLSPKAAESEPRHCRFSAPPPSTGPMCIRTIGPRL
FCWATGNERLGDFLKIFYLSGLSLCPGAALSPQLHSVAMRLRRPLAPEQIADAILVNLFAESALVAICNSLVLALTCIDQMLREPGRQSAHGVRGIL
ITSCREHAAGADVEVRCVVYP

>**ORF9**, position 13316-11682, length 545 aa, proposed function: FAD dependent oxidoreductase
MRRYIEEIAADEIIAADICIIGSGPAAFALALPFIKGVPTGTPLKVVMLESAPNVPPIRQALDGAECDVQSAFYKGDQAGVLLYEKRSGSIVSRSVG
YLGTAWPSGEPAFPGRVRAYGGGTHHYHWGGWSWPLEEHDLGGRAFHDGVAWPIPFAELKHWYDQVLTGVVRLNNTEFDNPDYWVRTYPDMGLAVMP
LSEASPLRTRILQFAPFDYAAQYGREIFASDHTDIFYGANAAEFDQVSDGDRKRISRLKVQALNADCTPGRRWYVEANTYVICAGGLESTRLLLLAG
LGDQGSKLGRTFMDHPYMGRMVKFNTTDAVPSGVRKFYMSPDQLLGGPPNHSNFIAGLVPSYKWLNENPKLGDFRILLGSEGFGPSPRTVGINTEPQ
PDHDSRITLTDELGTDPFGQPLMRVDWRTRTIDGDNADTQTLRATIDTLRAVLVDELKYATNFEVIDPAYAETEWTAWRDRSDGSVVGPGLHPMGST
RMSREPGAGVVDQNLRVHNTTNLYVSASSVFPTAGYQNPTNTVCALSARLAQHLIDTSES

>**ORF10**, position 13881-16613, length 911 aa, tryptophan-5-halogenase
MHTPSFSLDTIELMALDSQSFVVRTQNDFVHVEGFSRDKITALLQALAHAATDAEIQEVFPEAGRPEGARLLRRLRAGAPSVAKPAVLYRSGASSKD
EAAPLQPRSIVILGNGALAAALQARLREQGIGRIRTVCVRSFVSLNDASFTKRANELYPQPLSSPSESSLDDFACVSDTVMETANEPMLAEVMQGSE
LVVCALENTYYQAFIEVSKATHALGLPILFVTANRQTCMVGPLSIPGVTVTFREAYEAMFHKMAPSVNPEGFLAYVHAQSLDLAHGQAVLQFVEQEL
RTLLSPRPDSVLFGHLVELRAGQVSRMAAVLPVGTAEAYDGEAGWAELRHAAVLRLAVTDGACTSHRSDRQVPHEHTAIKSICVVGGGTAGYLSALT
LRRRMPHLKISLLESSRVPIIGVGEATTPSLPALLHGVLDMDIVDFFTQVKPTFKMGIRFEWGGTTPEHYFNYPFGAHGQLLEAMTYQGHIRDYSLG
SMLMSADKMAIFRTGDGAYESRLHYGAISYAYHIDNQRFVSWLQREAKRAGLHHIDANIQDVVTDDSGQNVVALLTDDGARHEYDFYVDCSGFRSLL
LEKALDSEWVPYDQTLFTDRAVTANVPHGGHIKPYTVAETMDHGWCWNIPMREDDHRGYVFSSAFCDEETAIKEMQAKNPGMSDTKLVTFRSGRHRH
FWKGNVVAIGNSYAFVEPLESTGVHMIAEEILAFIGNFPQSTHDDANRAALNRYMAAYWDELRWFLGLHFRFNTKLDTPFWKACRADVDVSGFDDCL
ATYRKCAPLSYRNHYLTFNRLWGDHGRDVLLMGQQVPANFLPPRESKQQWLRRVELARQAVEMAVDQAEAIRLLESNPDVLRQLVTDDGAWIHTVGE
LFSSETALYSDVYAARLGHMKKPAGFGGAQNVILPSTL

>**ORF11**, position 17641-16649, length 331 aa, tryptophanyl-tRNA synthetase
LGIEWHMASAIHPPRLICGSEPRGQLGIGTYAMFVHEQLRLQDQYPGDTFVLIADYAALTHAQPASLVRQTAILGAIYLALGLDVRTSVIYRQSDVP
QLFELFWLVAGLLPPQLAHGSPQAAHLPQNLAEELYPALMAADILGLRASIVNQFTISYERLDYACQIAKEINRRLDQDLFPIPRVVGNFPQPPVTH
TSAQDLFAGPSVFGEASEFAFWLSALDHYGKTHHLVADMAAWFDAIATRLVDMDPLRQLQAACPEGHLLEDRLVRNELAALMVERLKPLHDTYLHLR
SDPHIVEKAMQAGALRVRSELKETVGLLRETLGFSRRGDR

>**ORF12**, position 18311-17634, lenght 226 aa, proposed function: tryptophan-2,3-dioxygenase
VSCEHDEYFFMRVLQMSECCFWGALSAVIAAIEATKFGCLEEAMAHMAQALEFLDLLQPITQVLKTIPPEHFWAFRDATDGSSAIQSRWFHLLHIYL
HGLDERKIAIFADIPEISDLIHYGKPGFVSLRSLLRGFAGQPPPELEAFADMASQLDASLYAYRCVHLGIVHRFLPPDAKQGTGGTSGIPYLRKHLT
RRLWPLESEYPQPDLEELNAIDMFQPLVGVGN

>**ORF13**, position 18424-20013, length 530 aa, IclR family transcriptional regulator
MGFWAKSLMVFNCFYHDSKPSIRDVAQQTGLSKSSVHRLKQAILRRDQHPESWFWETREGQQWLIRLVVATLYTFGLKRGVGAETLSEFLLRLGLQS
QIGCSPSALRHLMQLLEQAIVETTQEWEREGVTQGQIGPIIGAVDETFLQCMMLVFMDLSSGYLLFEETAQDRTYDTWYARVTARLETLGTRVLYLV
SDRAKPLVKLGKTGLGCLSVPDLFHLIREFVKGYGPAICLRLRQARQALKAARERLEQGQNHDPKSAVGLQTQREVEACQAEVERWEGIRHTYRDHL
MRLSLMVHPWQIEEARPQSSQQVKQRLEAEIEAIETFMETTGLPMKKQMLEKVRKQLADVSALVDLWWQEVHDSAQSQVILTPMWHQWIDTVLLPLM
YWQAQVSRTREPRRKAELVKALQASQGAFEAHALTQYLAPELIEGWKQWAMEHVRAFQRASSAVEGRNGYLSQMQHNHRGLPKRRYQVWSALYNFDC
HAADRSTPASRFFRHNFPDLFEAVLAQIDELPRPRERRQSKVLTV

>**ORF14**, position 24857- 20745, length 1371 aa, predicted function: a NRPS module (C-A-T-TE)
MSTTAELLAQLSSLGIKLWDDDGYLGYSAPKGALTPALRAELAAHKADILELLRQSERAAPAPSVGRAPEPRFPLSYGQQALWVLYQQAPESPAFNM
GVSLHIQTPLEVDALRWALQSLAGRHAALRTTFSIQDGLLTQDIQASHDLVLETVDAAALTWDDLMEQVYRASQRPFDLERGPVFRATLFRAAPPSH
ILLLNLHHIAGDAASFGILMADLQALYTQKTRGQAADLPALETTYADYVRWESSLLASSAGARLAEYWTQQLAGELPVLDLPTDRPRPAVRTYQGAS
YTFTLSPALSAQLKTLVQAEQTTLFSLLLAAFQLLLHRYTGQDEVWVGSPSGSGRSQPAFAGIVGYLINPLVFRARFPHEPELTFRQFLQQARQTVF
EALEHQAYPFPLLVQQLQPVRHASYTPLFQTFFVFDSSRLDTASQVEGDLPMTPLGLDQMEGQFDLTLTLTDGERLLGALRYNADLFDTATVARMAG
HFEPLLAGIVTNPDQPLHTLPLLTEAEQHQLLVAWNDTHVPFEPGACVHHWVEQQAQRTPHAMAIEYEGQQLTYDAFNQRANQLAHFLRQRGVGPET
RVGICLERSVEMIVALLGICKSGGAYVPLDPAYPRDRLAFMMQDSQMRLLVTQTSLLEQLPIPDGLPQMVCLDRDWPHIATMPSDNPVVQVAAEQLA
YIIYTSGSSGAPKAVMVPHRGLRNLVSWYHRVFDVNASDHTMQFANLIFDASALDIWPTLCRGATLHIVKPELMSAPQRLQDWLIQLGITIAHTPTV
ITEQLMAMTWPENATLRLLITGGEQLHVPGTASPLPFEVVNNYGPTENSVASTWYVVPPDSPAIPPIGRALDNQSMYVVDGHLQPVPVGVAGELYVG
GVGVSRGYLNQPGLTAEKFVPNPFAQGRLYRTGDLVRYLPDGNLEFLGRIDNQIKLRGFRIELGEIESALRQYPGVWECVVAVREQTAGSSKQLVAY
LVPDAVDPSIKDDQATEIRAHLSHTLPDYMVPTAYVILDALPLTASGKVDMRALPAPERADLQLHIDFAPPRTPTETRLAAMWTQVLGVDRIGIHDH
FFELGGHSLLATRLLAEIETQLGKKLPLSSLFEAPTIAELAPLLIEAPPPQPWSPVVPIQPTGNNPNLFCLPGGAAHVFYFWDLARRLGAEQPFYGL
QPPGYEEDTAPLNRMADFVSLFLENLQQIQPQGPYLLGGHSSGAYIAMALVLALQRQGQAVPAMVILDTIPPGMIPDGGPHVMRNSLADVVRIAKEG
LADRLPIDIETLETLSGDAAWQYVVDGYKQINFLPPQAGVAQLKRMHDMYTGVMQAIVDYRPEDHYEGQLLFFQTHEITIDHPEDMAAGWQGSAPIP
SAFIPFLETIRVC
>**ORF15**, position 25457-30154, length 1566 aa, predicted function: a PKS module (KS-AT-KR-ACP)
VQHRAAMRIVRLEMTENLYHIAIVGLAGRFPGANTIETFWENLRNGRESISFFSGEELEQAGVNPALIHHPDYVKAGAILDHIEQFDAPFFGFYPKE
AAILNPQHRLFLECAWEALERAGYMPEQYRGRIGVYAGAGANNYLSSFHLPRGASGALDPFQVMLGNDRDFLATQVAYKLNLTGPSVAVQTACSTSL
VAVHLACQSLLNHECDMALAGGVNINVPQHQGYLYVEGGIGSHDGHCRAFDAQAQGTVGGNGAGIVVLKRMADVLADGDTVLAVIKGTAINNDGATK
VGFTAPSLEGQAEVIADALAMADVAPETITYIEAHGTGTALGDPIEIGALTYAFQASTGKTGFCKIGAVKTNVGHLNTAAGVTGLIKTTLALQHQML
PPTLHFETPNAQIDFDNSPFAVNTTLTPWETPGYPRRAGVSSFGIGGTNAHVIVEEAPEAPASAPSPRPWQLLVLSARTESALEVMTRHLHDHLRQH
PGLNLADVAYTLQVGRKPFAHSRTLVCQTLDEAVTALEARPPDQVLTTYQEARSRPVVFMFSGQGSQYPNMGRGLYQDEPYFRDIVDQCADALLPHL
GRDLREVIYPSEIAPETAAQQLSQTALTQPALFVIEYALAKLWQHWGIEPHAMIGHSIGEFVAACLAGVFSLQDALFLVATRGRLMQAQASGDMLAV
PLAVEALEPFILDIDGEAHDGSLSLAVINGPVQCVVSGPADAIAHLESRLAEQGVAARRLHTSHAFHSTMMEPALEPFIAAFAGITLNAPRIPYLSN
LTGTWITGEQATDPAYWANQLRQTVNFSAGLEELVKTPDWILLEVGPGTVLRSLALQHPARDAAQTVLASLPHVTDQRSAQQAALQTLGQLWSHRVA
VDWDGYYAAEARRRVLLPTYPFERQRHWIDGAATSARPQADADKRLPLPDWFAIPSWKRSQPVALLPQLDLKAHVFTWLVFTDTCGLGTQLAQALQQ
RGQTVITVMAGEAFSPQDEGGYTLNPTQIADYRALFRALSSAGQLPSRIVHCWHVTQTGDEQAPPLSEPDFDSFYGLLFLAQALGEQNLSEPVNISV

IANQLQAVTGDESLVAEKATLLGPSRVISQEYPQMACRSIDIVLPQTTSVPQPLTEQLLGELITDSADAIVAYRGPHRWVQSFEPVTLNPPSVPPIR
PGGVYLITSGLSDTGLVMAYYLAQTAQAKLVLLDPAPFPDQATWPEALASPASSDETRRKIEQLQAIEALGADIFVVQASVTDLDAMQAALQQARQQ
FGPLHGVLHIATSRGSGLMQLKTSEQAQEVLASRVTGTRVLERVLQETPLDFLALFGSNVSIAGGLGQVDECAANAFLDAYAHSRSMAVQAPYPVVA
IDWSGWGWDDHVEQSLTEMPQIQNQVRQLRERYGITPEEAGQAFERILASGQPQIIVSTQSLQAWIDQQNTLTSSSFMDQLDPVPPGPNMPLSDEAY
EAPVNDVKVRIAEVWQEVFGVERIGRHDNFFDLGGHSILAIQLVSRTREALQLDELPLSALFETPTIAGLAATAGAGEPDEADIEEIEAVLEEIEGL
SEEELRALLAEMSE

>**ORF16**, position 30163-31380, length 406 aa, predicted function: flavin utilizing monooxygenase (MOX)
MNELSKRLENLSPEQRELLARRLRAKGVNLPGAEAAADALPQARHVPEARRAQGMQFSLLFFSGDGTTETQGKYDFLLECARFADAHGFRAIWTPER
HFQAFGGLFPNPSVLSAALAMVTERLQIRAGSVVVPLHNPIRLAEEWAVVDNLSNGRVAISLATGWNRGDFTIMPAHFDDRRAIAFDNLATIRRLWR
GECVSFAGVDGEAFEVTILPKPLQPELPVWVTASGSPQTWIDAGAIGANILTAIGRNPDDLVPSITSYREAREKHGHDPQAGIVSVMMHTFLGEDEA
LVKQQVRGPLSGYLKTFIAQDEVLSREETGVEAQQISERDREEMINFAFERYFQQAALFGTPQKCARMVEWMQAVDVDEIACLVDFGADHDAVRASF
QHLSQLREQFSPDHEKQV

>**ORF17**, position 31386-32639, length 418 aa, predicted function: C domain of NRPS
MSELSEKIASLSPEQRAVFDLKLKALKAKKQQAPKAEPPKIIRQTVDGPLPLSLDQERLWFINQFAPENAAYNIHTATRLIGNLDMAVLEKSFNAVI
QRHDVLRTSFVAENGQPKQVIAPSLQLSIPIIDLQHLSQAEQDAEAMRHASEEAEAPFDLSQAPLMRVWLLKMSLTEHVLYFNQHHIITDWWSAKLL
FQEMATFYEALLSGQLPVLPEPYQYKDFVLWERERVENQDMARSLAYWRKHLAEGTFQLDLPIAHRRPVEQTFGGRRQWVTIPPGLAHDLKAFSRRE
NVTLFITLTAAFKMLLFRYTGQTDITLGTPLANRSQVELERVLGFLITMLVLHTDLSGALSFKELVHKVRQVILEAYAHQDVPFAKILEIVQLARDW
SRNPIFQFSFIFIDEFYYRTDSLNRFRIDI

>**ORF18**, position 32700-33794, length 365 aa, predicted function: transposase
MSDNLRRYRAIHEALRQWYPGDPSGNRARHLTPLAAMISGIVASQSSQVPRIASKVPNGTKPESRVKRFTRWLDNEHVTETVYFLPYAQLLLVGLGL
ETLVLAIDGSTVGRGCVALMVSVVYKGRALPLAWLVREGKKGHFPEQMHIELIEQVQLLVPPGVEVVLVGEGEFDGIDLQRTLHGEEADGDWSYVCR
TALNTIASLDGQDFRLDVMGTLIQPGKWVALEPVYVTSAAYGPVTTICWWAKGYDAPIYLVTNMRDAEQACRYYQKRFRIETFFSDQKSRGFNLHKS
HMSNAERLSPLLMASCLAYIWMVYLGDLCERDGWREVIHRRSRCDLSLFQLGLRLLDYFIDEALPIPVQFHVTI

