Investigation of polyketide biosynthetic pathways in the sponge *Theonella swinhoei* and the beetle *Paederus fuscipes*

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Tu Anh Nguyen

aus

Ha Noi, Viet Nam

Bonn (2009)

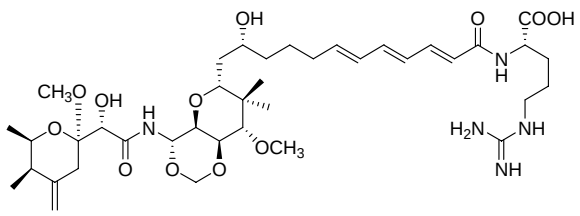
Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Gutachter:
1. Prof. Dr. Jörn Piel
2. Prof. Dr. Gabriele M. König
3. Prof. Dr. Christa E. Müller
4. Prof. Dr. Hans-Georg Sahl

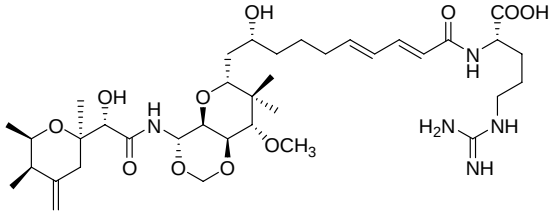
Tag der Promotion: 7. 12. 2009

Abstract

Onnamides are natural products that possess strong antitumor activity and were isolated from the marine sponge *Theonella swinhoei*. Previous data suggested that an as-yet unculturable symbiotic bacterium living in association with the sponge is the actual onnamide producer. In order to gain insight into the biosynthetic pathway of onnamides in nature, their biosynthetic genes had to be identified. Onnamides are encoded by the onnamide (*onn*) gene cluster, which contains giant *trans*-AT polyketide synthase (PKS) and nonribosomal peptide synthase (NRPS) genes. PKSs and NRPSs are constituted of modules, each module consisting of domains. In *trans*-AT PKSs, the acyltransferase (AT) domains are present as discrete genomic regions, outside of the multimodular enzymatic core.



Onnamide A



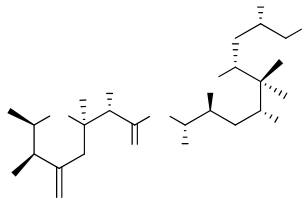
Onnamide B

For the purpose of covering the entire gene cluster, *onn* genes were cloned from the metagenomic sponge DNA consisting of DNA of the sponge and of its complex symbiotic community. A portion of the *onn* cluster was previously identified by screening a fosmid library constructed from the metagenomic DNA but was found to encode a truncated PKS. The present study deals with the isolation of the remaining *onn* genes from the 400,000 clone library of the sponge *T. swinhoei*. Based on the genomic region isolated previously, an iterative primer walking strategy was applied, which enabled cloning of the remaining regions of the *onn* gene cluster. According to a recently developed method for the prediction of substrate specificities of *trans*-AT PKSs, the isolated gene cluster is responsible for synthesis of onnamide B. Isolation of the entire cluster was hampered by the finding that the metagenomic DNA contained numerous variants of the onnamide PKS. We hypothesize that these PKS variants belong to different symbiont strains that jointly generate the structurally diverse library of onnamides present in *T. swinhoei*. Based on the phylogenetic analysis of *ftsZ*, a non-PKS taxonomic marker gene located immediately downstream of the *onn* genes,

evidence was obtained that the onnamide producer is a member of the bacterial phylum *Chloroflexi*. In addition to these studies, the adenylation (A) domains of two NRPS modules of the *onn* gene cluster were expressed to provide the basis for functional studies on onnamide biosynthesis.

The onnamide genes represent the first biosynthetic gene cluster that as isolated from a sponge. With the *onn* genes in hand, not only fundamental insights into the biosynthetic pathway of onnamides can be obtained but also strategies can be developed for the sustainable production of rare marine drug candidates by microbial heterologous expression systems.

Previous studies by Piel *et al.* in 2004 provided evidence for the presence of another *trans*-AT PKS cluster in the total DNA of *T. swinhoei*. This hypothesis was corroborated by a more detailed bioinformatic analysis conducted in the present study on short PCR products amplified from the sponge metagenome. To isolate the cluster, specific primers were used to screen again the fosmid library of the sponge. From this screening, two positive fosmids were identified and completely sequenced. Analysis of this sequence verified that the newly isolated genes belong to an as-yet uncharacterized polyketide. According to the analysis, some structural features of this polyketide were predicted. However, frame-shifts present in several genes suggested that the pathway is not functional and likely represents an evolutionary relic. Nevertheless, the results confirmed that PKS functions can be predicted even when starting from a short partial sequence amplified by PCR.



Pederin

pPD7E4 is a cosmid that was previously isolated from the genome of an uncultivated *Pseudomonas* symbiont of the beetle *Paederus fuscipes*. The complete sequence showed that this cosmid contained a small gene cluster with both PKS and non-PKS genes. In particular, one of the non-PKS genes was found to be a halogenase gene. This finding suggests that the gene cluster from pPD7E4 does not belong to the known *ped* genes responsible for pederin biosynthesis, but rather encodes biosynthesis of a halogenated compound. In nature, organohalogens often possess bioactivities.

Therefore, in an attempt to identify the compound, the cosmid pPD7E4 was heterologously expressed applying the homologous recombination. After several genetic modifications, the heterologous expression of pPD7E4 inside *Pseudomonas putida* KT2440 was successful at the RNA level, setting the stage for the characterization of the unknown compound.

Contents

List of Figures

List of Tables

From deep in my heart, this thesis is dedicated to my father now in heaven who passed away when I was in Germany….

Acknowledgements

I wish to express my deepest gratitude to my supervisor, Prof. Dr. Jörn Piel for his valuable guidance, support and encouragement during my study. It has been a great opportunity to learn from his knowledge and experience.

I also would like to express my sincere gratitude to Prof. Dr. Gabriele M. König for her refereeing and evaluation of my research.

I would like to display my appreciation for the committee members, Prof. Dr. Christa E. Müller and Prof. Dr. Hans-Georg Sahl, for their critical reading and helpful contributions.

My special thanks go to Dr. Cristian Gurgui for his invaluable instruction for my experiments; for his endless kindness, willingness, for sharing difficulties not only in my work but also in daily life in Germany and for proof-reading my thesis. I am grateful to Dr. Sonia van der Sar for proof-reading my thesis.

I am very thankful to my colleagues in the laboratory such as Kathrin Reinhardt - the first “Bonner” I met - who picked me up at Bonn train station; and the others, Dr. Katja Maria Fisch, Dr. Jing He, Dr. Katrin Zimmermann, Marija Avramović, Holger Niederkrueger, Daniel Flachshaar, Thomas Hochmuth, Jana Moldenhauer, Minna Eklund, Agus Robert Uria,….They have been supportive and helpful. Their friendship made my life in Bonn during these three years easier and memorable.

I would like to thank the chiefs’ School of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Viet Nam, especially Prof. Dr. Le Quan Nghiem - President of School of Pharmacy, for the opportunity of applying for this scholarship and also their spiritual support. My special thanks go to Prof. Dr. Nguyen Van Thanh for his help and encouragement. I would also like to thank my colleagues in the Department of Microbiology, at the School of Pharmacy.

I am especially grateful to my mother who has helped me look after my son, so that I felt secure while studying abroad. She is the one person that has encouraged me the most during my studies. I

must thank my little son, *Nguyen Anh Duy*, for putting up with his mother not being by his side as he grew up. I take deep pleasure in thanking my sister's family and my husband for all that they have done for me. My family is my real spiritual encouragement.

I am thankful to my friends: Loan, Thanh, Thuy, Tam, etc. in Viet Nam and Khanh, Tuan, Thong, etc. in Germany. They are valuable friends.

Finally, I would like to show my sincere gratitude to the Viet Nam Ministry of Education and the Training and the German Academic Exchange Service (DAAD) for awarding me a scholarship which has enabled me to carry out my studies in Germany.

Chapter 1
Introduction

1.1 Natural products

Natural products are compounds produced by living organisms in nature. They usually have diverse chemical structures which lead to a diverse range of bioactivities; therefore, they have been the most important candidates in the development of new drugs.[1-3] Since 1981, natural products have occupied more than 40% of the new chemical entities and surprisingly, about 70% of antibiotics and 60% of anticancer drugs that are used in medical treatment today are natural products or of natural origin.[4] To date, approximately one third of the best selling drugs originate from natural sources or have been developed based on leading structures provided by nature.[5]

Marine natural products

There are three major sources of natural products: terrestrial plants and animals, microorganisms and marine macro-organisms. Currently, among these three main sources, the marine sources - particularly the invertebrates - have provided numerous promising compounds.[5, 6] Furthermore, the ocean covers more than 70% of the earth' surface[7] and contains approximately 80% of world's plant and animal species, but up till now, only about 10% of these species have been explored with respect to their biological activity.[8, 9] Due to their pharmacological potential, by 1975 there were already three field studies on marine natural product chemistry including marine toxins, marine

biomedicinals and marine chemical ecology.[1] Marine natural product chemistry has developed rapidly in the last decades and has resulted in the isolation of about 14,000 metabolites.[5] Several of these metabolites possess unique structures that are not found in terrestrial organisms and have become a rich source for screening pharmaceutical products.[10-15]

In contrast to terrestrial macro-organisms, where plants are the major resource of natural products, the leading role in the marine environment is carried out by invertebrates such as sponges, cnidarians, red algae, echinoderms, with the most prolific source of marine natural products being sponges.[16, 17]

Sponges are animals of the phylum Porifera. They are classified into three classes according to skeletal type, Calcareous sponges - *Calcarea*, Demosponges - *Demospongiae*, and Glass sponges - *Hexactinellida* (Figure 1). There are about 8,000 currently described sponge species.[16] The class Calcarea is entirely marine and not common. Demospongiae is the largest class involving 90% of modern sponges, with species proliferating in marine, brackish and freshwater at all depths. The class Hexactinellida is also entirely marine and mostly exists in deep-water.[18] A sponge can be a competitor, symbiont, a host of a symbiont, consumer or prey.[19]

Calcarea class Demospongiae class Hexactinellida class

Figure 1. Sponge morphology. (BIOADAC)

Numerous natural products from sponges or their analogs have been pharmacologically assayed *in-vitro* or *in-vivo*,[11, 20-23] and several of them are currently tested for preclinical and clinical evaluation or have entered the pharmaceutical market (Table 1). The mechanism of action of these compounds are varied *e.g.* targeting ion channel, enzyme inhibition (such as methionine

aminopeptidase, protein kinase or phospholipase A_2); interfering with microtubule formation; interacting with DNA; inducing oxidative stress and stimulation of the immune response.[13, 24, 25]

Table 1. Some natural products of sponges on market or clinical trial.[7, 13, 26-29]

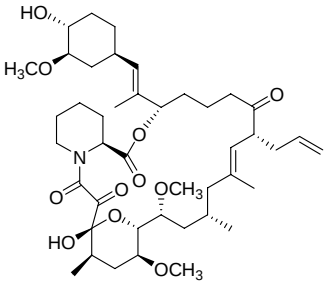
Bioactivity	**Compound**	**Source**	**Status**
Antivirus	Analogs of spongeothymidine, spongeouridine	*Cryptotethia crypta*	Market (Ara A)
Inflammation, asthma	IPL512,602	*Petrosia contignata*	Phase II
	Topsentins	*Topsentia genitrix*, *Hexadella* sp., and *Spongosorites ruetzleri*	Preclinical
Anticancer	KRN7000	*Agelas mauritianus*	Phase I
	LAF389	*Jaspidae* sp	Phase I
	Dictyodendrins	*Dictyodendrilla verongiformis*	Preclinical
	Hemiasterlins	*Auletta* sp., *Siphonochalin* sp.	Preclinical
	Lasonolides	*Forcepia* sp.	Preclinical
	Laulimalide	*Cacospongeia mycofijiensis*	Preclinical
	Peloruside A	*Mycale hentscheli*	Preclinical
	Salicylihalimides A and B	*Haliclona* sp.	Preclinical
Anti-Alzheimer	Debromohymenialdisine (DBH)	*Stylotella aurantium*	Phase I
Antifungal	Bengazole, bengamide	*Pachastrissa* sp.	Preclinical
Antibacterial	Psammaplin A	*Aplysinella rhax*	Preclinical
Anti-malaria, tuberculosis, HIV	Manzamine A	*Haliclona* sp.	Preclinical

Researches on natural products can be based on the sources or the chemical structures of compounds or based on both. Presently, along with investigations into sponges, several studies have been focused on substances belonging to the polyketide and nonribosomal peptide groups due to their striking diversity in structures as well as pharmacological properties.

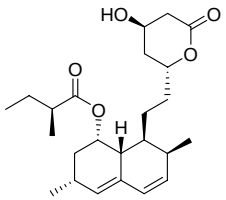
1.2 Polyketides and nonribosomal peptides

Polyketides (PKs) are compounds assembled from short acid precursors; whereas nonribosomal peptides (NRPs) are built from amino acids.[30] PKs and NRPs can be found in a wide range of

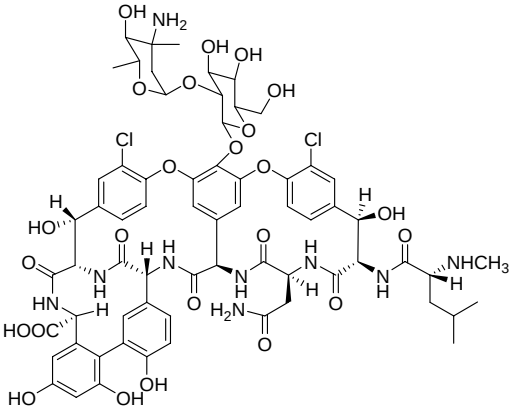
organisms. These include bacteria, fungi, plants, dinoflagellates, mollusks, sponges and insects.[31, 32] They possess a broad spectrum of valuable bioactivities such as antimicrobial, antifungal, antiparasitic, antitumor, cytostatic, anticholesterolemic and agrochemical.[33] Not surprisingly, they also fulfill significant clinical, economical and industrial roles in the pharmaceutical industry. Among them are many well-known PKs and NRPs that are in commercial use. These are the PKs such as the macrolide antibiotics erythromycin and azithromycin, the polyene antibiotics amphotericin, the tetracycline family of antibiotics, the immunosuppressant FK506, the anticholesterol agent lovastatin, the anticancer doxorubicin; and the NRPs such as the antibiotics vancomycin and gramicidin and the cytostatic bleomycin and the immunosuppressant cyclosporin A (Figure 2).[34-36]



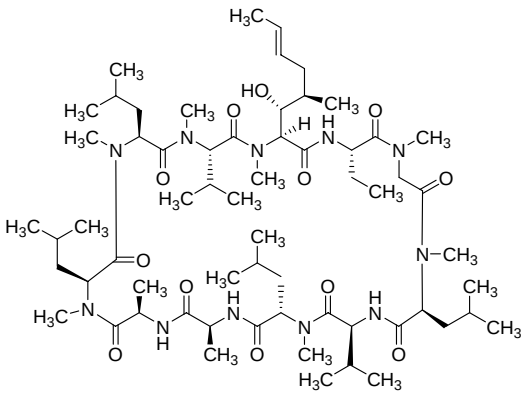
(1)



(2)



(3)



(4)

Figure 2. The polyketides (1) FK506 from the bacterium *Streptomyces tsukubaensis* (2) lovastatin from the fungus *Aspergillus terreus* and the nonribosomal peptides (3) vancomycin from the bacterium *Amycolatopsis orientalis*, and (4) cyclosporin A from *Tolypocladium* spp.

PKs and NRPs are synthesized by sequential reactions catalyzed by one or more specialized enzymes called polyketide synthases (PKSs) or nonribosomal peptide synthases (NRPSs), respectively.

1.3 Polyketide synthases

There are several types of PKSs. In this section, only the structure and function of the PKSs that synthesize bacterial complex PKs, also referred to as type I PKSs, are explained in more detail. PKSs are giant multimodular, multifunctional enzymes that contain coordinated groups of active sites.[32] The number and the order of biosynthetic modules as well as the type of catalytic domains present in each module are responsible for varied structures of PKSs.

1.3.1 PKS structure

Each PKS module is typically organized into several domains with distinct functions. In a textbook PKS, the minimum set of domains of the elongation module include a ketoacyl synthase or ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) which are separated by short spacer regions (Figure 3).[37] Optionally, the polyketide structure can be modified by additional domains such as a keto reductase (KR) which reduces keto groups to hydroxy groups, a dehydratase (DH) is responsible for removal of H_2O, and an enoyl reductase (ER) that reduces double bonds to single bonds (see below). Every module is usually responsible for the incorporation of individual building blocks into the growing polyketide chain by attachment of malonyl-CoA or a small derivative.[38]

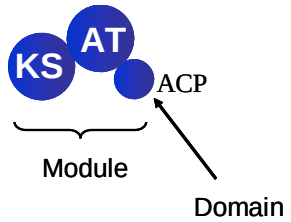


Figure 3. Domains and minimal module of PKSs.

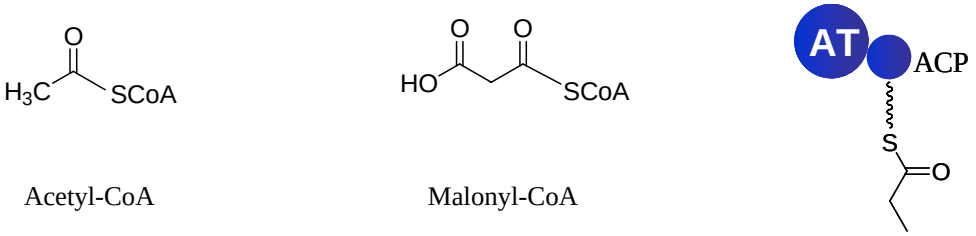
From N-terminus to C-terminus, the order of modules and domains of a complete PKS are usually as follows:

- Initiation or starting module: usually AT-ACP- or KS-AT-ACP-
- Elongation or extending module: -KS-AT-[KR-DH-ER]-ACP-
- Termination or releasing module: -TE

1.3.2 Polyketide synthesis

The polyketide synthesis process consists of three stages.[37]

a) In the initiation stage, the carboxy group of the starter unit, usually acetyl-CoA or propionyl-CoA, is bound to a 5'-phosphopantetheinyl SH group of the ACP (Figure 4).



Acetyl-CoA Malonyl-CoA

(a) (b) (c)

Figure 4. (a) The starter unit (b) the extender unit (c) the initiation stage of polyketide synthesis.

b) Elongation stage: the polyketide chain passes from the ACP to the KS of the following module, is elongated with one unit and optionally further modified depending on the domains present (Figure 5). The PKS chain is thus transferred from module to module until its final structure is achieved. The elongation group, usually malonyl-CoA or its derivatives is loaded onto the current ACP catalyzed by the AT.

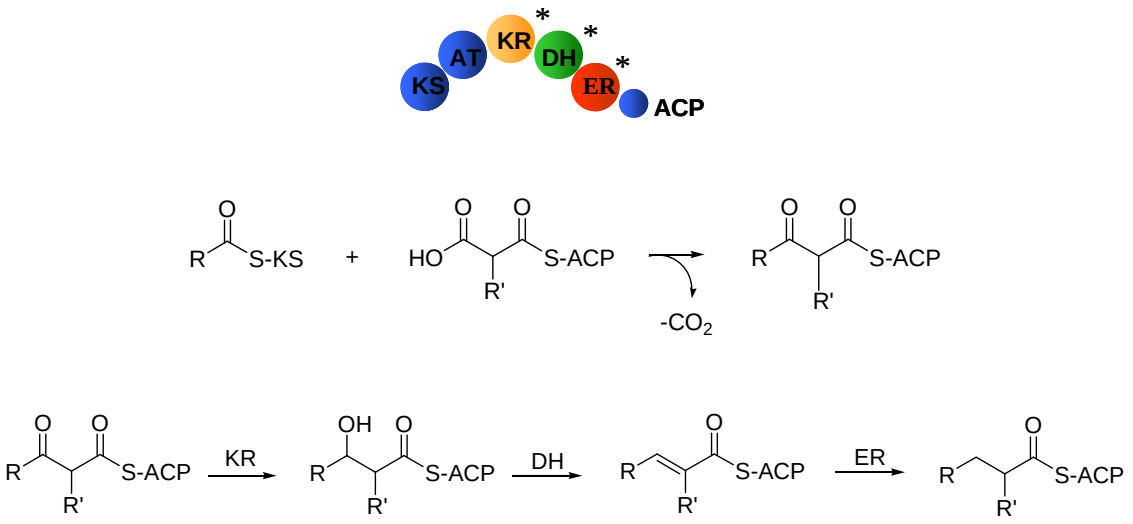


Figure 5. The elongation stage of polyketide synthesis. The domains with asterisk are optional domains.

c) Termination stage: the TE domain hydrolyzes the completed PKS chain from the ACP of the previous module (Figure 6).

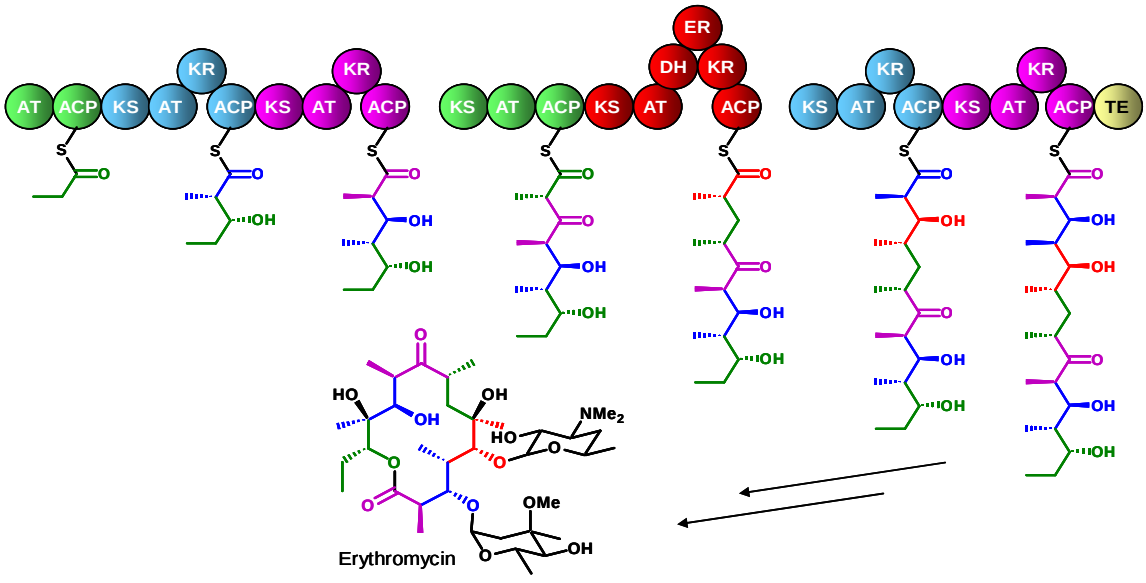


Figure 6. The biosynthetic pathway of erythromycin.

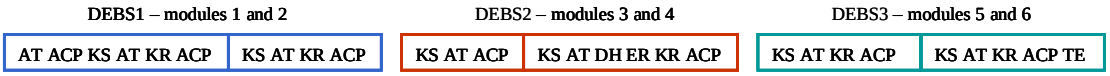
1.3.3 PKS classification

1.3.3.1 Types of PKSs

Based on the molecule architectures of known PKSs, PKSs can be categorized into three types - type I, type II and type III PKSs.[39, 40]

As described above, type I PKSs are composed of one or more very large multifunctional proteins which are constituted by multidomains; for example the deoxyerythronolide B synthase (DEBS) involved in biosynthesis of the antibiotic erythromycin (Figure 7).[30]

(a)



(b)

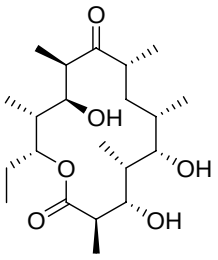


Figure 7. (a) Deoxyerythronolide B synthase and (b) Deoxyerythronolide B.

Type II PKSs consist of several monofunctional proteins. An example is the actinorhodin synthase (Figure 8).[41]

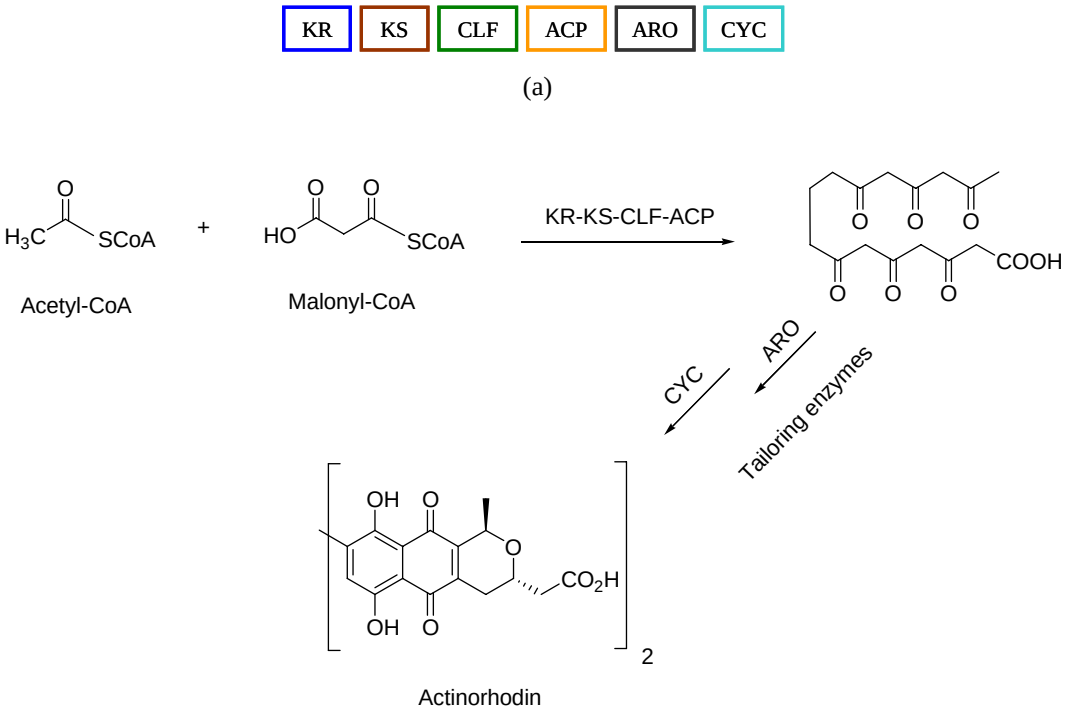


Figure 8. (a) Actinorhodin biosynthesis proteins. (CLF) chain-limiting factor, (ARO) aromatase, (CYC) cyclase and (b) actinorhodin synthetic pathway.

For type III PKSs, the most well-known members of this type are the enzymes synthesizing chalcones (Figure 9) and stilbenes in plants or polyhydroxy compounds in bacteria.[42] Different from type I and type II, these proteins consist of only one functional unit that does not have an ACP domain. Therefore, they use acyl-CoA directly as substrates.[40]

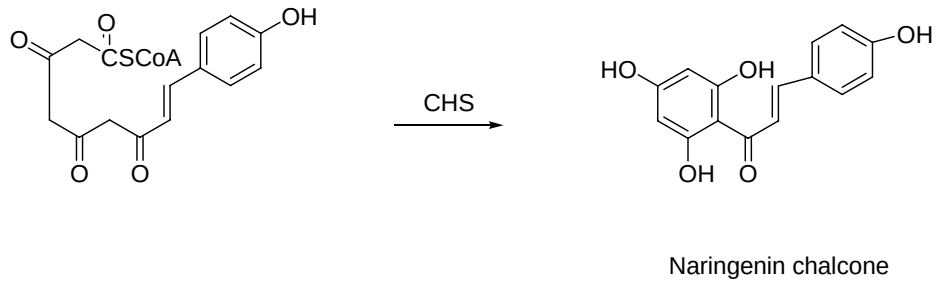


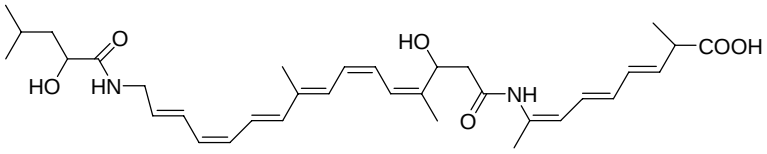
Figure 9. Chalcone synthetic pathway. (CHS) chalcone synthase.

1.3.3.2 Hybrid PKS-NRPS

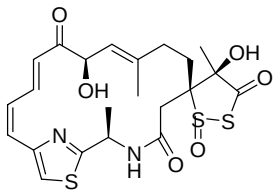
Mixed type I PKS-NRPS systems can also occur. Depending on the direction of the growing chain, the order can be PKS-NRPS, in which the polyketide intermediates are extended by amino acid extender units, or vice versa NRPS-PKS, in which the peptide intermediates are extended by carboxylic acid extender units.[40] Examples for known hybrid PKs-NRPs are chivosazols, jamaicamides, onnamides, pederin, etc.

1.3.3.3 *Cis* and *trans*-AT PKSs

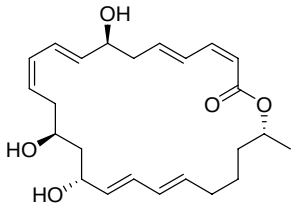
As described in section 1.3.1, the modules of a PKS chain normally contain AT domains that exist in each module. This type of PKS is called *cis*-AT PKS. Recently, an architecturally unusual type of PKS that are smaller than regular PKSs has been reported.[43] In the latter group, the modules are devoid of AT domains and are thus termed *trans*-AT PKS. Instead, ATs are encoded by discrete genes and responsible for loading the extender unit malonyl-CoA to the AT-less modules.[44] In contrast to *cis*-AT PKSs, where the polyketide structure can be directly predicted from the domain architecture, the resulting polyketide of *trans*-AT PKSs cannot be predicted by this colinearity rule, this owing to the presence of unusual domain sets, seemingly redundant modules and novel domain types.[45, 46] The biosynthetic mode of *trans*-AT PKSs is therefore significantly different from those of the *cis*-AT PKS group.[43, 44] Phylogenetic data suggest that *trans*-AT PKSs have evolved independently from *cis*-AT PKSs.[35] *Trans*-AT PKSs have been found in many bacterial groups, from Gram-negative α and δ-proteobacteria, to Gram-positive bacteria with low and high GC content (Figure 10).[47]



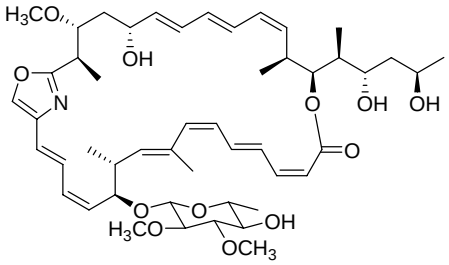
Bacillaene from *Bacillus* spp.



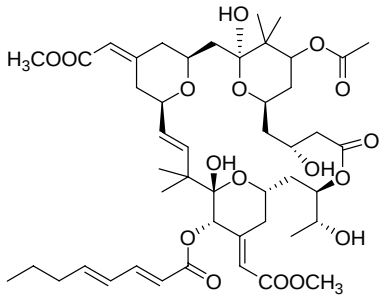
Leinamycin
from *Streptomyces atroolivaceus*



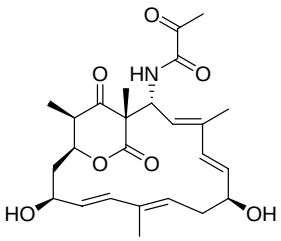
Macrolactin
from *Bacillus amyloliquefaciens*



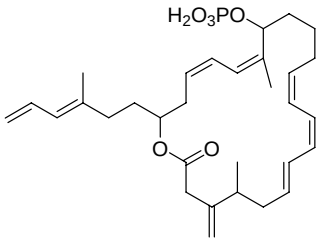
Chivosazol A
from *Sorangium cellulosum*



Bryostatin 1
from the symbiont of *Bugula neritina*



Lankacidin C
from *Streptomyces rochei*



Difficidin
from *Bacillus amyloliquefaciens*

Figure 10. Some polyketides generated by *trans*-AT PKSs.