>**ORF19**, position 33891-36881, length 997 aa, predicted function: a NRPS module (A-MT-A-T)
VSRFDMTLIAEDHGEELITYIEYNNELFDAESMDRFASHFQTLLEQMVATPDQPIEQLPMLTEAERQQVLLEWNDTRADYPRNSCLHHLFEEQVQRT
PEATALVFDAPDAASIKLTYRELNERANQLAHYLMQQGVGPETLVGICAERSIEMVVGLYGILKAGGAYVPLLPTYPQERLAVMMADAQVPVLLTQA
HLCEQLPAHEAQTLCLDADWRRVISPMPVENPDSGIQAEHLAYVIYTSGSTGKPKGAMNAHRGVVNRLLWMQDAYGLRADDHILQKTPFSFDVSVWE
FFWPLFTGARLVVAQPEGHQDPSYLVSAIRNHQITTLHFVPPMLSIFLEAPEVETCDTLRRVICSGQALPYELQKRFFSKLSAELHNLYGPTEAAID
VTYWACQPESEQALVPIGRPVANTQIYLLDEHLAPVPIGLPGELHIGGVQVGRGYHHRPELTEAQFIPNPFAEDQERSPTLYKTGDLARWLPDGSID
FIGRLDDQAKLRGFRIELGEIESVLNAHPAVQETLAMVREDEQDQRLVAYLVPDLETAYPIRQQLHLQHDGTVTEDQTSYDLPNGMTLYYQNRSETD
FMYREIFEDHCYLRHGISLPQGSCVFDVGANVGLFSLFVHQACPNAQVYAFEPIPDIFDLLRINTALYGVNANLFAGGLASTNRQDRFTYYPHVSIL
SGRFADAAAEQEVVRTFLLNEDTAHIPEAMLQDMLEERLTSQDITCEFKTLSAVIQENQIDTIDLLKVDVEKSELDILQGIAPHDWPKIKQMVLEVH
DDAAGSFQKVTQLLDTHGFEYVIEQEEMLKDTNLYSIYAKQPASAQPASAQSAATPAVAAWSSVTRLTDDLRTFLQQKLPSYMIPSAFMMLETLPAT
PNGKVDRNALPDPDASFMRQRYEPPGTPTEAQLAVIWQTVLGIEQVGRQDDFFTLGGHSLLAPQVILQVGHAFEIELPLRQLLSNPTLSELAQRIDS
TRETLAGLQDMTSIGDDEADELEELVL

>**ORF20**, position 36881-40948, length 1356 aa, predicted function: a NRPS module (C-A-OX-T)
MTPIETFLAELNQRDIKLWAEEENLRIKAPKGALTQELRQQLADRKAEILSFLRPVTRLPVVQPAPDERYEPFHLTDVQYAYWLGRQSAFELGNIST
HVYFELENTDLDLPRLTQAWQRVIERHEMLRAIVLPSGQQQILDIEDLPPYQIDLSGLTAQDEAAVQEGLMSVREEMSHQVLEADQWPLFDIRATRY
AGTTRLHISLDALTVDATSIGMILREWGAWYQHPSTTLKPLDLSYRDYLLTEKALEETDLYRRSERYWRNRLDTLPPAPALPFAQDPATLAEPRFTR
RQAALSPPLWQALQDRAMKENLSPANVILAAFAEILTAWCKTPHFTLNQTLFNRLPLHPQVNDIVGDFTSLMLLEIDNREPGSMLARAATIQEQLWQ
DLEHRYYSGLRVMRDLASQQPSGSALMPIVFTNVLGMKEDANLGGDDSPDAALGKEVYTISQTSQVYLDHAITEWNGHLAYTWDAVEELFPAGMLDE
MFGAYGAFLKRLAADDTAWTETHRQLVPLAQLQQREAINATQAPISSELLHTLFIAQVEAHSSDPAVIATERTLTYGELYQRAAQVGNWLRERGARP
NTLVAVVMEKGWEQVVGVLGILMSGAAYLPVDPALPTERQHYLLEQAQSTLALTQSHLEQSLTWPADPALECLSVDTAPLNVSPSDAPIAFESVQTP
TDLAYVIYTSGSTGLPKGVVIDHRGAVNTVLDINQMFNVTAQDRVLALSALNFDLSVYDIFGLLAVGGTIVMPAEEGRRDPGHWALLMEQHGVTLWD
TVPALMQMLVEYQAGWPLNAPLRLVMMSGDWIPLTLPDRIRTLWPEAQLMSLGGATEASIWSIYYPIGDIDPAWKSVPYGKPMTNQTFHVLNPQLEP
CPTWVPGDLYIGGIGLALGYWRDEEKTNASFITHPHTGERLYKTGDLGCYLPDGNIEFLGREDFQVKVRGHRIELGEIEVNLVRHPAVKEAVVTAVG
KPLEDKQLVAYIVADATENGALSANAQIDQAVYGDANGTLSEPGERAAFKLEQLGLRRLNGNRSVVQFSRPAVDDGQTNAYLRRQSYRQFLSTPMPL
EKLGEWLACLSPHTFAQAVLPKYRYGSAGSLYPVQTYLYIKPDRVAGLAGGPYYYHPADYQLIQLSTTAAIGRDAHGGPNAEIFEHSAFSLFLVGEL
AAIEPMYGAAARDFCLLEAVYMSQLLMEEAPGYEIGLCPIGGLNFDPLRPAFELGDSQVLLHSFVGGAITPEQMRSLPQPEGAPQSLEEDIKAYLGQ
KVPSYMVPNFYVSLEALPLSTNGKVNRQALPVPDLTAFEEDVVAPENETEEQIAAIAREVLSLPQISVTTNFFDLGANSVLMVQMYNQLREALGQ

Supplementary for Chapter 3.4

S3.4.1 Protein sequences of the partial NRPS gene cluster identified from genome sequencing of candidates "*Entotheonella*" spp. cell pellet of *T. swinhoei* chemotype Y

>**ORF1**, position 26292-22648, lenght 1215 aa, predicted function: A_{Orn}-T, specificity code of OrnA: DVGEIGSIDK (100%)
YLAWEDAYQLRVAATSHLQMASFSFDVFSGDVVRALCSGGKLVLCPQDLLLAPKELYALMQREQVDCAEFVPAVLRHLLDFLEATEQSLDFMRLLIV
GSDLWHMGEYQKSLDASGPETRHINSYGTTEATIDSSYFERAFEDVLADASSSGLVPIGRPFPNTQLYVLDAHLQPVPVGIPGELYVGGTGLARGYF
NHGELTAERFVPSPFMSHQPGARLYKTGDRARYLPDGNIEYINRTDTQVKLRGFRIELGEIETRLTQHDDVRESVVIAHEVAEGQQVLIAYLVPMGE
PSPAIAELRQHLSEKLPDYMIPSVFVPLQAFPLTPNGKVDRRALPIPEQIRVETVFEPPYTPTEDVLAAVWAEVLRLDRVSRHDNFFELGGHSLLAT
QVVSRVREAFQIEVPVRTVFEAPTIAK

>**ORF2**, C-A_{Pro}-T-C-A_{Leu}-T, specificity code of ProA: DVQFIAHVAK (90%), Leu-ArgA: DVYFVGAVIK (60%)
LGQALETAQQGEHPLPLPPPMQPVPREGDLPLSFAQQRLWFLDQFEGSSATYNTLEAFNLRGRLDIPALEKALHALIRRHEVLRTTFPSQGGQPLQQ
IHPELSLPFSVLDLRHLAEPEQIAEVNRLLAQDAGQPFDLVHGPLLRVALYVRGDTEYVLLVNMHHIISDVWSMGVWWRELDALYQSAVSEPDSPAS
ASLPELPLQYADFAQWQRTWLTGEVLERQLAYWEQHLADAPVRLELPTDHTRSLVQTFRGDMERFELAPPLRADLRALSRRCGASLFMTLYGAFAVL
MSRYTGQDDLVIGTPIANRHYRDIEPLIGFFLNTLALRTDLSGDPGFVDLLAQVRQVTLDAYAHQDIPFEQLVDELAVERNLSHSPLFQVLFVWQDL
PQPVALGEMQMVPLDLTTVTSKFDLTFFLGEAEPEAGGGLRGLVEYNTDLFERATICRMIGHFQQLLHGILDNPDQAVRQLPLLTEAERLQLLVEWN
DTETLDSSDLCLHHMFEQQVARTPDAIAVVFEQHQLTYRELNRRANRLAHHLLALGVGPDVLVGLYVERSLEMVVGLLAILKAGGAYVPLDPAYPEE
RLSYMLADARVQVLLTQEGLCSGLSVPAETITLAPDAMAERLSQQSRENPSSSAGPQNLLYVIYTSGSTGRPKGVAMRHQPLANLVRWQLAQANGAS
PARTLQFTPLSFDVSCQEIFTTWCAGGTLVLITDEIRQDGDALLAYLNAQRIEQLFLPFIALQHLAESVTYQSPPEALRDVITAGEQLHITPALLTG
LANPIAVCTITMAQLRAMWPVPIALHNHYGPTESHVASAYSLTGAVSDWPATPPIGQPVANTQLYILDRVYQPVPVGVTGELYLGGECLARGYLHRP
DLTAERFILNPFEAGRLYKTGDLARYRADGQIEFLGRADDQVKLRGFRIELGEIEAVLAQHPDVRETAVVVREDEGHPKQLVAYVVPHEAPPSFIRD
LRDHLSNTVPDYMVPAAFVLLDTFPLTPSGKVDRRALPAPQQTGLEGGYEAPLNPIEELLAAIWCEVLGLTQIGRHDNFFDLGGHSLLATQVISRIR
DTFEVALPVRTVFESASIAKLALALETARQTEHPVPVPPPIRAVARKGDVPLSFAQQRLWFLDRMIGPSAAYNFTGAFRVHGLLDVKVLEPALNEIV

RRHEVFRTTFPSVNGEPVQRIAPGTTLPLVVVDLQAPERRAELNHLVSHEMAQPFDLARGPLLRVILYQLDEREQVVLVNMHHIISDGWSMRVWWCE
LDILYRAFVVGDPSPLPELALQYADFSEWQRGWLSGQTLEQQLAYWRQQLADAPTLLELPTDHPRPTVQSFNGAVACFEIEASKVSQLRGLSRSCGV
SLFMTLYGALAVLMSRYSGQEDLVIGSPIANRHYQEIEPLIGFFVNTLPLRVDLSGGPTFTELLVRVRQITLEAHTYQDVPFEQLVVEVEVERNLSH
PPLFQVVLAWQELFPESATLGDLPVTLMEELDLAPAKFDLVLYLDQPQSEDGERLRGQFEYNTDLFERATITRMISHFEQLLDGIIADPEQGVHTLP
LLTQAERHQLLDTRNATQTTYSRTLCAHQLIEAQAQLRPMAQAVLHPPPGPNASLTYGALNQRANQLAHYLRKLGVGPEVIVAVYLDRSLDMIISLL
AILKAGGAYTPIDPGYPPERVKFILDDTQTPVLLTQTNLAEPLVDAQVEVICVDRDWPVVAQESRANPVNVITPDNMAYLIYTSGSTGRPKGVQVTH
GSLLNLIFWLHRLYDMTDADRIPSCSGLSFDVSVYDIWPCLTAGATLCLSVPDVLFSPLQLQDWLITNRITKGFVVTALAERLLSLEWPPDTPLQLM
VTGGEKSHRVPAPHHPFQYYNAYGPAENTVITTCGLVPPNEKTDTIPSIGRPMDNVQVHILDRQLQLVPTGIYGDLCIGGDNLTRGYLRRPGLTAER
FLPNPFSPNPGARMYTTGDLVRYLPNGEIEFAGRMDHQIKIRGFRIELGEIESVLASHAQVQAATVLAWDMGSSQSDKRLVAYLVLDPEQPPTAEAL
RDYLKAHLPEYMVPAVFVPLEAMPLTPNGKVDRRALPSPDDIQPILNEHYVSPQTALEKTIAEVWQEVLSVEKVGIHDNFFDIGGHSLSAIRIQTGL
QTALAEDIPITDLFRYATVASLAEHLSQKQPEPSSWESSQTRSQARRAAMQQRLDRQRGRAHS

>**ORF3**, position 18537-13849, length 1563 aa, predicted function: a PKS module KS-AT-KR-ACP
LIVKEGAPIHETEPIHETEEDGMTHSETTENFDSNSNLNFYDIAVVGLSGRFPGATSMAAFWANLCNGVESITHFSDQELEAAGVNPALLKRPDYVK
AGAAPDDIDLFDAGFFGFYPKEAEILNPQNRLFLECAWEVLEIVGYIPETYSGRIGVYAGAGLNGYLTALQATADLDSFQVMIGNEKDFLATQVAYK
LNLTGPGISVQTACSTSLVAVHLACQSLLNRECDMALAGGVTINAAYNQGYLYVEGGIMSPDGHCRTFDAKAQGTVGGSGVGIVALKRLTDAVADGD
TVLAVIKGTAVNNDGAAKVGFTAPGLDGQANVITDALTMAEVEPETISYIEAHGTGTALGDPIEIAALTHAFRATTAKTGFCKIGSVKSNIGHLDAA
AGVTGLIKTVLALHHQVLPPTLHFEQANPKIDFANSPFVVNTALTGWQTHGQPRRAGVSSFGIGGTNAHAILEEAPVVPVASSPSRPWQLLILSART
STALETMTGNLQHHLQHHSELNLADVAYTLQVGRKRFDHSRLLVCQTLDDTIAVLETRPPDRMLTTSQEMRFRPVVFMFSGQGAQYPDMGRELYQRE
PIFRDTVDKYAAYLEPHLGLDIREMIYPDGAQRGEAAQQLSQTSYTQPALFMIEYALAQLWRSWGIEPQAMIGHSIGEYVAACLAGVLSPEDALILV
AARGRLMQEQPRGDMLAVPLTEEAIQPFLDASLSVAAINGPTQCVVSGPADAMTQLESRLSEQGIEGTRLQTSHAFHSEMMTPMLESFRATVASVTL
NPPQIPYLSNLTGTWMTPGQXTDPTYWAQHVRQTVRFADGLDELVKEPEWILLEVGPGTTLRTLAQQYPGRETGQTVLASLPHARDSRPAMQAMLQT
LGQLWCHGAAVDWNGFYAAEARRRVVLPTYPFERQRYWIDSQPMAQAPQLEQKLPLSEWFSIPSWKRSKPLELLPQAGSPQHWLVFVDACGLGTQLV
QRLQQRGQTVITVTAGDEFGQHDAFTYTLNPSEKGDYQALFRALQTGETQPNTIAYLWGVTPAAEVVEGTPDFYSVLFLAQALGERYLTDPIRMAVI
ANQSQAVTGEEPLVPEKSTLIGPCRVIPQEYPHITCRHIDIVLQKAQNHGTEWLVDQLLTELVTDAPDPVIAYRGQHRWAQTFEPVELGQRDSALRP
QGVYLITGGLSDTGLVLAHYLAQTAQAKLVLTDTMPFPDRVEWPGWLETAVAEDETRHKVEQLQAMEALGAEVIVLQADETELTQMQEAVEHARQRF
GPLNGVIHIAASRGGGLMQLKSPEQAHQILAPRVIGTQILTTVLHDTPLDFLALFGSQVSVTGGFGLIDECAANAFLDAYAHSHGPQCPFPVVAIDW
ASWHWDQHFEQLAAGVPQLQEGIKLLRETYGITPDEAGEAFERILSSGQPQIIVSTQDLQTAIEQQNAVTSTNVSGQLDTVQPAAPLDRDFAAPINE
TEQRVAAVWQDTFGIEPIGRDDNFFDLGGHSLLAIQLISRLRDALHLDDLPLSSLFESPTVAELAAITTARQEGSDATDTIAALLEEIEGLSDEEIE
AALAEEMEDDT

>**ORF4**, position 13821-11965, length 619 aa, predicted function: a NRPS module C-A$_{Ala}$-T, Ala (90%) DIYNNALTYK
MNDLSDKLAALSPKQRALFELRLKKLNEQQKPPSPQITRRNLTGSCPLSLDQERLWFINRIAPGNAAYNIYFAIRLTGSLDAAILKASINEAIKRHE
ALRTTFGEQNGQPVQHIAPTLTVELPIIERLNLSRAEQDEVAQALATEQAQAPFDLTRGPLVRPVLLKLSDTEHVLIVTQHHIITDWWSSQQLHAEI
VSIYDALVAGHSPDLPDVPFQYSDFVMWEREHLESKDLEDSLAYWRKHLAGGSFKLDLPIANRRPGLQTFQGRRQPLSFPPALAQGLRALSQRENST
MFMTLTAAFYALLFRYTNQTDITLGTPLANRSRVELEQVYGFLITMLVLHIRLSGELSFQELLGRVRTALLEAYTHQDVPFTKILDVVQXERDWSRN
PLFQFSFIFLTETEISLESKALQQSAIEYDPMVSRFDMTLIAWDRGDAIAGCIEYNTDLFTPEAMTRFADHFQTLMESIVAAPDQPIVQLPMLTEAE
RRQLLVEWNETQLDYPTDLCMHHLFEAQAARTPDAVAVVCGSERLTYHELNRQAGQLADHLRALGVGPDVLVGLCVHRSLDMVVGLLGILKAGGAYV
ALDPEYPRERQAFMLRDARLTSCSPNRVWLILCPPVDLLTQQSLVDTLPTSGLTVLCLDTPEAATVPEMASGKPSPAVTPDHLAYILYTSGSTGQPK
GVAIEHRSAVAMLSWSQAEFSPDEVAGMLASTSICFDISVYELFLPLSCGGTVILVDNALALHHQVPTDVRVTLVNTVPSAMRELVRIGGVPPSVKV
VNLAGEPLSNTLVQAVYEQEHVERVYNLYGPSEDTTYSTFLVAEKGAEENPTVGRPIANSQVYIVDEHMQPVPILVPGELYLGGAGLARGYLNQPEM
TQDRFIDNPFGQGRLYKTGDLARYLPDGQIEFLGRMDYQVKLRGFRIELGEIETILSQHEAVRETVVVVREDQPGDPRLVAYVVPNGQAADDEDAEK
LSPALRVHLRQKLPDYMVPSAFVALDSLPLTPNGKIDRLALPAPEGIGLETSYVAPDTPTEETLAAIWAEVLGVPRVGRHDDFFELGGHSLLAIQIM
LRTSAAFEVELPLHYLLHHPNIAEFAQRIDATQQALANLQDTEAATEGEREEVLL

>**ORF5**, position 10502-6294, length 1403 aa, predicted function: a NRPS module Cy-A-OX-T, Cys(90%) DLYNMSLIWK
MKPIEALLDELSQLNIKLQAEGENLRLNAPSGVLTPALRQELAERKAEILSFLRPVTTLPQVQPDPAARHQPFPLTDIQQAYWIGRNSAFELGDIST
HVYFELESPNLDVLRLNRAWLKVIERQDMLRAVVLPNGEQQILEQVPAYRIADADLTDQDEAVVAQTLISIRQDMSHQVFAPDQWPLFDIRTSRYDG
RVRLHLSFDSLIVDAWSYGLLLEEWSRFYREPDLVLPPLELSFRDYLIAEKSLEKTDLYERSRQYWLNRMDTIPPAPKLPFAQDPSQIGQPRFIRRE
HQLDRPLWQALKQHAAQVNLLPGNVILAAFAEILSVWCERPHFTLNLTMASRLPLHPQVGDLVGDFTAITLLEIDNSTPGTFMERARRIQDRLWQDL
DHRYFSGLRVLRELMRRQKGGGNALMPIVFTNPLGIEKSDFAQXHEEEGTPADQETPFGENVYTISQTSQVYLDHSLAEWEGSLLFNWDSIADIFPE
GMLDEMFEAYRGFLTRLATDDAAWTEPERQLVPLQQLKQRAAINDTAGSVSEAMLHTLFLSQVEGHADDCAVIAPQRILTYRELYHQAKQVGHWLRQ
EGATPNRLVAVVMEKGWEQIVAVLGVLMSGAAYLPIDPELPTERQHYLLDQGEVDLALTQSRLDRQLTWPADIRRLCVDTAQLNFSGDPEPVQTAND
LAYVLYTSGSTGLPKGVAVPHRGPVNTLLDINRRCEVTAQDRVLALSALNFDLSVYDVFGLLAVGGTIVMPAPEGRRDPAHWADLMVNHGVTLWNSV
PALMQMLVEYQSERPANAPLRTVMMSGDWIPVSLPERIRERWPSVQLMGLGGPTETSIWSNYYPIEQVDPAWKSIPYGKPMTNQTLHVLNGLLEPCP
VWVPGELYIGGRGLALGYWRDEEKTQASFITHPRTGERLYRSGDLGRYLPNGNIEILGREDFQVKIRGYRIELGEIEANLAKHPEVKEALVSAVGDP
KGDKQLVAYVVPAARTDDAVLDQAAYELVAEDGVLTDPLERAAFKLDWPGIRRLEGQRSAVKLSQPSVNEAAYLSRQSYRQFLTDPVPLEKLGAWLA
CLKPQTFAESILPKYRYGSAGGLYPVQLYLYIKPDRVNGLDGGTYYYHPFDHQLIPLSTVTRIGRELHADVNRANFDQSAFSVFLVGELAAIEPMYG
TRAHEFCLLEAGYMSQLLMEEATAYDLGLCPVGGLDFAPLRDDFELGDSHVLLHSLVGGAIAPEQMQTLPQPGAPSRSLEEDIRADLTQKLPTYMVP
GVYVMLEALPLTPNGKVDRQALPMPGTIETFETMVAAENDIEEALASMIQSVLERDQVSVTGNFFDLGASSIQLVQIYNHLREAFDQELSVTDIFRL
PTIRLLARHLSCAEPDAPIAQQQGEQRAQARLAARRPPRPRREGD

>**ORF6**, position 6394-2206, length 544 aa, proposed function: a PKS module (KS-AT-DH-KR-ACP)
VPSRTHRSHSSRASNERRPASPPGGHRVRAVKETNMAPIEPEFFEEVAIIGLSGRFPGAVNIEQFWQNLRDGVEALSWFSDDDVAAHHSPHIRQAPN
FVKAGFILDDADVEHFDAAFFDVSPRQAQWTDPQQRLFLECAWEALEHAGLDVTTYAGLIGVFAGASVSTYYLYHLLTAESTSAAYDLQMWMGNEKD
YLATQTAYKLNLQGPCLSVQTACSTGLVAVHLACQSLLNYECDVALAGGVTIKVPQKSGYFYEEGAIFSPDGHCRSFDAKAQGTVFGSGLGLVVLKR
LSQALEDRDHILAVIKGSAINNDGSQKVGYTAPSLNGQANVISQAITMARVEPETITYIEAHGTGTPIGDPIEIAALTQVFRAHTERKGFCALGSVK
TNVGHLESAAGIAGLIKTVLALQHKWLPSSLHFDTPNPEIDFDNSPFYVNTQSAAWQTNGYPRRAGVSSFGIGGTNAHLVLEEAPELHVPETQSTRP
LHVLTLSAKTETALAERARQYQDYLASQTGAHLPDICFTANTGRAHFRHRAAVVTATESLKQFIRNKQFAEPHIKLISNIKGDVASKGVATPAYWVR
HVRQPVRFAAGMRTLEQQGYHLFVELGPQPTLLGMGAQCLPEGHGRWLPSLRQGQADWLPLLTSLAELYVQGVSVDWTGFDRDVPRSKVALPTYPFQ
RQRYWFTASPRSLQPVASDVSSHPLLGQRLRLADAGKTRFQVQIRPESPAYLADHRVFGRAVLPATAYLEMALAAGREVFQHDRLTVEELFIQQALI
LADGWQAFKGSDGSQSRGRRVCLATVQPQLSALQVDCPEEIEMTAFYQPRDGTGGVEYGPYFQALTKLYRGANQALGRVRLPSDLIADAAAYHLHPV
LLDACLQITSALLTGESDAVYLPSGLERLTLFRSDTYTELWAHARASSNIDTTVVIDFDLFDANGERIAELRGSAFRRVGQQTLMRNPAMDWLYDIT
WQPMPYSPSSVGSVQSGSWLVLAGETGLGKTLAARLEARGQRCVLVSASQADQTGLADLLRGSMGAGQPPCQGVVYLWGVDTDSTPAELSLPELTHR
LCVQVLDLVQAIIQAEMTPQLWLVTQGAQAVGKTDRLQVQQAPLWGLGRTIAREHPELQCVCLDLPADATSEDNAHTLFEALSGTDAENQIAYRQGT
RYVARLVRQRESERSGIPLPIHETGGYLITGGLGGLGLRVAQWLVAEGARQLVLCGRQGAATPAARQAVEALQQIGAKVVILQADVAVQNDVVRMLT
AAQELAPLRGIIHAAGVLDDGILLQQSGARFATVMTPKVAGSWHLHQLTEIIPPRFFLCVFPRPHPCWERRGRGIMLRPTRLWMPSRITAAPWDCPG

References

Adam, W., M. Lazarus, B. Boss, C.R. Saha-Möller, H-U. Humpf, and P. Schreier, 1997. Enzymatic resolution of chiral 2-hydroxy carboxylic acids by enantioselective oxidation with molecular oxygen catalyzed by the glycolate oxidase from spinach (*Spinacia oleracea*). *J. Org. Chem.* 62(22):7841-7843.

Aicher, T.D., R. K.R. Buszek, F.G. Fang, C.J. Forsyth, S.H. Jung, Y. Kishi, M.C. Matelich, P.M. Scola, D.M. Spero, and S.K. Yoon. Total synthesis of Halichodrin B and Norhalichondrin B. *J. Am. Chem. Soc.* 114(8):3162-3164.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Altschul, S.F. *et al*., 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res* 25:3389-3402.

Amann, R.I., W. Ludwig, and K.-H. Schleifer, 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59(1):143-169.

Amsler, C.D., J.B. McClintock, and B.J. Baker, 2001. Secondary metabolites as mediators of trophic interactions among antarctic marine organisms. *Amer. Zool*. 41:17–26.

Andrianasolo E.H., L. Haramaty, R. Rosario-Passapera, K. Bidle, E. White, C. Vetriani, P. Falkowski, and R. Lutz, 2009. Ammonificins A and B, hydroxyethylamine chroman derivatives from a cultured marine hydrothermal vent bacterium, *Thermovibrio ammonificans*. *J. Nat. Prod*. 72:1216–1219.

Andrianasolo, E.H., H. Gross, D. Goeger, M. Musafija-Girt, K. McPhail, R.M. Leal, S.L. Mooberry, W.H. Gerwick. *Org. Lett.* 7:1375 (2005).

Aparicio, J.F., I. Molnar, T. Schwecke, A. König, S.F. Haydock, L.E. Khaw, J. Staunton, and P.F. Leadley, 1996. Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* 169:9-16.

Arillo, A., G. Bavestrello, B. Burlando, M. Sará, 1993. Metabolic integration between symbiotic cyanobacteria and sponges: a possible mechanism. *Mar. Biol.* 117:159-162.

Azam, F. and A.Z. Worden, 2004. Microbes, molecules, and marine ecosystems. *Science* 303:1622-1624.

Baerga-Ortiz, A., B. Popovic, A.P. Siskos, H.M. O`Hare, D. Spiteller, M.G. Williams, N. Campillo, J.B. Spencer, and P.F. Leadley, 2006. Directed mutagenesis

alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. *Chem. Biol.* 13:277-285.

Banik, J.J., and S.F. Brady, 2010. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opinion Microbiol.* 13:603-609.

Basran, J., I. Efimov, N. Chauhan, S.J. Thackray, J.L. Krupa, G. Eaton, G.A. Griffith, C.G. Mowat, S. Handa, and E.L. Raven, 2011. The mechanism of formation of *N*-formylkynurenine by heme dioxygenases. *J. Am. Soc. Chem. Soc.* 133:16251-16257.

Becker, J.E., R.E. Moore, and B.S. Moore, 2004. Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization. *Gene* 21:35-42.

Beer, S. and M. Ilan, 1998. *In situ* measurements of photosynthetic irrandiance responses of two Red Sea sponges growing under light conditions. *Mar. Biol.* 131:613-617.

Bej, A.K., M.H. Mahbubani, R.M. Atlas, 1991. Amplification of nucleic acids by polymerase chain reaction (PCR) and other methods and their applications. *Critical Rev. Biochem. Mol. Biol.* 26(3/4):301-334.

Bellwood, D.R. and T.P. Hughes, 2001. Regional-scale assembly rules and biodiversity of coral reefs. *Science* 292: 1532–1535.

Bergmann W. and R.J. Feeney, 1950. The isolation of a new thymine pentoside from sponges. *J. Am. Chem. Soc.* 72:2809–2810

Bergmann W. and R.J. Feeney, 1951. Contributions to the study of marine products, 32: the nucleosides of sponges. I*. J. Org. Chem.* 16:981–987

Bergmann W., A.N. Swift, 1951. Contributions to the study of marine products, 30: Component acids of lipids of sponges. I*. J Org Chem* 16:1206–1221

Bewley, C.A. and D.J. Faulkner, 1994. Theonegramide, an antifungal glycopeptides from the Philippine litthistid sponge *Theonella swinhoei*. *J. Org. Chem.* 59(17):4849-4852.

Bewley, C.A., N.D. Holland, N.D., and D.J. Faulkner, 1996. Two classes of metabolites from Theonella swinhoei are localized in distinct populations of bacterial symbionts. *Experientia* 52:716-722.

Bewley, C.A. and D.J. Faulkner, 1998. Lithistid sponges: star performers or hosts to the stars. *Angew. Chem. Int. Ed*. 37(16): 2162–2178.

Bouchet, P., 2006. The magnitude of marine biodiversity. In C.M. Duarte, eds. *The exploration of marine biodiversity: scientific and technological challenges*. Fundación BBVA 33-63.

Bevitt, D.J., J. Cortes, S.F. Haydock, and P.F. Leadlay, 1992. 6-deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.* 204:39-49.

Birch, D.E., L. Kolmodin, W.J. Laird, N. McKinney, J. Wong, K.K.Y. Wong, G.A. Zangenberg, and M.A. Zoccoli, 1996. Simplified hot start PCR. *Nature* 381:445-446.

Birnboim, H.C. and J. Doly, 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513-1523.

Bonnefond, L., T. Arai, Y. Sakaguchi, T. Suzuki, R. Ishitani, and O. Nureki, 2011. Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proc. Natl. Acad. Sci. USA* 108:3912-3917.

Bowling, J.J., A.J. Kochanowska, N. Kasanah, and M.T. Hamann, 2007. Nature`s bounty – drug discovery from the sea. *Expert Opin. Drug Discov.* 2(11):1505-1522.

Brady, S.F., and J. Clardy, 2000. Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J. Am. Chem. Soc.* 122:12903-12904.

Brady, S.F., C.J. Chao, J. Handelsman, and J. Clardy, 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* 3(13):1981-1994.

Brady, S.F., and J. Clardy, 2005. Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. *Angew. Chem. Int. Ed. Engl.* 44:7063-7065.

Brady, S.F., 2007. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nature Prot.* 2:1297-1305

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

Briggs, J.C., 1994. Species diversity: land and sea compared. *Syst. Biol.* 43(1):130-135.

Brown, M.J.B., J. Cortes, A.L. Cutter, P.F. Leadley, and J. Staunton, 1995. A mutant generated by expression of an engineered DEBS1 protein from the erythromycin-producing polyketide synthase (PKS) in Streptomyces coelicolor produces the triketide as a lactone, but the major product is the nor-Analogue derived from acetate as starter acid. *J. Chem. Soc., Chem. Commun.* 1517-1518. DOI**:** 10.1039/C39950001517.

Brown, T.A., 2006. Gene cloning & DNA analysis, an introduction - 5th edition. Blackwell Publishing Ltd.

Brück, W.M., S.H. Sennett, S.A. Pomponi, P. Willenz and P.J. McCarthy, 2008. Identification of the bacterial symbiont *Entotheonella* sp. in the mesohyl of the marine sponge *Discodermalide. The ISSM J.* 1-5.

Brune, I., N. Jochmann, K. Brinkrolf, A.T. Huser, R. Gerstmeir, B.J. Eikmanns, J. Kalinowski, A. Puhler and A. Tauch, 2007. The IclR-type transcriptional repressor LtbR regulates the expression of leucine and tryptophan biosynthesis genes in the amino acid producer *Corynebacterium glutamicum*. *J. Bacteriol.* 189:2720–2733.

Bucior, I. and M.M. Burger, 2004. Carbohydrate-carbohydrate interaction as a major force initiating cell-cell recognition. *Glycoconjugate J.* 21(3-4):112-123.

Bull, A.T., A.C. Ward, and M. Goodfellow, 2000. Search and discovery strategies for Biotechnology: the paradigm shift. *Microbiol. Mol. Biol. Rev*. 64:573-606.

Calderone, C.T., W.E. Kowtoniuk, N.L. Kelleher, C.T. Walsh and P.C. Dorrestein, 2006. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* USA 103:8977-8982

Calderone, C.T., D.F. Iwig, P.C. Dorrestein, N.L. Kelleher, and C.T. Walsh, 2007. Incorporation of nonnethyl branches by isoprenoid-like logic: multiple β-alkylation events in the biosynthesis of myxovirescin A1. *Chem. Biol.* 14:835-846.

Calvin, N.M. and P.C. Hanawalt, 1988. High efficiency transformation of bacterial cells by electroporation. *J. Bacteriol.* 170:2796-2801.

Cane, D.E., C.T. Walsh, and C. Khosla, 1998. Mutations harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* 282:63-68.

Cane, D.E. and C.T. Walsh, 1999. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.* 6(12):R319-R325.

Cardellina II, J.H. and B.S. Moore, 2010. Richard E. Moore (1933-2007). Editorial. *J. Nat. Prod.* 73:301-302.

Carmely, S., and Y. Kashman, 1985. Structure of swinholide A, a new macrolide from the marine sponge, *Theonella swinhoei*. *Tetrahedron Lett.* 26(4):511-514.

Carvalho,R., R. Reid, N. Viswanathan, H. Gramajo and B. Julien, 2005. The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. *Gene* 359:91-98.

Caffrey, P., 2003. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* 4(7):654-657.

Caffrey, P., 2005. The stereochemistry of ketoreductase. *Chem. Biol*. 12(10):1060-1062.

Challis, G.L., J. Ravel, and C.A. Townsend, 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* 7:211-224.

Chan, Y.A., A.M. Podevels, B.M. Kevany, and M.G. Thomas, 2008. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* 26:90-114.

Chang, Z., P. Flatt, W.H. Gerwick, V-A. Nguyen, C.L. Willis, and D.H. Sherman, 2002. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296:235-247.

Chang, Z.X., N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Y. Jia, D. H. Sherman and W. H. Gerwick, 2004. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscule*. *J. Nat. Prod.* 67(8):1356–1367.

Chassy, B.M., A. Mercenier, and J. Flickinger, 1988. Transformation of bacteria by electroporation. *TIBTECH* 6:303-309.

Chachaty, E. and P. Saulnier, 2000. Bacterial DNA Extraction for Polymerase Chain Reaction and Pulsed-Field Gel Electrophoresis. The Nucleic Acid Protocols Handbook I:33-36, DOI: 10.1385/1-59259-038-1:33

Chen X-H, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A.W. Strittmatter, G. Gottschalk, R.D. Süssmuth, and R. Borriss, 2006. Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* 188(11): 4024-4036.

Chen,X.H., A. Koumoutsi, R. Scholz, A. Eisenreich, K. Schneider, I. Heinemeyer, B. Morgenstern, B. Voss, W.R. Hess, O. Reva,H. Junge, B. Voigt, P.R. Jungblut, J. Vater, R. Sussmuth,H. Liesegang, A. Strittmatter, G. Gottschalk, and R. Borriss, 2007. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat. Biotechnol.* 25(9):1007-1014.

Chen, X-H., A. Koumoutsi, R. Scholz, and R. Borriss, 2009. More than anticipated – production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. *J. Mol. Microbiol. Biotechnol.* 16:14-24.