1.4 Nonribosomal peptide synthetases

NRPs are peptides that are not synthesized by ribosomes but by modular NRPSs.[48] Each NRPS module is responsible for the introduction of one amino acid into the growing peptide chain and is separated by short spacer regions of about 15 amino acids. The elongation module is minimally comprised of condensation (C), adenylation (A), and thiolation (T) domains, the latter also being called peptidyl carrier protein (PCP) domain.[48] Optionally, further domains such as formylation

(F), cyclization (Cy), oxidation (Ox), reduction (Red) and epimerization (E) domains can be present. The basic architecture of an NRPS is analogous to that of PKS, which is made up of a series of initiation, elongation and termination modules.[48-50]

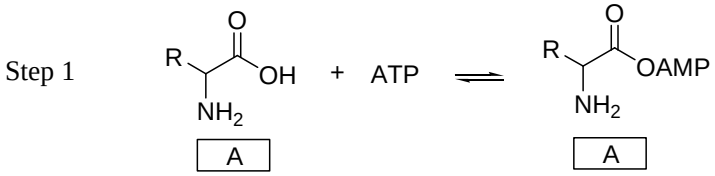
Similar to complex polyketide formation, nonribosomal peptide synthesis also involves three stages.[48]

a) Initiation stage: the first amino acid is activated by the A domain and is loaded onto the PCP. In each following module, the specific amino acid is loaded onto the corresponding PCP. This cycle is repeated depending on the NRPS structure.

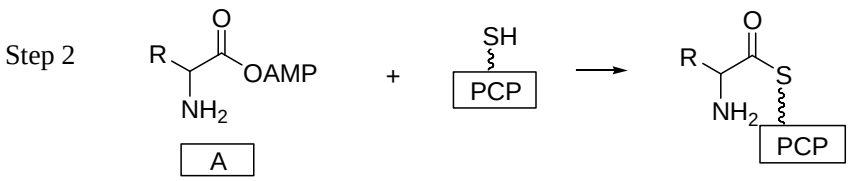
b) Elongation stage: the C domain catalyzes the formation of amide bond between the previous module with amino group of the current module. The peptide chain is thus elongated.

c) Termination stage: the TE domain hydrolyzes the completed polypeptide chain from the PCP of the previous module.

Because the A domain is responsible for recruiting specific amino acids for the growing peptide chain, it is referred to as the "gate keeper" unit of every module of an NRPS.[51] The A domain fulfills its duty through a two-step process. In the first step, the A domain specifically recognizes one amino acid from the pool of available substrates and activates given amino acids through the formation of a corresponding aminoacyl adenylate using ATP.[36, 52-54]



The activated aminoacyl moiety is then covalently transferred to the sulfhydryl group of a phosphopantetheine cofactor located on the PCP domain of the same module.



The A domain of NRPSs is comprised of about 500 amino acids. Most of its highly conserved sequences, A1 to A10 (Table 2 and Figure 11), are found to surround the active site where the substrate binds.[36]

Table 2. Conserved motifs of the A domain (x is a variable amino acid).[36]

Core	Conserved sequence	Core	Conserved sequence
A1	LTYxEL	A6	GELxIxGxGLARGYW
A2	LKAGxAYVPID	A7	YKTGDQ
A3	LAYxxYTSGTTGxPKG	A8	GRxDxQVKIRGx-RVELEEVE
A4	FDxS	A9	LPxYMIP
A5	NxYGPTE	A10	NGKIDR

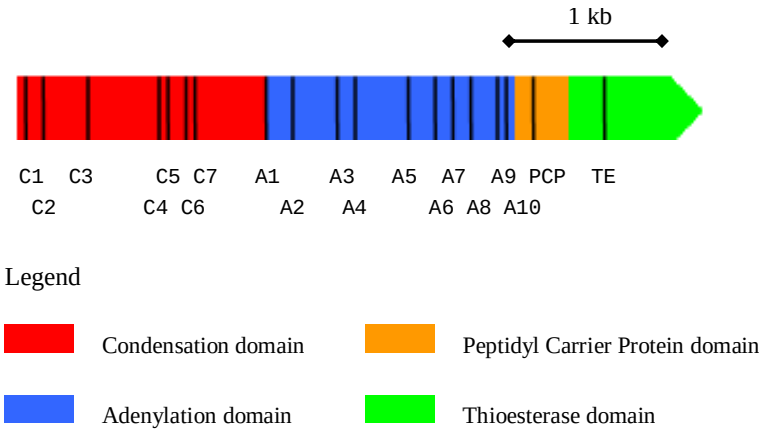


Figure 11. Organization of NRPS modules (source: D. Schwarzer *et al.*, Marburg).[36]

Around 100 amino acids that lie between the core motifs A4 and A5 form a pocket that accommodates the sidechain of the possible substrates. In this pocket, ten residues are identified to be crucial for substrate recognition, forming a binding pocket where the substrate is adenylated. A study on the A domain of phenylalanine (PheA), these ten residues were revealed to be located at positions 235, 236, 239, 278, 299, 301, 322, 330, 331 and 517.[55-58] Due to a high similarity between A domains of NRPSs, consensus sequences of various substrate-binding pockets have been identified and referred to as ‘code’ of NRPSs. Therefore, based on the ‘nonribosomal code’, the substrates of binding pockets can be predicted.[57]

1.5 The symbiont hypothesis

Symbiosis was defined by Anton de Bary as “the living together of unlike organisms”. In a symbiotic association, the larger partner is called the host and the smaller is the symbiont.[59] There are three relationship modes between the host and the symbiont, *mutualism* - the host and the symbiont require each other for their existence -, *parasitism* - the parasite takes unfair advantage of the host - and *commensalism* - one benefits from symbiosis and the other is not affected. However, in nature, there are some interrelated symbiotic systems that are difficult to classify into one of the above three groups.[60]

Often, secondary metabolites play an important role in symbiosis. The function of these compounds are usually for predator defense, interspecies competition, coloring, etc.[59] Correspondingly, they often have long-term effects on the living organisms.

The hypothesis in which microbial symbionts may be the producers of natural products isolated from hosts is known as the symbiont hypothesis. This assumption was raised owing to the discovery that many natural products often remarkably resemble compounds isolated from microorganisms in terms of their chemical structure.[5, 61] Such as the case of the antitumor polyketide rhizoxin, it is not biosynthesized by *Rhizopus* fungus itself, but by the endosymbiotic bacteria of the genus *Burkholderia*.[62, 63] Importantly, results from various research groups indicate that symbiotic bacteria are likely to be the true owners of several compounds that are extracted from marine invertebrates.[64] For example, the cyanobacterial symbiont *Prochloron didemni* of the ascidian *Lissoclinum patella* has been shown to be the actual source of patellamide-type cyclic peptides,[62, 65-69] while the anti-cancer macrolides of the bryostatin family from the bryozoan *Bugula neritina* likely originate from a rod-shaped bacterium, the γ-proteobacterium *Candidatus* Endobugula sertula.[70-74]

The following sections will focus on the symbionts of insects, sponges and their secondary metabolites.

1.5.1 Insects and their symbionts

Insects are the most diverse group of animals on the earth. They are a major group of the phylum Arthropoda and belong to the class Insecta. Among these, the beetles possess the largest number of described species comprising 40% of known insect species. There are approximately 360,000 species of beetles, and they live in a variety of different environments.[75] Their roles in the ecosystems are varied; they can be harmful, useful or both. Beetles and other insects have been well studied with respect to their symbionts, such as bacteria and fungi.[75] Symbionts of insects provide nutrients to their hosts and metabolites as chemical defense weapons.[76] Among the secondary metabolites of insects, there are a lot of compounds that show significant bioactivities such as antitumor, anti-inflammatory and antiviral.[43, 77] Symbiotic bacteria have been demonstrated to be the actual producers of some of these secondary metabolites and derivatives. As outlined below, Piel *et al.* found that the potent anticancer compound pederin from the beetle *Paederus fuscipes* is a product of a symbiotic bacterium belonging to the genus *Pseudomonas*.[43, 78]

1.5.2 Sponge symbiosis and secondary metabolites

1.5.2.1 Microbial sponge symbiosis

Sponges filter bacteria from seawater as food and often harbor diverse microbial communities including bacteria, dinoflagellates, diatoms and archaea.[79-81] Many sponges have a large number of bacteria within their tissues (Figure 12), often in high density - occupying up to 40-60% of the total biomass of the sponge (equivalent to 10^8 to 10^{10} bacteria per gram).[68, 82, 83] Bacterial symbionts can be located both intercellulary and intracellulary. Some photosynthetic symbionts, such as cyanobacteria, live close to the sponge's surface where there is abundance of light. Generally, sponge microbial population can be grouped in three categories: exosymbionts, endosymbionts and intracellular symbionts (Figure 13). Most sponge-specific microbes exist in the extracellular matrix (known as sponge mesohyl) and the others are in the intracellular space. In addition, there are also some nonspecific bacteria resembling those in the surrounding environment.[79]

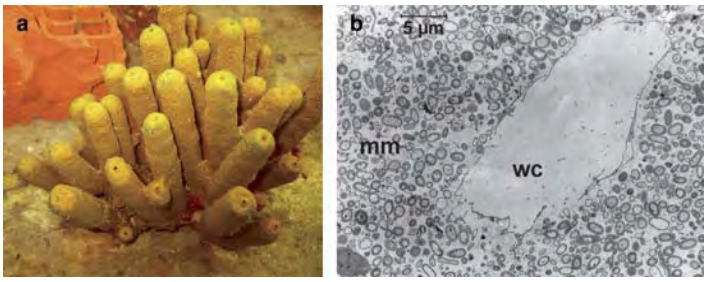


Figure 12. (a) The sponge *Aplysina aerophoba* and (b) symbionts of *Aplysina aerophoba* (mm) mesohyl matrix (wc) a water channel (photography by K. Bayer and M. Wehrl, Würzburg, respectively).[84]

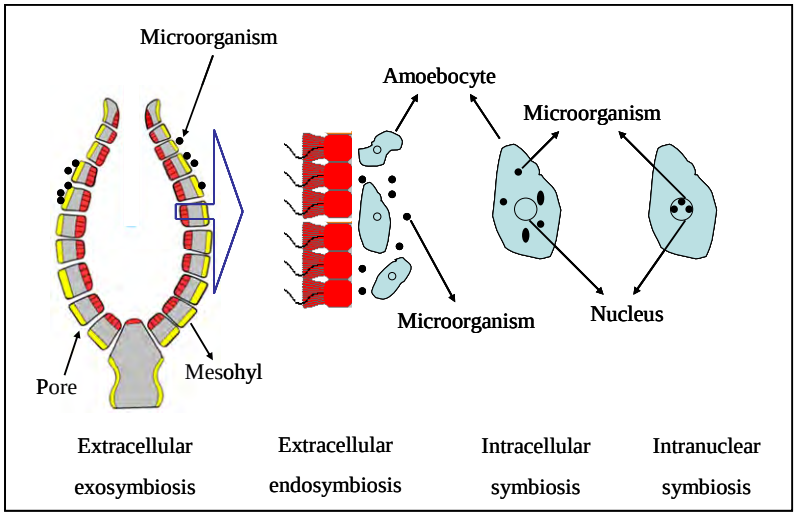


Figure 13. Locations of symbiont bacteria on sponges (source: Y. K. Lee *et al*., Seoul).[79]

Microbiologists have identified a large diversity of bacterial associations in sponges. This complexity is affected by many factors such as the sponge's ecosystem, its structure, etc. Microscopic studies have determined the variation of bacterial communities between different sponges, and each sponge seems to possess an exclusive bacterial association.[79] During genetic studies, *ca*. 100 species of microorganisms from sponge symbiosis were discovered.[82] These were distributed into 14 phyla of bacteria and 2 phyla of archaea *e.g*. Acidobacteria, Chloroflexi, Actinobacteria, Proteobacteria, Cyanobacteria, Nitrospira, Bacteriodetes, etc.[81, 85, 86] Symbiotic bacteria play distinct roles in the lives of their host sponges, for instance, stabilization of the sponge skeleton, processing of waste products, transferring and providing nutrients or producing secondary

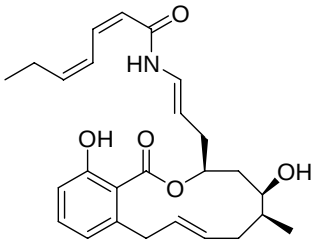
metabolites.[80, 84, 87-90] However, it is difficult to study these bacteria because most symbionts are as yet unculturable.

In detail, an examination by Hentschel, Moore and coworkers showed the complexity of the symbiotic bacterial systems of the two sponges *A. aerophoba* and *T. swinhoei* with almost 160 different bacteria of diverse taxa.[64, 85] By transmission electron microscopy (TEM), four distinct types of cells were found in *T. swinhoei*. These were comprised of eukaryotic sponge cells, unicellular heterotrophic bacteria (mean concentration, 20% of the sponge tissue), unicellular cyanobacteria (15% of the sponge tissue) and filamentous heterotrophic bacteria (40% of the sponge tissue).[79, 91, 92] All these symbiont bacteria were located extracellularly.[91] Normally, cyanobacteria exist rarely in sponges; however in tropical sponges, cyanobacteria protect the sponge from excessive solar radiation.

1.5.2.2 Sponge secondary metabolites

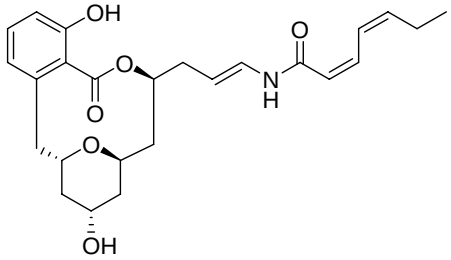
Table 3. Some sponge metabolites that are similar to bacterial metabolites.[64]

Sponge metabolite	Source	Bacterial metabolite	Source
Arenastatin A	*Dysidea arenaria*	Cryptophycin	*Nostoc* sp.
Discodermide	*Discodermia dissoluta*	Alteramide A	*Alteromonas* sp.
Jaspamide	*Jaspis* spp.	Chondramide D	*Streptomyces* sp.
Keramamide A	*Theonella* sp.	Ferintoic acid A	*Chondromyces crocatus*
Mimosamycin	*Petrosia* sp.	Mimosamycin	*Microcystis aeruginosa*
Motuporin	*Theonella swinhoei*	Nodularin	*Nodularia spumigena*
Renieramycin E	*Reniera* sp.	Safracin B	*Pseudomonas fluorescens*
		Saframycin A	*Streptomyces lavendulae*
Salicylihalamide A	*Haliclona* sp.	Apicularen A	*Chondromyces* sp.
Swinholide A	*T. swinhoei*	Tolytoxin	*Tolypothrix* sp.
Misakinolide A	*Theonella* sp.	Scytophycin C	*Scytonema* sp.



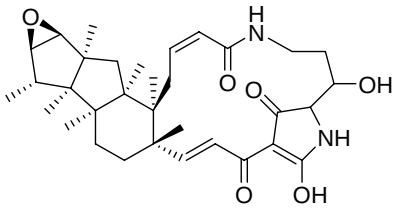
Salicylihalamide A

from the sponge *Haliclona* sp.



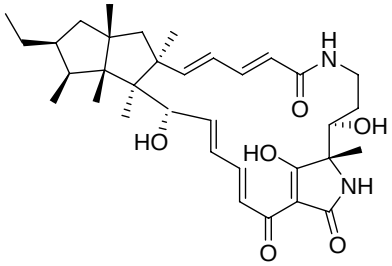
Apicularen A

from the bacterium *Chondromyces* sp.



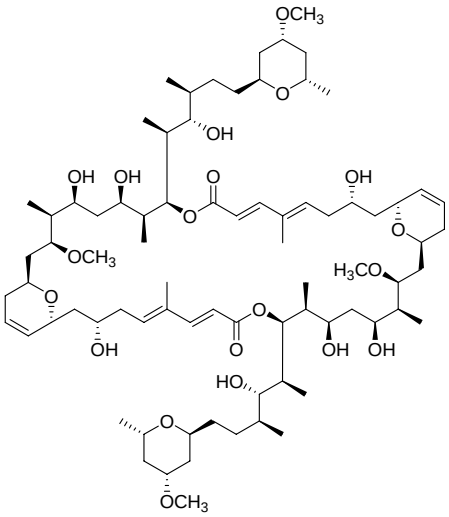
Discodermide

from the sponge *Discodermia dissoluta*



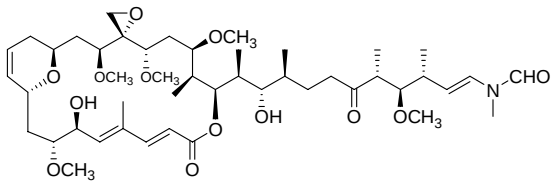
Alteramide A

from the bacterium *Alteromonas* sp.



Swinholide A

from the sponge *Theonella swinhoei*



Tolytoxin

from the bacterium *Tolypothrix* sp.

Figure 14. Some sponge metabolites that are similar to bacterial metabolites.

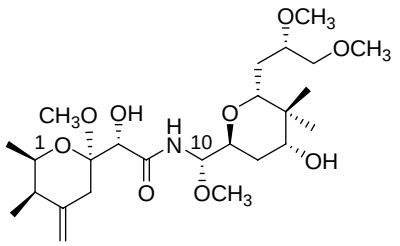
The secondary metabolites of sponges belong to diverse structural classes. They can be derivatives of amino acids or nucleosides, macrolides, cyclic peroxides, alkaloids, polyketides or peptides, etc.[16, 30, 93-95] Many of them have been hypothesized or shown to be efficient chemical weapons against predators.[16, 96, 97] Noticeably, some of these metabolites exhibit impressive structural similarities to known microbial metabolites (Table 3 and Figure 14).[5, 61] Therefore, it has been suggested that many sponge secondary metabolites are products of the complex symbiont bacterial associations.[61, 64, 80, 86, 98] For example, in the case of the sponge *T. swinhoei*, the unicellular bacterial fraction was found to be the location of highest swinholide A concentration, whereas theopaulamide was mainly found in the fraction containing filamentous bacteria.[91] Several bacteria isolated from sponges were shown to be producers of natural products, such as antimicrobial lipopeptides found in *Bacillus subtilis* and *B. pumilus* from the sponge *A. aerophoba*.[99] Or the peptide antibiotic andrimid isolated from *Vibrio* strain M221 from the sponge *Hyatella*.[100] However, since these compounds have so far not been found in the sponges, it is uncertain whether these bacteria are real symbionts or accidental contaminators.

A bacterial biosynthesis is particularly likely for complex polyketides and peptides resembling NRPs, since the type I PKSs and NRPSs are so far virtually exclusively known from microorganisms.[33, 101] The only example of a proven symbiont origin exists to date for the pederin family, an important group of antitumor polyketides.

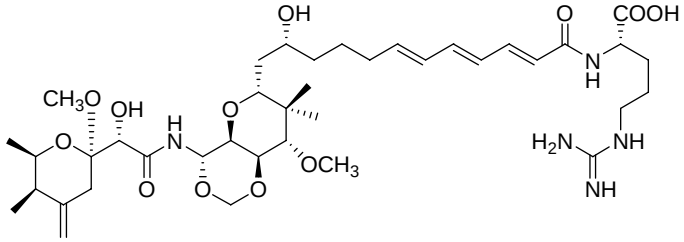
1.6 Pederin family

The pederin family consists of compounds that exhibit a similar identical heterocyclic core structure with a side chain that has a variable polyketide or amino acid terminus.[78] The first member of this group was isolated from the beetle *P. fuscipes* and given the name pederin by Pavan and Bo.[102] Independently, the Quilico group in Italy determined its correct molecular formula ($C_{25}H_{45}O_9N$) in 1961 and its structure was proposed in 1965 by Matsumoto group in Japan.[102] In nature, this substance is believed to help adult female beetles protect their larvae against spiders.[76] *In-vivo*, pederin is able to block protein and DNA synthesis[103] and extend the life of mice bearing a variety of tumors.[43] Pederin from *Paederus* and *Paederidus* of the family Staphylinidae occupied an exclusive position in the natural product world until two other marine natural products with striking similarity in structure with pederin were identified in 1988. These were mycalamide A that was isolated from a sponge *Mycale* sp. by Perry *et al.*[104] and onnamide A that was obtained from a

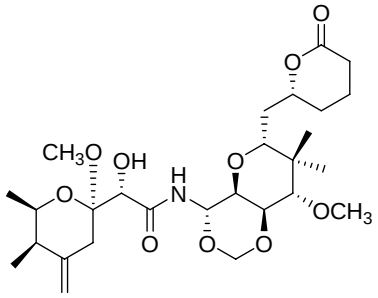
sponge *Theonella* sp. by Sakemi *et al.*[105] In contrast to the understanding of pederin's role in the beetle, the ecological function of the pederin-type compounds in sponges has not been understood yet. It is likely that similar to other cytotoxins from sponges, they play a role in defense against predation or epibionts.[16] Today, the pederin family has grown to become a large group including 36 members, pederin, the 13 onnamides, mycalamides A-D, psymberin, icadamides, and the theopederins A-L, etc.[78, 106] All members of this family are scarce. Therefore to obtain these compounds, total synthesis[102] and biotechnology methods[78, 86] have been used.



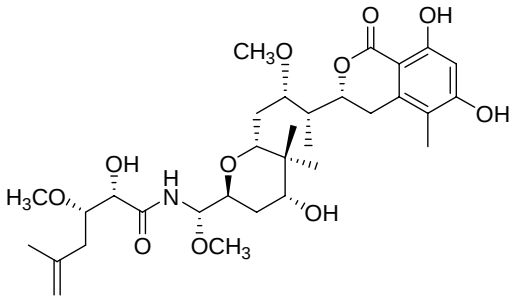
Pederin from the beetle *P. fuscipes*



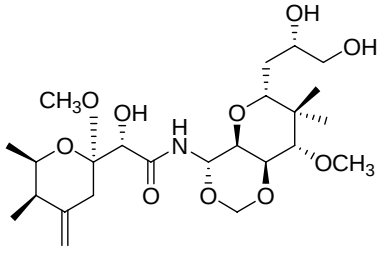
Onnamide A from the sponge *T. swinhoei*



Theopederin C from the sponge *T. swinhoei*



Psymberin from sponges *Psammocinia* aff. *bulbosa*



Mycalamide A from sponges *Mycale* sp.

Figure 15. Some members of the pederin family from terrestrial beetles and marine sponges.

Though biological properties of the individual members of the pederin family have not been evaluated systematically, several of them exhibit highly potent bioactivities. This characteristic is due to the presence of the pederic acid subunit O(1) to C(10).[107] For example, mycalamide A was tested *in-vivo* against the P388 murine leukemia cell line and various solid tumor model systems including Lewis lung, M5076, and Burkitt's lymphoma.[11, 107] Moreover, mycalamide A is not only an antitumor and antiviral agent but also immunosuppressive. It is able to block T-cell activation in mice with an effect that is 10fold more potent than FK-506 and 1,000fold more potent than cyclosporin A.[102]

1.6.1 The beetle *P. fuscipes* and the previous genetic work on pederin

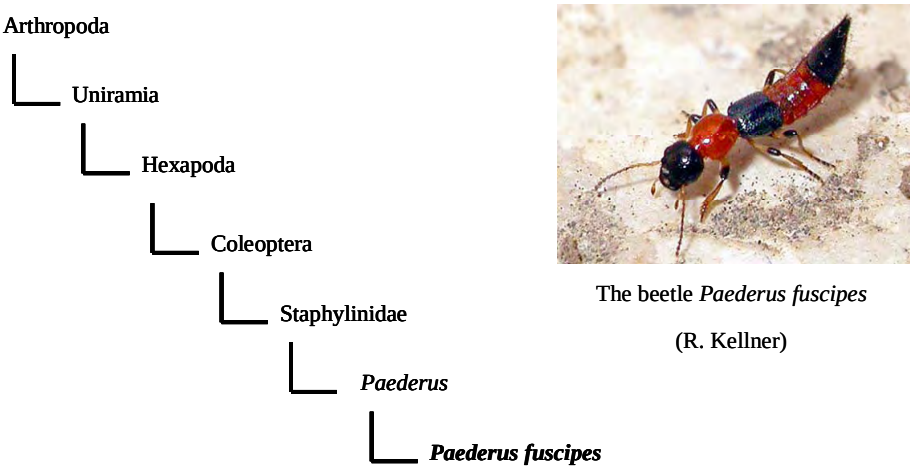


Figure 16. The beetle *P. fuscipes* and its taxonomy.

Rove beetles, *Paederus fuscipes*, are rather small with a length of approximately 0.7 cm (Figure 16). They are often found in tropical and temperate countries.[108] These beetles cause skin and eye dermatitis (also being called *Paederus* dermatitis) in humans due to their toxin, pederin.[76, 109] To date, within the pederin family, pederin is still the only member found from a terrestrial source; whereas the remaining members were isolated from sponges.[43] These beetles are an ideal model for studies on natural products as secondary metabolites of symbiosis because of their simple symbiosis system.[43, 78]

In our laboratory, genetic studies on the pederin (*ped*) gene cluster from *P. fuscipes* were carried out in order to gain insights into the biosynthetic pathway and the producer of not only pederin but other pederin-like compounds. Using degenerate primers based on the universally conserved motifs of KS domains, the first PKS genes encoding for pederin were amplified from the metagenomic DNA of *P. fuscipes*.[43] A cosmid library including 80,000 clones was constructed from the total DNA of *P. fuscipes*, which was sufficient for coverage of the symbiont genome. This DNA library was screened with specific primers that were designed based on the cloned *ped* genes. In addition, about 95% of the *P. fuscipes* symbiont genome was sequenced, resulting in isolation of the complete *ped* gene cluster.[101] The modular architecture of the pederin PKS is almost in agreement with pederin structure. Remarkably, in contrast to the biosynthetic genes of other bacterial secondary metabolites, 18 putative *ped* genes are located on three distinct genome regions, *pedLMNOPQR*,[101] *pedIJK*[78] and *pedABCDEFGH*[43] (Figure 17 and Figure 20). The presence of transposase gene fragments on both sides of each *ped* region suggests that the unclustered architecture arose from gene transposition. In particular, there was no AT domain in the PKS modules, instead two AT genes were found on the upstream region of *pedF*. This was one of the first examples of a *trans*-AT PKS. An unexpected feature was the presence of *pedH*, in which a second NRPS module predicted to be responsible for incorporation of an arginine residue was found. This finding did not match with pederin structure and will be discussed in Chapter 3.

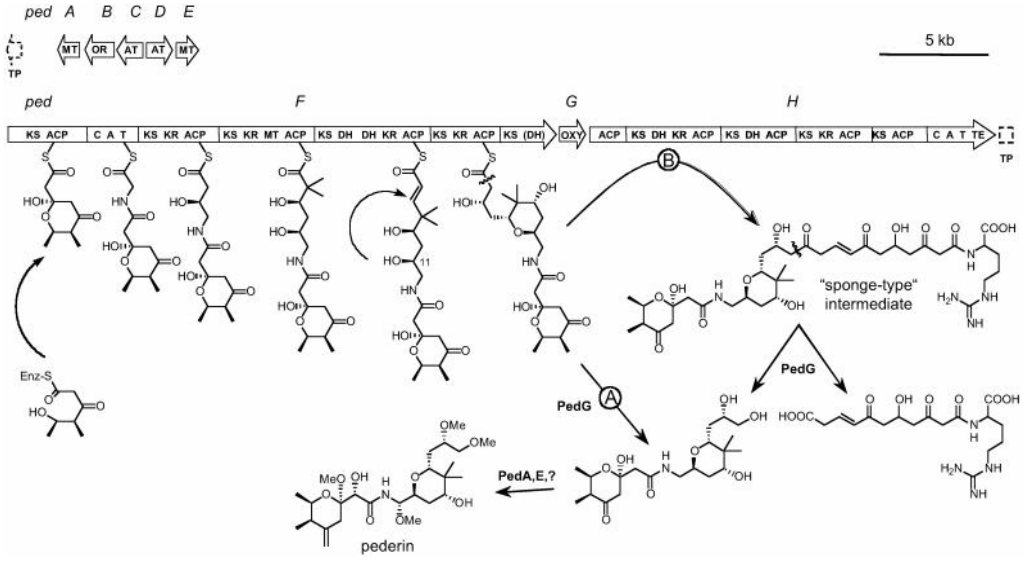
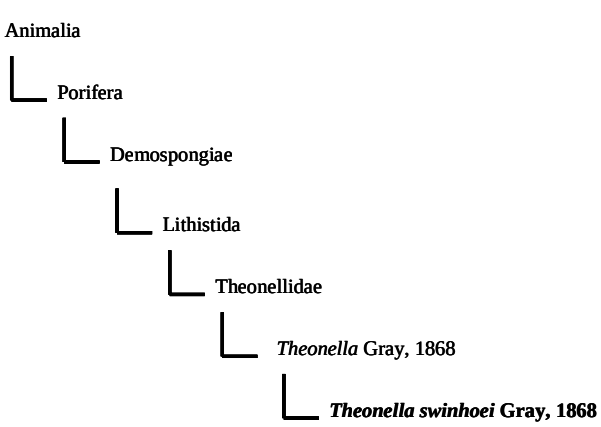
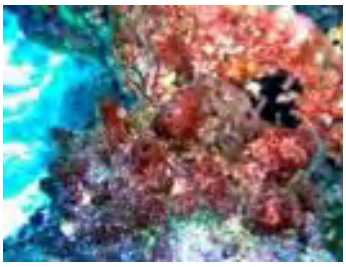


Figure 17. Proposed pederin biosynthetic pathway.[43]

In addition, several genes unrelated to *ped* genes that exhibited high similarities to those from *P. aeruginosa* were also identified.[43] This was in agreement with studies by Kellner *et al.*, who discovered that pederin was accumulated by the female beetles only if they possessed a specific endosymbiont *Pseudomonas* sp.; whereas pederin was not found in beetles without endosymbionts.[110] Furthermore, PCR-based analysis of 16S rRNA showed that only the *P. fuscipes* beetles containing high pederin content were the host of unculturable bacterium related to *P. aeruginosa*.[78, 111, 112]

1.6.2 The sponge *T. swinhoei* and the previous genetic work on onnamides

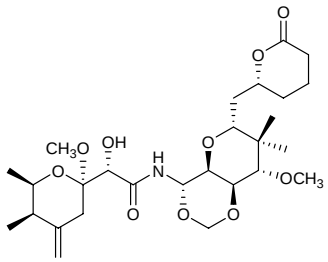




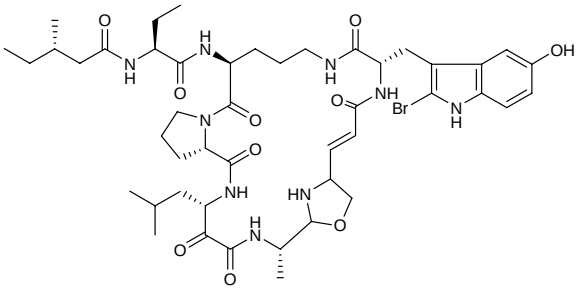
The sponge *Theonella swinhoei*
(Yoichi Nakao)

Figure 18. The sponge *T. swinhoei* and its taxonomy.

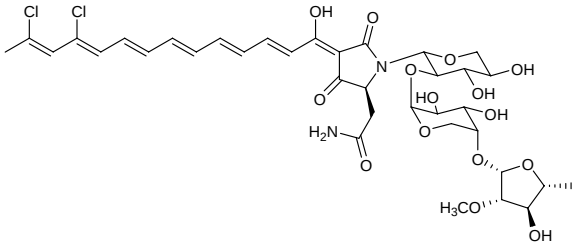
The sponge *T. swinhoei* belonging to the order Lithistida is typically found in deep water or caves of tropical and sub-tropical ocean (Figure 18).[91, 96] The sponge *T. swinhoei* has proven to be an extraordinarily rich source of natural products (Figure 19).[50, 91, 105] It is from the sponge *T. swinhoei*, that 22 pederin-type compounds have been reported so far.[106] Most of the secondary metabolites of *T. swinhoei* reveal biological activities, in particular antiviral, antifungal or antitumor.[113, 114] Some typical compounds are the macrolide swinholide A, a compound that is cytotoxic to a variety of cancer cell lines; theonegramide, an antifungal bicyclic glycopeptide, the mixed PK-NRP theopederins with antifungal activity and onnamides possessing a potential antitumor property.



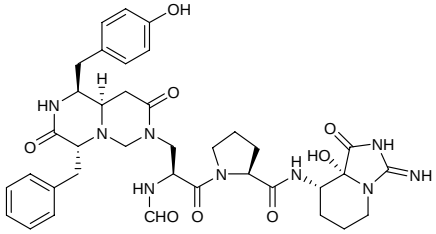
Theopederin A



Orbiculamide A



Aurantoside A



Pseudotheonamide A2

Figure 19. Some natural products from *T. swinhoei*.

Specimens of the sponge *T. swinhoei* that were studied in our laboratory included two different chemotypes collected in Japan. A yellow type contains the pederin family of compounds and other polyketides, and nonribosomal peptides such as the aurantoside series, the pseudotheonamides and orbiculamide A. A white type however possesses compounds unrelated to pederin.

Onnamide A was the first onnamide isolated from a Japanese sponge *T. swinhoei* that was collected from the Okinawan Sea.[105] As in the case of pederin and mycalamide A and B, onnamide A showed a strong ability to inhibit protein synthesis at an IC_{50} in a subnanomolar concentration.[115] It induced activation of p38 mitogen-activated protein kinase (this protein is involved in the activation of PAI-1 (plasminogen activator inhibitor-1) gene expression) and c-Jun NH_2-terminal kinase (JNK), which is a stress-activating protein kinase.[116] It is cytotoxic to various cell lines and significantly induces apoptosis in HeLa cells within 24 hours. Furthermore, onnamide A is known as ribotoxic by binding with ribosome, thereby inhibiting ribosomal function.[11, 116] These results suggest onnamides as drug candidates in cancer therapy.

The sponge *T. swinhoei* was chosen for our experiments because it is an ideal study model. The same strategy that was used for the study on pederin was applied; whereby genes from the complex sponge symbionts could be accessed more rapidly. Using degenerate KS primers, several PKS gene fragments were amplified from the sponge metagenome. Phylogenetic analysis of the PCR amplicons identified one as being involved in the onnamide biosynthesis. Based on this amplicon, specific primers were designed in order to screen 60,000 clones of a fosmid library of *T. swinhoei* and one positive clone was identified. This fosmid was completely sequenced and a partial *onn* gene cluster was found. As expected, its domain architecture displayed a high similarity to that of pederin (Figure 20).[101] This achievement raised the expectation that resulting from cloning of the entire onnamide biosynthetic genes, a renewable onnamide source can be generated by heterologous expression in culturable bacteria. The present study therefore deals with the isolation of the remaining *onn* genes.

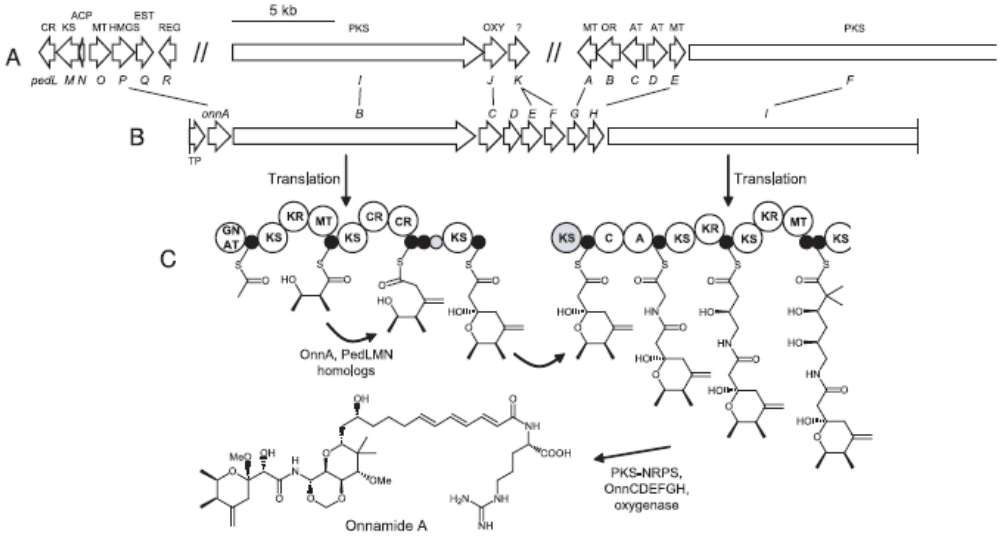


Figure 20. Comparison of the isolated *onn* genes with the *ped* system and a partial proposed onnamide A biosynthesis.[101] (A) Map of the *ped* cluster from the *P. fuscipes* symbiont. Double slashes separate the three genome regions. (B) Map of the *onn* cluster and correlation to ped homologs. (C) Protein products of the PKS-NRPS genes *onnB* and *onnI* and proposed biosynthetic pathway leading to onnamide/theopederin-type compounds. Each circle represents one domain. Domains in gray lack active site motifs and are presumably nonfunctional. CR: CR superfamily, EST: esterase, REG: regulator, OXY: oxygenase, OR: oxidoreductase, TP: transposase, C: condensation domain, A: adenylation domain.

1.7 The metagenomic approaches in natural product research

1.7.1 Difficulties in natural product exploitation

Although marine life is in theory an unlimited source of pharmaceutical candidates, the exploitation of the marine resource is so far difficult because of a series of supply problems. The minute concentration of active compounds from marine organisms is the first obstacle, these compounds often occupy less than 10^{-6}% of respective wet weight of an invertebrate.[5] Demonstrating this fact is the case of the promising antitumor agent, bryostatin 1. In order to isolate 1.4 gram of bryostatin 1, an estimated one metric ton (wet weight) of the bryozoan *Bugula neritina* is required.[86] Likewise, using up to one metric ton of the sponge *Lissodendoryx* sp. will only yield approximately 300 mg of halichrondin, a cytostatic agent.[117] This means, if halichondrin is applied in cancer therapy, the annual demand will fall in the range of about 1 - 5 kg. Correspondingly, 3,000 - 15,000 metric tons of sponge will need to be collected annually.[117] This demand-supply ratio will effectively destroy the ecological balance of the marine environment. Directly or indirectly, a great number of species will be affected, symbiotic relationships will change, and some organisms could disappear and eventually become extinct. Scientifically, harvesting and processing of such enormous amounts of material are impossible, and economically not feasible.

The disadvantages in obtaining bioactive compounds from insects are similar to those cited in the processes involving marine organisms. To extract pederin, an antitumor agent, 25 million of the *P. fuscipes* beetles (*ca.*100 kg) were collected by hand.[102] Time and labor costs were high and it was also a human health risk, due to the toxicity of pederin.

The total syntheses of these natural compounds such as ET-743, halichondrin B, bryostatin, were established successfully but economically, they are also not feasible, due to the complexities of the chemical structures which lead to low synthetic efficiency.[5]

To date, efforts in cultivating these organisms as well as their symbiotic microbes that are implicated to be the producers of pharmacologically active products, have not gained significant achievements because so far we lack the knowledge of their specific living conditions.[80]

1.7.2 Metagenomics

Microbiologists have achieved excellent results in studying individual microorganisms. As mentioned, there has however been very little success in research on the diverse microbial symbionts. In spite of many efforts, 99% of the symbiont microbes in the ecosystem have not been able to be successfully cultivated in the laboratory.[118-121] That means many important microbial metabolic pathways and their products are as yet untapped. In order to overcome these obstacles, a metagenomic strategy based on the genomic analysis of microbial population has had widespread use since the last decade.[4, 120] Metagenomics has been an important breakthrough in revolution of natural product researches.[16, 122-124]

The first molecular experiments following this way were carried out by N. R. Pace who explored the diversity of 16S rRNA sequences from uncultured organisms in environmental sample.[125] In 1998, the concept of metagenomics was named by J. Handelsman. This was based on *meta*-analysis and genomics of the environment sample.[126] She said “Metagenomics lets us see into the previously invisible microbial world, opening a frontier of science that was unimaginable until recently." This approach allows us to address the questions “who’s there?”, “what are they doing?” and “how are they doing it?”[126] To date, many research groups apply the metagenomic strategy for studies on natural products. By this approach, biosynthetic genes of various bioactive compounds were successfully isolated from natural sources, *e.g*. the patellamide gene cluster from the tunicate *Lissoclinum patella*[67], the bryostatin genes from the bryozoan *Bugula neritina*[72], and the erdacin gene cluster from desert soil.[127]

Metagenomic analysis is a multi-step process (Figure 21). The metagenomic DNA is isolated directly from all the microbes living in a particular sample. It is followed by DNA library construction with a suitable cloning vector and a host strain, typically *Escherichia coli*.[128, 129] The library of metagenomic DNA is employed for screening bioactivities or screening target function genes by PCR with specific primers.[120] One important application of the metagenomic approach is that the cloned biosynthetic genes from the total DNA can then be heterologously expressed in a culturable bacterium in order to produce target compounds in large scales.[80, 128, 130, 131]

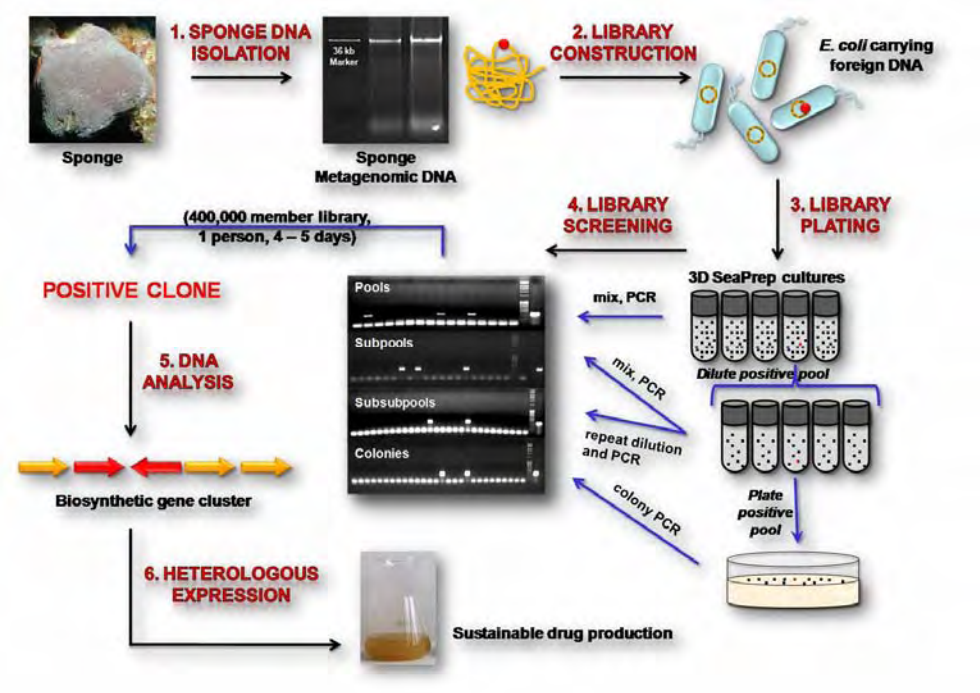


Figure 21. Metagenomic strategy applied for study on natural products from sponges.

Chapter 2
Aims of the study

2.1 Isolation of the remaining onnamide gene cluster from the sponge *T. swinhoei*

Onnamide was the first sponge-derived compound, for which a bacterial biosynthesis was shown, and the first, for which biosynthetic genes have been isolated in a metagenomic strategy. The compound is therefore a model that helps us to understand how renewable source of marine drug candidates can be generated by microbial expression systems. Therefore and due to the potent cytotoxicity of onnamides, studies on onnamide biosynthetic genes from the sponge *T. swinhoei* were effortfully carried out in our lab. Applying the same strategy used for research on the biosynthetic gene cluster of pederin,[43] a partial *onn* gene cluster was isolated and completely sequenced.[101] The primary aim of this study, therefore, was to isolate the remaining *onn* genes from the sponge metagenomic DNA. Based on the previously sequenced region, the goal was to clone contiguous parts of the *onn* gene cluster by applying the primer walking method. A fosmid library of *T. swinhoei* consisting of about 400,000 clones and covering 32 Gbp of the sponge genome should be subjected to PCR-screening. Cloning of the entire *onn* gene cluster should enable us not only to understand the biosynthetic pathways of onnamides, but also to get further insight into pederin biosynthesis.

Furthermore, previous PCR experiments had suggested the presence of a second *trans*-AT gene cluster in the *T. swinhoei* metagenome. The fosmid library of *T. swinhoei* should therefore be screened and the cluster was sequenced after isolation.

Finally, the A domains of NRPS modules from the *onn* (if isolated) and the *ped* gene cluster should be expressed. These overexpressions should be performed to prepare for the elucidation of the roles of the A domains in the corresponding systems as well as for the heterologous expression of the entire *onn* and *ped* gene cluster.

2.2 Heterologous expression trials with the cosmid pPD7E4 from the beetle *P. fuscipes*

Partial screening of the genome of the *Paederus fuscipes* symbiont had revealed the presence of a small second PKS-NRPS cluster of unknown function. To elucidate the role of these genes, another aim of this work was to perform bioinformatic analysis in order to predict the structure and to conduct first expression and analytic studies to detect the putative compound.

Chapter 3

Isolation of the remaining onnamide gene cluster from the metagenomic DNA of the sponge *Theonella swinhoei*

3.1 Introduction

In a previous work by Piel *et al*. (2004),[101] from the total DNA of the sponge *T. swinhoei*, a 60,000 clone fosmid library was constructed and subsequent screening yielded a single positive clone (pTS1E4) containing a partial onnamide gene cluster inside (Figure 20 and Figure 22). This was most likely due to the small size of the fosmid library.

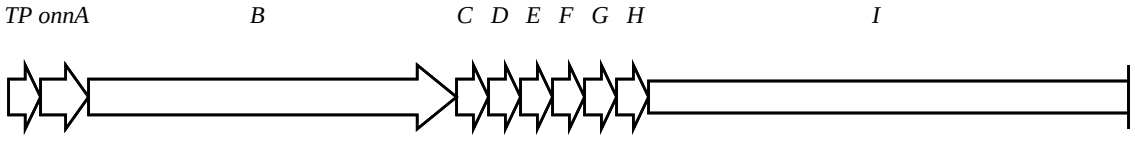


Figure 22. The identified genes of pTS1E4 isolated by Daniel Butzke. Ten putative genes including nine *onn* genes, *onnA*-*onnI* were located inside a 36.5 kb region of pTS1E4, in which *onnI* was incomplete. *TP* is a tranposase gene.

To overcome this limitation, another larger fosmid library was generated to cover more of the sponge metagenome. In constructing the new library, a new protocol for isolation of total sponge DNA was developed, in which CTAB was used to precipitate the polysaccharides. In addition, a

rapid PCR-based isolation procedure with a three- dimensional semi-liquid format for screening of the rare clones was applied.[132] With these approaches, *ca.* 800,000 clones covering 32 Gbp of DNA from the sponge *T. swinhoei* were prepared. About 400,000 of the clones were used for screening of the library. Based on the sequence of the fosmid pTS1E4, one downstream fosmid (pTSSH1) containing *onn* genes was isolated by Siniša Hrvatin, together with one upstream fosmid that did not contain any *onn* gene. Based on spot sequences of pTSSH1, two further fosmids, pTSSH2 and pTSSH3, were cloned from the fosmid library by Siniša Hrvatin and Cristian Gurgui and completely sequenced. First step of the current work was to analyze the sequence data and test whether it can be correlated with the onnamide structure.

3.2 Results and discussion

3.2.1 Analysis of the complete sequences of the fosmid pTSSH2 and pTSSH3

The software Lasergene (DNA star), BLAST from NCBI (http://www.ncbi.nlm.nih.gov/blast) and FramePlot (http://watson.nih.go.jp/~jun/cgi-bin/frameplot.pl) were used to analyze pTSSH2 and pTSSH3 with a size of 40,926 nucleotides and 37,357 nucleotides, respectively. The ‘open reading frames’ (ORFs) of the putative genes were identified between the putative start codons ATG or GTG and the stop codons TAA, TAG, or TGA. Results of the analyses are shown in Table 4 and Table 5.

Table 4. Analysis of the complete pTSSH2 sequence.

Reading frame	Similarity		Identity	Position	Length (aa)
	Protein	Origin			
+2	OnnI	Symbiont bacterium of *T. swinhoei*		1-19099	6,366
-1	ABC transporter	*Burkholderia* sp. 383	300/605 (49%)	19190-24270	1,693
-1	Periplasmic solute binding prot	*Herpetosiphon aurantiacus*	172/331 (51%)	24427-25443	339
-1	Fe^{2+}/Zn^{2+} uptake regulation protein	*Sphaerobacter thermophilus*	44/128 (34%)	25499-25900	134
-2	Hypothetical protein Rxyl_0285	*Rubrobacter xylanophilus*	96/298 (32%)	26166-27359	398
-1	2-Phosphonopropionate transporter	*Ralstonia eutropha* H16	124/424 (29%)	27363-28646	428
-1	Predicted ring-cleavage extradiol dioxygenase	*Catenulispora acidiphila*	70/144 (48%)	28903-29379	159
-1	Leucyl aminopeptidase	*Thermosinus carboxydivorans*	231/485 (47%)	29475-30989	505
-1	Hypothetical protein	*Dehalococcoides* sp.	68/182 (37%)	31221-31832	204
-1	Cell division membrane protein-like protein	*Clostridiales bacterium*	23/69 (33%)	31854-32858	335
-1	Transcriptional regulator	*Desulfotomaculum acetoxidans*	51/105 (48%)	32883-33212	110
-1	Pyrroline-5-carboxylate reductase	*Roseiflexus* sp. RS-1	125/266 (46%)	33584-34387	268
-1	4-Hydroxythreonine-4-phosphate dehydrogenase	*Thermosinus carboxydivorans*	165/329 (50%)	34539-35555	339
-1	Ribonucleotide-diphosphate reductase alpha subunit	*Dehalococcoides* sp. CBDB1	422/787 (53%)	35883-38261	793
-1	Transcriptional regulator	*Listeria grayi* DSM 20601	77/152 (50%)	38888-39523	212
-2	Cell division protein FtsZ	*Moorella thermoacetica*	179/324 (55%)	39924-40926	334

Table 5. Analysis of the complete pTSSH3 sequence.

Reading frame	Similarity		Identity	Position	Length (aa)
	Protein	Origin			
+2	OnnI	Symbiont bacterium of *T. swinhoei*		1-16447	5,482
+1	Ankyrin repeat protein, putative	*Trichomonas vaginalis* G3	115/388 (29%)	19222-21953	910
+3	Hypothetical protein	*Pelodictyon luteolum*	144/315 (45%)	23170-24291	374
-2	Major facilitator transporter	*Roseiflexus castenholzii*	118/376 (31%)	24502-26022	507
-2	Luciferase-like	*Mycobacterium* sp	67/165 (40%)	26175-27116	314
+3	NHL repeat	*Mesorhizobium* sp. BNC1	122/262 (46%)	27190-28164	325
+1	Propionyl-CoA carboxylase	*Ruegeria pomeroyi*	145/332 (43%)	28191-29547	452
+2	Cysteine desulfurase	*Pseudomonas mendocina* ymp	180/273 (65%)	30871-31761	297
+3	Nucleoside diphosphate kinase	*Kangiella koreensis*	99/141 (70%)	32359-32921	187
+1	Predicted Fe-S-cluster redox enzyme	*Hahella chejuensis* KCTC 2396	146/292 (50%)	32918-33888	323
+1	Histidyl-tRNA synthetase	*Cellvibrio japonicus* Ueda107	129/222 (58%)	35076-35928	284
+2	Hypothetical protein ABO_1858	*Alcanivorax borkumensis* SK2	69/190 (36%)	36384-35357	324

In addition to downstream parts harboring non-PKS genes, an addition portion of the gene *onnI* was found in upstream portions of both pTSSH2 and pTSSH3. A PKS region of ~ 19 kb on pTSSH2 comprised an overlapping fragment of ~ 7.6 kb with *onnI* from pTS1E4 and also 11.4 kb of a new *onnI* region. On pTSSH3, a 16.4 kb PKS region included ~ 2 kb of an overlapping fragment with *onnI* from pTS1E4 and ~ 14.4 kb of a new *onnI* region (Table 6).

Table 6. PKS regions of the fosmid pTSSH2 and pTSSH3.

Fosmid	**Size** [bp]	**PKS region**		
		Overlapping with *onnI* [bp]	**New *onnI* region** [bp]	**Total length of the *onnI* region** [bp]
pTSSH2	40,926	7,624	11,475	19,099
pTSSH3	37,357	2,086	14,361	16,447

Analysis programs such as Pfam, BLASTP, as well as the consensus sequences of PKS domains were used to determine the functional domains of the new PKSs on pTSSH2 and pTSSH3 (Figure 23).

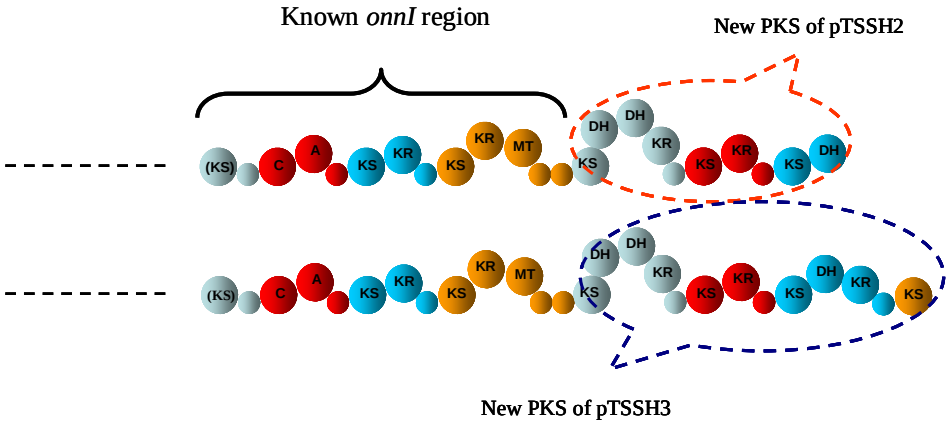


Figure 23. New PKS domains of pTSSH2 and pTSSH3.

Noticeably, although PKS domain organization of pTSSH2 and pTSSH3 was identical to each other in part (Figure 23); there was no homology between their regions outside of the *onn* genes (Table 4 and Table 5). In addition, the PKS termini were different. One explanation for this finding is that multiple variants of the *onn* gene cluster are present in the symbiont consortium and responsible for producing individual PKSs. An alternative is that we had isolated fosmids with chimeric inserts, *i.e.* artifacts.

Remarkably, the domain chains of the *onn* genes on pTSSH2 and pTSSH3 were analogous to that of *pedF* of the *ped* gene cluster. This was expected owing to the similarity in structure between pederin and onnamides. Nevertheless, the overall nucleotide sequence identity between *onnI* on

pTSSH2, pTSSH3 and *pedF* was rather low, only 47% and 48%, respectively (data not shown). This indicates that the two pathways separated from each other a considerable time ago, but that the PKS architecture has been preserved due to a beneficial role in symbiont or host survival.

The previous isolated fosmid pTSSH1 had not been completely sequenced. Therefore, it was checked how large the overlap between its *onn* region and that of pTSSH2 and pTSSH3 is. Since the PKS region on pTSSH3 was more extended than that on pTSSH2 (Figure 23), the checking experiments were only carried out on four regions of pTSSH3 using PCR. Three PKS modules were present inside pTSSH3, therefore, the end region of each module, as well as the KS domain belonging to the second module, were subjected to PCR analysis. The first fragment of 975 bp encoded a part of a KR and an ACP domain (Figure 24); the second, 544 bp fragment a partial KS domain; the third, 442 bp fragment included also a part of a KR and an ACP domain and the fourth, 554 bp fragment encoded the last incomplete KS domain on pTSSH3. Subsequently, four corresponding primer pairs were designed, namely pTSSH3-1, pTSSH3-2, pTSSH3-3, and pTSSH3-4.

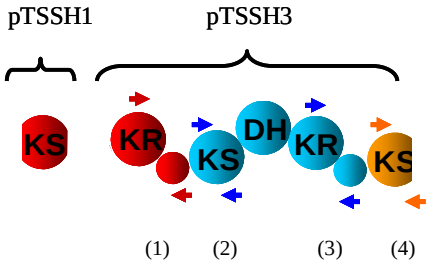


Figure 24. Primers for checking the continuity of *onn* genes on pTSSH3 (1) 1st fragment (2) 2nd fragment (3) 3rd fragment and (4) 4th fragment. Arrows represent primers of the four fragments.

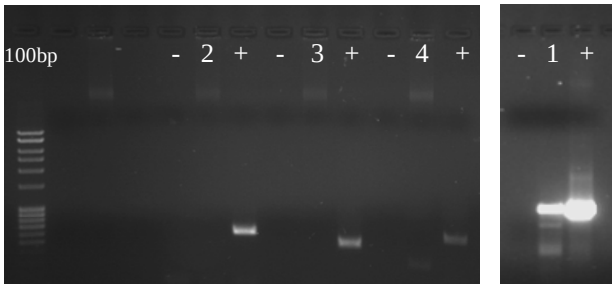


Figure 25. Checking the continuity of *onn* genes on pTSSH3. Agarose gels of electrophoresed PCR are shown. (-) negative control (+) positive control with pTSSH3 (1) 1st fragment (2) 2nd fragment (3) 3rd fragment and (4) 4th fragment.

PCRs were set up with DNA of pTSSH1 as the template. Only the first fragment was amplified (Figure 25). This meant that the corresponding KR-ACP region on pTSSH3 overlapped with pTSSH1. The three remaining regions were not present on pTSSH1 and they were thus the newly isolated *onn* genes. This confirmed that the PKS genes on pTSSH3 were continuous with those of the fosmid pTSSH1 and that all PKS possessed a different terminus. The hypothesis that chimeric inserts were isolated was therefore unlikely, since this rarely occurs three times in a row. Moreover, it would be a rare coincidence if all PKS fragments precisely ended with an intact domain. The most likely reason for the isolation of three distinct PKSs was therefore that multiple variants are present in *T. swinhoei*. This was unexpected and severely complicated the search for the entire gene cluster.

As mentioned, several genes unrelated to the *onn* system were also found from the metagenomic DNA of *T. swinhoei* including its symbiont community. These non-PKS genes on the two fosmids pTSSH2 and pTSSH3 (Table 4 and Table 5) were the best candidates to investigate the organismal origin of onnamides by phylogenetic analysis. As shown in Table 4, there was a gene encoding for the FtsZ protein on the outside region of the *onn* genes on pTSSH2. FtsZ is a bacterial tubulin-like protein that is essential for cell division and highly conserved among prokaryotes (Figure 26). Phylogenetic studies often use this protein to determine taxonomy.[133, 134]

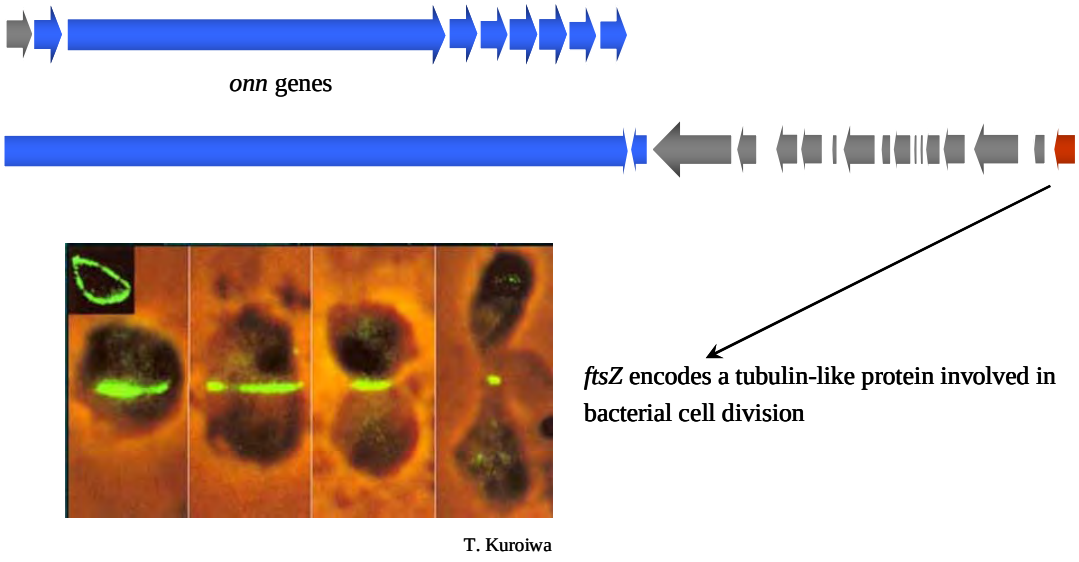


Figure 26. The *ftsZ* gene on the fosmid pTSSH2. Blue: the *onn* genes, grey: non-PKS genes, red: the *ftsZ* gene. The picture shows the FtsZ (green color) involved in bacterial cell division.

According to the alignment with 23 closest FtsZ homologues of other microorganisms, a phylogenetic tree was inferred. In this unrooted tree, the bacterial phyla Firmicutes, Cyanobacteria, and Chloroflexi were organized into three large clades, in accordance with their true taxonomy. The onnamide producer, represented by the owner of FtsZ from pTSSH2, was associated with the latter phylum and its node was at a deep position (Figure 27).

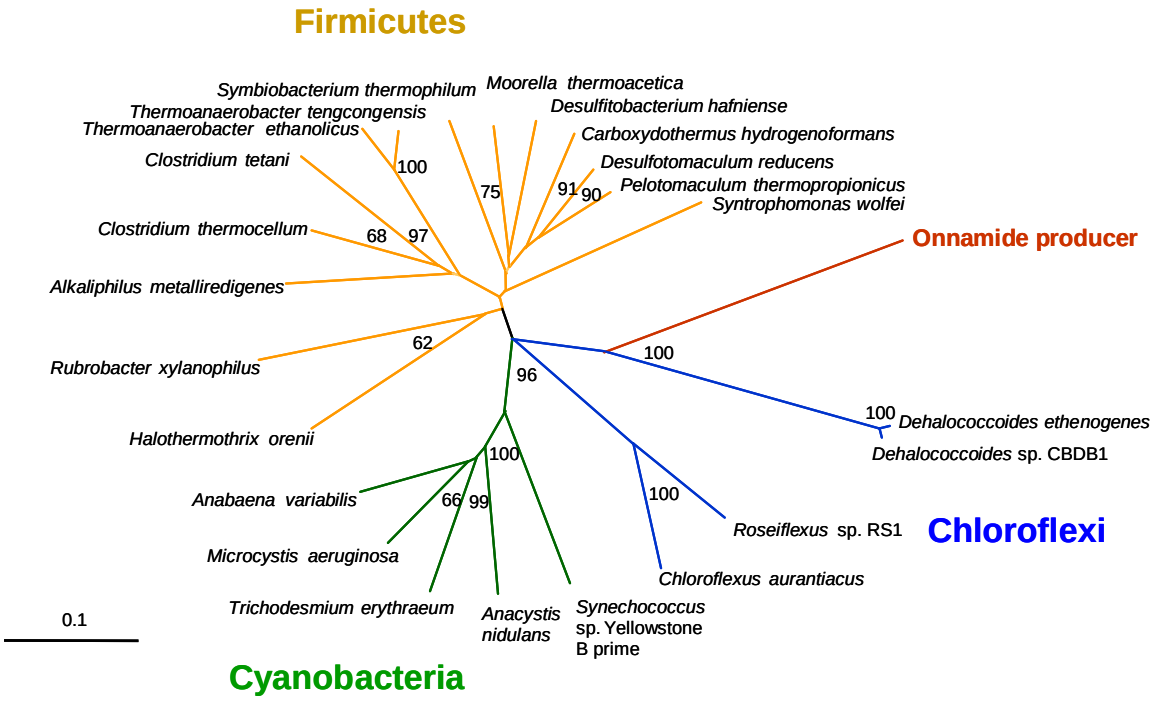


Figure 27. Neighbor-joining tree of the closest FtsZ relatives.

This analysis provided evidence that the onnamide producer belongs to the *Chloroflexi*, formerly known as green non-sulfur bacteria, are a phylum of eubacteria that are typically filamentous. These bacteria are autotrophs, which means they sustain themselves by producing organic molecules from carbon dioxide and other raw inorganic materials derived from the environment under activation of sunlight. Therefore, they are also classified in a group called photosynthesis bacteria.[135] These bacteria have been detected in a wide range of terrestrial, aquatic and are ubiquitously present in bacteriosponges. Together with acidobacteria, they are the major members of sponge symbiotic populations, 23% and 22%, respectively.[86] This seemed to be in agreement with the results obtained by Bewley *et al.* that major metabolites of *T. swinhoei* do not seem to be produced by cyanobacteria or sponge cells.[91]

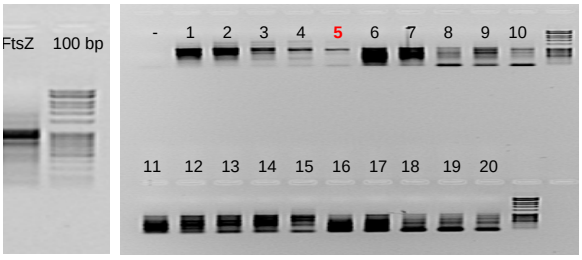
3.2.2 *ftsZ* gene cloning

The availability of the *ftsZ* gene not only on pTSSH2 but also in the symbiont system of the sponge *T. swinhoei* was an important factor that helped us to pinpoint the real producer of onnamides. The goal of this experiment is to obtain insight into the presence of multiple *ftsZ* genes from closely related strains. This could explain the presence of multiple closely related variants of the onnamide genes. For this reason, the *ftsZ* gene from the metagenomic sponge DNA had to be cloned for sequencing and then analysis.