Cheng, Y.Q., G.L. Tang, and B. Shen, 2002. Identification and localization of the gene cluster encoding biosynthesis of the antitumor macrolactam leinamycin in Streptomyces atroolivaceus S-140. *J. Bacteriol.* 184:7013-7024.

Cheng, Y.Q., G.L. Tang, and B. Shen, 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. USA* 100:3149-3154.

Chockalingam, P.S., L.A. Jurado, and H.W. Jarrett, 2001. DNA affinity chromatography. *Mol. Biotechnol.* 19:189-199.

Chou, Q., M. Russell, D. Birch, J. Raymond, and W. Bloch, 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* 20:1717-1723.

Chung, L. W., X. Li, H. Sugimoto, Y. Shiro, and K. Morokuma, 2008. Density functional theory study on a missing piece in understanding of heme chemistry: The reaction mechanism for indoleamine 2,3-dioxygenase and tryptophan 2,3-dioxygenase. *J. Am. Chem. Soc.* 130:12299–12309.

Cichewicz, R.H., F.A. Valeriote, P. Crews, 2004. Psymberin, a potent sponge-derived cytotoxin from Psammocinia distantly related to the pederin family. *Org. Lett*. 6:1951-1954.

Clark, J.M., 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucl. Acids Res.* 6(20):9677-9686.

Cragg, G.M. P.G. Grothaus, and D.J. Newman, 2009. Impact of natural products on developing new anti-cancer agents. *Chem. Rev.* 109(7):3012-3043.

Colabroy, K.L. and T.P. Begley, 2005. Tryptophan catabolism: identification and characterization of a new degradative pathway. *J. Bacteriol.* 187(22):7866-7869.

Coleman, J.E., E.D. de Silva, F. Kong, and R.J. Anderson, 1995. Cytotoxic peptides from the marine sponge *Cymbastela* sp. *Tetrahedron* 51(39):10653-10662.

Committee on Biological Diversity in Marine Systems, National Research Council, 1995. Understanding marine biodiversity. *National Academy Press, Washington, D.C.* 114 pages.

Conti E., T. Stachelhaus, M.A. Marahiel, P. Brick (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin sythetase. *EMBO J.* 16:4174-4183.

Cortes, J., S.F. Haydock, G.A. Roberts, D.J. Bevitt, and P.F. Leadley, 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea. Nature* 348:176-178.

Cortes, J. K.E. Wiesmann, G.A. Roberts, M.J. Brown, J. Staunton and P.F. Leadley, 1995. Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* 268(5216):1487-1489.

Cosmina, P., F. Rodriguez, F. de Ferra, M. Perego, G. Venema, D. van Sinderen, 1993. Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol*. 8:821-831.

Costello, M.J., M. Coll, R. Danovaro, P. Halpin, H. Ojaveer, and P. Miloslavich, 2010. A census of marine biodiversity knowledge, resources, and future challenges. *PloS ONE* 5(8): e12110.

Courtois S., C.M. Cappellano, M. Ball, F-X. Francou, P. Normand, G. Helynck, A. Martinez, S.J. Kolvek, J. Hopke, M.S. Osburne, P.R. August, R. Nalin, M. Gue´rineau, P. Jeannin, P. Simonet, and J-L Pernodet, 2003. Recombinant Environmental Libraries Provide Access to Microbial Diversity for Drug Discovery from Natural Products. *App. Env. Microbiol.* 69(1):49-55.

Cragg, G.M., P.G. Grothaus, and D.J. Newman, 2012. Plant Bioactives and Drug Discovery: Principles, Practice, and Perspectives. The 4th edition, edited by Valdir Cechinel-Filho. John Wiley & Sons, Inc.

Davies, C., R.J. Heath, S.W. White, and C.O. Rock, 2000. The 1.8 angstrom crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure* 8:185–195.

Dawlaty, J., X. Zhang, M.A. Fischbach, and J. Clardy, 2010. Dapdiamides, tripeptides antibiotics formed by unconventional amide ligases. *J. Nat. Prod.* 73:441-446.

DeLong, E.F., L.T. Taylor, T.L. Marsh, and C.M. Preston, 1999. Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent *in situ* hybridization. *Appl. Environ. Microbiol*. 65(12):5554-5563.

Demirev, A.V., C-H. Lee, B.P. Jaishy, D-H. Nam, and D.D.Y. Ryu, 2005. Substrate specificity of nonribosomal peptide synthetase modules responsible for the biosynthesis of the oligopeptide moiety of cephabacin in *Lysobacter lactamgenus*. *FEMS Microbiol. Lett.* 255:121-128.

DiGuistini S., N.Y. Liao, D. Platt, G. Robertson, M. Seidel, S.K. Chan, T.R. Docking, I. Birol, R.A. Holt, M. Hirst, E. Mardis, M.A. Marra, R.C. Hamelin, J. Bohlmann, C. Breuil, and S.J.M. Jones, 2009. *De novo* genome sequence of a filamentous fungus using Sanger 454 and illumine sequence data. *Genome Biol.* 10(9):R94.1-12.

Donadio, S., M.J. Staver, J.B. McAlpine, S.J. Swanson, and L. Katz, 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science* 252:675-679.

Dong, C., F. Flecks, S. Unversucht, C. Haupt, K-H. van Pée, J.H. Naismith, 2005. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* 306:2216-2219.

Donoghue, M., S.J. Lemery, W. Yuan, K. He, R. Sridhara, S. Shord, H. Zhao, A. Marathe, L. Kotch, J. Jee, Y. Wang, L. Zhou, W. Adams, V. Jarral, A. Pilaro, R. Lostritto, J. Gootenberg, P. Keegan, and R. Pazdur, 2012. Eribulin Mesylate for the treatment of patients with refractory metastatic breast cancer: use of a “physicans` choice” control arm in a randomized approval trial. *Clin. Cancer Res.* DOI:10.1158/1078-0432.CCR-11-2149.

Doi, M., T. Ishida, M. Kobayashi, I Kitagawa, 1991. Molecular conformation of swinholide A, a potent cytotoxic dimeric macrolide from the Okinawan marine sponge *Theonella swinhoei:* X-ray crystal structure of its diketone derivative. *J. Org. Chem.* 56:3629-3632.

Douvere, F. and D. Laffoley, 2010. Marine world heritage: the time is now. Protecting the `best of the best` in the ocean. *World Heritage* 56:18-25. http://whc.unesco.org/en/marine-programme/

Dower, W.J., J.F. Miller, and C.W. Ragsdale, 1988. High efficiency transformation of E. coli by high voltage electroporation. *Nucl. Acids Res.* 16(13):6127-6145.

Droege M, Hill B (2008) The genome sequencer FLX (TM) system-longer reads, more applications, straight forward bioinformatics and more complete data sets. *J. Biotechnol*. 136:3–10.

Drummond, A.J., B. Ashton, M. Cheung, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, T. Thierer, and A. Wilson, 2011. Geneious v5.5.

Du, L.C. and L. Lou, 2010. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* 27:255-278.

Du, L., C. Sánchez, and B. Shen, 2001. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic Eng.* 3:78-95.

Du, L., C. Sánchez, M. Chen, D.J. Edwards, and B. Shen, 2000. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus*. *FEMS Microbiol. Lett.* 189:171-175.

Duitman, E.H., L.W. Hamoen, M. Rembold, G. Venema, H. Seitz, W. Saenger, F. Bernhard, R. Reinhardt, M. Schmidt, C. Ulrich, T, Stein, F. Leenders, and J. Vater, 1999. The mycosubtilin synthetase of Bacillus subtilis ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci.* 96:13294-13299.

Dubilier, N., C. Bergin, and C. Lott, 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev.* 6:725-740.

Edwards, D.J., B.L. Marquez, L.M. Nogle, K. McPhail, D.E. Goeger, M.A. Roberts, and W.H. Gerwick, 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscule. Chem. Biol.* 11:817–833.

Eerkes-Medrano, D.I. and S.P. Leys, 2006. Ultrastructure and embryonic development of a syconoid calcareous sponge. *Invertebrate Biol.* 125(3):177-194.

Efimov, I., J. Basran, S.J. Thackray, S. Handa, C.G. Mowat, and E.L. Raven, 2011. Structure and reaction mechanism in the heme dioxygenases. *Biochemistry* 50:2717-2724.

El-Sayed, A.K., J. Hothersall, S.M. Cooper, E. Stephens, T.J. Simpson, and C.M. Thomas, 2003. Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* 10:419-430.

Elsaesser, R., and J. Paysan, 2004. Liquid gel amplification of complex plasmid libraries. *Biotechniques* 37(2):200-202.

Engelhardt K, K. F. Degnes, S.B. Zotchev, 2010. Isolation and Characterization of the Biosynthetic Gene Cluster for Thiopeptide Antibiotic TP-1161. *Appl. Environ. Microbiol*. 76(21):7093-7101.

Erol, O., Schaberle,T.F., Schmitz,A., Rachid,S., Gurgui,C., El Omari,M., Lohr,F., Kehraus,S., Piel,J., Muller,R. and Konig,G.M., 2010. Biosynthesis of the myxobacterial antibiotic corallopyronin A. *Chembiochem* 11 (9), 1253-1265.

Esteves, A.I.S., M. Nicolai, M. Humanes, and J. Goncalves, 2011. Sulfated polysaccharides in marine sponges: extraction methods and anti-HIV activity. *Mar. Drugs* 9(1):139-153.

Faulkner, D.J., 2000. Highlights of marine natural products chemistry (1972-1999). *Nat. Prod. Rep.* 17:1-6.

Faulkner, D.J., 1994. Marine Natural Products. *Nat. Prod. Rep.* 11:355-394.

Fawaz, M.V., M.E. Topper, and S.M. Firestine, 2011. The ATP-grasp enzymes. *Bioorg. Chem.* 39(5-6):185-191.

Fenical, W., 1993. Chemical studies of marine bacteria: *Chem. Rev.*1673-1683.

Fenical, W. and P.R. Jensen, 2006. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* 2:666–673.

Fenical, W., P.R. Jensen, M.A. Palladino, K.S. Lam, G.K. Lloyd, and B.C. Potts, 2009. Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Biorg. Med. Chem.* 17(6):2175-2180.

Feling, R.H., G.O. Buchanan, T.J. Mincer, C.A. Kauffman, P.R. Jensen, W. Fenical, 2003 Salinosporamide A: a highly cytotoxic proteasome inhibitor from a noval microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed*. 42(3):355-357.

Feng, Z., J. Qi, T. Tsuge, Y. Oba, T. Kobayashi, Y. Suzuki, Youji Sakagami, and M. Ojika, 2005. Construction of a bacterial artificial chromosome library for a myxobacterium of the Genus Cystobacter and characterization of an antibiotic biosynthetic gene cluster. *Biosci. Biotechnol. Biochem*. 69(7):1372-1380.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evol.* 39:783-791.

Fewer, D.P., J. Osterholm, L. Rouhiainen, J. Jokela, M. Wahlsten, and K. Sivonen, 2011. Nostophycin Biosynthesis Is Directed by a Hybrid Polyketide Synthase-Nonribosomal Peptide Synthetase in the Toxic Cyanobacterium *Nostoc* sp. Strain 152. *Appl. Environ. Microbiol.* 77 (22):8034-8040.

Fieseler, L. U. Hentschel, L. Grozdanov, A. Schirmer, G. Wen, M. Platzer, S. Hrvatin, D. Butzke, K. Zimmermann, and J. Piel, 2007. Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl. Environ. Microbiol.* 73, 2144–2155.

Findlow S.C., C. Winsor, T.J. Simpson, J. Crosby, and M.P. Crump, 2003. Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochem.* 42(28):8423-8433.

Finking, R. and M.A. Marahiel, 2004. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* 58:453-488.

Fisch, K.M., C. Gurgui, N. Heyke, S.A. van der Sar, S.A. Anderson, V.L. Webb, S. Taudien, M. Platzer, B.K. Rubio, S.J. Robinson, P. Crews, and J. Piel. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol*. 5(7):494-501.

Flachshaar, D. and J. Piel, 2008. Symbiosis: topics in chemical biology. Wiley Encyclopedia of Chemical Biology. John Wiley & Sons, Inc. 13 p.

Foulston, L.C. and B.J. Bibb,. 2010. Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proc. Natl. Acad. Sci.* U.S.A. 107(30):13461-13466.

Freeman, M.F., C. Gurgui, M.J. Helf, B.I. Morinaka, A.R. Uria, N.J. Oldham, H-G. Sahl, S. Matsunaga, and J. Piel. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Sience* DOI:10.1126/science.1226121.

Friedrich, A.B., H. Merkert, T. Fendert, J. Hacker, P. Proksch, and U. Hentschel, 1999. Microbial diversity in the marine sponge *Aplysina cavernicola* (formerly *Verongia cavernicola*) analyzed by fluorescence in situ hybridization (FISH). *Mar. Biol.* 134(3):461-470.

Friedrich A.B., I. Fischer, P. Proksch, J. Hacker, and U. Hentschel, 2001. Temporal variation of the microbial community associated with the Mediterranean sponge *Aplysina aerophoba*. *FEMS Microbiol. Ecol.* 38(2-3):105–113.

Fujimori D.G., S. Hrvatin, C.S. Neumann, M. Strieker, M.A. Marahiel, and C.T. Walsh, 2007. Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci.* USA 104(42):16498–16503.

Fusetani, N., S. Matsunaga, H. Matsumoto, and Y. Takebayashi, 1990. Bioactive marine metabolites. 33. Cyclotheonamides, potent thrombin inhibitors, from a marine sponge *Theonella* sp. *J. Am. Chem. Soc*. 112(19):7053-7054.

Fusetani, N., Y. Nakao, and S. Matsunaga, 1991. Nazumamide A, a thrombin-inhibitory tetrapeptide, from a marine sponge, Theonella sp. *Tetrahedron Lett.* 32(48):7073-7074.

Fusetani, N., T. Sugawara, S. Matsunaga, and H. Hirota, 1991. Orbiculamide A: a novel cyclic peptide from a marine sponge *Theonella* sp. *J. Am. Chem. Soc.* 113(20):7811-7812.

Fusetani, N., T. Sugawara, and S. Matsunaga, 1992. Theopederins A-E, potent antitumor metabolites from a marine sponge, *Theonella* sp. *J. Org. Chem.* 57(14):3828-3832.

Fusetani, N. and S. Matsunaga, 1993. Bioactive sponge peptides. *Chem. Rev.* 93:1793-1806.

Fusetani, N., T. Sugawara, S. Matsunaga, and H. Hirota, 1991. Orbiculamide A: a novel cytotoxic cyclic peptide from a marine sponge Theonella sp. *J. Am. Chem. Soc.* 113(20):7811-7812.

Fusetani, N, 1996. Bioactive substances from marine sponges. *J. Toxicol*. –Toxin Review 15(2):157-170.

Fusetani, N., T. Sugawara and S. Matsunaga, J. Org. Chem., 1992, 57, 3828–3832.

Fusetani N., Y. Nakao, S. Matsunaga, 1991. Eight new cytotoxic metabolites closely related to onnamide A from two marine sponges of the genus *Theonella*. *Tetrahedron Lett*. 48(39):8369-8376.

Fusetani, N., 2010. Antifungal peptides in marine invertebrates. *ISJ* 7:53-66.

Gaitatzis, N., B. Silakowski, B. Kunze, G. Nordsiek, H. Blocker, G. Hofle, and 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 (15):13082-13090.

Gamble, W.R., N.A. Durso, R.W. Fuller, C.K. Westergaard, T.R. Johnson, D.L. Sackett, E.H. Hamel, J.H. Cardellina, and M.R. Boyd, 1999. Cytotoxic and tubulin-interactive hemiasterlins from *Auletta* sp. and *Siphonochalina* spp. Sponges. *Bioorg. Med. Chem.* 7(8):1611-1615.

Gatto Jr, G.J., S.M. McLoughlin, N.L. Kelleher, and C.T. Walsh, 2005. Elucidating the substrate specificity and condensation domain activity of FkbP, the FK520 pipecolate-incorporating enzyme. *Biochem.* 44(16):5993-6002.

Gehret, J.J., L. Gu. W.H. Gerwick, P. Wlpf, D.H. Sherman, and J.L. Smith, 2011. Terminal alkene formation by the tioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J. Biol. Chem.* 286(16):14445-14454.

Gehring, A.M., I. Mori, R.D. Perry, and C.T. Walsh, 1998. The nonribosomal peptide synthetase HMWP2 forms a thiozoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis. *Biochem.* 37:11637-11650.

Gerwick, W.H., R.C. Coates, N. Engene, L. Gerwick, R.V. Grindberg, A.C. Jones, and C.M. Sorrels, 2008. Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3(6):277-284.

Gerwick, W.H., P.J. Proteau, D.G. Nagle, E. Hamel, A. Blokhinr and D.L. Slates, 1994. Structure of Curacin A, a Novel Antimitotic, Antiproliferative, and Brine Shrimp Toxic Natural Product from the Marine Cyanobacterium *Lyngbya majuscule. J. Org. Chem.* 59(6):1243-1245.

Giessen, T.W. and M.A. Marahiel, 2012. Ribosome-independent biosynthesis of biologically active peptides: application of synthetic biology to generate structural diversity. *FEBS Lett.* Doi:10.1016/j.febslet.2012.01.017.

Gillespie, D.E, S. F. Brady, A.D. Bettermann, N.P. Cianciotto, M.R. Liles, M.R. Rondon, J. Clardy, R.M. Goodman, and J. Handelsman, 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.*, (68):4301–4306.