Based on the 883 nucleotide sequence of *ftsZ* gene on pTSSH2, specific *ftsZ* primers were designed to amplify *ftsZ* from the total sponge DNA. For PCR analysis, both $MgCl_2$ concentration and annealing temperature of primers were optimized in order to obtain a clean PCR product. This required the preparation of twenty PCR reactions (Table 7).

Table 7. Optimization of PCR conditions with the *ftsZ* primers.

Sample	Mg^{2+}	Annealing temperature [^oC]	Sample	Mg^{2+}	Annealing temperature [^oC]
1	2 mM	54	11	4 mM	54
2	2 mM	56	12	4 mM	56
3	2 mM	58	13	4 mM	58
4	2 mM	60	14	4 mM	60
5	**2 mM**	**62**	15	4 mM	62
6	3 mM	54	16	5 mM	54
7	3 mM	56	17	5 mM	56
8	3 mM	58	18	5 mM	58
9	3 mM	60	19	5 mM	60
10	3 mM	62	20	5 mM	62



(a) (b)

Figure 28. (a) Checking the *ftsZ* primers with the total sponge DNA (b) Optimization of PCR conditions with the *ftsZ* primers. Agarose gels of electrophoresed PCR are shown.

The PCR product PTA1 amplified in the 5th PCR reaction gave the sharpest band, hence a $MgCl_2$ concentration of 2 mM and an annealing temperature at 62 ^{o}C were chosen for the next steps (Figure 28). PTA1 was then purified from agarose gel and prepared for TA cloning with the vector pBluescript SK II (-) (section 7.2.12). Plasmid DNAs from five white clones were isolated and sent for sequencing.

The five sequences pTA1-1 – pTA1-5 were analyzed by BLAST homology searches and the results showed that all proteins encoded by these amplicons were closest homologues of FtsZ of *Dehalococcoides ethenogenes*, a member of the phylum Chloroflexi with an identity ranging from 56% to 58%. These five nucleotide sequences were subsequently aligned with the *ftsZ* sequence of pTSSH2 using the BioEdit and ClustalW programs. The cloned FtsZ amplicons and that of pTSSH2 shared high but not complete identities of ≥ 96% (Table 8 and see Appendices). These results demonstrate the existence of multiple closely related *ftsZ* variants inside the metagenomic DNA of *T. swinhoei*. This might explain the presence of closely related PKS variants, which could belong to distinct, diversified strains.

Table 8. Alignment of FtsZ sequences.

Seq A	Name	Seq B	Name	Score
1	pTA1-1	2	pTA1-2	96
1	pTA1-1	3	pTA1-3	97
1	pTA1-1	4	pTA1-4	96
1	pTA1-1	5	pTA1-5	96
1	pTA1-1	6	FtsZ of pTSSH2	98
2	pTA1-2	3	pTA1-3	97
2	pTA1-2	4	pTA1-4	97
2	pTA1-2	5	pTA1-5	96
2	pTA1-2	6	FtsZ of pTSSH2	97
3	pTA1-3	4	pTA1-4	97
3	pTA1-3	5	pTA1-5	97
3	pTA1-3	6	FtsZ of pTSSH2	97
4	pTA1-4	5	pTA1-5	96
4	pTA1-4	6	FtsZ of pTSSH2	96
5	pTA1-5	6	FtsZ of pTSSH2	96

Based on the above results, surprisingly, the producer of onnamides is different from that of pederin, which is *Pseudomonas* sp. Although the chemical structures of these two compounds are analogous, their producers belong to two completely distinct bacteria phyla, Chloroflexi and Proteobacteria. It can be presumed that there could have been a horizontal gene transfer (HGT) between the two bacterial species. In nature, HGT is a common phenomenon within genetic material transferred among bacteria, even between very distantly related ones (Figure 29). This process allows adaptation to new environments and inheritance through different generations as the cell divides.[136]

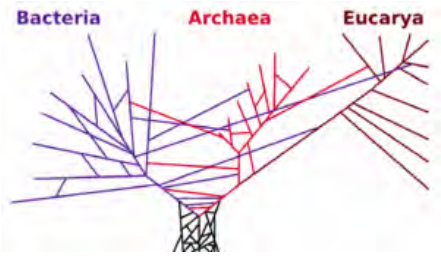
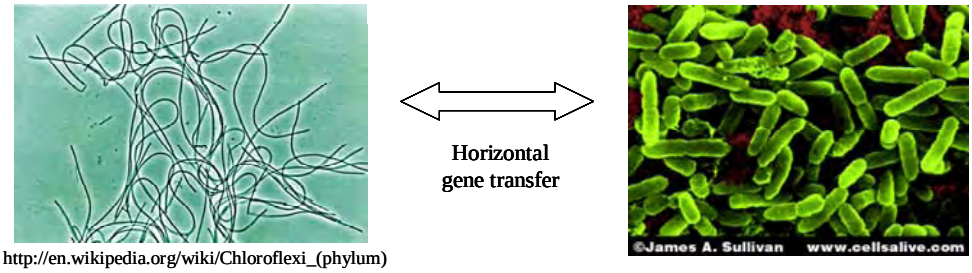


Figure 29. High frequency of HGT between organisms
(http://en.wikipedia.org/wiki/Horizontal_gene_transfer)



http://en.wikipedia.org/wiki/Chloroflexi_(phylum)

(a) (b)

Figure 30. Horizontal gene transfer between (a) Chloroflexi, the onnamide producer and (b) *Pseudomonas* sp., the pederin producer.

3.2.3 Isolation of fosmids, pTSTA4, pTSTA5 and pTSTA6 from the fosmid library of *T. swinhoei*

Onnamide A terminates with an arginine residue, which should be incorporated by an NRPS module. So far, no such module was found, suggesting that the complete onnamide cluster had not been cloned yet. Therefore, isolation of the remaining *onn* gene cluster had to be continued.

As shown in Figure 23, the new PKS fragment on pTSSH3 was larger than that on pTSSH2. Consequently, pTSSH3 was used for the primer walking method (section 7.2.6). The pTSSH3-4 primers derived from the last KS domain of pTSSH3 (Figure 24) were reused to screen the 400,000 clone *T. swinhoei* fosmid library for clones encoding new, further elongated PKS variants.

The rapid PCR-based isolation procedure for obtaining positive clones from fosmid libraries was applied [132] (for a detailed description on this procedure, see section 7.2.4 and 7.2.5).

Screening row and column superpools

As discussed previously, due to the presence of different variants of onnamide-type compounds in *T. swinhoei*, more positive clones had to be isolated in an effort to gain insights into the diversity of onnamide derivatives. For this reason, three positive superpools consisting, each of about 9,000 clones, were picked in the first PCR screening. From each superpool, one positive was selected in every subsequent step. To avoid obtaining coincidental clones with those of already isolated fosmids, the positive pools had to be different from those previously screened for the isolation of pTSSH2 (pool 1G5) and pTSSH3 (pool 4E8). Therefore, the column superpool 14 and two of the row superpools 1E and 3H were selected (Figure 31).

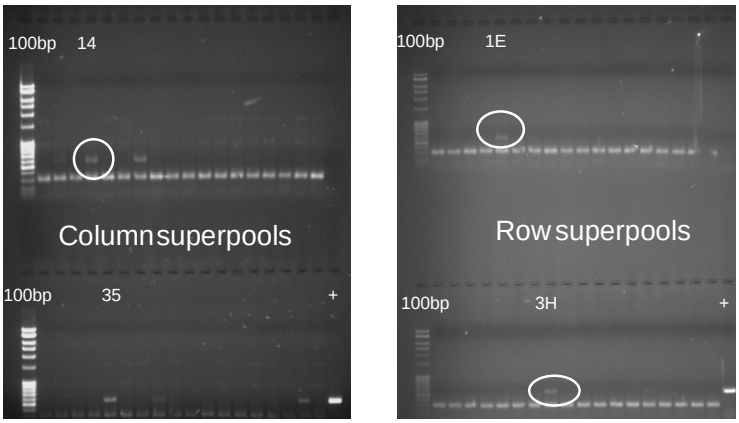


Figure 31. Screening superpools with pTSSH3-4 primers. Agarose gels of electrophoresed PCR are shown. The positive control (+) was the PCR product of pTSSH3.

Screening pools

The nine pools from which each positive superpool was derived were PCR-screened. From the superpools 14 and 1E, only one positive signal appeared whereas from the superpool 3H, there

were three positive signals (Figure 32). Three positive pools were picked to prepare subpools, namely 14G, 1E7 and 3H8.

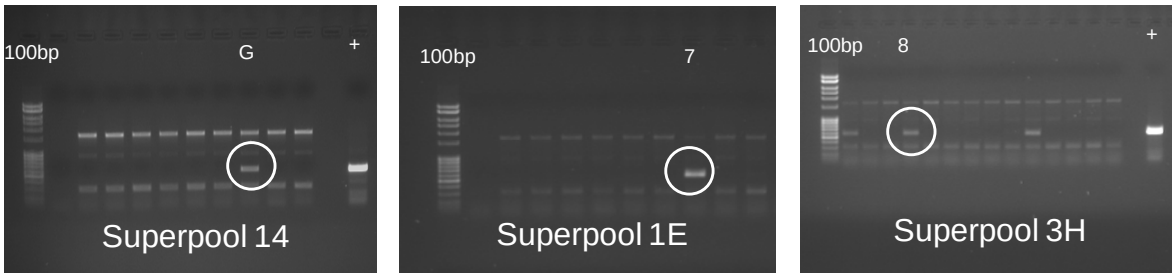


Figure 32. Screening pools with pTSSH3-4 primers. Agarose gels of electrophoresed PCR are shown.

Screening subpools

A dilution 10^4 times from each positive pool was prepared to construct thirty subpools containing about 100 cfus/ml. Therefore, a total of ninety PCRs of subpools constituted from the three pools were carried out to search for positive ones. Several positive signals were detected and three of them were chosen for the following screening step. They were 14G-17, 1E7-11and 3H8-20 (Figure 33).

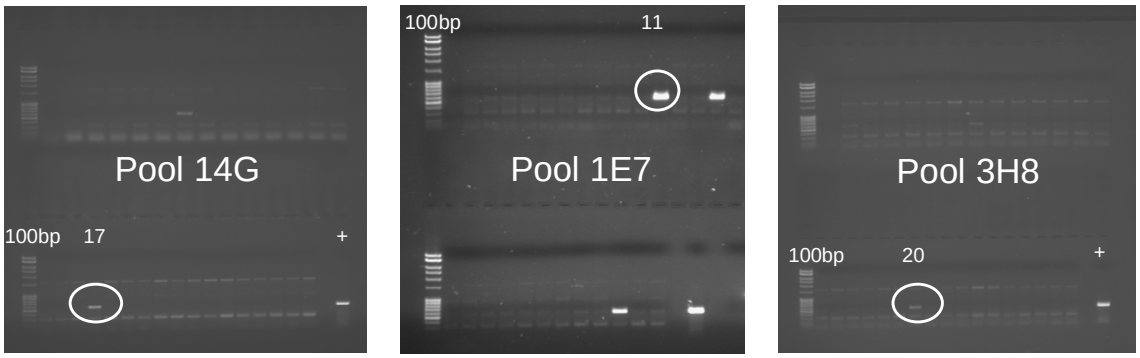


Figure 33. Screening subpools with pTSSH3-4 primers. Agarose gels of electrophoresed PCR are shown.

Screening subsubpools

Each positive subpool was diluted 10^5fold in order to obtain twenty subsubpools consisting of *ca.* 15 cfus/ml. Subsequently, a total of sixty subsubpools derived from the three subpools were subjected to PCR screening. Several positive signals were present, especially in the subpool 14G-17. The three chosen subsubpools were 14G-17-11, 1E7-11-7 and 3H8-20-19 (Figure 34).

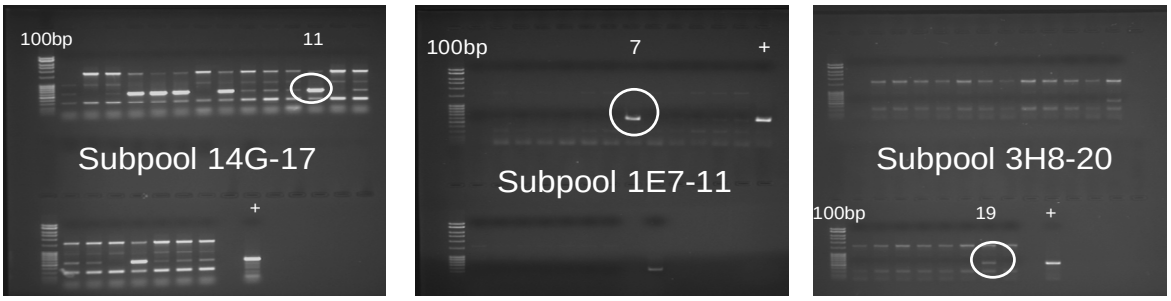


Figure 34. Screening subsubpools with pTSSH3-4 primers. Agarose gels of electrophoresed PCR are shown.

Screening colonies

After 10^5 dilutions of each positive subsubpool were made, 100 µl aliquots containing about 200 cfus were spread on LB agar plate. Colonies from plates were picked for colony PCR. Three positive clones were identified, pTSTA4 **(**1E7-11-7-3**),** pTSTA5 (3H8-20-19-1), and pTSTA6 **(**14G-17-11-1) (Figure 35).

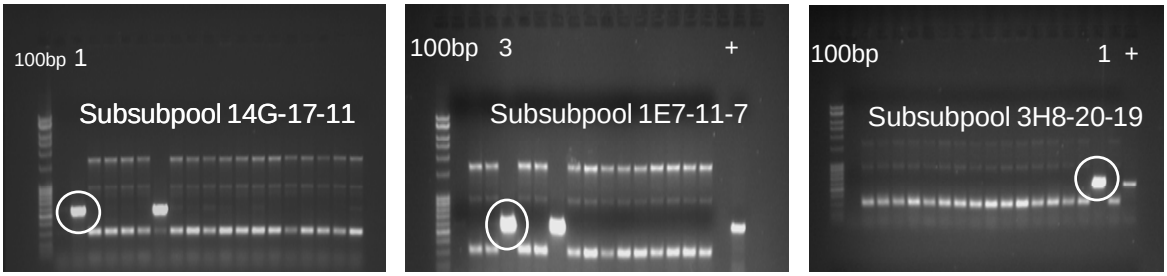


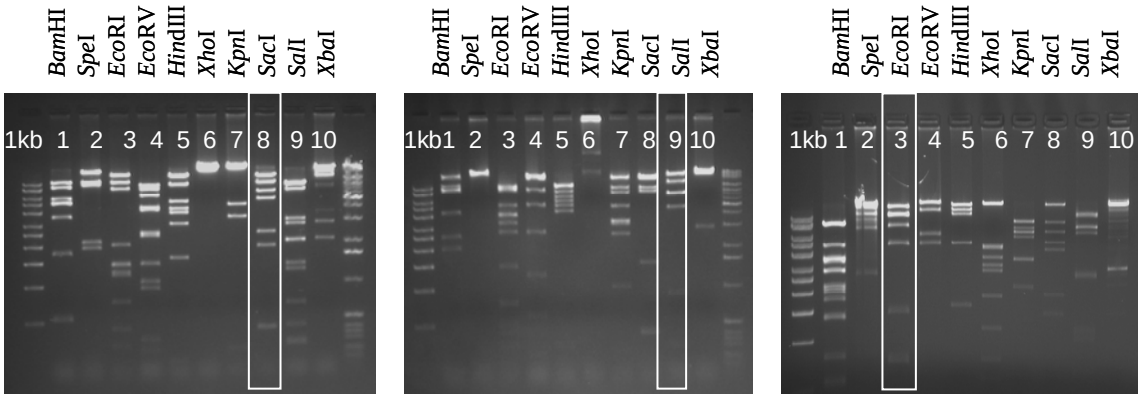
Figure 35. Screening colonies with pTSSH3-4 primers. Agarose gels of electrophoresed PCR are shown.

Subcloning of the fosmids, pTSTA4, pTSTA5 and pTSTA6 for end-sequencing

Each positive fosmid contained a DNA insert of ~ 35-40 kb. In order to obtain primary genetic information, fosmid DNA from the positive clone was digested into a series of smaller fragments that were subcloned and subjected to end-sequencing.

Plasmid DNAs of three positive clones were extracted after inducing with CopyControl induction solution (section 7.2.8). Their restriction maps were obtained by digesting isolated DNA with ten restriction enzymes belonging to the multi cloning site (MCS) of pBluescript SK II (-) (section

7.2.11.1 and Table 49). The most suitable enzymes that sheared DNA into distinguishable fragments were chosen (Figure 36).



(a) (b) (c)

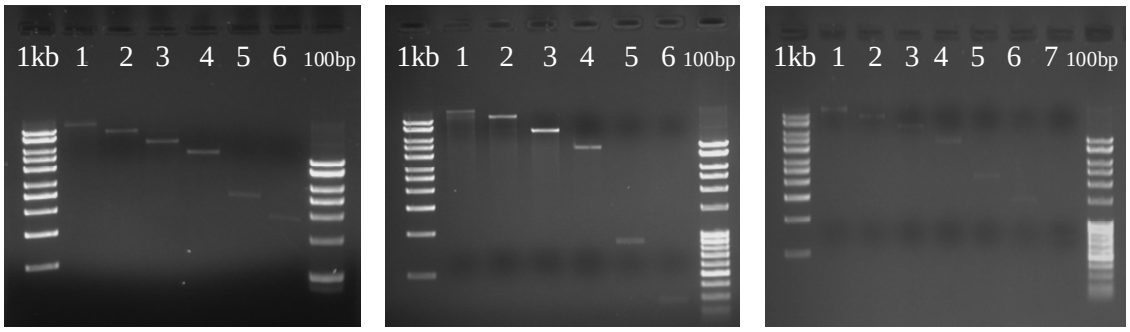
Figure 36. Agarose gel electrophoresis of digestion map of (a) pTSTA4 (b) pTSTA5 and (c) pTSTA6.

The digestion maps of the three fosmids were different from each other and were also different from those of pTSSH2 and pTSSH3. This indicated that the isolated clones contained distinct fosmids from the sponge fosmid library. Based on the separation and the size of digested fragments, *Sac*I (No. 8), *Sal*I (No. 9) and *Eco*RI (No. 3) were chosen for fragmentation of pTSTA4, pTSTA5, and pTSTA6, respectively (Table 9).

Table 9. Digested fragments of pTSTA4, pTSTA5 and pTSTA6.

Fosmid	**Fragment**						
	No. 1 [kb]	**No. 2** [kb]	**No. 3** [kb]	**No. 4** [kb]	**No. 5** [kb]	**No. 6** [kb]	**No. 7** [kb]
pTSTA4	12	10	7	6	2.5	2	
pTSTA5	12	11	7	5	0.9	0.3	
pTSTA6	12	10	8	5	2.5	1.5	0.8

The digested fragments were recovered from agarose gels as described in section 7.2.11.1 and 1 µl of each purified DNA fragment was checked for size and concentration on gel (Figure 37).



(a) (b) (c)

Figure 37. Agarose gel electrophoresis of digested fragments of (a) pTSTA4 (b) pTSTA5 (c) pTSTA6.

The cloning vector pBluescript SK II (-) was digested with the corresponding restriction enzyme *Sac*I, *Sal*I or *Eco*RI and then treated as described in section 7.2.11.2 and 7.2.11.3. The purified DNA fragments were cloned into the vector (section 7.2.11.4) to obtain constructs that were transformed into *E. coli* XL1 blue (section 7.2.14). This yielded a series of colonies containing plasmids which were isolated and sent for end-sequencing using T7 and T3 primers. In addition, in order to get more genetic data, the whole fosmids were also end-sequenced. Since the genomic fragments were all cloned inside pCC1FOS, the end regions of each positive construct pTSTA4, pTSTA5 and pTSTA6 were sequenced with pCC1/pEpiFOS primers (Table 47). The ligation of the two largest fragments (12 kb) of pTSTA4 and pTSTA6 and pTA52 (11 kb) of pTSTA5 with the cloning vector, however, failed. In fact, it is difficult to clone a large DNA insert into pBluescript SK II (-) and the success rate is therefore often very low. Vecscreen of BLAST was used to remove vector sequence from end-sequences, after which the clean sequences were analyzed by BLAST homology searches. Results are shown in Table 10, Table 11, and Table 12.

Table 10. Analysis of pTSTA4 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA42	**Fragment**	**10,000**				
	T3	886	Trans-sialidase [*Trypanosoma cruzi* strain CL Brener]	24/71 (33%)	612-809	13-73
	T7	748	Flavin-dependent oxidoreductase [*Blastopirellula marina* DSM 3645]	93/216 (43%)	17-661	57-272
pTA43	**Fragment**	**7,000**				
	T3	829	Putative lipase [*Methylibium petroleiphilum* PM1]	42/113 (37%)	312-644	65-177
	T7	793	Modification methylase, putative [*Roseobacter denitrificans* OCh 114	22/62 (35%)	601-437	5-66
pTA44	**Fragment**	**6,000**				
	T3	845	Putative beta-glucosidase [*Arthrobacter nicotinovorans*]	26/77 (33%)	682-467	50-125
	T7	751	Similar to Polyketide synthase modules and related proteins [*Geobacter uraniumreducens* Rf4]	78/133 (58%)	1-399	2931-3042
pTA45	**Fragment**	**2,500**				
	T3	913	PilT protein, N-terminal [*Dechloromonas aromatica* RCB]	85/132 (64%)	216-611	1-132
	T7	807	Hypothetical protein FB2170_13046 [*Flavobacteriales bacterium* HTCC2170]	50/111 (45%)	364-696	18-128
pTA46	**Fragment**	**2,000**				
	T3	901	Nonribosomal peptide synthase [*Pseudomonas fluorescens* Pf-5]	135/208 (64%)	172-795	432-636
	T7	779	RhiE protein [*Burkholderia rhizoxina*]	118/219 (53%)	729-73	594-808
pTSTA4	pCC1-FP	809	Modification methylase, putative [*Roseobacter denitrificans*]	22/62 (35%)	587-423	5-66
	pCC1-RP	709	Polyketide synthetase pksP [similarity] - *Bacillus subtilis*	66/150 (44%)	15-464	4259-4395

Table 11. Analysis of pTSTA5 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA51	**Fragment**	**12,000**				
	T3	1,081	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str. 168]	61/185 (32%)	664-110	1140-1321
	T7	1,274	OnnI [symbiont bacterium of *Theonella swinhoei*]	245/344 (71%)	103-1128	4360-4703
pTA53	**Fragment**	**7,000**				
	T3	317	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str.168	45/104 (43%)	2-313	1536-1639
	T7	835	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str. 168]	54/185 (29%)	587-33	1140-1321
pTA54	**Fragment**	**5,000**				
	T3	822	AMP-dependent synthetase and ligase [*Herpetosiphon aurantiacus* ATCC 23779]	97/142 (68%)	426-1	2-143
	T7	826	Putative adenylate cyclase [*Sinorhizobium meliloti* 1021]	30/88 (34%)	426-686	266-353
pTA55	**Fragment**	**900**				
	T7	901	AMP-dependent synthetase and ligase [*Herpetosiphon aurantiacus* ATCC 23779]	153/296 (51%)	6-878	142-433
pTSTA5	pCC1-FP	910	OnnI [symbiont bacterium of *Theonella swinhoei*]	47/165 (28%)	78-572	4358-4522
	pCC1-RP	862	dJ34F7.1.1 (tenascin XB isoform 1) [*Homo sapiens*]	141/587 (24%)	206-691	248-789

Table 12. Analysis of pTSTA6 spot sequences.

Name		Size [bp]	Protein	Identity	Position of query	Position of subject
pTA62	**Fragment**	**10,000**				
	T3	500	ppe family protein [*Mycobacterium bovis* AF2122/97]	38/114 (33%)	174-488	36-147
	T7	389	Hypothetical protein DDBDRAFT_0217012 [*Dictyostelium discoideum* AX4]	73/287 (25%)	34-295	691-952
pTA63	**Fragment**	**8,000**				
	T3	793	Putative oxidoreductase [*Mycobacterium vanbaalenii* PYR-1]	112/210 (53%)	52-681	4-213
	T7	467	Polyketide synthase [*Micromonospora griseorubida*]	24/72 (33%)	13-216	3650-3718
pTA64	**Fragment**	**5,000**				
	T3	834	Unnamed protein product [*Tetraodon nigroviridis*]	10/22 (45%)	464-399	58-79
	T7	824	Hypothetical protein GSPATT00021279001 [*Paramecium tetraurelia* strain d4-2]	21/58 (36%)	277-334	1855-1912
pTA66	**Fragment**	**1,500**				
	T3	836	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str. 168]	91/170 (53%)	509-3	528-695
	T7	865	Beta-ketoacyl synthase [*Geobacter uraniumreducens* Rf4]	113/254 (44%)	5-745	2069-2322
pTA67	**Fragment**	**800**				
	T7	712	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str. 168]	161/233 (69%)	68-766	706-938
pTSTA6	pCC1-FP	630	Y25C1A.3 [*Caenorhabditis elegans*]	60/267 (22%)	66-332	82-348

BLASTX results showed that all of the three newly isolated fosmids contained PKS genes. The PKS gene fragments of pTSTA5 were present in the largest region ~ 19 kb (Table 11), while PKSs of pTSTA4 were encoded in a region < 8 kb (Table 10) and that of pTSTA6 in a region < 10.3 kb (Table 12). PKSs of these fosmids were aligned with those of various microorganisms, the highest identities ranged from 28% to 71%. It was shown that the phylogeny of KS domains in *trans*-AT PKSs correlates with substrate specificities.[47] In order to predict substrates for the sequenced KSs, according to the three closest homologs that belonged to the *trans*-AT system, their corresponding KS domains were determined and placed in a Bayesian cladogram of full-length KS domains from *trans*-AT PKSs (Figure 68). This had to be done for prediction of the roles of the isolated PKS

sequences in the chain reaction of the onnamide biosynthesis (Table 13). There was no KS domain identified in the end-sequences of pTSTA5.

Table 13. KS domains of pTSTA4 and pTSTA6 and their predicted substrates. (Misc) denotes instances where the closest BLASTP hits belong to different clades.

Fosmid	Fragment		Corresponding KS	Identity	Clade	Predicted substrate
	Name	Direction				
pTSTA4	pTA46	T3	BaeKS5	43%	IX	α, β double bond
			BTKS3	43%	IX	
			DszKS7	43%	IX	
		T7	BaeKS6	43%	I	Misc
			DifKS9	46%	V	
			DifKS11	49%	IX	
pTSTA6	pTA66	T3	BaeKS13	41%	X	Non-elongating KSs
			DszKS10	47%	X	
			MlnKS5	38%	IX	
	pTA67	T7	BaeKS13	65%	X	Non-elongating KSs
			DszKS10	57%	X	
			DszKS8	56%	X	

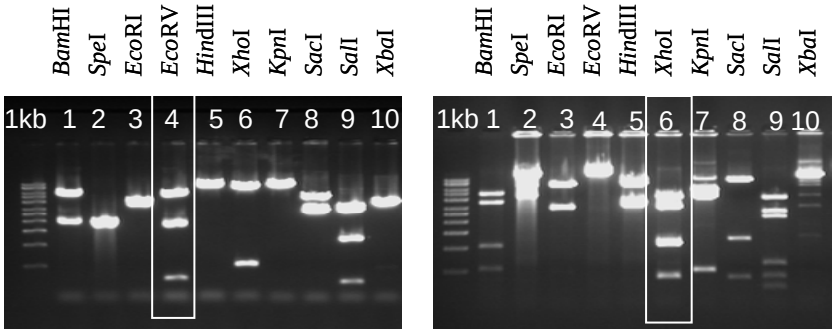
The data in Table 13 shows that in pTSTA4, all three homologs of the KS domain from the T3 end of the fragment pTA46 are similar to KSs belonging to clade IX with an identity of 43%. These KSs recognize unsaturated substrates, which are thus expected for onnamide biosynthesis.

In pTSTA6, the KSs of pTA66 (T3 end) and pTS67 fell into clade X in which KSs lack the conserved motif HGTGT and therefore they do not elongate polyketide chains.

The next step was to identify the terminal regions of the isolated PKS genes. This was necessary to determine whether the complete PKS was cloned. For this reason, fragments carrying a PKS portion on one side and a non-PKS section on the other side, namely pTA44 from the fosmid pTSTA4 and pTA63 from the fosmid pTSTA6, had to be subcloned and end-sequenced.

3.2.4 Subcloning fragments of pTA44 and pTA63

The DNA of pTA44 and pTA63 was isolated from *E. coli* XL1 blue clones. Digestion maps of pTA44 and pTA63 were constructed with the ten previously used restriction enzymes (Table 49). The most appropriate restriction enzymes were *Eco*RV (No. 4) for pTA44 and *Xho*I (No. 6) for pTA63. They were selected to digest DNA into smaller fragments for subcloning (Figure 38).



(a) (b)

Figure 38. Agarose gel electrophoresis of digestion map of (a) pTA44 and (b) pTA63.

Table 14.. Digested fragments of pTA44 and pTA63.

Plasmid	**Fragment**			
	No. 1 [kb]	**No. 2** [kb]	**No. 3** [kb]	**No. 4** [kb]
pTA44 (6 kb)	6	2	0.3	
pTA63 (8 kb)	3.5	3	1.2	0.4

The digested fragments were purified from agarose gel (section 7.2.11) and ligated with the vector pBluescript SK II (-) at the corresponding sites (section 7.2.11.4). After introduction into competent *E. coli* XL1 blue cells, plasmids were isolated and end-sequenced using T7 and T3 primers. The obtained sequences were then subjected to BLAST analysis. Results are shown in Table 15 and Table 16.

Table 15. Analysis of pTA44 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA44-1	**Fragment**	**6,000**				
	T3	942	Putative beta-glucosidase [*Arthrobacter nicotinovorans*]	26/77 (33%)	676-461	50-125
	T7	758	Regulating synaptic membrane exocytosis 2 isoform a [*Homo sapiens*]	33/102 (32%)	484-179	996-1088
pTA44-2	**Fragment**	**2,000**				
	T3	901	Beta-ketoacyl synthase [*Geobacter uraniumreducens* Rf4]	76/129 (58%)	1-387	2935-3054
	T7	879	Peptidase S1 and S6, chymotrypsin/Hap [*Stenotrophomonas maltophilia* R551-3	63/184 (34%)	26-523	177-355
pTA44-3	**Fragment**	**300**				
	T7	740	PksE [*Lechevalieria aerocolonigenes*]	35/102 (34%)	381-677	307-400

Table 16. Analysis of pTA63 spot sequences

Name		Size [bp]	Protein	Identity	Position of query	Position of subject
pTA63-1	**Fragment**	**3,500**				
	T3	743	Leucyl-tRNA synthetase [*Moorella thermoacetica* ATCC 39073]	109/215 (50%)	735-109	465-677
	T7	553	RNA polymerase, sigma-24 subunit, ECF subfamily [*Frankia* sp. EAN1pec]	30/53 (56%)	259-104	36-88
pTA63-3	**Fragment**	**1,200**				
	T3	849	Sigma-70, region 4 type 2 [*Nocardioides* sp. JS614]	39/57 (68%)	1-183	112-172
	T7	629	Polyketide synthase [*Micromonospora griseorubida*]	21/52 (40%)	61-216	3670-3718
pTA63-4	**Fragment**	**400**				
	T7	348	Leucyl-tRNA synthetase, class Ia [*Chloroflexus aurantiacus* J-10-fl]	29/63 (46%)	281-995	865-927

Unexpectedly, no further information about the PKS sequences was obtained when pTA44 and pTA63 were subcloned. In these two plasmids, there were mostly non-PKS genes. PKS genes were present in small fragments, 2 kb fragment of pTA44, and 1.2 kb fragment of pTA63 (Table 15 and Table 16).

The expected final extension module of the *onn* gene cluster in the isolated three fosmids (pTSTA4, pTSTA5 and pTSTA6), an NRPS, was not observed. It was thus necessary to continue to clone further *onn* downstream regions by metagenome walking. A 453 bp fragment from the end of the PKS region of pTA44 was used for designing primers in order to expose its adjacent sequence from the sponge fosmid library. Nevertheless, PCR signals from the superpool screening step to the colony screening step were not clearly specific, even though different PCR conditions, as well as hot-start PCR, and touchdown PCR were used.

Conclusion

Using the specific primers based on the last KS domain of pTSSH3, three positive clones containing the *onn* genes were isolated. Because a border between PKS genes and non-PKS genes could not be identified on the PKS fragments of pTSTA5, this fosmid was not used for the subsequent isolation trials of the rest of the onnamide regions.

In contrast, from pTSTA4 and pTSTA6, the fragments pTA44 and pTA63, respectively, contained both PKS and non-PKS genes. This characteristic was very important to provide information for the isolation of the downstream *onn* gene regions, by applying the primer walking method. According to the alignment between the analyzed sequences and the sequences obtained by BLAST searches, positions of the PKS fragments on pTSTA4 and pTSTA6 were predicted as shown in Figure 39.

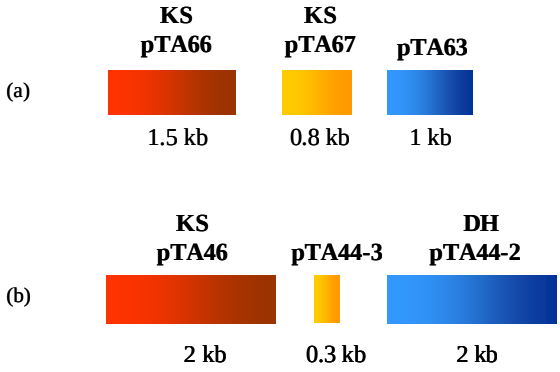


Figure 39. Positions of the PKS fragments on (a) pTSTA6 and (b) pTSTA4.

3.2.5 Isolation of fosmids, pTSTA8, pTSTA9 and pTSTA10 from the fosmid library of *T. swinhoei*

As mentioned, PCR with primers derived from a fragment of pTA44 could not identify any positive clone inside the sponge fosmid library. Therefore, the immediate upstream KS domain (encoded on pTA46) containing the conserved motifs HGTGT and NAHVVLEE was used for designing primers (named pTSTA4-KS primers) for isolation of the remaining *onn* genes. A PCR product with a size of 542 bp should be obtained in case the desired genes are present.

The same procedure was used as described previously (section 3.2.3). In order to avoid coincidental results with previously isolated fosmids, positive pools other than 14, 1E, and 3H of pTSTA4, pTSTA5, and pTSTA6 had to be chosen.

Screening row and column superpools

When screening the row superpools, the outcome was the same as that of the previous screening using the pTSSH3-4 primers; therefore, only the positive column superpools were chosen. They were 25, 53 and 59 (Figure 40).

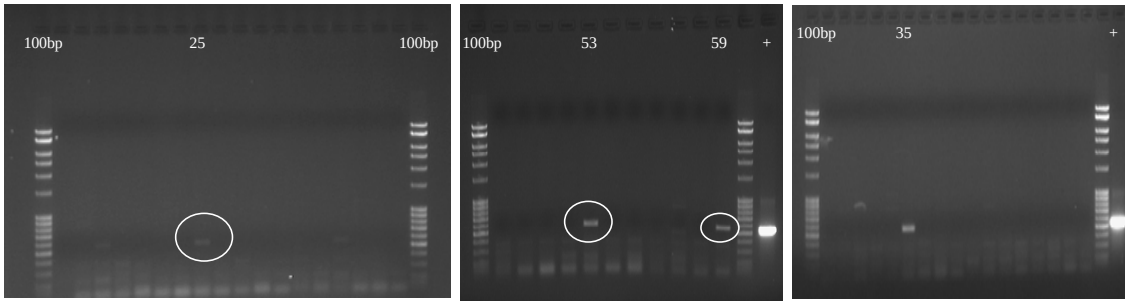


Figure 40. Screening column superpools with pTSTA4-KS primers. Agarose gels of electrophoresed PCR are shown. The positive control (+) was the PCR product of pTSTA4.

Screening pools

PCR-screening of the nine pools constituting the positive column superpool 53, yielded no positive signal. Instead two positive pools from the superpool 59 were selected together with one positive pool from the superpool 25. They were 25A, 59C and 59E (Figure 41).

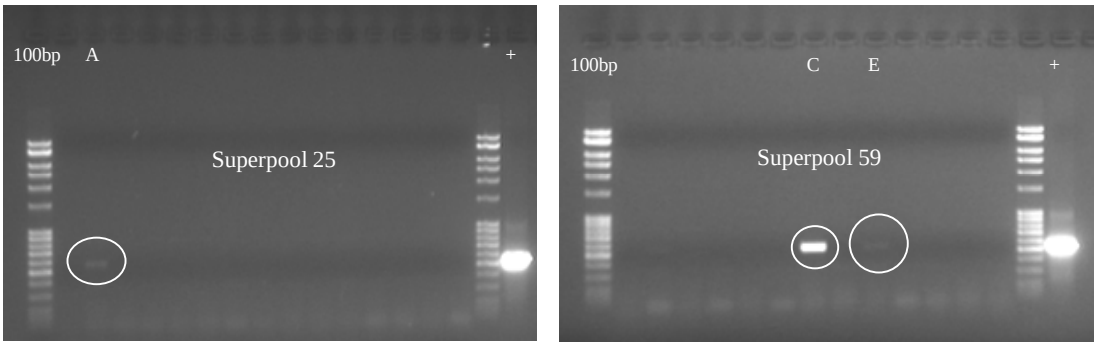


Figure 41. Screening pools with pTSTA4-KS primers. Agarose gels of electrophoresed PCR are shown.

Screening subpools

When applying a normal dilution (10^4fold) in preparation for subpools, only one positive subpool from the pool 59C was detected. Therefore, the dilution ratios of subpools derived from the pools 25A and 59E had to be readjusted. In order to increase the probability that a positive clone is present, the dilution was reduced from 10^4 to 10^3 times. However, PCR signals were still quite faint. Three positive subpools were chosen, namely 59C-27, 25A-18 and 59E-1 (Figure 42).

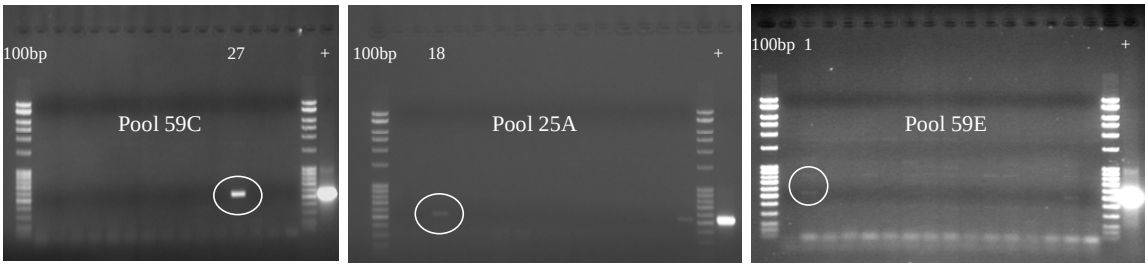


Figure 42. Screening subpools with pTSTA4-KS primers. Agarose gels of electrophoresed PCR are shown.

Screening subsubpools

A 10^5-fold dilution of subpools in order to obtain subsubpools yielded no signal upon PCR from the subpool 59E-1. For this reason, its dilution was decreased 100 times (from 10^5 to 10^3folds). Nevertheless, only a weak band was identified as shown in Figure 43. Three positive subsubpools were picked, namely 59C-27-3, 25A-18-16 and 59E-1-18.

Figure 43. Screening subsubpools with pTSTA4-KS primers. Agarose gels of electrophoresed PCR are shown.

Screening colonies

To perform this step, a 10^5-fold dilution was applied for all the three positive subsubpools. Colonies on the LB plates were subjected to colony PCR. Three positive clones were identified, and named pTSTA8 (59C-27-3-23), pTSTA9 (25A-18-16-14) and pTSTA10 (59E- 1-18-21) (Figure 44).

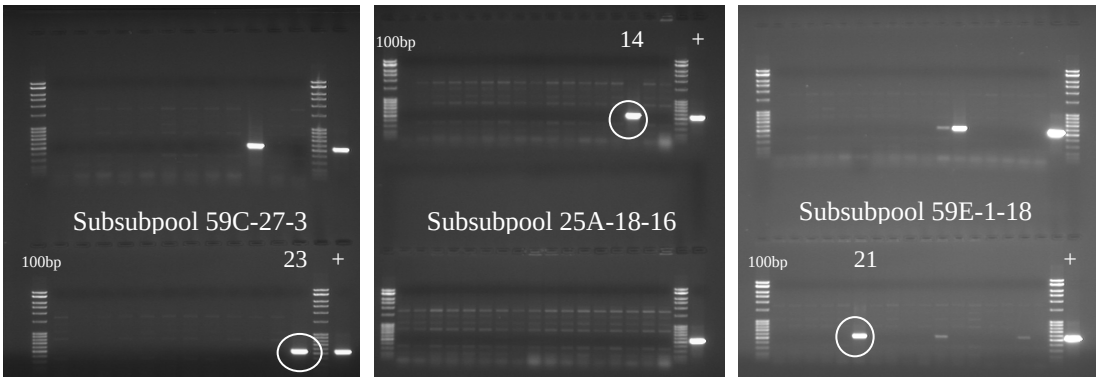
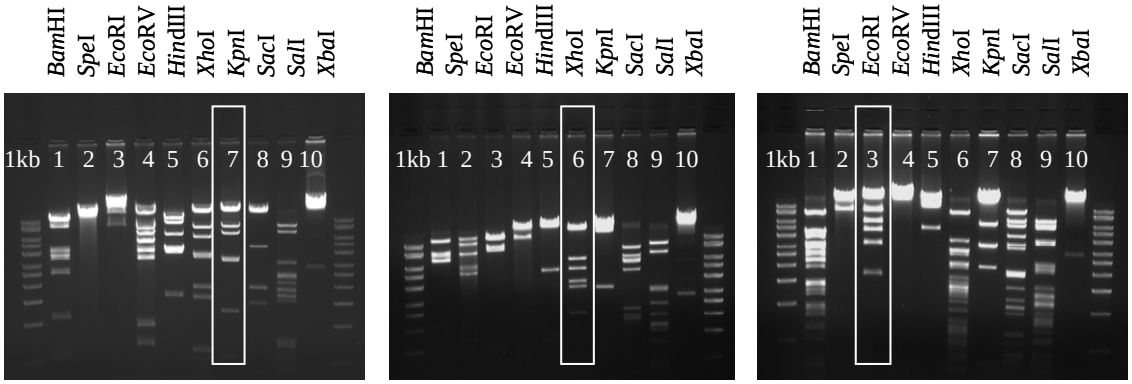


Figure 44. Screening colonies with pTSTA4-KS primers. Agarose gels of electrophoresed PCR are shown.

Subcloning of the fosmids, pTSTA8, pTSTA9 and pTSTA10 for end-sequencing

Digestion maps of the three isolated positive clones were created using the ten restriction enzymes shown in Table 49. These maps were completely different from those of the previously cloned fosmids, meaning they contained new sequence information (Figure 45). According to the digestion patterns, *Kpn*I (No. 7) was the best choice for fragmentation of pTSTA8, *Xho*I (No. 6) for pTSTA9 and *Eco*RI (No. 3) for pTSTA10 (Figure 45 and Table 17), respectively.



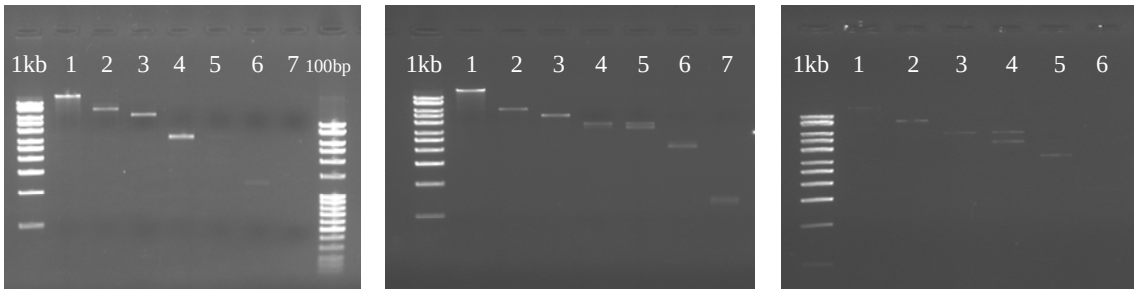
(a) (b) (c)

Figure 45. Agarose gel electrophoresis of digestion map of (a) pTSTA8 (b) pTSTA9 (c) pTSTA10.

Table 17. Digested fragments of pTSTA8, pTSTA9 and pTSTA10.

Fosmid	**Fragment**						
	No. 1 [kb]	**No. 2** [kb]	**No. 3** [kb]	**No. 4** [kb]	**No. 5** [kb]	**No. 6** [kb]	**No. 7** [kb]
pTSTA8	12	7	6	3.5	1.8	1.2	0.4
pTSTA9	12	6	5	4	3.5	2.2	0.7
pTSTA10	12	8	6	5	3.5	1.7	

After recovering DNA fragments from agarose gel using the QIAquick gel extraction kit (section 7.2.11.1), 1 µl of each 15 µl DNA sample was checked on agarose gel (Figure 46).



(a) (b) (c)

Figure 46. Agarose gel electrophoresis of digested fragments of (a) pTSTA8 (b) pTSTA9 (c) pTSTA10.

In preparation for sticky-end ligation, the vector pBluescript SK II (-) was digested with the corresponding restriction enzymes and treated as described in sections 7.2.11.2 and 7.2.11.3. The

new constructs were transformed into electrocompetent *E. coli* XL1 blue cells. The ligation of the largest DNA fragment of ~ 12 kb was successful in all three cases. A series of plasmids isolated from the transformed colonies were sent for end-sequencing using T7 and T3 primers. In addition, the fosmid pTSTA8 was also end-sequenced with the forward and reverse primers of pCC1/pEpiFOS. The end-sequences were then subjected to BLAST analysis. Results are shown in Table 18, Table 19, and Table 20.

Table 18. Analysis of pTSTA8 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA81	**Fragment**	**12,000**				
	T3	663	ABC-type oligopeptide transport system, periplasmic component [*alpha proteobacterium* BAL199]	87/209 (41%)	23-646	405-608
	T7	932	Mixed type I polyketide synthase - peptide synthetase [symbiont bacterium of *Paederus fuscipes*]	200/313 (63%)	929-3	2970-3282
pTA82	**Fragment**	**7,000**				
	T3	428	Hypothetical protein MpopDRAFT_2373 [*Methylobacterium populi* BJ001]	16/39 (41%)	117-1	117-155
pTA84	**Fragment**	**3,500**				
	T3	224	Unnamed protein product [*Tetraodon nigroviridis*]	15/43 (34%)	149-21	1138-1180
	T7	433	SopA [*Klebsiella pneumoniae*]	68/105 (64%)	117-428	200-304
pTA85	**Fragment**	**1,800**				
	T3	962	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str. 168]	172/302 (56%)	1-876	2241-2538
	T7	1057	OnnI [symbiont bacterium of *Theonella swinhoei*]	120/320 (37%)	969-13	2490-2792
pTA87	**Fragment**	**400**				
	T3	671	Beta-ketoacyl synthase [*Geobacter uraniumreducens* Rf4]	85/181 (46%)	92-634	4898-5075
pTSTA8	pCC1-FP	750	Beta-ketoacyl synthase [*Geobacter uraniumreducens* Rf4]	109/247 (44%)	730-2	1806-2046
	pCC1-RP	641	L-carnitine dehydratase/bile acid-inducible protein F [*Paracoccus denitrificans* PD1222]	105/201 (52%)	4-606	126-325

Table 19. Analysis of pTSTA9 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA91	**Fragment**	**12,000**				
	T3	796	Beta-ketoacyl synthase [*Geobacter uraniumreducens* Rf4]	99/261 (37%)	783-4	1783-2042
	T7	591	Hypothetical protein Acid_0126 [*Solibacter usitatus* Ellin6076]	75/134 (55%)	62-463	164-293
pTA92	**Fragment**	**6,000**				
	T3	628	Aminodeoxychorismate lyase [*Solibacter usitatus* Ellin6076]	77/202 (38%)	11-616	28-221
	T7	716	Hypothetical protein M23134_03563 [*Microscilla marina* ATCC 23134]	60/176 (34%)	688-185	148-320
pTA93	**Fragment**	**5,000**				
	T3	567	Peptidase M16 domain protein [*Anaeromyxobacter dehalogenans* 2CP-1]	90/165 (54%)	498-4	37-201
	T7	676	Adenosylmethionine--8-amino-7-oxononanoate transaminase [*Agrobacterium tumefaciens* str. C58]	30/60 (50%)	127-297	361-420
pTA94	**Fragment**	**4,000**				
	T3	621	Putative signal transduction histidine kinase [*Acidothermus cellulolyticus* 11B]	43/128 (33%)	1-384	30-157
	T7	813	Integrase family protein [*Methylobacterium extorquens* PA1]	73/136 (53%)	277-681	14-149
pTA95	**Fragment**	**3,500**				
	T3	632	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase [*Limnobacter* sp. MED105]	79/128 (61%)	389-6	18-145
	T7	666	Site-specfic recombinase, phage integrase family protein [*Sulfitobacter* sp. NAS-14.1]	98/209 (46%)	662-39	122-329
pTA96	**Fragment**	**2,200**				
	T3	473	Holliday junction resolvase YqgF [*Opitutaceae bacterium* TAV2]	25/51 (49%)	385-233	51-101
	T7	679	Esterase/lipase/thioesterase family active site[*Nitrosospira multiformis* ATCC 25196]	46/153 (30%)	24-479	143-292
pTA97	**Fragment**	**700**				
	T3	750	Esterase/lipase/thioesterase family active site [*Nitrosospira multiformis* ATCC 25196]	45/155 (29%)	712-257	143-292
	T7	867	Esterase/lipase/thioesterase family active site [*Nitrosospira multiformis* ATCC 25196]	46/153 (30%)	24-479	143-292

Table 20. Analysis of pTSTA10 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA10-1	**Fragment**	**12,000**				
	T3	1,063	Acetate--CoA ligase [*Escherichia coli* SECEC SMS-3-5]	185/185 (100%)	1014-460	468-652
	T7	1,072	AhpC/TSA family protein [*Colwellia psychrerythraea* 34H]	46/122 (37%)	444-79	37-158
pTA10-2	**Fragment**	**8,000**				
	T3	926	Hypothetical protein Acid_3333 [*Solibacter usitatus* Ellin6076]	71/124 (57%)	37-408	225-342
	T7	271	Hypothetical protein [*Oryza sativa* Japonica Group]	18/43 (41%)	216-103	54-96
pTA10-3	**Fragment**	**6,000**				
	T3	926	Ribosomal protein S4, bacterial and organelle form [*Chlorobium limicola* DSM 245]	96/182 (52%)	790-245	24-203
	T7	737	50S ribosomal protein L18 [*Carboxydothermus hydrogenoformans* Z-2901]	69/116 (59%)	36-374	1-116
pTA10-4	**Fragment**	**5,000**				
	T3	994	DNA-directed RNA polymerase subunit alpha [*Solibacter usitatus* Ellin6076]	172/256 (67%)	3-770	68-323
	T7	715	High molecular weight glutenin subunit [*Triticum aestivum*]	32/92 (34%)	48-320	25-101
pTA10-5	**Fragment**	**3,500**				
	T3	1,046	PI-irrepressible alkaline phosphatase [uncultured marine bacterium Ant29B7]	88/193 (45%)	630-64	33-224
	T7	1,065	Conserved hypothetical protein [*Anaeromyxobacter* sp. K]	37/105 (35%)	297-602	39-143
pTA10-6	**Fragment**	**1,700**				
	T3	812	PhoA [uncultured bacterium]	84/263 (31%)	1-762	242-502
	T7	735	Hypothetical protein Acid_3333 [*Solibacter usitatus* Ellin6076]	125/228 (54%)	709-32	2-229

PKSs were found inside the fragments pTA81, pTA85, pTA87 of the fosmid pTSTA8 (Table 18) and inside the fragment pTA91 of the fosmid pTSTA9 (Table 19). As shown in Table 20, no PKS sequence was identified in the fosmid pTSTA10. Additionally, six PCRs were carried out with DNA templates that were plasmids sent for end-sequencing and using the pTSTA4-KS primers. A positive signal was detected with the fragment pTA10-3. This showed that information of the end-sequence is not sufficient for detailed sequence analysis because important data can be missed.

3.2.6 Subcloning fragments of pTA81

Both fosmids pTSTA8 and pTSTA9 possessed a boundary region including PKS genes and non-PKS genes in the largest fragments, pTA81 and pTA91. The pTA91 was fragmented in order to be subcloned for the purpose of obtaining more detailed information about its PKS region. However, no further data could be obtained.

Meanwhile, from the digestion map performed by the ten restriction enzymes (Table 49), *Xho*I (No. 6) was used for shearing pTA81 into four smaller fragments 10 kb, 8 kb, 4 kb, and 2 kb (Figure 47). After applying the methods in section 7.2.11, plasmid DNA from the white colonies was isolated and sent for end-sequencing using T7 and T3 primers. The end-sequences of these four fragments were then analyzed by BLAST homology searches. The result is shown in Table 21.

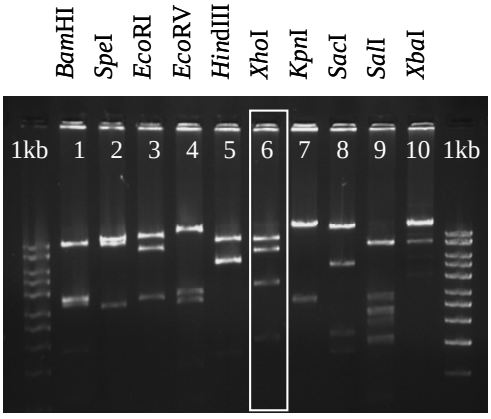
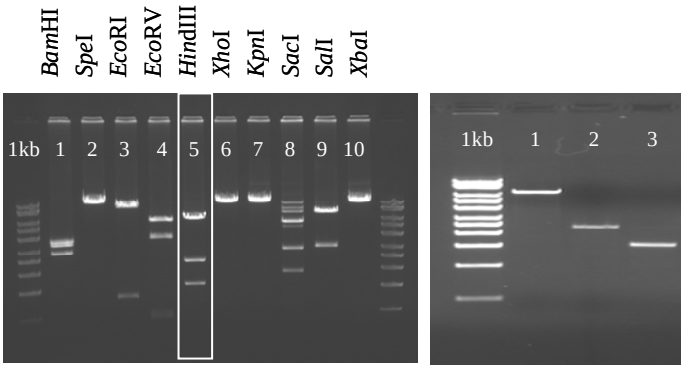


Figure 47. Agarose gel electrophoresis of digestion map of pTA81.

Table 21. Analysis of pTA81 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA81-1	**Fragment**	**10,000**				
	T3	960	Integral membrane protein [*Arthrobacter* sp. FB24]	47/217 (21%)	149-733	132-347
	T7	997	Mixed type I polyketide synthase - peptide synthetase [symbiont bacterium of *Paederus fuscipes*]	172/261 (65%)	772-2	3023-3283
pTA81-2	**Fragment**	**8,000**				
	T3	1,026	Putative reverse transcriptase	19/36 (52%)	440-547	17-107
	T7	438	Hypothetical protein NP2326A [*Natronomonas pharaonis* DSM 2160]	32/108 (29%)	433-113	121-222
pTA81-3	**Fragment**	**4,000**				
	T3	1,256	Carotenoid oxygenase [marine gamma proteobacterium HTCC2143]	208/334 (62%)	2-982	117-443
	T7	316	Carotenoid oxygenase [marine gamma *proteobacterium* HTCC2143]	34/63 (53%)	6-185	118-178
pTA81-4	**Fragment**	**2,000**				
	T3	994	Hypothetical protein R2601_02503 [*Roseovarius* sp. HTCC2601]	57/187 (30%)	145-660	162-334
	T7	1,283	Deacetylases, including yeast histone deacetylase and acetoin utilization protein [marine gamma *proteobacterium* HTCC2080]	147/232 (63%)	13-708	71-302

BLASTX results indicated that only the fragment pTA81-1 contained PKS genes as well as owned the boundary region including PKS section and non-PKS section (Table 21). Thus, pTA81-1 continued to be the subject of further subcloning experiments. A digestion map of pTA81-1 with ten previously used restriction enzymes (Table 49) was set up and *Hin*dIII (No. 5) was chosen to cut pTA81-1 into three fragments 7.5 kb, 2.5 kb and 1.7 kb (Figure 48). Applying the methods described in section 7.2.11, then plasmid DNAs from the transformed colonies were isolated and sent for end-sequencing using T7 and T3 primers. At the end, the end-sequences were subjected to BLAST analysis. The result is shown in Table 22.



(a) (b)

Figure 48. Agarose gel electrophoresis of (a) digestion map of pTA81-1 (b) digested fragments of pTA81-1.

Table 22. Analysis of pTA81-1 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA811-1	**Fragment**	**7,000**				
	T3	878	Nonribosomal peptide synthase [*Pseudomonas fluorescens* Pf-5]	118/236 (50%)	42-740	719-950
	T7	993	Nonribosomal peptide synthase [*Pseudomonas fluorescens* Pf-5]	136/196 (69%)	602-24	5382-5577
pTA811-2	**Fragment**	**2,500**				
	T3	1,303	Nonribosomal peptide synthase [*Pseudomonas fluorescens* Pf-5]	70/247 (28%)	707-21	3-238
	T7	1,150	Polyketide synthase [*Bacillus subtilis subsp. subtilis* str. 168]	26/93 (27%)	367-107	1225-1317
pTA811-3	**Fragment**	**1,700**				
	T3	1,032	Nonribosomal peptide synthase [*Pseudomonas fluorescens* Pf-5]	166/218 (76%)	721-68	514-731
	T7	1,023	Nonribosomal peptide synthase [*Pseudomonas fluorescens* Pf-5	71/217 (32%)	1-651	248-450

Finally, in the last subcloning, the spot sequence data showed that all the three fragments of pTA81-1 contained PKS genes (Table 22). According to the BLASTX results, these three fragments seemed to be assembled continuously. They were aligned with the upstream regions of a NRPS from *Pseudomonas fluorescens Pf-5*, beginning at the 3rd nucleotide and ending at the 5577th nucleotide with an identity ranging from 28% to 76%. According to this alignment, the order of the three fragments inside the fosmid pTSTA8 is the following: pTA811-2, pTA811-3 and pTA811-1.

Subsequently, the three closest homologs of the KS sequences on pTSTA8 that belonged to the *trans*-AT system were identified by BLAST analysis. Their substrates were then predicted through comparison with the Bayesian cladogram in Figure 68 to deduce substrate specificities.

Table 23. KS domains of pTSTA8 and their predicted substrates. (Misc) denotes instances where the closest BLASTP hits belong to different clades.

Plasmid	Direction	Corresponding KS	Identity	Clade	Predicted substrate
pTA81	T7	DifKS13	61%	IX	α, β double bond
		BaeKS5	61%	IX	
		DszKS7	58%	IX	
pTA85	T3	BTKS9	57%	XI	Misc
		BaeKS5	53%	IX	
		BaeKS9	54%	VII	
	T7	OnnKS5	37%	XVI	Misc
		GUKS8	42%	V	
		BaeKS11	34%	V	
pTA81-1	T7	BaeKS5	63%	IX	α, β double bond
		DszKS7	59%	IX	
		BTKS9	63%	XI	
pTA811-1	T3	GUKS9	43%	IX	α, β double bond
		BaeKS6	43%	I	
		DifKS11	50%	IX	
	T7	GUKS9	64%	IX	α, β double bond
		DszKS7	60%	IX	
		BTKS3	67%	IX	
pTA811-3	T3	DszKS10	48%	X	Non-elongating KSs
		BaeKS5	45%	IX	
		BTKS10	44%	X	

Conclusion

The analysis results in Table 23 show that almost all of the homologous KSs with the KS domain in the three fragments, pTA81, pTA81-1, and pTA811-1 (only two exceptions), belonged to the clade IX of which substrate is olefinic. Therefore, PKS genes of these fragments are responsible for formation of double bonds in the onnamide structure. In addition, the consensus sequences of a DH domain in pTA811-2, an ACP domain in pTA811-3, and a KR domain in pTA85 were also found.

However, the final genes of the *onn* gene cluster, such as an NRPS gene for arginine terminus or a thio-esterase domain for PKS cleavage still could not be identified.

PKS genes of the fosmid pTSTA9 were located on a fragment of about 5 kb, whereas PKS fragments of the fosmid pTSTA8 were distributed on a larger region of about 13.4 kb. Based on the above data, the arrangement of PKS fragments on pTSTA8 was predicted and pTA811-1 was at the end (Figure 49).

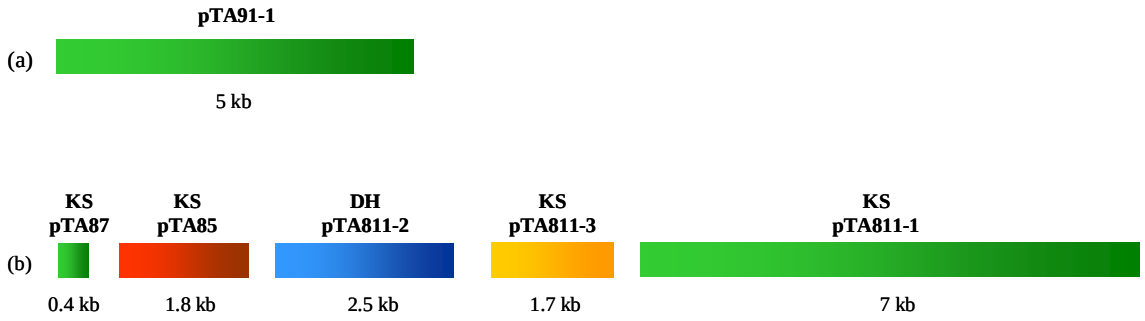


Figure 49. PKS fragments on (a) pTSTA9 and (b) pTSTA8.

3.2.7 Isolation of fosmid pTSTA11 from the fosmid library of *T. swinhoei*

A 434 bp sequence belonging to the last KS domain of the fosmid pTSTA8 was used for primer design (named pTSTA8-KS primers), in order to reveal the remaining region of the *onn* gene cluster. The same primer walking strategy was used to screen the fosmid library for new onnamide downstream regions.

Screening row and column superpools

When using the pTSTA8-KS primers to screen the superpools, PCR signals were very faint despite of some efforts of optimization. However, the column superpool 35 was the only one yielding positive signals with all three primer pairs tested: pTSSH3-4 primers, pTSTA4-KS primers, and pTSTA8-KS primers (Figure 31, Figure 40, and Figure 50). Hence, clones from the superpool 35 were expected to contain not only new PKS regions but also PKSs from pTSTA4 and pTSTA8. Following this presumption, the superpool 35 was subjected to the subsequent screening (Figure 50).

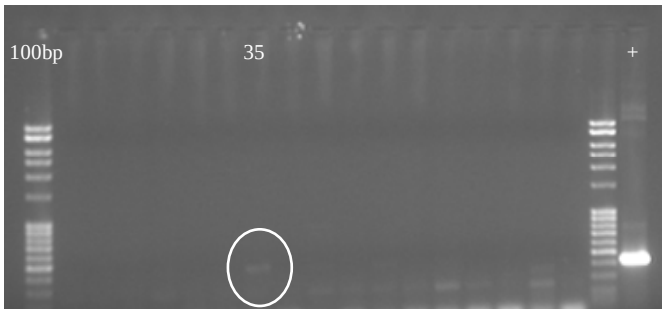


Figure 50. Screening superpools with pTSTA8-KS primers. Agarose gels of electrophoresed PCR are shown. The positive control (+) was the PCR product of pTSTA8.

Screening pools

The PCR product from the superpool 35 was rather faint. Therefore, PCR with 1.5 µl of DNA template and 40 cycles was carried out in the following steps in order to improve amplification. However, when screening the nine pools constituting superpool 35, only one weak band was detected on gel, namely the pool 35H (Figure 51).