Goffredi, S.K., V.J. Orphan, G.W. Rouse, L. Jahnke, T. Embaye, K. Turk, R. Lee and R.C. Vrijenhoek, 2005. Evolutionary innovation: a bone-eating marine symbiosis. *Environ. Microbiol.* 7(9):1369–1378.

Gourmelon, C., J.S. Frenel, and M. Campone, 2012. Clinical evidence for the role of eribulin mesylate in the treatment of breast cancer. *Clinical Investigation* 2(2):207-213.

Grassle, J.F. and N.J. Maciolek, 1992. Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *Amer. Nat.* 139:313-341.

Gray, J.S., 1997. Marine biodiversity: patterns, threats and conservation needs. *Biodiver. Conservation* 6:153-175.

Guenzi, E., G. Galli, I. Grgurina, D.C. Gross, and G. Grandi, 1998. Characterization of the syringomycin synthetase gene cluster. J. Biol. Chem. 273:32857-32863

Guerrero-Ferreira, R.C. and M. K. Nishiguchi, 2007. Biodiversity among luminescent symbionts from squid of the genera *Uroteuthis*, *Loliolus* and *Euprymna* (Mollusca: Cephalopoda). *Cladistics* 23:497–506.

Gu, L., T.W. Geders, B. wang, W.H. Gerwick, K. Hånkansson, J.L. Smith, and D.H. Sherman, 2008. GNAT-like strategy for polyketide chain intitiation. *Science* 318:970-974.

Gu, L., B. Wang, A. Kulkarni, J.J. Gehret, K.R. Lloyd, L. Gerwick, W.H. Gerwick, P. Wipf, K. Hakansson, J.L. Smith, and D.H. Sherman, 2009. Polyketide decarboxylative chain termination preceded by *o*-sulfonation in curacin A biosynthesis. *J. Am. Chem. Soc.* 131, 16033–16035.

Gulder, T.M., M.F. Freeman, and J. Piel, 2011. The catalytic diversity of multimodular polyketide synthases: natural product biosynthesis beyond textbook assembly rules. *Top Curr. Chem.* DOI: 10.1007/128_2010_113.

Gunasekera, S.P., S.A. Pomponi, and P.J. McCarthy, 1994. Discobahamins A and B, new peptides from the Bahamian deep water marine sponge *Discodermia* sp. *J. Nat. Prod.* 57(1):79-83.

Gurgui, C. and J. Piel, 2010. Metagenomic approaches to identify and isolate bioactive natural products from microbiota of marine sponges. Methods in Molecular Biology, Springer Science + Business Media LLC.

Haefner, B, 2003. Drugs from the deep: Marine natural products as drug candidates. *Drug Discov. Today* 8:536-544.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95-98.

Hall, N., 2007. Advanced sequencing technologies and their wider impacts in microbiology. *The J. Experiment. Biol.* 209:1518-1525.

Hamada, T., S. Matsunaga, M. Fujiwara, K. Fujita, H. Hirota, R. Schmucki, P. Güntert, and N. Fusetani, 2010. Solution structure of polytheonamide B, a highly cytotoxic nonribosomal polypeptide from marine sponge. *J. Am. Chem. Soc.* 132(37):12941-12945.

Hay, M.E. and W. Fenical, 1996. Chemical ecology and marine biodiversity: insights and products from the sea. *Oceanography* 9(1):10-20.

Haydock S., J. Aparicio, I.T.S Molnar, L. Khaw, A. König, A. Marsden, I. Galloway, J. Staunton, P.F. Leadley, 1995. Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* 374:246-248.

Hamilton, G. A., 1969. Mechanisms of two- and four-electron oxidations catalyzed by some metalloenzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* 32, 55–96.

Helda, M., S. Norland, E. S. Erichsen, R-A. Sandaa, A. Larsen, F. Thingstad, and G. Bratbak, 2012. Mg^{2+} as an indicator of nutritional status in marine bacteria. *The ISME J.* 6:524-530.

Henstchel, U., J. Hopke, M. Horn, A.B. Friedrich, M. Wagner, J. Hacker, and B.S. Moore, 2002. Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* 68(9):4431-4440.

Hentschel, U., K. M. Usher, M.W. Taylor, 2006. Marine sponges as microbial fermenters. *FEMS Microbiol. Ecol.* 55(2):167-177.

Harada, K-i, K. Fujii, T. Shimada, and M. Suzuki, 1995. Two cyclic peptides, anabaenopeptins, a third group of bioactive compounds from the cyanobacterium Anabaena flos-aquae NRC 525-17. *Tetrahedron Lett*. 36(9):1511-1514.

Hertweck, C., 2009. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int.* Ed. 48:4688-4716.

Hirata, Y. and D. Uemura, 1986. Halichondrins – antitumor polyether macrolides from a marine sponge. *Pure & Appl. Chem.* 58(5):701-710.

Hoffmann, D., J.M. Hevel, R.E. Moore, and B.S. Moore, 2003. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster of Nostoc sp. GSV224. *Gene* 311():171-180.

Hoffmann, F., R. Radax, D. Woebken, M. Holtappels, G. Lavik, H.T. Rapp, M-L. Schläppy, C. Schleper, M.M.M. Kuypers, 2009. Complex nitrogen cycling in the sponge *Geodia barretti*. *Environ. Microbiol.* 11(9):2228-2243.

Hohn, B., 1979. *In vitro* packaging of λ and cosmid DNA. *Methods Enzymol.* 68:299-309.

Hollenhorst, M.A., S.B. Bumpus, M.L. Matthews, J.M. Bollinger Jr, N.L. Kelleher, and C.T. Walsh, 2010. The nonribosomal peptide synthetase enzymes DdaD tethers N(beta)-fumaramoyl-l-2,3-diaminopropionate for Fe(II)/alpha-ketoglutarate-dependent epoxidation by DdaC during dapdiamide antibiotic biosynthesis. *J. Am. Chem. Soc.* 132:15773-15781.

Hopwood, D.A., 1997. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* 97:2465-2497.

Hoyer, K.M., C. Mahlert, and M.A. Marahiel, 2007. The iterative Gramicidin S Thioesterase catalyzes peptide ligation and cyclization. *Chem. Biol.* 14(1):13-22.

Hu, Y., V. Phelan, I. Ntai, C.M. Farnet, E. Zazopoulos, and B.O. Bachmann, 2007. Benzodiazepine biosynthesis in Streptomyces refuineus. *Chem. Biol.* 14:691-701.

Hugenholtz P., B.M. Goebel, N.R. Pace, 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180(18):4765-4774.

Hrvatin, S. and J. Piel, 2007. Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *J. Microbiol. Meth.* 68:434-436.

Huson, D.H., A.F. Auch, J. Qi, and S.C. Schuster, 2007. MEGAN: analysis of metagenomic data. *Genome Res.* 17:377–386.

Irschik,H., M. Kopp, K.J. Weissman, K. Buntin, J. Piel, and R. Muller, 2010. Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by *trans*-AT PKS. *Chembiochem* 11(13):1840-1849.

Ishibashi, M., R.E. Moore, G.M.L. Patterson, C. Xu and J. Clardy, 1986. Scytophycins, cytotoxic and antimycotic agents from the cyanophyte *Scytonema pseudohofmanni*. *J. Org. Chem*. 51(26): 5300-5306.

Ishikawa, J. and K. Hotta, 2006. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol. Lett.* 174(2):251-253.

Itagika, F., H. Shigemori, M. Ishibashi, T. Nakamura, T. Sasaki, and J. Kobayashi, 1992. Keramamide F, a new thiazole-containing peptide from the Okinawan marine sponge *Theonella* sp. *J. Org. Chem.* 57(20):5540-5542.

Itoh, H., S. Matsuoka, M. Kreir, and M. Inoue, 2012. Design, synthesis and functional analysis of danyslated polytheonamide mimic: an artificial peptide ion chanel. *J. Am. Chem. Soc.* 134:14011-14018.

Iverson, V., R.M. Morris, C.D. Frazar, C.T. Berthiaume, R.L. Morales, E.V. Armbrust, 2012. Untailing genomes from metagenomes revealing an uncultured class of marine euryarchaeota. *Science* 335:587-590.

Iwamoto, M., H. Shimizu, I. Muramatsu, S. Oiki, 2010. A cytotoxic peptide from a marine sponge exhibits ion channel activity through vectorial-insertion into the membrane. *FEBS Lett.* 584(18):3995-3999.

Jackson, C.R., J.P. Harper, D. Willoughby, E.E. Roden, P.F. Churchill, 1997. A simple, efficient method for the separation of humic substances and DNA from environmental samples. *App. Env. Microbiol.* 63:4993-4995.

Jena Bioscince, 2011. Thermophylic polymerase: a user guide (http://www.jenabioscience.com).

Jensen, P.R., R. Dwight, and W. Fenical, 1991. Distribution of actinomycetes in near-shore tropical marine sediments*. Appl. Environ. Microbiol*. 57:1102-1108.

Jensen, P.R. and W. Fenical, 1994. Strategies for the discovery of secondary metabolites from marine bacteria. *Annu. Rev. Microbiol.* 48:559-584.

Jensen, K., H. Niederkrüger, K. Zimmermann, A.L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Pöplau, C. Kohlhaas, C.A. Townsend, M. Oldiges, C. Hertweck, and J. Piel. Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* 19:329-339.

Joint, I., M. Mühling, and J. Querellou, 2010. Culturing marine bacteria – an essential prerequisite for biodiscovery. *Microbial Biotechnol.* 3(5):564-575.

Jones, A.S. and R.T. Walker, 1963. Isolation and analysis of the deoxyribonucleic acid of *Mycoplasma mycoides* var. *capri.* Nature 198:588-589.

Julien,B., Tian,Z.Q., Reid,R. and Reeves,C.D., 2006. Analysis of the ambruticin and jerangolid gene clusters of Sorangium cellulosum reveals unusual mechanisms of polyketide biosynthesis. *Chem. Biol.* 13 (12), 1277-1286.

Kadi, N., D. Oves-Costales, F. Barona-Gomez, and G.L. Challis, 2007. A new family of ATP-dependent oligomerization-marcocyclization biocatalysts. *Nat. Chem. Biol.* 3:652-656.

Kao, C.M., L. Katz, and C. Khosla, 1994. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* 265:509-512.

Kao, C.M., G. Luo, L. Katz, D.E. Cane, C. Khosla, 1995. Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase.. *J. Am. Chem. Soc*. 117, 9105-9106.

Kato, Y., N. Fusetani, S. Matsunaga, K. Hashimoto, R. Sakai, T. Higa, Y. Kashman. 1987. Antitumor macrodiolides isolated from a marine sponge *Theonella* sp.: structure revision misakinolide A. *Tetrahedron Lett*. 28(49):6225-6228.

Karner, M.B., E.F. DeLong, and D.M. Karl, 2009. Archaeal dominance in the mesopelagiczoneof the Pacific Ocean. *Nature* 409:507.510.

Kaufman, B., S. Richards, and D.A. Dierig, 1999. DNA isolation method for high polysaccaride *Lesquerella* species. *Industrial Crops and Products* 9:111-114.

Keller, S., T. Wage, K. Hohaus, M. Hölzer, E. Eichhorn, K-H van Pée, 2000. Purification and partial characterization of tryptophan-7-halogenase (PrnA) from Pseudomonas fluorescens. *Angew. Chem. Int. Ed. Engl.* 39:2300-2302.

Kellner, R.L.L. and K. Dettner, 1996. Differential efficacy of toxic pederin in deterring potential arthropod predators of *Paederus (Coleoptera:* Staphylinidae) offspring. *Oecologia* 107:293-300.

Kellogg, D.E., I. Rybalkin, S. Chen, N. Mukhamedova, T. Vlasik, P.D. Siebert, and A. Chenchik, 1994. TaqStart Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *Biotechniques.* 16:1134-1137.

Keating, T.A. and C.T. Walsh, 1999. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. Curr. Opinion Chem. Biol. 3:598-606.

Keating-Clay, T.A. and R.M. Stroud, 2006. The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* 14(4):737-748.

Keating-Clay, T.A., 2007. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* 14(8):898-908.

Keatinge-Clay, T.A., 2008. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* 384(4):941-953.

Kennedy, J., J.R. Marchesi, and A.D.W. Dobson, 2007. Metagenomic Approach to exploit the Biotechnological Potential of the Microbial Consortia of Marine Sponges. *Appl. Microbiol. Biotechnol*. DOI 10.1007/s00253-007-0875-2.

Kho, E., D.K. Imagawa, M. Rohmer, Y. Kashman, C. Djerassi, 1981. Sterols in marine invertebrates. 22. Isolation and structure elucidation of conicasterol and theonellasterol, two new 4-methylene sterols from the Red Sea sponges *Theonella conica* and *Theonella swinhoei*. *J. Org. Chem.* 46:1836-1839.

Khosla, C., 1997. Harnessing the biosynthetic potential of modular polyketide synthases. *Chem. Rev.* 97:2577-2590.

Khosla, C. Y. Tang, A.Y. Chen, N.A. Schnarr, and D.E. Cane, 2007. Structure and mechanism of the 6-Deoxyerythronolide B synthase. *Annu. Rev. Biochem.* 76:195-221.

Kleiner, M., C. Wentrup, C. Lott, H. Teeling, S. Wetzel, J. Young, Y-J. Chang, M. Shah, N.C. VerBerkmoes, J. Zarzycki, G. Fuchs, S. Markert, K. Hempel, B. Voigt, D. Becher, M. Liebeke, M. Lalk, D. Albrecht, M. Hecker, T. Schweder, and N. Dubilier, 2011. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Prod. Nat. Acad. Sci. USA*. www.pnas.org/cgi/doi/10.1073/pnas.1121198109

Kirby, K.S., 1964. Isolation and fractionation of nucleic acids. *Prog. Nucleic Acid Res. Mol. Biol.* 3:1-31.

König, G.M., S. Kehraus, S.F. Seibert, A. Abdel-Lateff, and D. Müller, 2006. Natural products from marine organisms and their associated microbes. *ChemBioChem* 7(2):229-238.

Kroutil, W., H. Mang, K. Edegger, and K. Faber, 2004. Biocatalytic oxidation of primary and secondary alcohols. *Adv. Synth. Catal.* 346:125-142.

Keatinge-Clay, A.T., 2007. A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. 14(8):898-908.

Keatinge-Clay, A.T. and R.M. Stroud, 2006. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* 14(4):737-748.

Kevany, B.M., D.A. Rasko, and M.G. Thomas, 2008. Characterization of the complete zwittermicin A biosynthesis gene cluster from Bacillus cereus*. Appl. Environ. Microbiol.* 75(4):1144-1155.

Kim,J.-S., R.S. Dungan, and D. Crowley, 2008. Microarray analysis of bacterial diversity and distribution in aggregates from a desert agricultural soil. *Biol. Fertil. Soils* 44:1003-1011.

Kimura, M., T. Wakimoto, Y. Egami, K.C. Tan, Y. Ise, and I. Abe, 2012. Calyxamides A and B, cytotoxic cyclic peptides from the marine sponge Discodermia calyx. *J. Nat. Prod.* DOI**:** 10.1021/np2009187

Kitagawa, I., M. Kobayashi, T. Katori, M. Yamashita, J. Tanaka, M. Doi, and T. Ishida, 1990. Absolute stereostructure of swinholide A, a potent cytotoxic macrolide from the Okinawan marine sponge *Theonella swinhoei*. *J. Am .Chem. Soc.* 112(9):3710-3712.

Kitamura, M., P.J. Schupp, Y. Nakano, D. Uemura, 2009. Luminaolide, a novel metamorphosis.enhancing macrolide for scleractinian coral larvae from crustose coralline algae. *Tetrahendron Lett.* 50(47):6606-6609.

Kleinkauf, H. and H. von Döhren, 1990. Nonribosomal biosynthesis of peptide antibiotics. *Eur. J. Biochem.* 192:1-15.

Kobayashi, J., F. Itagaki, H. Shigemori, M. Ishibashi, K. Takahashi, M. Ogura, S. Nagasawa, T. Nakamura, and H. Hirota, 1991. Keramamides B .appx. D, novel peptides from the Okinawan marine sponge *Theonella* sp. *J. Am. Chem. Soc*. 113(20):7812-7813.

Kobayashi, J., F. Itagaki, I. Shigemori, T. Takao, and Y. Shimonishi, 1995. Keramamides E, G, H, and J, new cyclic peptides containing an oxazole or a thiazole ring from a *Theonella* sponge. *Tetrahedron* 51(9):2525-2532.

Koetsier, M.J., P.A. Jekel, H.J. Wijma, R.A.L. Bovenberg, and D.B. Janssen, 2011. Aminoacyl-coenzyme A synthesis catalyzed by a CoA ligase from Penicillium chrysogenum. *FEBS Lett.* 585:893-898.

Kohli, R.M. and C.T. Walsh, 2003. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun*. 3:297-307.

Kopp, M., H. Irschik, S. Pradella, and R. Müller, 2005. Production oft he tubulin destabilizer disorazol in Sorangium cellulosum: biosynthetic machinery and regulatory genes. *ChemBioChem* 6(7):1277-1286.

Korman, T.P., J.A. Hill, T.N. Vu, and S-C. Tsai, 2004. Structural analysis of actinorhodin polyketide ketoreductase: cofactor binding and substrate specificity. *Biochemistry* 43(46):14529-14538.

Kowalczyk, J.J., S.E.R. Schiller, M. Spyvee, H.Yang, B.M. Seletsky, C.J. Shaffer, V. Marceau, Y. Yao, G. Kuznetsov, K. Tendyke, D. Liu. C. Rowell, B.A. Littlefield, and E.M. Suh, 2005. Synthetic analogs of the marine natural product hemiasterlin: optimization and discovery of E7974, a novel and potent antitumor agent. *Proc. Amer. Assoc. Cancer Res.* 46:282-b.

Koumoutsi, A., X-H. Chen X-H, A. Henne, H. Liesegang, G. Hitzeroth, P. Franke, J. Vater, R. Borriss, 2004. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J. Bacteriol.* 186:1084–1096.

Kubenak, J., P.R. Jensen, P.A. Keifer, M.C. Sullards, D.O. Collins, and W. Fenical, 2003. Seaweed resistance to microbial attack: a targeted chemical defense against marine fungi. *Proc. Natl. Acad. Sci.,* USA 100(12):6916-6921.

Kurahashi, M. and A. Yokota, 2007. Endozoicimonas elysicola gen. nov., sp. nov., a gamma-proteobacterium isolated from the sea slug Elysia ornate. *Syst. Appl. Microbiol.* In press.

Kuznetsov, G., K. TenDyke, M.J. Towle, H. Cheng, J. Liu, J.P. Marsh, S.E.R. Schiller, M.R. Spyvee, H. Yang, B.M. Seletsky, C.J. Shaffer, V. Marceau, Y. Yao, E.M. Suh, S. Campagna, F.G. Fang, J.J. Kowalczyk, and B.A. Littlefield, 2009. Tubulin-based antimitotic mechanism of E7974, a novel analogue of the marine sponge natural product hemiasterlin. *Mol. Cancer Ther.* 8(10):2852-2860.

Kwan, D.H., Y. Sun, F. Schulz, H. Hong, B. Popovic, J.C.C. Sim-Stark, S.F. Haydock, and P.F. Leadley, 2008. Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. *Chem. Biol.* 15:1231-1240.

Lambalot, R.H., A.M. Gehring, R.S. Flugel, P. Zuber, M. LaCele, M.A. Marahiel, R. Reid, C. Khosla, and C.T. Walsh, 1996. A new enzyme family – the phosphopantetheinyl transferases. *Chem. Biol.* 3:923-936.

Lau, J., H. Fu, D.E. Cane, and C. Khosla, 1999. Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochem.* 38(5):1643-1651.

Lau, J., D.E. Cane and C. Khosla, 2000. Substrate specificity of the loading didomain of the erythromycin polyketide synthase. *Biochemistry* 39:10514-10520.

Lautru, S., M. Gondry, R. Genet, and J.L. Pernodet, 2002. The albonoursin gene cluster of *S. noursei*: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.* 9(12):1355-1364.

Leal, M.C., J. Puga, J. Serôdio, N.C.M. Gomes, R. Calado, 2012. Trends in the discovery of new marine natural products from invertebrates over the last two decades – where and what are bioprospecting. *PloS ONE* 7(1):1-15.

Lee, O.O., P.Y. Chui, Y.H. Wong, J.R. Pawlik, and P-Y. Qian, 2009. Evidence of vertical transmission of bacterial symbionts from adult to embryo in the Caribbean sponge *Svenzea zeai. App. Environ. Microbiol*. 75(19):6147-6156.

Lee, Y.K., J-H. Lee, and H.K. Lee, 2001. Microbial Symbiosis in Marine Sponges. *The J. Microbiol.* 39(4): 254-264.

Li Y.M., J.C. Milne, L.L. Madison, R. Kolter, C.T. Walsh, 1996. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* 274:1188-1193.

Llewellyn, N.M. and J.B. Spencer, 2007. Enzymes line up for assembly. *Nature* 448(16):755-756.

Long, P.F., W.C. Dunlap, C.N. Battershill, and M. Jaspars, 2005. Shotgun cloning and heterologous expression ofthe patellamide gene cluster as a strategy to achieving sustained metabolite production. *ChemBioChem* 6:1760-1765.

Magarvey, N.A., Z.Q. Beck, T. Golakoti, Y. Ding, U. Huber, T.K. Hemscheidt, D. Abelson, R.E. Moore, and D.H. Sherman, 2006. Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins, potent anticancer agents from cyanobionts. *ACS Chem. Biol.* 1(12):766-779.

Marchuk, D., M. Drumm, A. Saulino, F.S. Collins, 1990. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucl. Acids Res*. 19(5):1154.

Margulies M *et al* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380

Marsden, A.F., P. Caffrey, J.F. Aparicio, M.S. Loughran, J. Staunto acyl acyln, and P.F. Leadley, 1994. Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* 263(5145):378-380.

Marahiel, M.A., T. Stachelhaus, and H.D. Mootz, 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* 97(7):2651-2673.

Marahiel, M.A. and L.O. Essen, 2009. Nonribosomal peptide synthetases mechanistic and structural aspects of essential domains. *Methods Enzymol.* 458:337-351.

Martín, J.F., 2000. a-aminoadipyl-cysteinyl-valine synthetases in *β*-lactam producing organisms. From Abraham's discoveries to novel concepts of non-ribosomal peptide synthesis, *J. Antibiot*. 53. 1008e1021.

Mattheus, W., Gao,L.J., Herdewijn,P., Landuyt,B., Verhaegen,J., Masschelein,J., Volckaert,G. and Lavigne,R., 2010. Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. *Chem. Biol*. 17 (2), 149-159

Matsunaga, S., N. Fusetani, K. Hashimoto, M. Walchli, 1989. Theonellamide F. a novel antifungal bicyclic peptide from a marine sponge *Theonella* sp. *J. Am. Chem. Soc.* 111(7):2582-2588.

Matsunaga, S., N. Fusetani, and Y. Kato, 1991. Aurantosides A and B: cytotoxic tetramic acid glycosides from the marine sponge *Theonella* sp. J. Am. Chem. Soc. 113(25):9690-9692.

Matsunaga, S., N. Fusetani, and Y. Nakao, 1992. Eight new cytotoxic metabolites closely related to onnamide A from two marine sponges of the genus Theonella. *Tetrahedron* 48(39):8369-8376.

Matsunaga, S. and N. Fusetani, 1995. Theonellamides A-E, cytotoxic bicyclic peptides, from a marine sponge *Theonella* sp. *J. Org. Chem.* 60(5):1177-1181.

McDaniel, R., A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach, and G. Ashley, 1999. Multiple genetic modifications of the erythromycin polyketide synthase

to produce a library of novel “unnatural” natural products. *Proc. Natl. Acad. Sci. USA* 96:1846-1851.

McGinnis, S., and T.L. Madden, 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucl. Acids Res.* 32:W20-W25.

McGuire, J.M., R.L. Bunch, R.C. Anderson, H.E. Boaz, E.H. Flynn, H.M. Powell, J.W. Smith, 1952. Ilotycin, a new antibiotic. Antibiotics & Chemotherapy 2:281.283.

Mellin, C., S. Delean, J. Caley, G. Edgar, M. Meekan, R. Pitcher, R. Przeslawski, A. Williams, and C. Bradshaw, 2011. Effectiveness of biological surrogates for predicting patterns of marine biodiversity: a global meta-analysis. *PLoS ONE* 6(6):e20141.

Menis, J. and C. Twelves, 2011. Eribulin (Halaven): a new, effective treatment for women with heavily pretreated metastatic breast cancer. Breast Cancer: Targets and Therapy 3:101-111.

Miller, D.A., L. Luo, N. Hillson, T.A. Keating, and C.T. Walsh, 2002. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem. Biol.* 9(3):333-344.

Mincer, T.J.; P.R. Jensen, C.A. Kauffman, and W. Fenical, 2002. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl. Env. Microbiol.* 68(10):5005-5011.

Mochizuki, S., K. Hiratsu, M. Suwa, T. Ishii, F. Sugino, K. Yamada, and H. Kinashi, 2003. The larger linier plasmid pSLA2-L of Streptomyces rochei has an unusually condensed gene organization for secondary metabolism. *Mol. Microbiol.* 48(6):1501-1510.

Moffitt, M.C. and B.A. Neilan, 2004. Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl. Environ. Microbiol.* 70(11):6353-6362.

Molina-Henares, A. J., T. Krell, M.E. Guazzaroni, A. Segura, and J. L. Ramos. 2006. Members of the IclR family of bacterial transcriptionalregulators function as activators and/or repressors. FEMS *Microbiol. Rev*. 30:157–186.

Moldenhauer, J., X-H Chen, R. Borriss, and J. Piel, 2007. Biosynthesis of the antibiotic Bacillaene, the product of a giant polyketide synthase complex of the *trans*-AT family. *Angew. Chem.* 119:8343-8345.

Molnár, I., J.F. Aparicio, S.F. Haydock, L.E. Khaw, T. Schwecke, A. König, J. Staunton, and P.F. Leadlay, 1996. Organization of the biosynthetic gene

cluster for rapamycin in *Streptomyces hygroscopicus*: Analysis of genes flanking the polyketide synthase*. Gene* 169: 9-16.

Montalvo, N.F. and R.T. Hill, 2011. Sponge-associated bacteria are strictly maintained in two closelyrelated but geographically distant sponge hosts. *Appl. Environ. Microbiol.* 77 (20), 7207-7216

Moore, R.E, 1996. Cyclic peptides and depsipeptides from cyanobacteria: a review. *J. Ind. Microbiol.* 16(2):134-143.

Moore, B.S. and C. Hertweck, 2002. Biosynthesis and attachment of novel polyketide synthase starter units. *Nat. Prod. Rep*. 19:70-99.

Mora, C., D.P. Tittensor, S. Adl, A.G.B. Simpson, and B. Worm, 2011. How many species are there on earth and in the ocean? *PloS Biol.* 9(8):e1001127.

Motamedi, H., A. Shafiee, S.J. Cai, S.L. Streicher, B.H. Arison, and R.R. Miller, 1996. Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520, *J. Bacteriol. 178*, 5243-5248.

Motamedi, H., S.J. Cai, A. Shafiee, and K.O. Elliston, 1997. Structural organization of a multifunctional polyketide synthase involved in biosynthesis of the macrolide immunosuppressant FK506. *Eur. J. Biochem.* 244:74-80.

Motamedi, H., and A. Shafiee, 1998. The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.* 256:528-534.

Mynderse, J.S. and R.E. Moore, 1978. Toxins from blue-green algae: structures of oscillatoxin A and three related bromine-containing toxins. *J. Org. Chem.* 43:2301-2303.

Müller, M., A. Krick, S. Kehraus, C. Mehner, M. Hart, F.C. Küpper, K. Saxena, H. Prinz, H. Schwalbe, P. Janning, H. Waldmann, and G.M. König, 2006. Brunsvicamides A-C: sponge-related cyanobacterial peptides with *Mycobacterium tuberculosis* protein tyrosine phosphatase inhibitory activity. *J. Med. Chem.* 49:4871-4878.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich, 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* Vol. LI:263-273.

Murakami, M., H.J. Shin, H. Matsuda, K. Ishida, and K. Yamaguchi, 1997. A cyclic peptide, anabaenopeptide B, from the cyanobacterium Oscillatoria agardhii. *Phytochem.* 44(3):440-452.

Nagai, M., A. Yoshida, and N. Sato, 1997. Additive effects of bovine serum albumin, dithiothreitol, and glyceron on PCR. *Biochem. Mol. Biol. Intl.* 44(1):157-163.

Nakao, Y., S. Matsunaga and N. Fusetani, 1995. Three more cyclotheonamides, C, D, and E, potent thrombin inhibitors from the marine sponge, *Theonella swinhoei*. *Bioorg. Med. Chem.* 3(8):1115-1122.

Nakao, Y., A. Masuda, S. Matsunaga, and N. Fusetani, 1999. Pseudotheonamides, serine protease inhibitors from the marine sponge *Theonella swinhoei*. *J. Am. Chem. Soc.* 121(11):2425-2431.

Namikoshi, M., K. L. Rinehart, R. Sakai, R. R. Stotts, A. M. Dahlem, V. R. Beasley, W. W. Carmichael, W.R. Evans, 1992. Identification of 12 hepatotoxins from a homer lake bloom of the cyanobacteria *Microcystis aeruginosa*, *microcystis viridis*, and *mycrocystis wesenbergii*: nine new mycrocystins. *J. Org. Chem. 57* (3):866–872.

Newman, D.J. and G.M. Cragg, 2004. Marine natural products and related compounds in clinical and advanced preclinical trials. *J. Nat. Prod.* 67(8):1216-1238.

Nguyen, T. K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, and J. Piel, 2008. Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol*. 26(2):225-233.

Nishijima,M., Matsuo,S., Kano,K., Kamino,K., Shizuri,Y. and Yamasato,K., 2005. Spongiobacter nickelotolerans gen. nov., sp. nov., a novel nickel tolerant bacterium isolated from marine sponge. *Unpublished*. Submitted to the GenBank (accession number AB205011).

Nunnery, J.K, E. Mevers, and W.H. Gerwick, 2010. Biologically active secondary metabolites from marine cyanobacteria. *Curr. Opin. Biotechnol*. 21(6):787-793.

Ortega, V. and J. Cortés, 2012. Potential clinical applications of halicondrins in breast cancer and other neoplasms. *Breast Cancer: Targets and Therapy* 4:9-19.

Parish, C.R., K.B. Jakobsen, D.R. Coombe, and A. Bacic, 1991. Isolation and characterization of cell adhesion molecules from the marine sponge, *Ophlitaspongia tenuis*. *Biochim. Biophys. Acta* 1073(1):56-64.

Partida-Martinez, L. P. and C. Hertweck, 2007. A Gene cluster encoding rhizoxin Biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. *ChemBioChem* (8):41–45. doi: 10.1002/cbic.200600393.

Perlova, O., K. Gerth, O. Kaiser, A. Hans, and R. Müller, 2005. Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *J. Biotechnol.* 121(2):174-191.

Persson, B., Y. Kallberg, U. Oppermann, and H. Jornvall, 2003. Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). Chem. Biol. Interact. 143-144:271-278.

Perry, N.B., J.W. Blunt, M.H.G. Munro, A.M. Thompson, 1990. Antiviral and antitumor agents from a New Zealand sponge, *Mycale* sp. 2. Structures and solution conformations of mycalamides A and B. *J. Org. Chem*. 55(1):223-227.

Pettit, G.R., 2004. Antineoplastic agents, 520. Isolation and structure of irchiniastatin A and B from the Indo-Pasific marine sponge *Ircinia ramose*. *J. Med. Chem*. 47:1149-1152.

Pfeifer, B.A., S.J. Admiraal, H. Gramajo, D.E. Cane, and C. Khosla, 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli. Science* 291(5509):1790-1792.

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.

Piel J., 2004. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* 21:519–538.

Piel, J., D. hui, N. Fusetani and S. Matsunaga, 2004a. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Env. Microbiol.* 6(9):921-927.

Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, and S. Matsunaga, 2004b. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci,* USA *101*, 16222-16227.

Piel, J., G. Wen, M. Platzer, and D. Hui, 2004c. Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *ChemBioChem 5*, 93-*98.*

Piel J., D. Butzke, N. Fusetani, D. Hui, M. Platzer, G. Wen, and S. Matsunaga, 2005. Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family. *J. Nat. Prod.* 68:472-479.

Piel,J., 2006. Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals. *Curr. Med. Chem.* 13(1):39-50.

Piel, J. 2010. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep*. 27:996-1047.

Pieper, R., G. Luo, D.E. Cane, and C. Khosla, 1995. Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthesis. *Nature* 378:263-266.

Pimentel-Elardo, S.M., L. Grozdanov, S. Proksch, and U. Hentschel, 2012. Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. *Mar. Drugs* 10(6):1192-1202.

Pistorius, D. and R. Müller, 2012. Discovery of the rhizopodin biosynthetic gene cluster in Stigmatella aurantica Sg a15 by genome mining. *ChemBioChem* 13(3):416-426.

Plaisance, L., M.J. Caley, R.E. Brainard, and N. Knowlton, 2011. The Diversity of coral reefs: what are we missing? *PloS ONE* 6(10):e25026.

Price, A.C., Y.M. Zhang, C.O. Rock, and S.W. White, 2001. Structure of beta-ketoacyl-acyl carrier protein synthase II from Streptococcus pneumonia. *J. Bacteriol.* 185(14):4136-4143.

Price, A.C., Y.M. Zhang, C.O. Rock, and S.W. White, 2004. Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. *Structure* (Camb) 12(3):417-428.

Proksch, P., 1994. Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs. *Toxicon* 32(6):639-655.

Proksch, P., R.A. Edrada, and R. Ebel, 2002. Drugs from the Sea – Current Status and Microbiological Implications. *Appl. Micrbiol. Biotechnol.* 59:125-135.

Qiagen. www.qiagen.com

Ramaswamy, A.V., C.M. Sorrels, and W.H. Gerwick, 2007. Cloning and biochemical characterization of the Hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* 70(12):1977-1986.

Rath, C.M., B. Janto, J. Earl, A. Ahmed, F.Z. Hu, L. Hiller, M. Dahlgren, R. Kreft, F. Yu, J.J.Wolff, H.K. Kweon, M.A. Christiansen, K. Håkansson, R.M. Williams, G.D. Ehrlich, and D.H. Sherman, 2011. Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chem. Biol.* 6(11):1244-1256.

Rattray, R.M., S. Perumbakkam, F. Smith and A.M. Craig, 2010. Microbiomic comparison of the intestine of the earthworm *Eisenia fetida* fed ergovaline. *Curr. Microbiol.* 60(3):229-235.

Rausch, C., T. Weber, O. Kohlbacher, W. Wohlleben, and D.H. Huson, 2005. Specificity prediction of adenylation domains in nonribosomal peptide

synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucl. Acids Res.* 33(18):5799-5808.

Rawlings, B.J., 1998. Biosynthesis of fatty acids and related metabolites. *Nat.Prod. Rep.* 15:275-308.