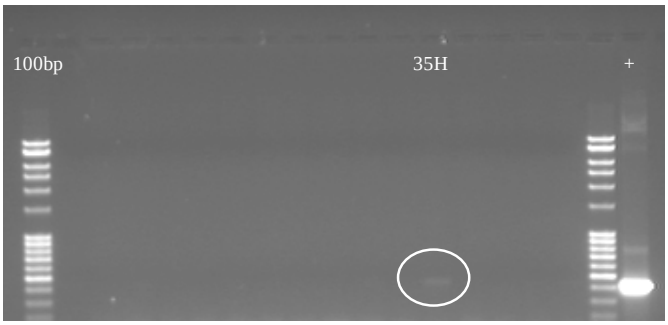


Figure 51. Screening pools with pTSTA8-KS primers. Agarose gels of electrophoresed PCR are shown.

Screening subpools

For the purpose of obtaining a better amplification, the dilution rate of subpools derived from the pool 35H was reduced tenfold than normal (10^4-fold dilution). As expected, PCR signals were stronger and four positive subpools were detected. The subpool 35H7 that yielded the strongest band was diluted further to subsubpools (Figure 52).

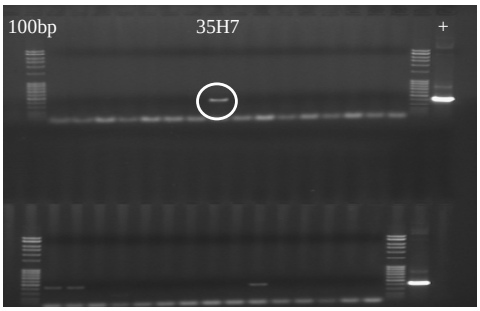


Figure 52. Screening subpools with pTSTA8-KS primers. Agarose gels of electrophoresed PCR are shown.

Screening subsubpools

The positive signal from the subpool 35H7 was strong enough, thus the normal 10^5-fold dilution was used to obtain subsubpools. Four positive subsubpools were detected and the subsubpool 35H7-20 was employed for colony PCR (Figure 53).

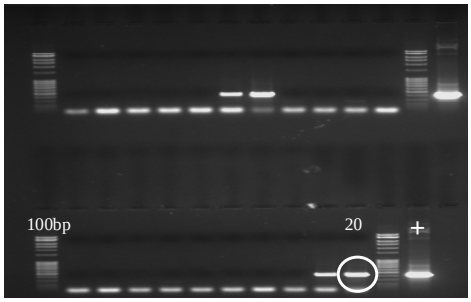


Figure 53. Screening subsubpools with pTSTA8-KS primers. Agarose gels of electrophoresed PCR are shown.

Screening colonies

10^5-fold dilutions of the subsubpool 35H7-20 were prepared and 100 µl aliquots were plated on LB agar to obtain individual colonies. Thirty colony PCRs were carried out and one positive clone was detected, which was named pTSTA11 (35H7-20-29) (Figure 54).

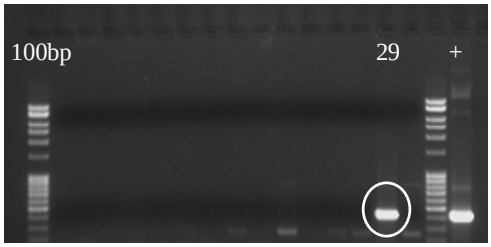
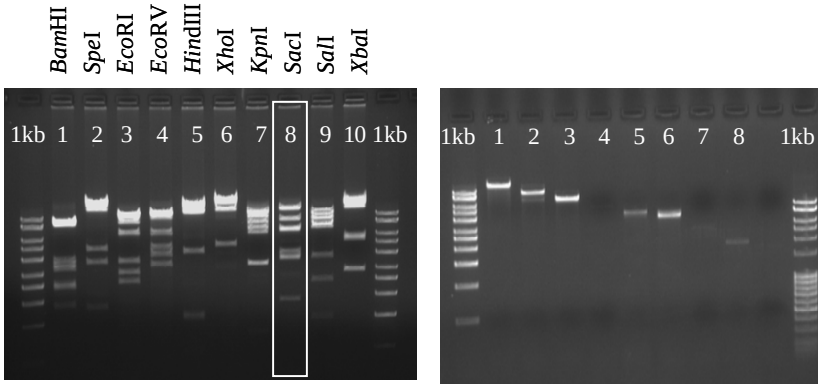


Figure 54. Screening colonies with pTSTA8-KS primers. Agarose gels of electrophoresed PCR are shown.

Subcloning of the fosmid pTSTA11 for end-sequencing

For the purpose of spot sequencing pTSTA11, this fosmid was fragmented. The previous ten restriction enzymes (Table 49) were used to obtain a restriction digestion map of pTSTA11. *Sac*I (No. 8) was selected to digest pTSTA11 into a series of eight smaller fragments 12 kb, 9 kb, 7 kb, 5 kb, 4 kb, 3.5 kb, 2.5 kb, and 1.8 kb (Figure 55). After cloning in pBluescript SK II (-) and electroporation into *E. coli* XL1 blue cells, following the methods in section 7.2.11, these plasmids were sent for end-sequencing using T3 and T7 primers. Their end-sequences were then analyzed using BLAST. The result is shown in Table 24.



(a) (b)

Figure 55. Agarose gel electrophoresis of (a) digestion map of pTSTA11 (b) digested fragments of pTSTA11.

Table 24. Analysis of pTSTA11 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA11-1	**Fragment**	**12,000**				
	T3	1,221	Hypothetical protein SS1G_08824 [*Sclerotinia sclerotiorum* 1980]	32/101 (31%)	873-1169	680-770

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
	T7	155	Sigma54 specific transcriptional regulator, Fis family [*Burkholderia phytofirmans* PsJN]	11/22 (50%)	67-132	73-94
pTA11-2	**Fragment**	**9,000**				
	T3	725	Putative peptidoglycan binding domain [*Shigella boydii*]	228/230 (99%)	723-34	287-516
	T7	282	Chromosome partition protein MukF [*Escherichia albertii* TW07627]	28/52 (53%)	125-274	143-193
pTA11-3	**Fragment**	**7,000**				
	T3	887	Hypothetical protein N9414_06264 [*Nodularia spumigena* CCY9414]	69/171 (40%)	764-279	307-471
	T7	1,165	Non-ribosomal peptide synthetase modules and related proteins [*Nostoc punctiforme* PCC 73102]	164/325 (50%)	6-977	1155-1479
pTA11-4	**Fragment**	**5,000**				
	T3	1,038	Rod shape-determining protein MreC [*Desulfotomaculum reducens* MI-1]	46/165 (27%)	303-785	15-175
	T7	1,227	Rod shape-determining membrane protein; cell elongation [*Pseudoalteromonas tunicata* D2]	89/194 (45%)	1046-465	157-345
pTA11-5	**Fragment**	**4,000**				
	T3	759	Predicted: similar to UNCoordinated family member (unc-51) [*Apis mellifera*]	20/58 (34%)	348-175	356-412
	T7	1,059	Ribonuclease, Rne/Rng family [*Acidobacteria bacterium* Ellin345]	107/261 (40%)	68-844	642-901
pTA11-6	**Fragment**	**3,500**				
	T3	599	Mixed type I polyketide synthase - peptide synthetase [symbiont bacterium of *Paederus fuscipes*]	31/113 (27%)	336-7	1579-4118
	T7	1,227	Thioester reductase [*Anabaena variabilis* ATCC 29413]	24/68 (35%)	210-7	9-75
pTA11-7	**Fragment**	**2,500**				
	T3	921	Glu-tRNAGln amidotransferase subunit A [*Clostridium butyricum*]	92/223 (41%)	845-192	272-483
	T7	1,227	Hypothetical protein Synpcc7942_2461 [*Synechococcus elongatus* PCC 7942]	73/174 (41%)	529-11	54-225
pTA11-8	**Fragment**	**1,800**				
	T3	1,171	RhiE protein [*Burkholderia rhizoxina*]	151/264 (57%)	871-80	549-808
	T7	1,127	RhiE protein [*Burkholderia rhizoxina*]	142/249 (57%)	226-969	296-544

As mentioned previously, the presence of an NRPS module being responsible for the terminal arginine residue is expected to be found at the end of the *onn* PKS (Figure 56).

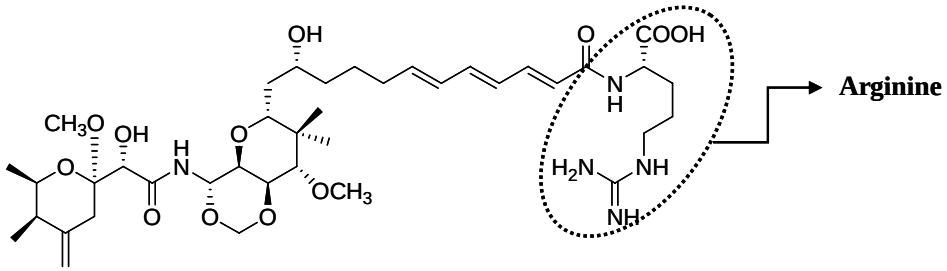


Figure 56. Arginine terminus of onnamide A.

The BLASTX indicated that three fragments from the fosmid pTSTA11contained PKS genes, namely pTA11-3, pTA11-6, and pTA11-8. Remarkably, the T7 end of pTA11-3 was similar to an NRPS module from *Nostoc punctiforme* (Table 24). This sequence was subsequently analyzed using the NRPS predictor program (http://www-ab.informatik.uni-tuebingen.de/toolbox/index.php). Arginine was predicted to be the specific amino acid of the A domain on pTSTA11 with an identity of 90% (see below).

10 amino acid code defined by Stachelhaus et al.
DVLDIGAID-
Alignments of the 10 amino acid code defined by Stachelhaus et al. to the database of known specificities
Score=34, identity=90 with subject O85168_m5__arg

```
DVLDIGAIDK
|| |:|||||
DVADVGAIDK
```

Based on the analysis of the spot sequences, the last part of the *onn* gene cluster seemed to finally be located inside the fosmid pTSTA11. For this reason, this fosmid was sent to *Leibniz Institute for Age Research, Fritz Lipmann Institute (FLI), Genome Analysis* for complete sequencing. The same analysis strategy was used as described previously (section 3.2.1) to analyze the 20,592 nucleotide sequence of pTSTA11.

Table 25. Analysis of the complete pTSTA11 sequence.

Reading frame	Similarity		Identity	Position	Length (aa)
	Protein	Origin			
+1	PKS			1-14976	4,992
-2	Putative transposase	*Burkholderia phymatum* STM815	183/400 (45%)	15675-16926	417
+1	Hypothetical protein N9414_06264	*Nodularia spumigena* CCY9414	118/319 (36%)	17210-18625	472
-1	Integrase, catalytic region	*Polaromonas* sp. JS666	40/60 (66%)	19707-19892	62
-2	Transposase	*Brucella abortus* S19	20/32 (62%)	20923-20592	100

The PKS-NRPS fragment occupied a giant portion of the fosmid pTSTA11. This region comprised 14,976 nucleotides ~ 4,992 amino acids (Table 25). There was an overlapping region of 1,926 nucleotides with *onnI* on the fosmid pTSSH3. This data suggests that the PKS-NRPS genes of pTSTA11 are contiguous with the cloned *onn* genes and that they are definitely the last genes of the *onn* gene cluster. In addition to the upstream clone of pTS1E4 that did not contain any candidate for *onn* genes, the downstream region of pTSTA11 also possessed non-PKS genes. In conclusion, the entire *onn* gene cluster is finally covered.

3.2.8 Discussion on the complete *onn* gene cluster

To obtain the complete *onn* PKS gene cluster, a 139.7 kb genomic region was cloned from the metagenomic library of the sponge *T. swinhoei* and completely sequenced. A total of ten fosmids related to the *onn* system were isolated from the 400,000 clone library: pTS1E4, pTSSH1-3, pTSTA4-6, pTSTA8-9, and pTSTA11. The onnamide biosynthetic genes were distributed on a 98.8 kb region belonging to the three fosmids, pTS1E4, pTSSH3, and pTSTA11; not to mention their partial presence on the other seven fosmids. Altogether, multiple related variants of PKS clusters were identified, which is unprecedented in polyketide research. The detection of numerous PKS variants is in agreement with the fact that a variety of onnamide-type compounds are found in *T. swinhoei* (Figure 57). These compounds differ with respect to the length of the polyketide termini, which closely corresponding parts to the presence of PKSs with distinct termini. The results suggest that different members of the symbiont community carry distinct gene clusters and that these

bacteria jointly synthesize a library of bioactive polyketides. A similar scenario was recently shown in the tunicate *Lissoclinum patella* being the source of several cycle peptide patellamides. In a study, multiple variants of the patellamide gene cluster were identified in a community of *Prochloron* spp. belonging to cyanobacterial symbionts.[67]

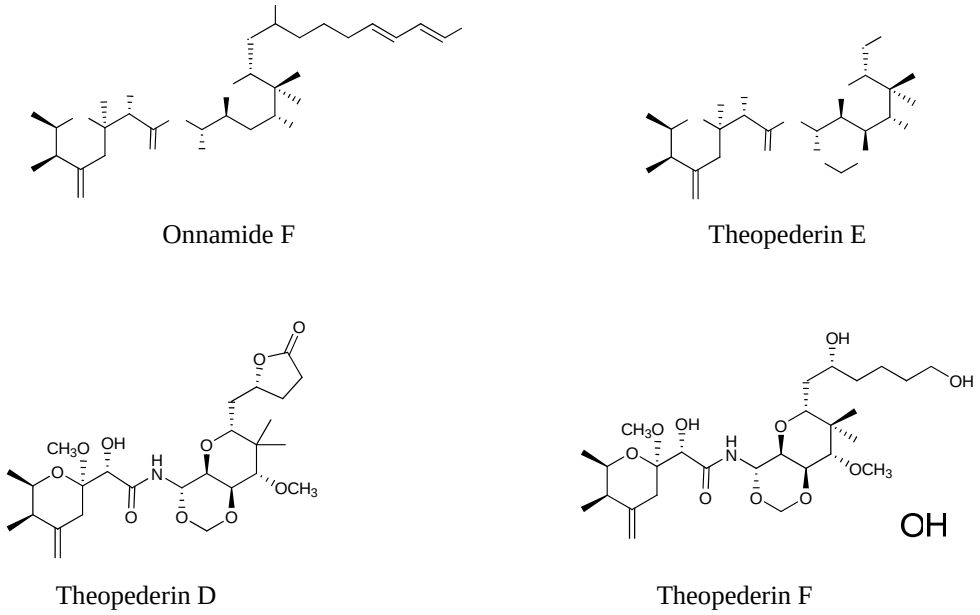


Figure 57. Some onnamide-type compounds found in the sponge *T. swinhoei*.

Ten putative genes, *onnA*-*onnJ* (Figure 59), were identified with a total size of 63,788 bp and an overall GC content of 54.09%. The distances between these genes were quite small, such as the case of a distance of 16 bp between *onnB* and *onnC*, 29 bp between *onnC* and *onnD*, 18 bp between *onnE* and *onnF*, and 29 bp between *onnI* and *onnJ*. This close spacing indicates that the transcribed mRNA is polycistronic, which is a characteristic of bacteria. In addition, the protein encoded by one of the non-*onn* genes was homologous to bacterial transposase (Table 25). Along with the phylogenetic tree of the *ftsZ* gene on pTSSH2 (Figure 27), these features revealed the probable bacterial producers of onnamides, namely members of the phylum *Chloroflexi*.

Interestingly, a direct comparison between PKS domain arrangement of the *onn* gene cluster and that of the *ped* gene cluster, pointed out only few minor differences *e.g.*, the presence of one more KR domain on *pedI*, and one more ACP domain on *onnI* (Table 26). However, the functions of these two domains in their corresponding system are not very clear; they are probably the remains

of an evolutionary event. Furthermore, there is also one additional KS domain in the last module of *pedH*. This KS, however, lacks the conserved HGTGT motif; its role is to pass the polyketide chain only, not to modify substrates. In addition, there is a reversed order between module 7 on OnnI (OnnI7) and module 1 on OnnJ (OnnJ1) of the *onn* genes and module 7 on PedF (PedF7) and module 1 on PedH (PedH1) of the *ped* genes (Table 26). However, the assignment of *pedH* in the pederin biosynthesis has not been elucidated to date.

Despite the similarity in domain organization, nucleotide identities between the *onn* and the *ped* genes were rather low, ranging from 45% to 65% (Table 26). Furthermore, no AT gene was identified in the sequenced regions of the *onn* gene cluster, whereas two AT genes were found in the pederin cluster. In addition, a homolog to oxygenase encoded by *pedG* was not found in the *onn* system. This is in agreement with the hypothesis that this oxygenase gene is required for the *ped* gene cluster for the cleavage of the intermediate product to form pederin (Figure 62). This is likely similar to the case of oxygenase encoded by *GhrO6* of the griseorhodin gene cluster, in which GhrO6 cleavage carbon-carbon bond to form spiroketal system.[137]

A tandem DH domain is present in the *onnI*, which seems to be responsible for the formation of the tetrahydropyrane ring with the geminal dimethyl group (Figure 60). Some tailoring genes required for post-PKS steps were found inside the *onn* gene cluster, such as the oxygenase gene *onnC*, three methyltransferase genes *onnD*, *onnG*, and *onnH*. There were also two genes, *onnE* and *onnF* encoding for unknown hypothetical protein, which are homologs of *pedK* (Figure 59 and Table 26). In contrast to the *ped* gene cluster in which genes are spread on three separate regions of the genome (*pedLMNOPQR*, *pedIJK*, and *pedABCDEFGH*), the *onn* gene cluster is located on only one region. The disjointed architecture of the pederin cluster might be due to the presence of transposase genes on each corresponding genomic region.

However, some genes required for onnamide biosynthesis still could not be identified in the sequenced region. For example, the genes responsible for the methylene ketal moiety, or additional genes required for formation of the exomethylene group, *e.g.*, stand-alone KS, ACP, and CR genes.

Table 26. Proteins encoded by putative genes on the *onn* gene cluster. The protein sequence of onnamide B can be referred in Appendices.

Protein	**Size (aa)**	**Function**	**Homologous protein**	**Identity (%)**
TP	140	Transposase	MM 1044	52
OnnA	420	HMGS	PedP	63
OnnB	4,376	PKS (domains: GNAT, ACP, KS, KR, MT, ACP, KS, CR, CR, ACP, ACP, KS, ACP)	PedI (domains: GNAT, ACP, KS, KR, MT, ACP, KS, CR, CR, ACP, ACP, KS, **KR**, ACP)	46
OnnC	386	Oxygenase	PedJ	65
OnnD	317	MT	PedA	50
OnnE	356	Unknown	PedK	50
OnnF	340	Unknown	PedK	55
OnnG	321	MT	PedA	53
OnnH	268	MT	PedE	52
OnnI	9,859	PKS (domains: KS, ACP, C, A, PCP, KS, KR, ACP, KS, KR, MT, ACP, **ACP**, KS, DH, DH, KR, ACP, KS, KR, ACP, KS, DH, KR, ACP, KS)	PedF (domains: KS, ACP, C, A, PCP, KS, KR, ACP, KS, KR, MT, ACP, KS, DH, DH, KR, ACP, KS, KR, ACP, KS)	49
OnnJ	4,148	PKS (domains: ACP, KS, DH, ACP, KS, KR, ACP, C, A, PCP, TE)	PedH (domains: ACP, KS, DH, KR, ACP, KS, DH, ACP, KS, KR, ACP, **KS**, ACP, C, A, PCP, TE)	45

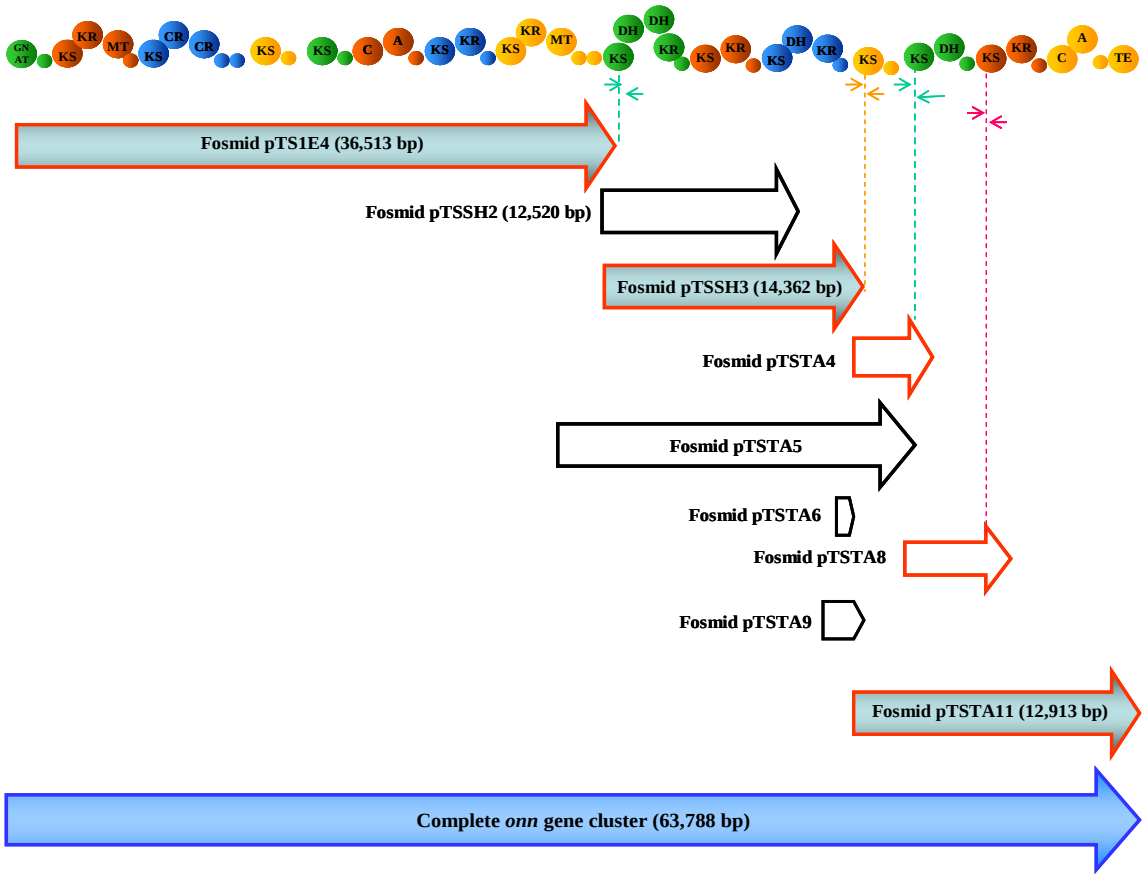


Figure 58 Positions of PKS regions of the isolated fosmids on the *onn* gene cluster. Small arrows together with dotted lines show the positions of primers on the corresponding fosmids. The fosmids with colored background were completely sequenced. Open arrows are partial sequences of large genes (only the 3' termini are characterized).

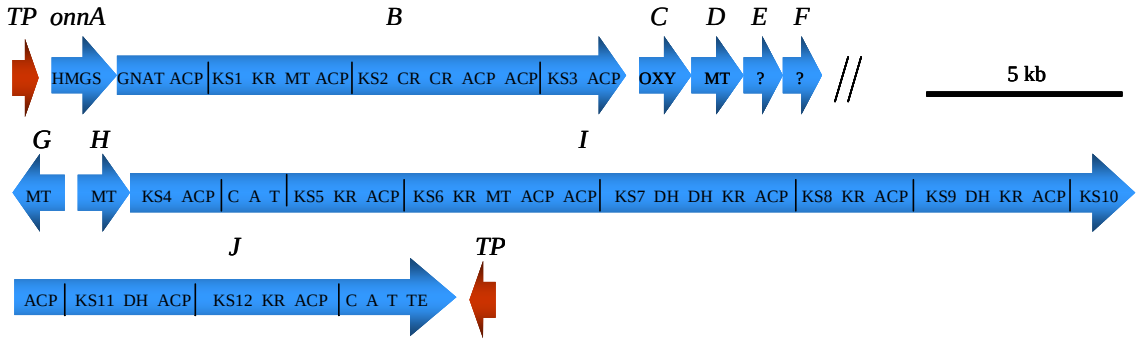


Figure 59. The complete *onn* gene cluster including ten putative genes *onnA*-*onnJ*, (HMGS) 3-hydroxy -3-methylglutaryl-CoA synthase, (GNAT) GCN5-related *N*-acetyltransferase, (*TP*) transposase gene.

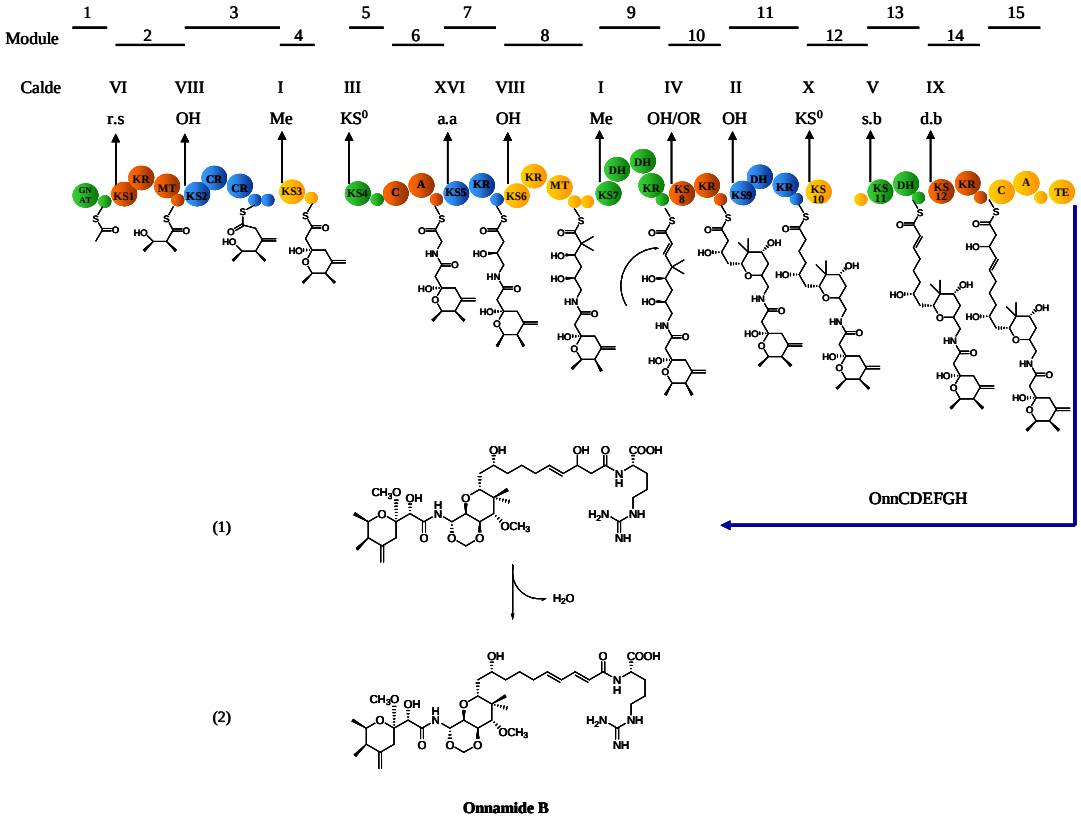


Figure 60. Proposed biosynthetic pathway of onnamide B encoded by the isolated *onn* gene cluster including 15 PKS modules. The deduced KS clades (in roman numbers) and substrate specificities are located above each KS domain. (r.s) related starters (OH) hydroxyl group (Me) methyl branch (a.a) amino acid (s.b) single bond. (1) intermediate metabolite and (2) onnamide B.

Twelve KS domains are present inside the onnamide PKS. Among them, the two KSs, KS numbered 4 and 10 (KS4 and KS10) on *onnI* (Figure 59) share homology to KS^0 (Figure 60), which is a non-elongating KS due to the lack of HGTGT motif. According to a phylogenetic analysis,[47] the KS11 on the module 13 should accept a single-bond substrate. Module 11 is lacking an ER domain that often performs the reduction of double bonds to single bonds. However it has recently been shown that in *trans*-AT PKSs, the ER function is often provided in *trans* by individual enzymes.[138, 139]

The product of this biosynthetic process would be (1) (Figure 60), which is strikingly similar to onnamide B. Its molecule is smaller than onnamide A by two carbon units (Figure 60). Onnamide B was firstly isolated from a sponge *Theonella* sp. and its structure was determined by S. Matsunaga

in 1992.[140] The structural difference of the predicted product to onnamide B might be due to a spontaneous dehydration that could occur in onnamide B biosynthesis or due to the fact that prediction of the last PKS extension step is difficult, since no downstream KS domains exist that can be used for structural prediction.

3.2.9 Expression of the adenylation domains of the *onn* and the *ped* gene clusters

The adenylation (A) domain of an NRPS module is responsible for specific selection and activation of a defined amino acid. Consequently, it plays an important role in determining the structure in a hybrid PK-NRP. Expression of A domains should help us gain insights into the biosynthetic pathways of pederin and onnamides as it is difficult to conduct other functional experiments with symbionts that cannot be cultivated.

There were two A domains on the module PedF2 and PedH6 of the *ped* gene cluster and also two A domains on the module OnnI2 and OnnJ4 of the *onn* gene cluster. Applying the NRPS predictor program (http://www-ab.informatik.uni-tuebingen.de/toolbox/index.php), the specific amino acids of these A domains were predicted (Table 27).

Table 27. Prediction of the amino acid specificity of the NRPS modules of the *onn* and the *ped* gene cluster based on known nonribosomal codes.

Module	**235**	**236**	**239**	**278**	**299**	**301**	**322**	**330**	**331**	**517**	**Similarity (%)**
Glycine consensus	D	I	L	Q	L	G	L	I	W	K	
OnnI2	D	I	L	Q	L	G	L	I	W	K	100
PedF2	D	I	L	Q	L	G	L	I	W	K	100
Arginine consensus	D	V	A	D	I/V	G	A	I	D	K	
OnnJ4	D	V	L	D	I	G	A	I	D	K	90
PedH6	D	A	E	D	I	G	A	I	T	K	70

In agreement with the structure of onnamide B, the prediction results showed that glycine and arginine were the specific amino acids of NRPS modules on the *onn* gene cluster (Figure 61). In contrast to an obvious role of the A domain on *onnJ*, the function of the A domain on *pedH* has not

been yet thoroughly understood due to the lack of an arginine residue in the pederin structure. It is possible that this A domain was still expressed to form an intermediate product of the pederin biosynthesis (Figure 62).

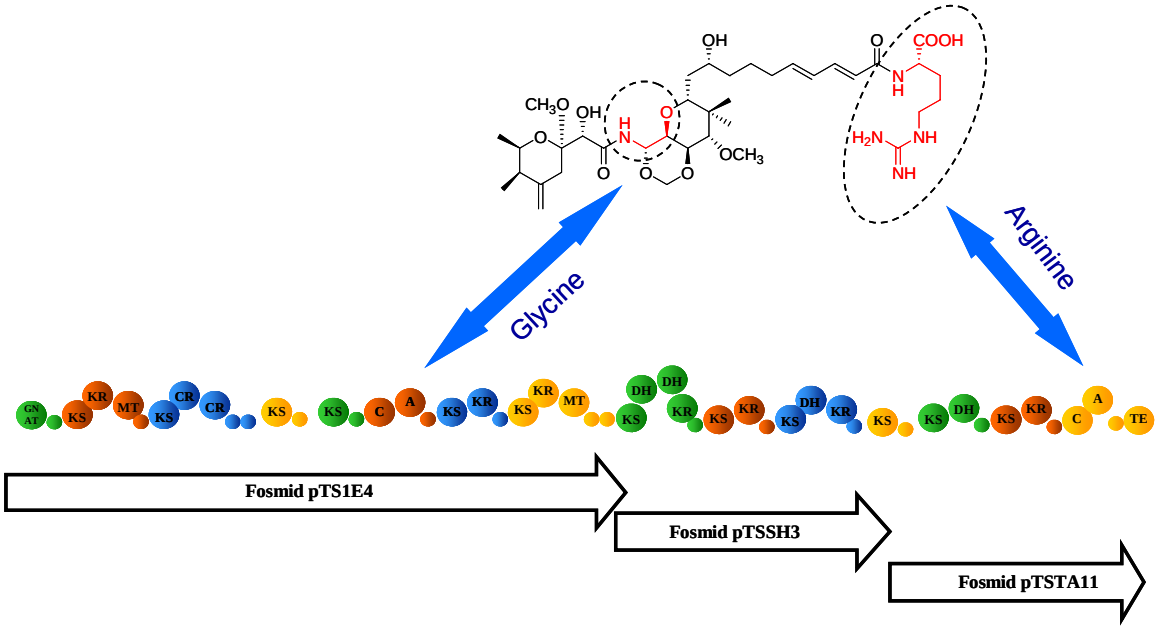


Figure 61 Two A domains inside the biosynthetic gene cluster of onnamide B. They could be responsible for incorporation of the two amino acid residues, glycine and arginine, into the onnamide B structure.