Rawlings, B.J., 2001. Type I polyketide biosynthesis in bacteria (Part B). *Nat. Prod. Rep. 18:231-281.*

Reeves, C.D., S. Murli, G.W. Ashley, M. Piagentini, C.R. Hutchinson, and R. McDaniel, 2001. Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochem*. 40(51):15464-15470.

Reid, R., M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D.V. Santi, C.R. Hutchinson, and R. McDaniel, 2003. A model of structure and catalysis of ketoreductase domains in modular polyketide synthases. *Biochemistry* 42(1):72-79.

Reiswig, H.M., 1975. Bacteria as food for temperate-water marine sponges. *Can. J. Zool.* 53:582-589.

Ribes, M., R. Coma, and J-M. Gili, 1999. Natural diet and grazing rate of the template sponge *Dysidea avara* (Demospongiae, Dendroceratida) throughout an annual cycle. *Mar. Ecol. Prog. Ser.* 176:179-190.

Rickwood, J., and B.D. Hames, 1982. Gel electrophoresis of nucleic acids. *IRL Press Ltd.*

Riesenfeld, C.S., P.D. Schloss, and J. Handelsman, 2004. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38:525-552.

Roberts, G.A., J. Staunton, and P.F. Leadley, 1993. Heterologous expression in *Escherichia coli* of an intact multienzyme component of the erythromycin-producing polyketide synthase. *Eur. J. Biochem.* 214(1):305-311.

Robbel, L., K.M. Hoyer and M.A. Marahiel, 2009. TioS T-TE – a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides. *FEBS J.* 276(6):1641–1653.

Rocha-Lima, C.M., S. Bayraktar, J. Maclntyre, L. Raez, A.M. Flores, A. Ferrell, E.H. Rubin, E.A. Poplin, A.R. Tan, A. Lucarelli, and N. Zojwalla, 2012. A phase 1 trial of E7974 administered on day 1 of a 21-day cycle in patients with advanced solid tumors. *Cancer* doi: 10.1002/cncr.27428.

Rounge, T.B., T. Rohrlack, A.J. Nederbragt, T. Kristensen, and K.S. Jakobsen, 2009. A genome-wide analysis of nonribosomal peptide synthetase gene

clusters and their peptides in a *Planktothrix rubescens* strain. *BMC Genomics* 10:396 doi:10.1186/1471-2164-10-396.

Röttig, M., M.H. Medema, K. Blin, T. Weber, C. Rausch, and O. Kohlbacher, 2011. NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity. *Nucl. Acids Res.* 39:W362-W367.

Rogers, S.O. and A.J. Bendich, 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol.* 5:69-76.

Rothberg J.M. and J.H. Leamon, 2008. The development and impact of 454 sequencing. *Nat Biotechnol* 26:1117–1124

Ruppert, E.E., R.S. Fox, and R.D. Barnes, 2004. Invertebrate Zoology: A functional Evolutionary Approach. Brooks/Cole-Thomson Learning, Inc. ISBN 0-03-025982-7.

Saitou, N. and M. Nei, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich, 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.

Sakai, R., T. Higa, and Y. Kashman, 1986. Misakinolide-A, an antitumor macrolide from the marine sponge *Theonella* sp. *Chem. Lett.* 1499-1502.

Sakemi, S., T. Ichiba, S. Kohmoto, G. Saucy, and T. Higa, 1988. Isolation and structure elucidation of onnamide A, a new bioactive metabolite of a marine sponge, *Theonella* sp. *J. Am. Chem. Soc*. 110:4851.

Sala E., and N. Knowlton, 2006. Global marine biodiversity trends. *Annu. Rev. Environ. Resour.* 31:93–122.

Sambrook, J. and D.W. Russell, 2001. Molecular cloning: a laboratory manual, third edition. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., S. Nicklen, and A.R. Coulson, 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463-5467.

Schafer,H., I.R. McDonald, P.D. Nightingale and J.C. Murrell, 2005. Evidence for the presence of a CmuA methyltransferase pathway in novel marine methyl halide-oxidizing bacteria. *Environ. Microbiol.* 7(6):839-852

Schaffer, M.L., and L.G. Otten, 2009. Substrate flexibility of the adenylation reaction in the Tyrocidine non-ribosomal peptide synthetase. *J. Mol. Cat. B*: Enzymatic 59:140–144.

Scheuermayer M., S. Pimentel-Elardo, L. Fieseler, L. Grozdanov, and U. Hentschel. 2007. Microorganisms of sponges: phylogenetic diversity and biotechnological

potential. *In* Frontiers in Marine Biotechnology, edited by Peter Proksch and Werner E.G. Müller.

Schmidt, E.W., 2008. Trading molecules and tracking targets in symbiotic interactions. *Nat. Chem. Biol.* 4:466-473.

Schmidt, E.W., J.T. Nelson, D.A. Rasko, S. Sudek, J.A. Eisen, M.G. Haygood, and 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(20):7315-7320.

Schmidt, E.W., C.A. Bewley, and D.J. Faulkner, 1998. Theopalauamide, a bicyclic glycopeptides from filamentous bacterial symbionts of the lithistid sponge, Theonella swinhoei from Palau and Mozambique. *J. Org. Chem*. 63:1254-1258.

Schmidt E.W., J.T. Nelson, D.A. Rasko, S. Sudek, J.A. Eisen, M.G. Haygood, 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.

Schmitt, S., H. Angermeier, R. Schiller, N. Lindquist, and U. Hentschel, 2008. Molecular microbial diversity survey of sponge reproductive stages and mechanistic insights into vertical transmission of microbial symbionts. *App. Env. Microbiol*. 74(24):7694-7708.

Schirmer, A., R. Gadkari, C.D. Reeves, F. Ibrahim, E.F. De Long, C.R. Hutchinson and R.C. Reid, 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge Discodermia dissoluta. *Appl. Environ. Microbiol.* 71(8):4840-4849.

Schneider, T.L., B. Shen, and C.T. Walsh, 2003. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* 42: 9722–9730.

Schneider, K., X-H. Chen, J. Vater, P. Franke, G. Nicholson, R. Borriss, and R.D. Süssmuth. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J. Nat. Prod.* 70(9):1417-1423.

Schwarzer D. and M.A. Marahiel (2001). Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* 88:93-101.

Schwarzer D., R. Finking, and M.A. Marahiel (2003). Nonribosomal peptides: from genes to products. *Nat. Prod. Rep*. 20:275-287.

Schweizer, E., and J. Hofmann, 2004. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol. Mol. Biol. Rev.* 68:501-517.

Smith, S., and S.C. Tsai, 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* 24(5):1041-1072.

Sieber, S.A. and M.A. Marahiel, 2003. Learning from nature`s drug factories: nonribosomal synthesis of macrocyclic peptides. *J. Bacteriol.* 185(24):7036-7043.

Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar, and U. Hentschel, 2011. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *The ISME J.* 5:61-70.

Simmons, T.L., R.C. Coates, B.R. Clark, N. Engene, D. Gonzales, E. Esquenazi, P.C. Dorrestein, and W.H. Gerwick, 2008. Biosynthetic Origin of Natural Products Isolated from Marine Microorganisms-Invertebrate Assemblages. *Proc. Natl. Acad. Sci.* 105(12):4587-4594.

Skropeta, D., 2008. Deep-sea Natural Products. *Nat. Prod. Rep*. 25:1131-1166.

Shaw-Reid, C., N.L. Kelleher, H.C. Losey, A.M. Gehring, C. Berg, and C.T. Walsh, 1999. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* 6(6):385-400.

Sheng, Y., V. Mancino, and B. Birren, 1995. Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucl. Acids Res.* 23(11):1990-1996.

Stachelhaus, T., H.D. Mootz, V. Bergendahl, M.A. Marahiel, 1998. Peptide bond formation in nonribosomal peptide biosynthesis: catalytic role of the condensation domain. *J. Biol. Chem.* 273(35):22773 –22781.

Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 6:493-505.

Staunton, J., P. Caffrey, J.F. Aparicio, G.A. Roberts, S.S. Bethell, and P.F. Leadley, 1996. Evidence for a double-helical structure for modular polyketide synthases. *Nature Structural Biol*. 3(2):188-192.

Staunton, J. and B. Wilkinson, 1997. Biosynthesis of erythromycin and rapamycin. *Chem. Rev*. 97(7):2611-2630.

Staunton, J. and K.J. Weissman, 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18:380-416.

Strom, S.L., 2008. Microbial Ecology of Ocean Biogeochemistry: A Community Perspective. *Science* 320:1043-1045.

Sogin, M.L., H.G. Morrison, J.A. Huber, D.M. Welch, S.M. Huse, P.R. Neal, J.M. Arrieta, and G.J. Herndl, 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci.* USA 103:12115-12120.

Stackebrandt, E., 2009. Phylogeny Based on 16SrRNA/DNA. eLS. DOI: 10.1002/9780470015902.a0000462.pub2.

Staley, J.T. and A. Konopka, 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Ann. Rev. Microbiol.* 39:321-346.

Sudek, S., N.B. Lopanik, L.E. Waggoner, M. Hildebrand, C. Anderson, H. Liu, A. Patel, D.H. Sherman, and M.G. Haygood, 2007. Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from “*Candidatus* Endobugula sertula”, the Uncultivated Microbial Symbiont of the Marine Bryozoan *Bugula neritina. J. Nat. Prod.,* 70:67-74.

Syn, C.K.C, and S. Swarup, 2000. A scalable protocol for the isolation of large.sized genomic DNA within an hour from several bacteria. *Anal. Biochem.* 278:86-90.

Talpir, R., Y. Benayahu, and Y. Kashman, 1994. Hemiasterlin and geodiamolide TA; two new cytotoxic peptides from the marine sponge Hemiasterella minor (Kirkpatrick). *Tetrahedron Lett*. 35(25):4453-4456.

Tamura K., J. Dudley, M. Nei, and S. Kumar, 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596-1599

Taylor, M.W., R. Radax, D. Steger, and M. Wagner, 2007. Sponge-Associated Microorganisms: Evolution, Ecological, and Biotechnological Potential. *Microbiol. Mol. Biol. Rev.* 71(2):295-347.

Terry, D.R., I. Spector, T. Higa, and M.R. Bubb, 1997. Misakinolide A is a marine macrolide that caps but does not sever filamentous actin. *J. Biol. Chem.* 272(12):7841-7845.

Teta, R., M. Gurgui, E.J.N. Helfrich, S. Künne, A. Schneider, G. van Echten-Deckert, A. Mangoni, and J. Piel, 2010. Genome mining reveals trans-AT polyketide synthase directed antibiotic biosynthesis in the bacterial phylum Bacteroidetes. *ChemBioChem* 11:2506-2512.

Thornburg, C.C., T.M. Zabriskie, and K.L. McPhail, 2010. Deep-sea hydrothermal vents: potential hot spots for natural products discovery. *J. Nat. Prod.* 73(3):489-499.

Thomas, I., C.J. Martin (née Rowe), C.J. Wilkinson, J. Staunton, and P.F. Leadley, 2002. Skipping in a hybrid polyketide synthase: evidence for ACP-to-ACP chain transfer. *Chem. Biol.* 7:781-787.

Thompson, J. D., D. G. Higgins, and T. J. Gibson, 1994. ClustalW: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.

Tillett D., E. Dittmann, M. Erhard, H. von Döhren, T. Börner, B.A. Neilan, 2000. Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide–polyketide synthetase system. *Chem. Biol*. 7(10):753-764.

Towe, K.M. and K. Rützler, 1968. Lepidocrocite iron mineralization in keratose sponge granules. *Science* 162(3850):268-269.

Tsai, S-C., H. Lu, D.E. Cane, C. Khosla, and R.M. Stroud, 2002. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* 41(42):12598-12606.

Tsai, S-C., and B.D. Ames, 2009. Structural enzymology of polyketide synthases. *Methods Enzymol.* 459:17-47.

Tsuda, M., H. Ishiyama, K. Masuko, T. Takao, Y. Shimonishi, and J. Kobayashi, 1999. Keramamides M and N, two new cyclic peptides with a sulfate ester from Theonella sponge. *Tetrahedron* 55(43):12543-12548.

Tuan, J.S., J.M. Weber, M.J. Staver, J.O. Leung, S. Donadio, and L. Katz, 1990. Cloning of genes involved in erythromycin biosynthesis from Saccharopolyspora erythraea using a novel actinomycete-*Escherichia coli* cosmid. *Gene* 90(1):21-29.

Udwary, D.W., L. Zeigler, R. N. Asolkar, V. Singan, A. Lapidus, W. Fenical, P. R. Jensen and B. S. Moore, 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. USA.* 104(25):10376–10381.

Uemoto, H., Y. Yahiro, H. Shigemori, M. Tsuda, T. Takao, Y. Shimonishi, and J. Kobayashi, 1998. Keramamides K and L, new cyclic peptides containing unusual tryptophan residue from *Theonella* sponge. *Tetrahedron* 54(24):6719-6724.

Uemura, D., K. Takahashi, T. Yamamoto, C. Katayama, J. Tanaka, Y. Okumura, and Y. Hirata, 1985. Norhalichondrin A: an antitumor polyether macrolide from a marine sponge. *J. Am. Chem. Soc.* 107(16):4796-4798.

Uria, A. and J. Piel, 2009. Cultivation-independent approaches to investigate the chemistry of marine symbiotic bacteria. *Phytochemistry Rev.* 8:401–414.

Uversucht, S., F. Hollmann, A. Schmid, K-H. Van Pée, 2005. FADH2-dependence of tryptophan-7-halogenase. *Adv. Synth. Catal.* 347:1163-1167.

Vacelet, J. and C. Donadey, 1977. Electron microscope study of the association between some sponges and bacteria. *J. Exp. Mar. Biol. Ecol.* 30(3):301.314.

Vacelet, J. and E. Duport, 2004. Prey capture and digestion in the carnivorous sponge *Asbestopluma hypogeal* (Porifera: Demospongiae). *Zoomorphology* 123:179-190.

Valenzano, C.R., R.J. Lawson, A.Y. Chen, C. Khosla, and D.E. Cane, 2009. The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc.* 131:18501-18511.

Valenzano, C.R., Y-O You, A. Gang, A. Keatinge-Clay, C. Khosla, and D.E. Cane, 2010. Stereochemistry of the dehydratase domain of the erythromycin polyketide synthase. *J. Am. Chem. Soc.* 132:14697-14699.

van Pée K-H, and Patallo, 2006. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbial. Biotechnol.* 70:631-641.

Van Soest, R.W.M., N. Boury-Esnault, J. Vacelet, M. Dohrmann, D. Erpenbeck, N.J. De Voogd, N. Santodomingo, B. Vanhoorne, M. Kelly, and J.N.A. Hooper, 2012. Global diversity of sponges (Porifera). *PLoS ONE* 7(4): e35105. doi:10.1371/journal.pone.0035105

Van Soest, R.W.M, N. Boury-Esnault, J.N.A. Hooper, K. Rützler, N.J. de Voogd, B. Alvarez de Glasby, E. Hajdu, A.B. Pisera, R. Manconi, C. Schoenberg, D. Janussen, K.R. Tabachnick, M. Klautau, B. Picton, M. Kelly, J. Vacelet, M. Dohrmann, 2012b. World Porifera database: www.marinespecies.org/porifera/porifera.php?p=taxdetails&id=558 on 2012-07-31.

Varadaraj, K. and D.M. Skinner, 1994. Denaturants or cosolvents improve the specificity of PCR amplification of a G+C-rich DNA using genetically engineered DNA polymerases. *Gene* 140:1-5.

Velasco, A.M., J.I. Leguina, and A. Lazcano, 2002. Molecular Evolution of the Lysin Biosynthetic Pathways. *J. Mol. Evol.* 55:445-459.

Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, and D. Rusch, J.A. Eisen, D. Wu, I. Paulsen, K.E.Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H.O Rogers, H.O Smith, 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74.

Vetting, M.W., S.S. Hegde, and J.S. Blanchard, 2010. The structure and mechanism of the Mycobacterium tuberculosis cyclodityrosine synthetase. *Nat. Chem. Biol.* 6:797-799.

Vogel, S., 1977. Current-induced flow through living sponges in nature. *Proc. Natl. Acad. Sci. USA* 74:2069-2071.

Yadav, G., R.S. Gokhale, and D. Mohanty, 2003. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* 328(2):335-363.

Yeh, E., S. Garneau, and T. Walsh, 2005. Robust in vitro activity of RebF and RebH, a two component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl. Acad. Sci. USA* 102:3960-3965.

Youssef T.A. and S.L. Mooberry, 2006. Hurghadolide A and swinholide I, potent actin-microfilament disrupters from the Red Sea sponge *Theonella swinhoei*. *J. Nat. Prod.* 69(1):154-157.

Yuzawa, S., W. Kim, and L. Katz, and J.D. Keasling, 2011. Heterologous production of polyketides by modular type I polyketide synthases in *Escherichia coli*. *Curr. Opin. Biotechnol.* 23:1-9.

Walsh, C.T., A.M. Gehring, P.H. Weinreb, L.E. Quadri, R.S. Flugel, 1997. Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr. Opin. Chem. Biol.* 1:309–315.

Walsh, C.T. and W. Zhang, 2011. Chemical logic and enzymatic machinery for biological assembly of peptidyl nucleoside antibiotics. *ACS Chem. Biol.* 6:1000-1007.

Wang, M. And S.J. Gould, 1993. Biosynthesis of capreomycin. 2. Incorporation of L-serine, L-alanine, and L-2,3-diaminopropionic. *J. Org. Chem.* 58(19):5176-5180.

Webster N.S, and R.T. Hill, 2001. The culturable microbial community of the Great Barrier Reef sponge Rhopaloeides odorabile is dominated by an α-Proteobacterium. *Mar. Biol.* (Berl) 138:843–851.