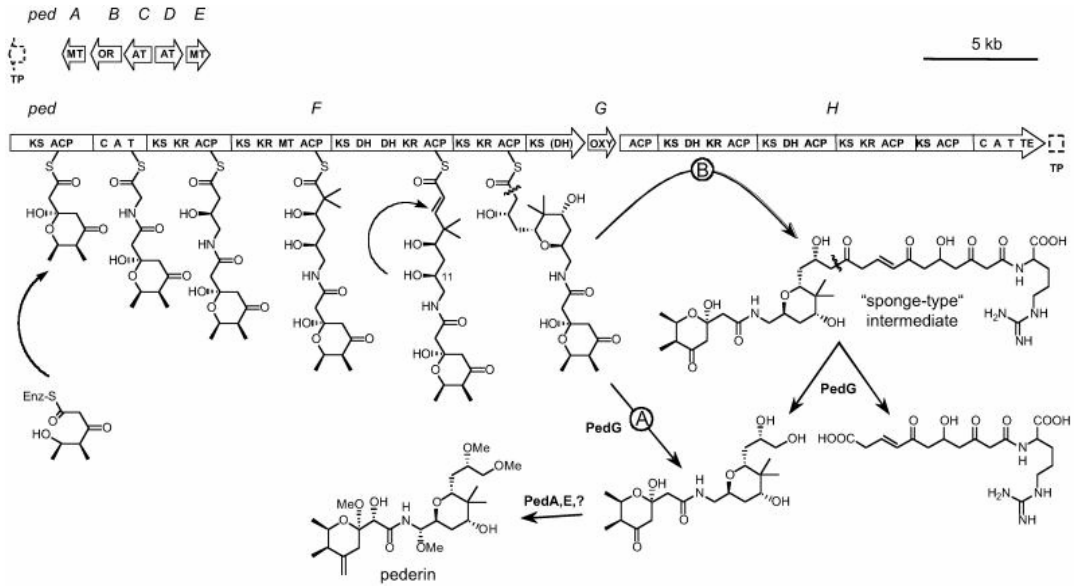


Figure 62. Proposed pederin biosynthetic pathway.[43]

3.2.9.1 Cloning the A domains

Four A domains of the *onn* and the *ped* system were the focus of our A domain expression. Among them, two A domains of the *ped* gene cluster were successfully cloned in our lab by Katrin Zimmermann (unpublished results).

For the purpose of covering completely the active sites of the A domain, a fragment of ~ 500 amino acids was required to be the subject of protein expression (section 1.4 and Table 2). We suspected that the position of the expressed fragment on the amino acid sequence of onnamide B influenced folding of the target protein. Therefore, in order to maximize chances for obtaining functional proteins, the two amino acid sequences of the A domains on the NRPS modules OnnI2 and OnnJ4 were aligned with those of hormaomycin (unpublished work of our group), jamaicamide[141] and pederin. Based on the alignment (Figure 63) and the length of the successfully expressed A domain of the jamaicamide NRPS, the nucleotide sequences of the two A domains that needed to be amplified were identified. The A domain of OnnI2 chosen for expression was from the position 25,180 to 26,821 and that of OnnJ4 was from the position 60,946 to 62,624 on the *onn* gene cluster. Subsequently, we amplified a fragment of 1,658 bp containing the A domain of *onnI* that was named PTA14 and a fragment of 1,694 bp containing the A domain of *onnJ* that was named PTA15 using the primers designed as described in section 7.2.15.1.

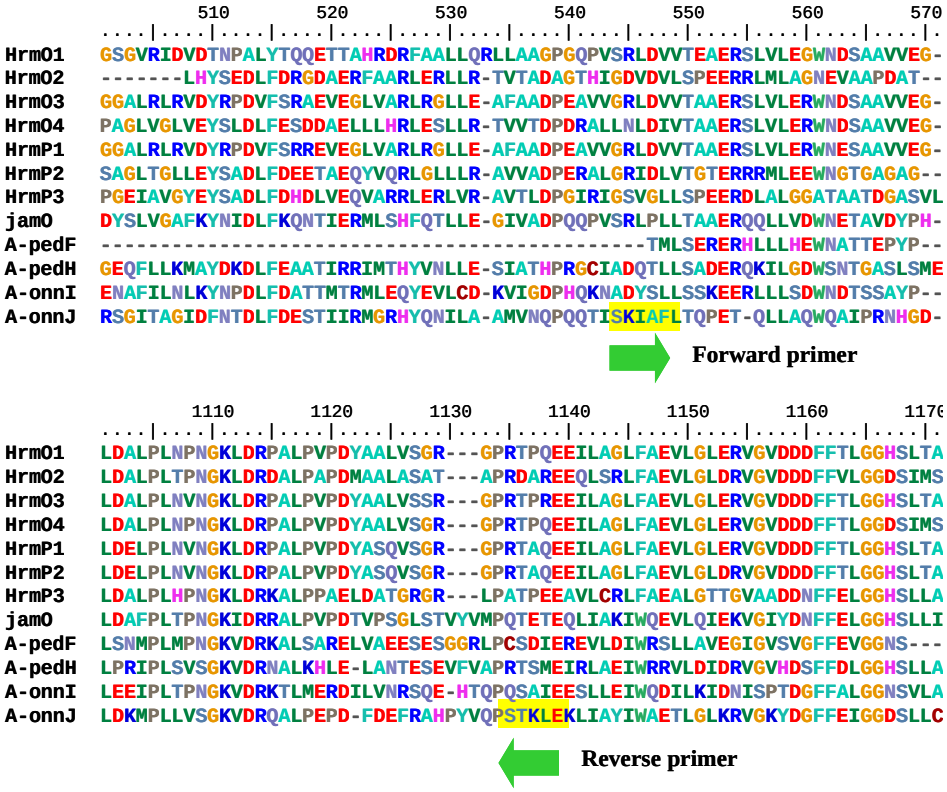


Figure 63. Alignment between two A domains of the *onn* gene cluster and seven A domains of the *hrm* gene cluster, one A domain of *jam* gene cluster, and two A domains of the *ped* gene cluster. The sequences with yellow background are primers used for amplification of the A domains of onnamide B.

His-tag vectors are often employed for protein expression because the His-tag is small, uncharged and usually not affect secretion as well as folding processes of the target protein. Furthermore, the structure and function of the target protein are often not altered when it is expressed with a His-tag. Moreover, it is simple to purify a His-tagged protein applying the Ni-NTA technology.[142, 143] Owing to these advantages, the vector pHis8-3 carrying an octahistidyl tag was chosen for expression of the A domains.

After ligation into pBluescript SK II (-) and sequencing in order to confirm the unchanged ORFs, plasmids pTA14 containing the A domain of *onnI* and pTA15 containing the A domain of *onnJ* were obtained. In order to produce cohesive terminus, plasmid DNAs and the expression vector pHis8-3 were digested with the corresponding restriction enzymes and purified. 1 µl of each DNA sample was checked on agarose gels (Figure 64). Subsequently, the pTA14 and pTA15 inserts were

ligated into the vector pHis8-3 to yield pTA16 containing the A domain of *onnI* and pTA17 containing the A domain of *onnJ*, respectively (section 7.2.15.2).

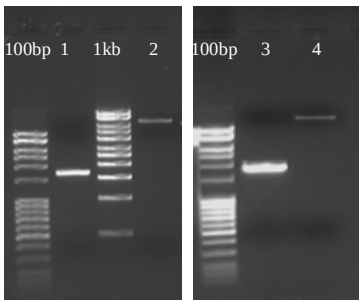
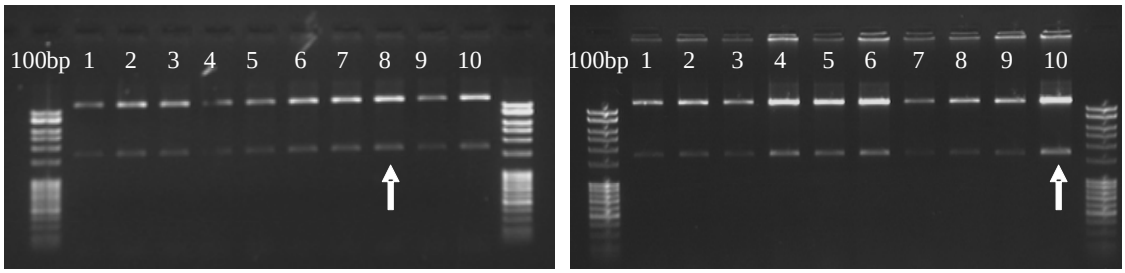


Figure 64. Agarose gel electrophoresis of (1) pTA14 (2) digested pHis8-3 vector with *Eco*RI and *Hin*dIII (3) pTA15 (4) digested pHis8-3 vector with *Bam*HI and *Sal*I.

Firstly, in order to ensure the stability of the recombinant DNA, pTA16 and pTA17 should be transformed into *E. coli* XL1 blue cells (section 7.2.15.3). From each transformation, ten separate colonies were picked to isolate plasmid DNAs that were checked by digestion with *Eco*RI and *Hin*dIII in the case of pTA16 and with *BamH*I and *Sal*I in the case of pTA17. The correct construct was found in all examined clones (Figure 65).

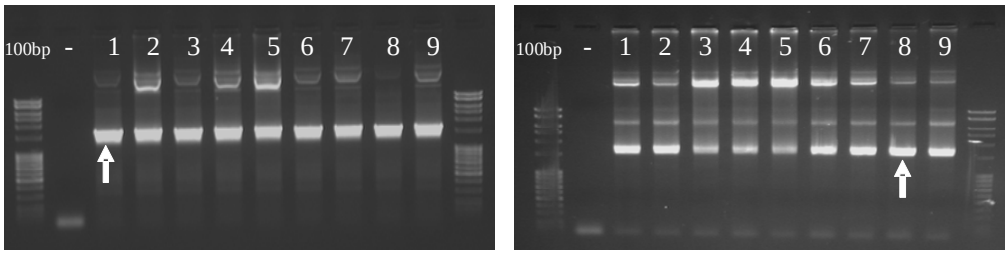


(a) (b)

Figure 65. Agarose gel electrophoresis of checking (a) pTA16 and (b) pTA17 in *E. coli* XL1 blue. Arrows show the correct clones containing pTA16 or pTA17 were selected for next steps.

E. coli is used widely for protein expression due to its advantages of low-cost, easy transformation and high protein yields. In particular, *E. coli* BL21 (DE3) produces highly stable proteins; it is thus a favorite host cell in protein expression. For this reason, pTA16 from colony 8 and pTA17 from colony 10 were randomly selected to be introduced into the host cells *E. coli* BL21 (DE3). Nine single colonies from each transformation were picked in order to isolate plasmid DNA in which the

presence of the A fragments was then checked by PCR. The target PCR products were present in all examined clones. Finally, colony 1 containing pTA16 and colony 8 containing pTA17 were cultivated to express the target proteins (Figure 66).



(a) (b)

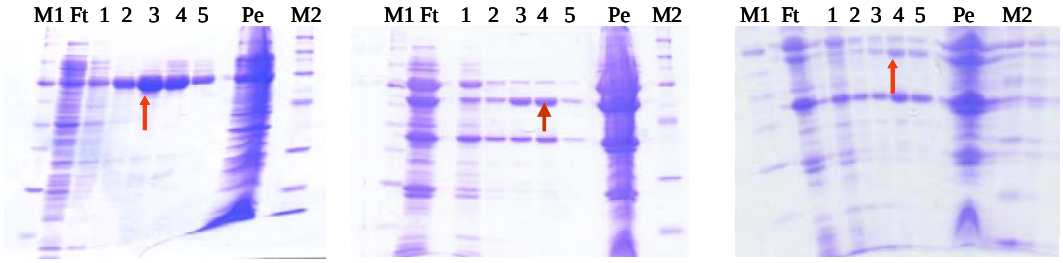
Figure 66. Agarose gel electrophoresis of checking (a) pTA16 and (b) pTA17 in *E. coli* BL21 (DE3). Arrows show the clones containing pTA16 or pTA17 were selected for overexpression.

3.2.9.2 Purification of the expressed proteins

The expressed proteins of the transformed *E. coli* BL21 (DE3) cells were purified using Ni-NTA protein purification system of Qiagen (section 7.2.15.6). However, the correct protein (62.7 kDa) encoded by pTA15 (containing the A domain of *onnJ*) could not be detected on SDS-PAGE gel. In contrast, the protein (61.3 kDa) encoded by pTA14 (containing the A domain of *onnI*) was present the most in the second elution fraction with 100 mM imidazole (Figure 67).

Coexpression of the A domains of the ped *gene cluster with the chaperone plasmid pKEJ7*

Two A domains on the NRPS modules PedF2 and PedH6 of the *ped* gene cluster were previously expressed using the expression vector pHis8-3 and the host cell *E. coli* BL21 (DE3). These expression trials, however, failed because protein inclusion bodies were formed. For this reason, in this section, coexpressions of these two A domains with the chaperone plasmid pKEJ7 were carried out. This plasmid affects the protein folding process and improves the recovery of protein in the soluble fraction.[144] The protein (63.8 kDa) encoded by the A domain on *pedF* (pKZ183-1) and the protein (63.5 kDa) encoded by the A domain on *pedH* (pKZ174-10) were co-expressed with the chaperone plasmid pKEJ7 as described in section 7.2.15.4. Successfully, the solubility of these target proteins significantly increased in comparison with the expression without the chaperon plasmid. Both proteins could be obtained the most in the 4th elution fraction with 150 mM imidazole (Figure 67).



(a) (b) (c)

Figure 67. SDS-PAGE of protein expression experiments with (a) pTA14 (b) pKZ183-1 and (c) pKZ174-10. (Ft) flow-through fraction, (1) 1st elution fraction with 30 mM imidazole, (2) 2nd elution fraction with 50 mM imidazole, (3) 3rd elution fraction with 100 mM imidazole, (4) 4th elution fraction with 150 mM imidazole, (5) 5th elution fraction with 200 mM imidazole and (Pe) pellet (M1) Prestained protein marker of NEB (M2) Protein marker Roti of Roth. Red arrows show the target proteins in the most soluble fractions.

Conclusion

Two NRPS genes of the *ped* gene cluster and one NRPS gene on *onnI* of the *onn* gene cluster were successfully expressed. In particular, the protein encoded by the A domain of *pedH* was present. This was an important result that could help us elucidate the assignment of *pedH* in the process of pederin biosynthesis for the next experiments.

These target proteins were soluble and their concentrations were rather high. Among them, the concentration of the protein encoded by pTA14 was the highest. Even though all the four A domains were expressed under the same conditions, the solubility of the expressed proteins was varied to a large extent. This shows that except for the known factors *e.g.*, the position of the target sequence on the corresponding biosynthetic gene cluster, the expression vector, the host cell as well as the cultivation conditions, there are also other important factors that influence the protein folding process. The chaperone plasmid pKEJ7 was recognized as an effective tool to improve protein solubility. However in addition to the target protein, the chaperone protein was also copurified and has to be separated off before biochemical studies can be conducted (Figure 67 and Figure 113).

Chapter 4
Isolation of an uncharacterized polyketide synthase gene cluster from the metagenome of the sponge *T. swinhoei*

4.1 Introduction

As mentioned in Chapter 1, the structure of polyketides generated by *cis*-AT PKSs can be predicted from the domain architecture and vice versa based on the colinearity rule. In contrast, to date, no rule for *trans*-AT PKSs existed until previously. This hampered biochemical and chemical studies when *trans*-AT PKS genes are discovered. This represents a particular problem for uncultivated bacteria, where knock-out studies cannot be performed. In order to overcome this obstacle, an attempt to find a feasible solution was carried out by Piel and coworkers. The objects of his study were 138 modules of 15 *trans*-AT PKSs submitted in the GenBank database before January 2007, including KS, MT, DH, KR, and ACP domains. A close relationship between KS phylogeny and structures of their substrates was identified (Figure 68). Thereby, applying this rule, polyketide structures can be predicted based on only their KS sequences, and the sequences of other PKS domains are not required.[47]

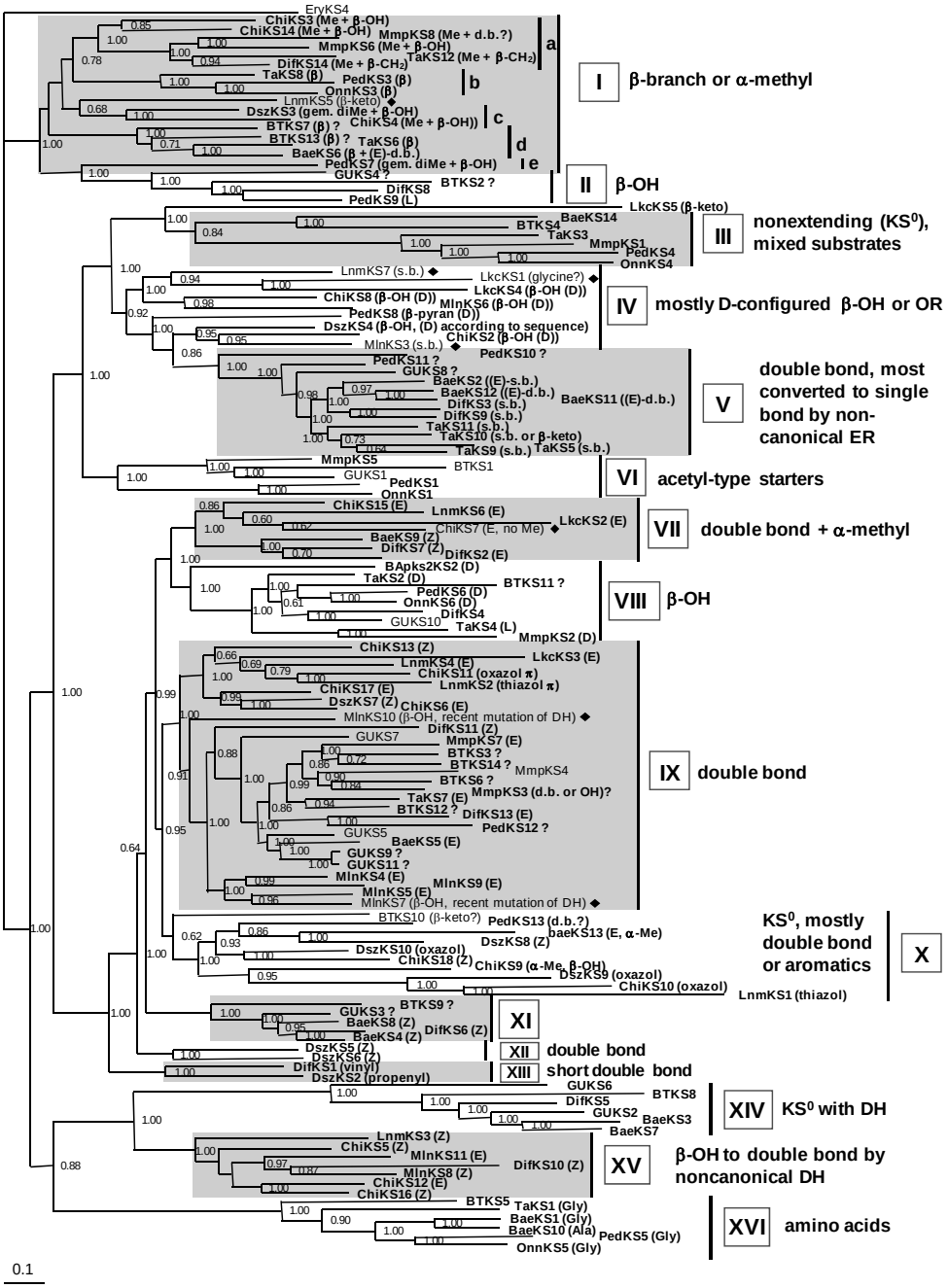


Figure 68. Bayesian cladogram of full-length KS domains from *trans*-AT PKSs.[47] KS numbering refers to their position within the corresponding gene cluster. KS clades are shown in roman numbers together with their main substrate type. Abbreviations: Bae, bacillaene; BT, uncharacterized PKSs from *B. thailandensis*; Chi, chivosazol; Dif, difficidin; Dsz, disorazol; GU, uncharacterized PKSs from *Geobacter uraniumreducens*; Lkc, lankacidin; Lnm, leinamycin; Mln, macrolactin; Mmp, mupirocin; Onn, onnamide; Ped, pederin; Ta, myxovirescin.

Another previous study by Piel *et al.* in 2004 revealed 21 distinct gene fragments from the metagenomic DNA of the sponge *T. swinhoei* using KS primers.[35] All of them showed the nearest homologies to KS domains of the type I PKS. When these sequences were then placed into a general PKS tree, only three sequences clustered into the *trans*-AT clade, namely sponge3, sponge7, and sponge8 (Figure 69).[35] Subsequent studies were carried out in an effort to investigate on the *trans*-AT members that were present inside the metagenome of *T. swinhoei*. Among them, the *onn* genes were successfully isolated using the sponge3 amplicon.

The present study focused on the sponge8 gene fragment because its primary genetic data revealed interesting information. The nucleotide sequence of the sponge8 amplicon was analyzed using BLASTP. Its three highest homologs were KS14 of the difficidin (DifKS14), KS3 of the disorazol (DszKS3), and KS14 of the chivosazol (ChiKS14). All these three KSs fell into the clade I in Figure 68, meaning their substrates are carbon-branched derivatives. These data indicate that genes on the sponge8 fragment do not encode for the unbranched polyketide onnamides.[47] This is a good illustration that the prediction rule is able to extract substrate information from short KS-PCR products, whereby screening a giant fosmid library of a complex metagenomic DNA is unnecessary.

Therefore, in an attempt to gain insights into the diverse metabolites of *T. swinhoei*, a *trans*-AT PKS gene cluster other than the *onn* system was isolated based on the sponge8 gene fragment. The same strategy used for screening the *onn* gene cluster in Chapter 3 was also applied in this case.

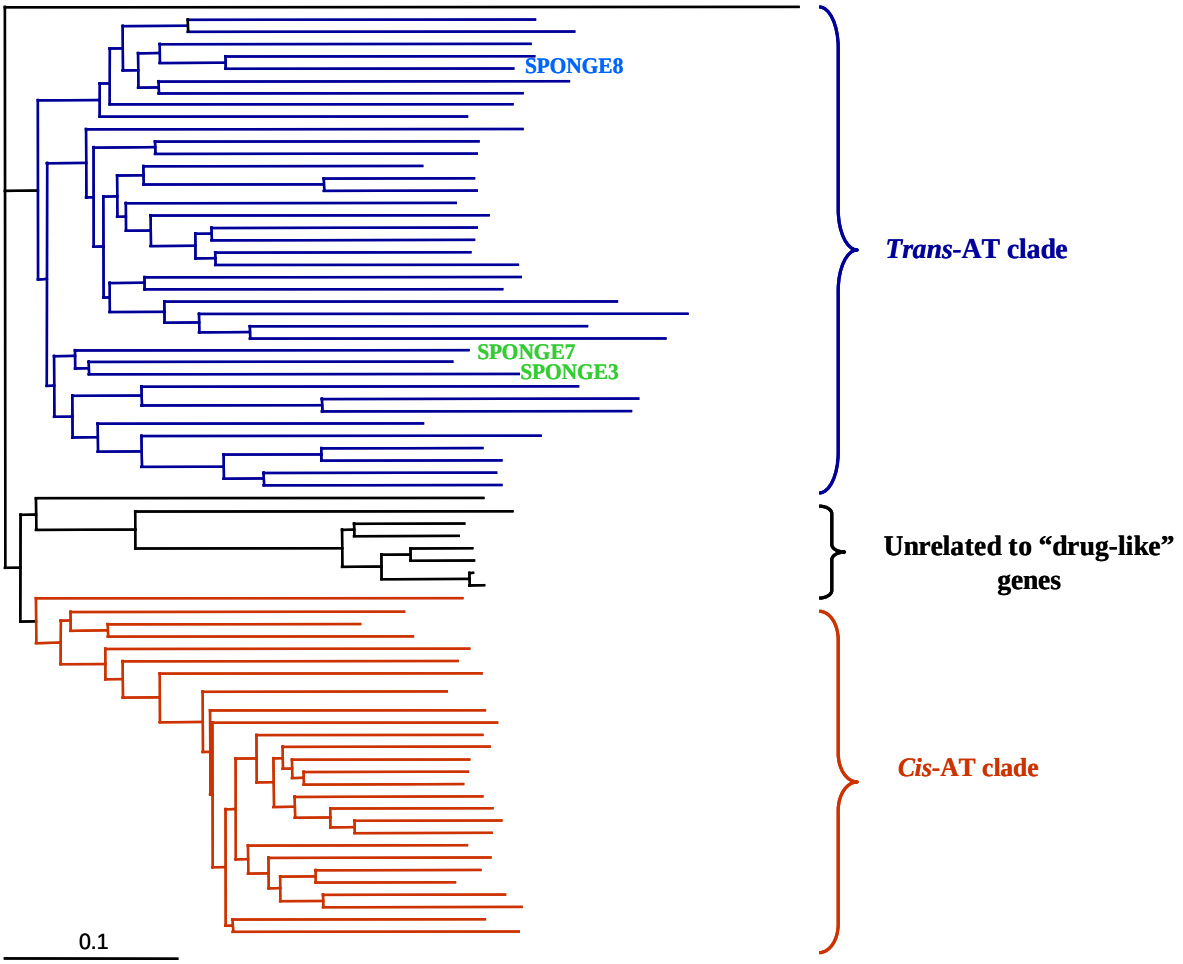


Figure 69. Location of the sponge8 gene fragment in the general PKS tree.[35]

4.2 Results and discussion

4.2.1 Isolation of fosmids, pTSTA2 and pTSTA3 from the fosmid library of *T. swinhoei*

A 425 bp fragment derived from the sponge8 amplicon containing the KS conserved motif CSSSL (sponge8 residues 83-88) was used to design the specific sponge8 primers.

Screening row and column superpools

One positive column superpool and one positive row superpool, 55 and 1A, were subjected to following steps (Figure 70).

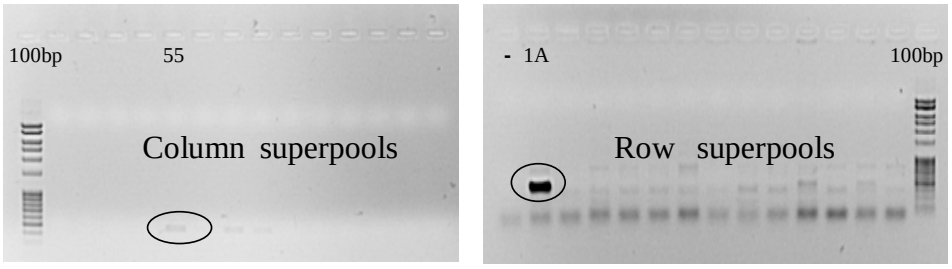


Figure 70. Screening superpools with sponge8 primers. Agarose gels of electrophoresed PCR are shown.

Screening pools

18 pools constituting the superpool 55 and 1A were employed for PCR screening. Two positive pools were picked, namely 55A and 1A3 (Figure 71).

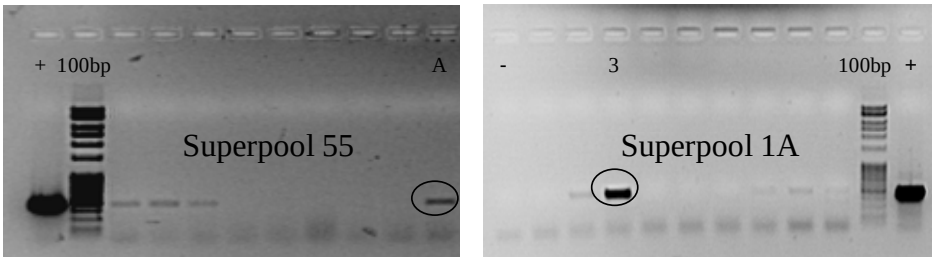


Figure 71. Screening pools with sponge8 primers. Agarose gels of electrophoresed PCR are shown. The positive control (+) was the PCR product of the total sponge DNA.

Screening subpools

10^4-fold dilutions of the positive pools were made in order to construct subpools. Thereafter, from each of thirty subpools, one positive subpool was used for preparation of the next round PCR. They were 55A-26 and 1A3-4 (Figure 72).

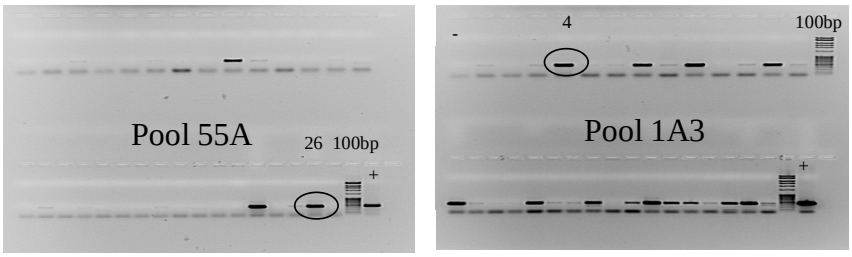


Figure 72. Screening subpools with sponge8 primers. Agarose gels of electrophoresed PCR are shown.

Screening subsubpools

There were several strong PCR signals from the screened subsubpools. One positive subsubpool from each set of twenty was diluted 10^5 times in preparation for colony PCR screening. They were 55A-26-3 and 1A3-4-6 (Figure 73).

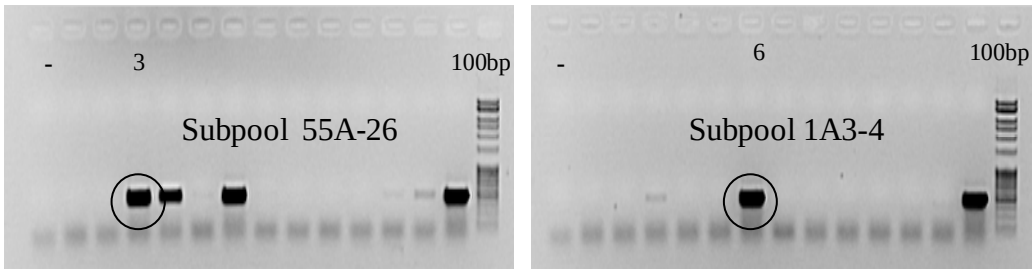


Figure 73. Screening subsubpools with sponge8 primers. Agarose gels of electrophoresed PCR are shown.

Screening colonies

After the final round of screening, two positive clones were identified, 55A-26-3-5 was named pTSTA2 and 1A3-4-6-13 was named pTSTA3, respectively (Figure 74).

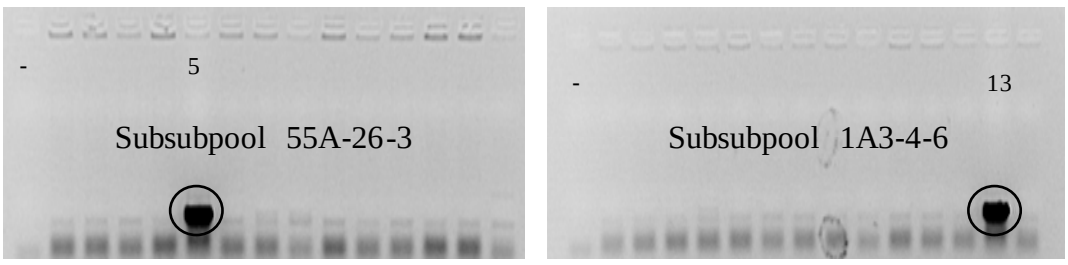
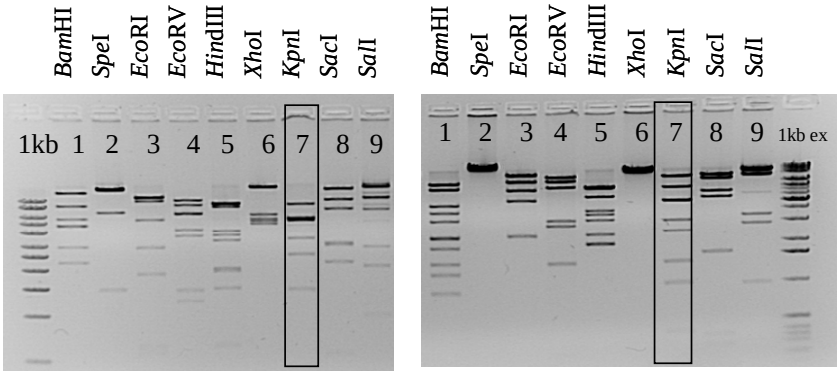


Figure 74. Screening colonies with sponge8 primers. Agarose gels of electrophoresed PCR are shown.