Wenzel, S.C., P. Meiser, T.M. Binz, T. Mahmud, and R. Müller, 2006. Nonribosomal peptide biosynthesis: point mutations and module skipping lead to chemical diversity. *Angew. Chem. Int. Ed. Engl.* 45(14):2296-2301.

Weinig, S., H-J. Hecht, T. Mahmud, and R. Müller, 2003. Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. *Chem. Biol.* 10:939–952.

Weisburg, W.G., S.M. Barns, D.A. Pelletier, and D.J. Lane, 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173(2):697-703.

Weissman, K.J., Hong, H., Oliynyk, M., Siskos, A.P., Leadlay, P.F. (2004) Identification of a phosphopantetheinyl transferase for erythromycin biosynthesis in *Saccharopolyspora erythraea. Chembiochem* **5**, 116–125.

Westenberg, M., S. Bamps, H. Soedling, I.A. Hope and C.T. Dolphin, 2010. Escherichia coli MW005: lambda Red-mediated recombineering and copy-number induction of oriV-equipped constructs in a single host. *BMC Biotechnol.* 10(27):1-8.

Wilkinson, C.J., E.J. Frost, J. Staunton, and P.F. Leadley, 2001. Chain initiation on the soraphen-producing modular polyketide synthase from Sorangium cellulosum. *Chem. Biol.* 8(12):1197-1208.

Witkowski, A., A.K. Joshi, and S. Smith, 2002. Mechanism of the *β*-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. *Biochem.* 41(35):10877-10887.

Witkowski, A., A.K. Joshi, and S. Smith, 2004. Characterization of the *β*-carbon processing reactions of the mammalian cytosolic fatty acid synthase: role of the central core. *Biochem.* 43(32):10458-10466.

Wommack, K.E., J. Bhavsar, and J. Ravel, 2008. Metagenomics: Read Length Matters. *App. Env. Microbiol.* 74(5):1453–1463.

Woyke T, H. Teeling, N.N. Ivanova, M. Huntemann, M. Richter, F.O. Gloeckner, D. Boffelli, L.J. Anderson, K.W. Barry, H.J. Shapiro, E. Szeto, N.C. Kyrpides, M. Mussmann, R. Amann, C. Bergin, C. Ruehland, E.M. Rubin, and Nicole Dubilier, 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950-955.

Whitman, W.B., D.C. Coleman, and W.J. Wiebe, 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci.* USA 95(12):6578-6583.

Wiley, P.F., K. Gerzon, E.H. Flynn, M.V. Sigal Jr., O. Weaver, U.C. Quarck, R.R. Chauvette, and R. Monahan, 1957. Erythromycin. X. Structure of Erythromycin. *J. Am. Chem.* 79(22):6062-6070.

Wild, J., Z. Hradecna, and W. Szybalski, 2002. Conditionally amplifiable BACs: switching from single copy to hig-copy vectors and genomic clones. *Genome Res.* 12:1434-1444.

Wilkinson, C.R. and P. Fay, 1979. Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* 279:527-529.

Wilkinson, C.R., 1978. Microbial associations in sponges I: ecology, physiology and microbial poupulations of coral reef sponges. *Mar. Biol*. 49:161-167.

Wilkinson, C.R., R. Garrone, and J. Vacelet, 1984. Marine sponges discriminate between food bacteria and bacterial symbionts: electron microscope radioautography and *in situ* evidence. *Proc. R. Soc. Lond*. B 220:519-528.

William, P.G., 2009. Panning for gold mining: marine bacteria as a source of new theraupetics. *Trends in Biotechnol*. 27(1):45-52.

Wilson, K., 2001. Preparation of genomic DNA from bacteria. *Curr. Protocols in Mol. Biol.* 2.4.1–2.4.5.

www.coml.org. First census shows life in planet ocean is richer, more connected, more altered than expected.

Xue, Q., G. Ashley, C.R. Hutchinson, and D.V. Santi, 1999. A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci.* USA 96:11740-11745.

Zadražil, S., J. Šatava, and Z. Šormová, 1974. Isolation procedure for bacterial DNA based on gel permeation chromatography on a sepharose column. *J. Chromatography A* 91:451-458.

Zehner, S., A. Kotzsch, B. Bister, R.D. Süssmuth, C. Méndez, J.A. Salas, K-H. van Pée, 2005. A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* 12(4):445–452.

Zhang, Y., S.A. Kang, T. Mukherjee, S. Bale, B.R. Crane, T.P. Begley, and S.F. Falick, 2008. Crystal structure and mechanism of tryptophan 2,3-dioxygenase, a heme enzyme involved in tryptophan catabolism and in quinolinate biosynthesis. *Biochemistry* 46(1):145-155.

Zhang, Y.M., B. Wu, J. Zheng, C.O. Rock, 2003. Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. *J. Biol. Chem.* 278(52):52935-52943.

Zhang, Y.M., J. Hurbert, S.W. White, and C.O. Rock, 2006. Roles of the active site water: histidine 303, and phenylalanine 396 in the catalytic mechanism of the

elongation condensing enzyme of *Streptococcus pneumoniae. J. Biol. Chem.* 281(25):17390-17399.

Zhang, W., I. Ntai, N.L. Kelleher, and C.T. Walsh, 2011. tRNA-dependent peptide bond formation by the transferase PacB in the biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. *J. Am. Chem. Soc.* 133:5240-5243.

Zhou, J., M.A. Bruns, and J.M. Tiedje, 1996. DNA recovery from soils of diverse composition. *App. Env. Microbiol.* 62:316-322.

Zimmermann, U., 1983. Electrofusion of cells: Principles and industrial potential. *Trends in Biotechnol.* 1(5):149-155.

ZoBell, C.E. and C.H. Oppenheimer, 1950. Some effects of hydrostatic pressure on the multiplication and morphology of marine bacteria. *J. Bacteriol*. 60(6):771-781.

ZoBell, C.E. and F.H. Johnson, 1949. The influence of hydrostatic pressure on the growth and viability of terrestrial and marine bacteria. *J. Bacteriol*. 57, 179-189.

ZoBell, C.E and R.Y. Morita, 1957. Barophilic bacteria in some deep sea sediments. *J. Bacteriol.* 73(4):563-568.

Curriculum Vitae

Agustinus Robert Uria was born on the 21st of July 1976 at Sienjo in Central Sulawesi, Indonesia. He attended basic education (elementary, junior, and senior high schools) in Central Sulawesi during 1983-1994. Then he pursued undergraduate studies at Sam Ratulangi University (UNSRAT) from 1994 to 1999. At the end of his study, he carried out a thesis research in Marine Biotechnology under the supervision of Dr. F.G. Ijong and Dr. I.F.M. Rumengan. While working on his thesis, he assisted Dr. Rumengan in teaching some undergraduate courses. He obtained a scholar degree in Marine Science on the 17th of July 1999.

On the recommendation of Dr. Rumengan, he joined the Research Group of Prof. Dr. Maggy Thenawidjaja Suhartono at Bogor Agricultural University (IPB) from August 1999 to March 2001. Upon joining the group, he initially learned enzyme purification and characterization from Ms. Budiasih Wahyuntari (a doctoral student) for around five months and then performed research on screening of protease-producing thermophilic bacteria and enzyme characterization under the supervision of Prof. M.T. Suhartono. While doing research, he attended a master program in Microbiology at IPB. After one semester of the program (September 2000-February 2001), he took a leave and joined the lab of Dr. Rumengan at UNSRAT, working as a research and teaching assistant from March 2001 to August 2002.

In 2002, he was awarded a Netherlands Fellowship Programme (NFP) scholarship to pursue a master study at Wageningen University and Research Center (WUR), the Netherlands during 2002-2004. At the end of his study, he did a master`s thesis in Microbiology (July 2003-January 2004) in the Research Group of Prof. Dr. John van der Oost. He obtained a master degree in Biotechnology from WUR on the 20th of February 2004. He then returned to Indonesia, working as a temporary lecturer at Sam Ratulangi University, where he taught subjects related to Marine Biotechnology in undergraduate classes from March to October 2004. In October 2004, he joined a research center in Jakarta (Research Center for Marine and Fisheries Product Processing and Biotechnology), where he worked with Dr. E. Chasanah and Mrs. N. Fawzya to isolate marine sponge-associated bacteria and study their hydrolytic enzymes.

On February 2007, he was elected to a Deutscher Akademischer Austausch Dienst (DAAD) scholarship to conduct doctoral research at the University of Bonn under the supervision of Prof. Dr. Jörn Piel. After German language courses in Jakarta and Marburg (April-September 2007), he started a PhD project in the Piel Group focused on natural product biosynthesis in uncultivated symbiotic bacteria. Since October 2011, he has been employed in the Piel Group as a “wissenschaftlicher Mitarbeiter”. During his PhD work, he was given opportunities by his supervisor to present his research progress in some conferences/meetings/workshops as well as to supervise bachelor/master students in internships/thesis.

Selected Publications:

- Freeman, M.F., C. Gurgui, M. Helf, B.I. Morinaka, A.R. Uria, N.J. Oldham, H-G. Sahl, S. Matsunaga, and J. Piel, 2012. Metagenome mining reveals ribosomal peptide biosynthesis of unprecedented structural scope. ***SCIENCE*** *DOI:10.1126/science.1226121.*
- Uria, A.R. and J Piel, 2009. Cultivation-independent approaches to investigate the chemistry of marine symbiotic bacteria. ***Phytochemistry Rev*****.** 8:401–414.
- Machielsen R., A.R. Uria, S.W.M. Kengen and J. van der Oost, 2006. Production and Characterization of a Thermostable Alcohol Dehydrogenase that belongs to Aldo-Keto Reductase Superfamily. ***Appl. Environ. Microbiol.*** 72(1):233-238.

Patent:

J. Piel, C. Gurgui, M.F. Freeman, A.R. Uria, and M. Helf, 2011. **European Patent Application** # 11180107.2., filed by Rheinische Friedrich-Wilhelms-Universität Bonn through PROvendis GmbH titled "Biosynthetic gene cluster for the production of peptide/protein analogs."

Degree Theses:

- Uria, A.R., 2012. Investigating natural product biosynthesis in uncultivated symbiotic bacteria of the marine sponge *Theonella swinhoei*. *PhD Dissertation in Microbiology*, University of Bonn. Supervisor: Prof. Dr. Jörn Piel.
- Uria, A.R., 2004. Purification and characterization of a NAD(P)-dependent alcohol dehydrogenase from the marine hyperthermophilic archaebacterium, *Pyrococcus furiosus* and generation of a DNA construct for its overexpression in the deep-sea hyperthermophile, *P. abysii*. *Master Thesis in Microbiology*, Wageningen University. Supervisors: Mr. Ronnie Machielsen (a PhD student) and Prof. Dr. John van der Oost
- Uria, A.R., 1999. Biochemical characteristics of thermophilic bacteria from Likupang coastal waters in North Sulawesi. *Scholar Thesis in Marine Biotechnology*, Sam Ratulangi University. Supervisors: Dr. Frans G. Ijong and Dr. Inneke F.M. Rumengan.

Selected Oral Conference Presentations (and Conference Proceedings):

- Analyzing the drug potential of uncultivated symbionts of the sponge *Theonella swinhoei. BMBF GenBioCom cluster-meeting*, on 21-22 Oct 2010 in Freudenstadt-Lauterbad, Schwarzwald, Germany.
- Insights into the Secondary Metabolism of an Unusual Uncultured Symbiont of Marine Sponges. *Die 39. Doktorandenworkshop, Naturstoffe: Chemie, Biologie und Ökologie*. Held on the 16th April of 2010 in Jena, Germany.
- Targeting Gene Sequences Encoding the Biosynthesis of Bioactive Metabolites from the Metagenomes of Marine Sponges. *Symposium of the World Ocean Conference* on 11-15 Mei 2009 in Manado, Indonesia.
- Alcohol dehydrogenases from marine hyperthermophilic microorganisms and their importance to the pharmaceutical industry. *International Seminar and Workshop on Marine Biodiversity & their Potential for Developing Bio-Pharmaceutical Industry,* on the 17-18th of May 2006 at Jakarta. Full paper was published in the conference proceeding (authors: Uria, A.R**.**, R. Machielsen, B.E. Dutilh, M.A. Huynen, and J. van der Oost, 2006).
- Properties of chitinase from a bacterial strain associated with marine sponge*. ASEAN Biochemistry Seminar and Workshop on Enzymes: Industrial and Medical Prospects,* on Feb. 6-7, Airlangga University. Published in the conference proceedings (authors: Uria, A.R. and E. Chasanah, 2006).
- Chitinase and chitosanase from microorganisms associated with marine sponges. *The 9th ASEAN Food Conference entitled: "Emerging Science and Technology in the Development of Food Industry in the ASEAN"* on the 8-10th of August 2005 in Jakarta. Published in the conference proceeding (authors: Uria, A.R. and E. Chasanah, 2005).
- Bioinformatic Analysis of NADP-dependent Alcohol Dehydrogenase from Marine Hyperthermophilic Bacteria. *Seminar and Workshop on Bioinformatics* held on the 17th March of 2005 at Research Center for Biotechnology, Indonesian Institute of Sciences (LIPI), Cibinong Science Center, Bogor. Uria, A.R., R. Machielsen, and J. van der Oost.

Selected Poster Conference Presentations (and Conference Proceedings):

- Uria, A. R., C. Rückert, J. Kalinowksi, S. Matsunaga, and J. Piel, 2011. Revealing the biosynthetic capacity of uncultured symbiotic bacteria. *International VAAM Workshop* on 28-30 September 2011 in Bonn, Germany.
- Uria, A. R., S. Matsunaga, J. Piel, 2010. Biosynthesis of the antitumor polyketide misakinolide A by an uncultivated symbiont of the marine sponge *Theonella swinhoei*. *VAAM symposium on symbiotic interaction*, on 7-8 Oct 2010 at Würzburg University.
- Uria, A. R., S. Matsunaga, and J. Piel, 2010. Insight into the Secondary Metabolism of *Entotheonella* sp., an Uncultivated Symbiont of Marine Sponges. *Die 22. Irseer Naturstofftage, Aktuelle Entwicklungen in der Naturstoff-Forschung*, on 22 -24 Februar 2010 in Irsee, Germany.
- Dewi, A.S., K. Tarman, A.R. Uria, 2008. Marine Natural Products: Prospects and Impacts on the Sustainable Development in Indonesia. *The Indonesian Student Scientific Meeting (ISSM) on Sustainable development in Indonesia*: *An interdisciplinary approach*. Held on 13–15 May 2008 at Delft University of Technology, Delft, The Netherlands. Published in the conference proceedings.

Student Supervision:

During PhD study, he was given responsibilities by Prof. Piel to supervise the internships of the following students:

- Ms.Samantha Keyser (an undergraduate student in Chemistry at Harvard University), 2010 (1.5 months). Construction of a fosmid library from the bacterium *Serratia ordorifera* and library screening for isolating a *trans*-AT gene cluster.
- Ms. Christina Poppel (a bachelor student in Chemistry at University of Bonn), 2010 (1.5 months). Generation and screening of a fosmid library from the bacterium *Dickeya dadantii* for isolating a *trans*-AT gene cluster.
- Ms. Silke Reiter (a master student in Molecular Biomedicine at University of Bonn), 2010 (1.5 months). Genetic screening of a metagenomic library for finding the remaining part of a Misakinolide A biosynthetic gene cluster.
- Mr. Birkam Pandey (a master student in Molecular Biotechnology at University of Bonn), 2011 (1.5 months). Filling in the gaps in the PKS/NRPS gene cluster and scaffolds.
- Mr. Mark Schultz (a master student in Molecular Biotechnology at University of Bonn), 2012 (2 weeks). Optimizing the PCR conditions for isolating the remaining parts of a partial gene cluster coding for the biosynthesis of keramamide from a sponge metagenomic library.

Before PhD study:

- In 2001: co-supervising 6 undergraduate students for internships in Microbiology at the lab of Dr. I.F.M. Rumengan, Sam Ratulangi University.
- In 2006: supervising 2 undergraduate students for a 2-month internship in Microbiology at Research Center for Marine and Fisheries Product Processing and Biotechnology, Jakarta.

Teaching Experience:

- During 2001-2002: teaching some subjects related to Biotechnology (such as Biochemistry, Genetics, and Genetic Engineering) at Faculty of Fisheries and Marine Science, Sam Ratulangi University. He worked as a teaching assistant of Dr. I.F.M. Rumengan.
- In 2004: teaching some subjects related to Marine Biotechnology at Faculty of Fisheries and Marine Science, Sam Ratulangi University.

Scholarships/award:

- Deutscher Akademischer Austausch Dienst (DAAD) scholarship from Juni 2007 to September 2011 for a doctoral study at University of Bonn, Germany.
- Netherlands Fellowship Programme (NFP) from Aug 2002 to Feb 2004 for a master study at Wageningen University and Research Centre, The Netherlands.
- National Scientific Event (PIMNAS) award for the category Scientific Writing in Natural Sciences. Held at IKIP Negeri Jakarta (Jakarta State University) in January 1999. Organized by General Directorate of Higher Education (DIKTI), Ministry of National Education of Indonesian Republic.

Selbständigkeitserklärung

Hiermit erkläre ich, die eingereichte Arbeit selbstständig verfasst und keine anderen Hilfsmittel und Quellen als die angegebenen benutzt zu haben. Diese Arbeit ist weder identisch noch teilidentisch mit einer Arbeit, die an der Rheinischen Friedrich-Wilhelms-Universität Bonn oder einer anderen Hochschule zur Erlangung eines akademischen Grades oder als Prüfungsleistung vorgelegt worden ist. Die Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn ist mir bekannt.

Agustinus Robert Uria