Subcloning of the fosmids, pTSTA2 and pTSTA3 for end-sequencing

After induction to high copy number, DNA of pTSTA2 and pTSTA3 was isolated and digested with nine restriction enzymes (Table 49) to obtain restriction maps. *Kpn*I (No. 7) was the best choice for fragmentation of both fosmids (Figure 75).



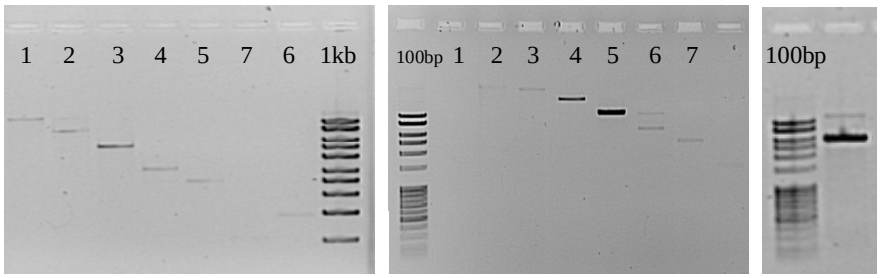
(a) (b)

Figure 75. Agarose gel electrophoresis of digestion map of (a) pTSTA2 and (b) pTSTA3. (1 kb ex) 1 kb extension ladder.

Table 28. Digested fragments of pTSTA2 and pTSTA3.

Fosmid	Fragment [kb]						
	No. 1 [kb]	**No. 2** [kb]	**No. 3** [kb]	**No. 4** [kb]	**No. 5** [kb]	**No. 6** [kb]	**No. 7** [kb]
pTSTA2	10	8	5	3	2.5	1.5	1
pTSTA3	10	9	8	5	3.5	2.5	1.5

The DNA fragments of the two above fosmids (Table 28) and the cloning vector pBluescript SK II (-) were treated according to the methods described in section 7.2.11. 1 µl of each DNA was checked on agarose gel (Figure 76).



(a) (b) (c)

Figure 76. Agarose gel electrophoresis of digested fragments of (a) pTSTA2 (b) pTSTA3 and (c) digested pBluescript SK II (-) with *Kpn*I.

Following the experiments described in section 7.2.11.4 and 7.2.14, the plasmid DNAs from white colonies of each transformation were isolated and sent for end-sequencing using T7 and T3 primers. Among them, the plasmid containing fragment 2 of pTSTA3 could not be identified. The fosmid pTSTA2 and pTSTA3 were also sent for end-sequencing using pCC1/pEpiFOS primers in an attempt to get more genetic information. The end-sequences from these two fosmids were then subjected to BLAST analysis.

According to the BLASTX results (Table 29 and Table 30), on pTSTA2, the fragment pTA24 overlapped with pTA25 and the same was observed in the case of fragments pTA31 and pTA33 on pTSTA3. In spite of this, PKS genes were still found in large regions, > 13 kb of pTSTA2 and ~ 19 kb of pTSTA3. The highest identities between PKSs of pTSTA2 and other PKSs ranged from 30% to 83%, whereas in the case of pTSTA3 they ranged from 30% to 84%. In order to analyze the KS domains, the three closest homologs belonging to *trans*-AT PKSs of each sequence were selected from BLASTP. Afterwards, these KSs were placed into the cladogram shown in Figure 68 for prediction of their substrates (Table 31).

Table 29. Analysis of pTSTA2 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA21	**Fragment**	**10,000**				
	T3	228	rCG48889, isoform CRA_a [*Rattus norvegicus*]	34/121 (28%)	110-228	370-485
	T7	150	Transposase, IS4 [*Polaromonas* sp. JS666]	10/20 (50%)	106-47	24-43

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA22	**Fragment**	**8,000**				
	T3	702	Modular polyketide synthase ketoacyl synthase domain [symbiont bacterium of *Theonella swinhoei*]	178/212 (83%)	645-10	4-215
	T7	805	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	91/183 (49%)	206-754	1074-1254
pTA23	**Fragment**	**5,000**				
	T3	702	Beta-ketoacyl synthase [*Burkholderia ubonensis* Bu]	76/201 (37%)	609-10	713-911
	T7	802	Hypothetical 29.4 kd integral membrane protein (possibly fragmentous) [*Rhizobium* sp. NGR234]	11/24 (45%)	210-281	97-120
pTA24	**Fragment**	**3,000**				
	T3	804	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	145/233 (62%)	778-83	3995-4227
	T7	901	Polyketide synthase[*Myxococcus xanthus* DK 1622]	128/221 (57%)	154-816	2391-2611
pTA25	**Fragment**	**2,500**				
	T3	801	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	146/235 (62%)	784-83	3993-4227
	T7	866	Polyketide synthase [*Myxococcus xanthus* DK 1622]	137/234 (58%)	153-854	2391-2624
pTA26	**Fragment**	**1,500**				
	T3	800	Non-ribosomal peptide synthase/polyketide synthase Ta1 [*Myxococcus xanthus* DK 1622]	72/238 (30%)	701-36	8702-8935
	T7	858	Beta-ketoacyl synthase [*Clostridium cellulolyticum* H10]	91/273 (33%)	72-857	4177-4437
pTA27	**Fragment**	**1,000**				
	T3	801	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	83/230 (36%)	707-24	3550-3764
	T7	895	Polyketide synthase *[Sorangium cellulosum* 'So ce 56']	78/179 (43%)	1-531	2442-2620
pTSTA2	pCC1-FP	756	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	86/249 (34%)	738-40	1433-1679
	pCC1-RP	758	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	77/176 (43%)	3-524	2445-2620

Table 30. Analysis of pTSTA3 spot sequences.

Name		Size [bp]	Homologous protein	Identity	Position of query	Position of subject
pTA31	**Fragment**	**10,000**				
	T3	744	Modular polyketide synthase ketoacyl synthase domain [symbiont bacterium of *Theonella swinhoei*]	80/95 (84%)	293-9	121-215
	T7	901	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	91/215 (42%)	201-839	16-228
pTA33	**Fragment**	**8,000**				
	T3	801	Modular polyketide synthase ketoacyl synthase domain [symbiont bacterium of *Theonella swinhoei*]	80/95 (84%)	293-9	121-215
	T7	901	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	93/230 (40%)	201-884	16-243
pTA34	**Fragment**	**5,000**				
	T3	785	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	83/230 (36%)	707-24	3550-3764
	T7	919	Beta-ketoacyl synthase [*Clostridium cellulolyticum* H10]	111/294 (37%)	63-914	4178-4465
pTA35	**Fragment**	**3,500**				
	T7	845	Unnamed protein product [*Tetraodon nigroviridis*]	21/107 (19%)	700-805	279-383
pTA36	**Fragment**	**2,500**				
	T3	782	Polyketide synthase *[Myxococcus xanthus* DK 1622]	107/206 (51%)	157-771	2391-2596
	T7	883	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	160/262 (61%)	857-72	3966-4227
pTA37	**Fragment**	**1,500**				
	T3	800	Non-ribosomal peptide synthase/polyketide synthase Ta1 [*Myxococcus xanthus* DK 1622]	72/238 (30%)	700-35	8702-8935
	T7	859	Beta-ketoacyl synthase [*Clostridium cellulolyticum* H10]	89/274 (32%)	73-858	4177-4437
pTSTA3	pCC1-FP	736	AMP-dependent synthetase and ligase [*Nostoc punctiforme* PCC 73102]	76/153 (49%)	10-468	502-643
	pCC1-RP	499	Polyketide synthase [*Sorangium cellulosum* 'So ce 56']	104/158 (65%)	499-26	4071-4228

Table 31. KS domains of pTSTA2 and pTSTA3 and their predicted substrates. (Misc) denotes instances where the closest BLASTP hits belong to different clades.

Fosmid	Fragment		Corresponding KS	Identity	Clade	Predicted substrate
	Name	Direction				
pTSTA2	pTA22	T7	BTKS3	45%	Id	α, β carbon-branch
			DszKS3	44%	Ic	
			OnnKS3	47%	Ib	
	pTA23	T3	MlnKS3	36%	V	Misc
			MlnKS6	35%	IV	
			DszKS4	34%	IV	
	pTA25	T3	OnnKS3	61%	Ib	α, β carbon-branch
			DifKS14	59%	Ia	
			BaeKS6	60%	Id	
	pTA26	T3	BaeKS5	28%	IX	Misc
			DifKS14	28%	Ia	
			MlnKS10	51%	IX	
		T7	MlnKS10	28%	IX	α, β double bond
			MlnKS6	29%	IV	
			GUKS3	29%	IX	
pTSTA3	pTA31	T3	DifKS14	65%	Ia	α, β carbon-branch
			OnnKS3	63%	Ib	
	pTA34	T7	DifKS6	36%	XI	Misc
			MlnKS6	33%	IV	
			DszKS6	32%	XII	
	pTA36	T7	OnnKS3	62%	Ib	α, β carbon-branch
			BaeKS6	62%	Id	
			DifKS14	60%	Ia	
	pTA37	T3	BaeKS5	30%	IX	Misc
			DifKS14	28%	Ia	
			DifKS6	64%	XI	
		T7	GUKS3	29%	XI	Misc
			MlnKS6	29%	IV	
			MlnKS10	28%	IX	

As predicted, one MT domain with the conserved motif GAGTG was present in the T7 end of the fragment pTA27 of pTSTA2. In addition, several KSs (*e.g.*, the KSs of pTA22, pTA25, and pTA36) fell into the clade I, which implicated that their substrates possessed carbon branches at either α or β-position. All the three closest homologs of the T7 end of pTA31 belonged to the clade XV that

was constituted by KSs in an extraordinary sequence KS-KR-ACP-KS_{XV}-DH-ACP-(ACP)-KR-KS_{IX}[47], which generates a diene moiety. Although the other PKS domains yielded homologs from various clades, they still fell into clades whose substrates were olefinic, such as clade V, IX, XI, and XIII. In conclusion, based on the prediction rule of *trans*-AT polyketide structures, the short gene fragments from the two fosmids pTSTA2 and pTSTA3 were analyzed and one was confirmed not to belong to the *onn* gene cluster. In particular, the prediction results also exposed some structural features of the polyketide generated by the PKS genes on these two fosmids. These attributes demonstrate that the new isolated PKS genes encode for an uncharacterized polyketide that has not been yet reported. The nucleotide sequence of pTA27 was submitted to GenBank under the accession number EU275209.1.

In order to gain insights into the isolated gene cluster, pTSTA2 and pTSTA3 were sent for complete sequencing. Because of unknown reasons, however, continuous sequences of both fosmids could not be obtained. There were two separate contigs from pTSTA2 and six contigs from pTSTA3 (Figure 77). Missing gaps, therefore, were required to be bridged by PCR and then cloned for sequencing. For this purpose, primers were designed and the PCRs were organized following the bridge diagram between two certain fragments of each fosmid (Figure 77).

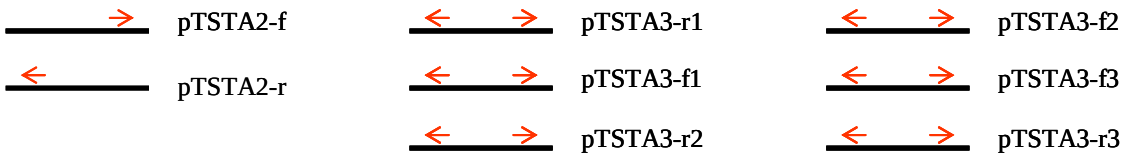


Figure 77. Positions of primers for identification of missing gaps from pTSTA2 and pTSTA3. The red arrows above each fragment represent primers.

Table 32. Organization of PCRs for identification of missing gaps from pTSTA3.

Reverse primer	Forward primer	Sample	Reverse primer	Forward primer	Sample
r1	f1	1	f3	r1	16
	r2	2		f1	17
	f2	3		r2	18
	f3	4		f2	19
	r3	5		r3	20
f1	r1	6	f2	r1	21
	r2	7		f1	22
	f2	8		r2	23
	f3	9		f3	24
	r3	10		r3	25
r2	r1	11	r3	r1	26
	f1	12		f1	27
	f2	13		r2	28
	f3	14		f2	29
	r3	15		f3	30

A series of thirty PCRs for finding the missing fragments of pTSTA3 (Table 32) and one PCR for finding that of pTSTA2 were carried out. Based on a deduction about the order of fragments on the fosmid pTSTA3, the PCR products numbered 4, 5, 9, 12, and 18 together with the gap2 from pTSTA2 were chosen for sequencing (Figure 78 and Table 33).

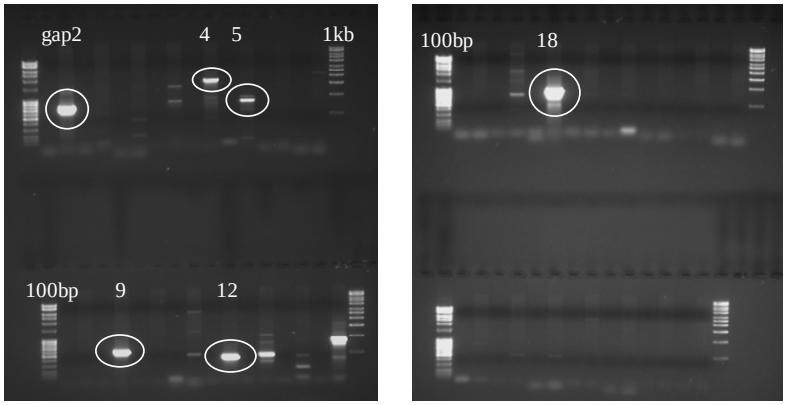


Figure 78. PCRs of gaps of pTSTA2 and pTSTA3. Agarose gels of electrophoresed PCR are shown.

Table 33. Plasmid names in finding gaps of pTSTA2 and pTSTA3.

PCR product	Gap2	4	5	9	12	18
Plasmid name	pTA18	pTA19	pTA20	pTA21	pTA22	pTA23

In order to clone for sequencing, these PCR products were purified from agarose gel (Figure 79) and ligated with the vector pBluescript SK II (-) applying the TA cloning method (section 7.2.12). Afterwards, plasmid DNAs from the transformed colonies were isolated and sent for sequencing using T7 and T3 primers.

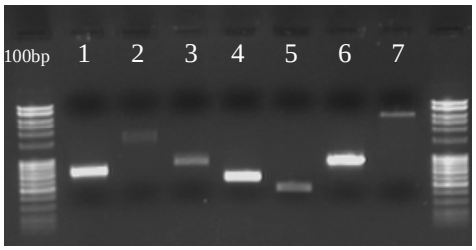


Figure 79. Agarose gel of electrophoresed purified PCR (1) PTA18 (2) PTA19 (3) PTA20 (4) PTA21 (5) PTA22 (6) PTA23 (7) digested pBluescript SK II (-) with *Eco*RV.

4.2.2 Discussion on the complete sequences of the fosmid pTSTA2 and pTSTA3

The Seqman program was applied to assemble all contigs obtained after sequencing. There was a 9,620 nucleotide overlapping fragment between the two complete sequences of pTSTA2 and pTSTA3, from position 23,612 to 33,231. The new contig comprised 54,842 nucleotides. This complete sequence was then analyzed applying the same method as that used for the previous complete sequences (section 3.2.1).

The sequenced region comprised ~ 45.5 kb of PKS genes with an overall GC content of 51.74%, and ~ 9.3 kb of non-PKS genes. Two ORFs were homologous to transposases of the bacteria *Nitrococcus mobilis* and *Nostoc* sp. (Table 34). In addition, ORFs were found to be preceded by putative Shine-Dalgarno sequence. According to these features, this unknown polyketide was also

predicted to be a product of a bacterium. Based on analyses from BLASTP, Pfam, as well as the conserved motifs, PKS domains of the new *pks* gene cluster were identified and shown in Figure 80.

Table 34. Analysis of the new *pks* gene cluster from *T. swinhoei.*

Reading frame	Similarity		Identity	Position	Length (aa)
	Protein	Origin			
+3	PKS	*Sorangium cellulosum* 'So ce 56'		340-13027	4,229
-3	Transposase	*Nitrococcus mobilis* Nb-231	20/75 (26%)	13016-13740	241
+1	PKS	*Sorangium cellulosum* 'So ce 56'	81/150 (54%)	13901-14545	245
-1	Transposase	*Nostoc* sp. PCC 7120	202/384 (52%)	14564-15764	400
+2	PKS	*Sorangium cellulosum* 'So ce 56'		16057-45391	9,778
+2	Putative heavy metal translocating ATPase	*Chondromyces crocatus*	322/687 (46%)	45632-47692	687
+1	Polyketide biosynthesis enoyl-CoA hydratase	*Myxococcus xanthus* DK 1622	103/190 (54%)	47936-48514	193
+1	Aldehyde oxidase and xanthine dehydrogenase	*Plesiocystis pacifica* SIR-1	329/706 (46%)	49070-51139	690
+1	PKS	*Clostridium cellulolyticum* H10	338/700 (48%)	52636-54842	735

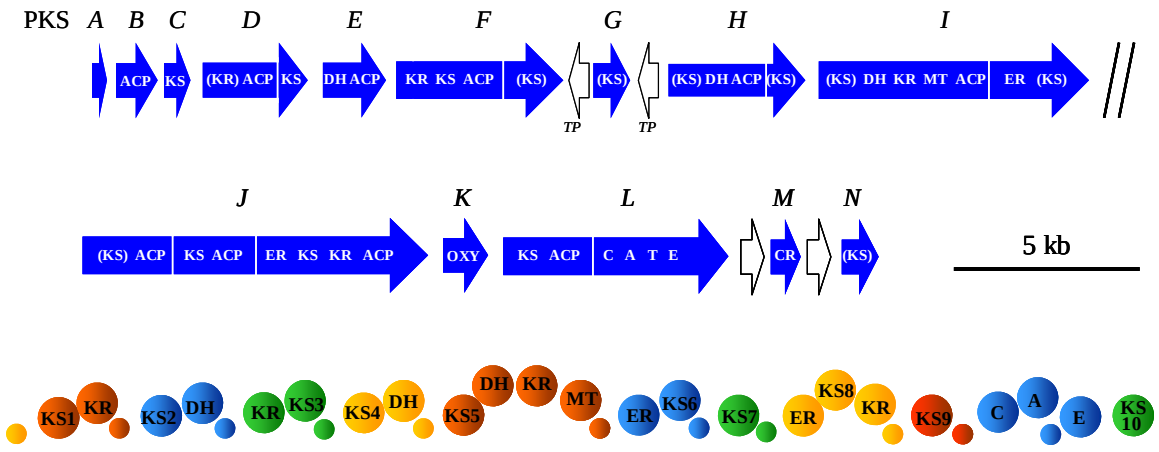


Figure 80. The newly isolated *pks* gene cluster.

pksA, which exhibits homology to an acyl-CoA dehydrogenase gene, is found at the beginning upstream end of the isolated region (see Appendices). In the initial step of each cycle, this enzyme catalyzes the formation of 2, 3-dehydroacyl-CoA (Figure 81).

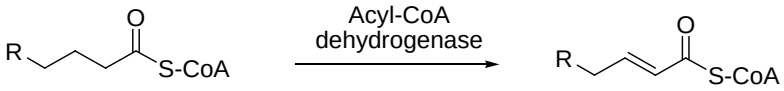


Figure 81. Catalysis of acyl-CoA dehydrogenase.

Two post-PKS tailoring genes, oxygenase and enoyl-CoA hydratase, a member of the crotonase superfamily were revealed (Figure 80). The role of these genes in polyketide biosynthesis could not be yet predicted. Interspersed between the *pks* genes are several transposase genes. This phenomenon is similar to the *ped* gene cluster.

One NRPS module was found on *pksL*. The A domain of this module was analyzed in order to identify its amino acid substrate using the NRPS predictor program (http://www-ab.informatik.uni-tuebingen.de/toolbox/index.php). An ornithine residue was predicted with an identity of 70%. In addition, an epimerization (E) domain was also recognized with the seven conserved core motifs E1-E7 (Table 35). This domain is responsible for racemization, converting L-ornithine into D-ornithine (Figure 82).[52] We searched MarinLit, a database of marine natural products, for

polyketides with D-ornithine residues, but no matching structure existed. Therefore a novel hybrid PK-NRP has been expected to be present in the sponge *T. swinhoei*.

Table 35. Conserved core motifs of the epimerization domain (x is variable amino acid).[52]

Core	Conserved sequence	Core	Conserved sequence
E1	PIQxWF	E5	RTVGWFTxxYP(YV)PFE
E2	HHxISDG(WV)S	E6	PxxGxGYG
E3	DxLLxAxG	E7	FNYLG(QR)
E4	EGHGRE		

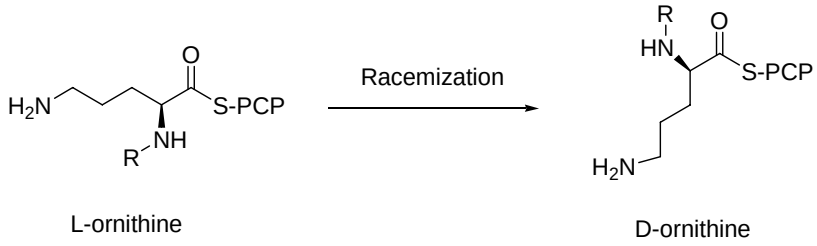


Figure 82. Amino acid ornithine. PCP: peptidyl carrier protein.

However, frameshifts in the middle of KS domains were observed on *pksF, G, H, I* and *J* (Figure 80). This suggests that the cluster is a nonfunctional evolutionary relic. Subsequently, the ten KS domains of the *pks* gene cluster were analyzed in order to predict their substrates using the cladogram in Figure 68 (Table 36). The first KS domain did not share any similarity to *trans*-AT PKS. It might be nonfunctional because of the absence of the motif HGTGT. The clade XV-IX pair of KSs (KS2 and KS3) was identified, as in the previous preliminary analysis, and the predicted domain architecture KS-KR-ACP-KS-DH-ACP-KR-KS was found in striking agreement. Therefore, a conjugated diene is probably present in the structure of this unknown polyketide. In addition, KS5 and KS10 belonging to the clade IX should introduce double bonds at the corresponding positions. The four KSs, KS4, KS6, KS7, and KS8 belonged to the clade I that recognizes substrates with carbon branches. But except for the KS6 on *pksI*, the other three KSs did not locate next to the MT domain (Figure 80). In modules involving such noncanonical PKSs, however, a keto group is often converted into a carbon branch at β-position. This was demonstrated in a study on the biosynthetic pathway of bacillaene.[145] In summary, a partial biosynthetic pathway

of a novel mixed polyketide-nonribosomal peptide was postulated. This is an unsaturated branched compound. At least seven double bonds, four methyl branches, one hydroxyl and one carbonyl group and skeleton of the amino acid D-ornithine (Figure 82) should be present in the revealed structure (Figure 83).

Table 36. Predicted substrates of the new *pks* gene cluster based on KS domains.

KS domain of *pks*	**Corresponding KS**	**Identity**	**Clade**	**Predicted substrate**
KS2	DifKS10	66%	XV	α, β double bond
	MlnKS11	58%	XV	
	MlnKS8	54%	XV	
KS3	BaeKS5	54%	IX	α, β double bond
	MlnKS4	54%	IX	
	DifKS13	49%	IX	
KS4	OnnKS3	63%	Ib	α, β carbon-branch
	DifKS14	70%	Ia	
	DszKS3	62%	Ic	
KS5	MlnKS4	48%	IX	α, β double bond
	MlnKS9	49%	IX	
	DifKS1	47%	XIII	
KS6	OnnKS3	64%	Ib	α, β carbon-branch
	BaeKS6	62%	Id	
	DszKS3	62%	Ic	
KS7	OnnKS3	56%	Ib	α, β carbon-branch
	DszKS3	55%	Ic	
	BaeKS6	54%	Id	
KS8	OnnKS3	58%	Ib	α, β carbon-branch
	DszKS3	57%	Ic	
	Mmp	56%	Ia	
KS9	DifKS3	44%	V	Single bond
	DifKS9	45%	V	
	DszKS4	42%	IV	
KS10	BaeKS5	67%	IX	α, β double bond
	DszKS7	58%	IX	
	BaeKS9	56%	VII	

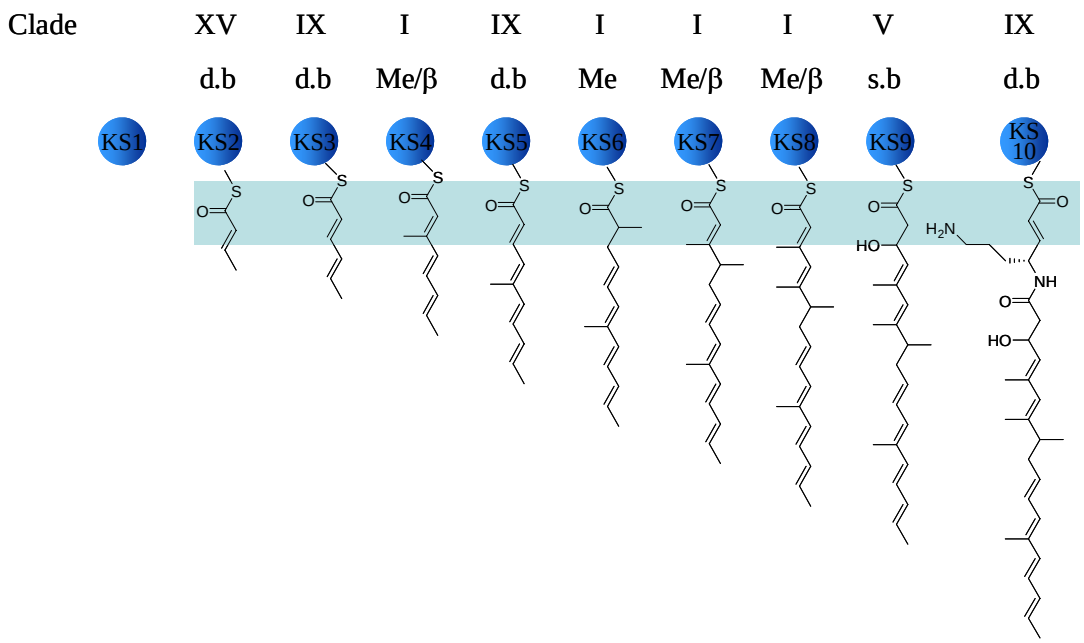


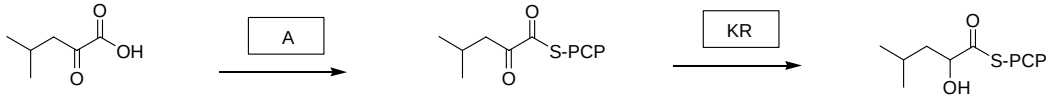
Figure 83. KS-based prediction of substrate specificities of the new *trans*-AT PKS. Deduced clades and substrate specificities are provided above each KS domain.

Chapter 5

Heterologous expression trials with the cosmid pPD7E4 from the metagenomic DNA of the beetle *Paederus fuscipes*

5.1 Introduction

Piel and coworkers obtained a partial genome sequence of the pederin producing symbiont of the beetle *P. fuscipes* (unpublished data). The data revealed the presence of an additional small PKS-NRPS cluster of unknown function. The entire cluster was located on the cosmid pPD7E4, which was isolated and used for these studies. An analysis of 37,450 nucleotide sequences revealed 10 putative genes (*pppA-pppK*) of a hybrid PKS-NRPS. One NRPS module was identified in *pppB* and analyzed using the NRPS predictor program. An aspartic acid residue was predicted to be the specific amino acid of the A domain of this NRPS. In addition, another A domain was also found in an extraordinary domain architecture A-KR-PCP of *pppE*. The function of this A domain is not to activate an amino acid as in a normal NRPS module due to lack of active site motifs. This is similar to *CesB* from the cereulide NRPS cluster, in which the A domain activates α-ketocaproic acid and then the KR reduces this α-keto acid to α-hydroxy caproic acid.[146]



α-ketocaproic acid α-hydroxy caproic acid

This gene cluster belonged to the *cis*-AT PKS group as one AT domain was present in *pppE* (Figure 84). In particular, pPD7E4 contained the tailoring gene *pppG*, with a length of 951 nucleotides, encoding for a halogenase enzyme. This sequence was analyzed by BLAST homology search and the closest homolog was found to be a chlorination enzyme from the *Burkholderia* genus (Table 37). These distinct features were an indication that there was a high possibility of finding a new PKS gene cluster, other than the *ped* genes of pederin, from the metagenome of *P. fuscipes*. It is also predicted that the compound generated by pPD7E4 is a novel halogenated natural product that should possess a promising bioactivity. For this reason, the cosmid pPD7E4 was chosen to be heterologously expressed.

The genes of pPD7E4 displayed typical bacterial attributes, not present in eukaryotic genes. For example, a nucleotide sequence comprised a high GC content of 57%, the putative genes are preceded by the Shine-Dalgarno boxes required for ribosomal binding. In addition, the genes are separated by small intergenic regions *e.g.*, a sequence of 28 bp between *pppB* and *pppC*, 22 bp between *pppE* and *pppG*. Moreover, the second ORF is similar to a tranposase of the bacteria *P. putida* (Table 37). All observed characteristics are in agreement with the fact that a symbiotic bacterium is the owner of pPD7E4. Also based on the biosynthetic genes, a partial structure of this unknown polyketide was predicted, however, the position of halogen atom could not be identified (Figure 84b).

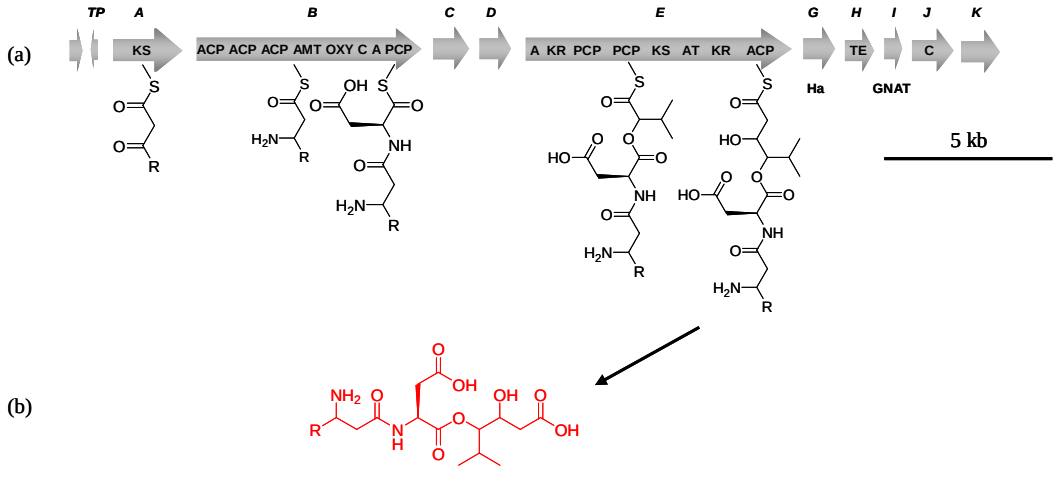


Figure 84. (a) The gene cluster of the cosmid pPD7E4. (AMT) Aminotransferase, transfers an amino group to the β-carbon[147, 148] (Ha) halogenase (b) predicted structure frame of the compound encoded by pPD7E4.

In addition, in bacteria, RNA transcription begins when RNA polymerase binds to DNA at a particular region, referred to as promoter. A promoter is typically located near the genes it regulates, on the upstream region of the same strand. It consists of two short sequences at -10 and -35 position upstream from the starting point of transcription. The consensus sequence at the -10 position named Pribnow box is TATAAT and at the -35 position is TTGACA. The optimal distance between these two sequences is 17 ± 1 nucleotides. However, this distance is not completely fixed because some nucleotides and distance can be changed.[149, 150] Therefore, in an attempt to detect a promoter being responsible for the transcription of the *ppp* gene cluster of the cosmid pPD7E4, the promoter prediction program (1999NNPP version 2.2) was used. A promoter at the position from 12,567 to 12,616 of pPD7E4, which was an immediate upstream region of the first gene *pppA*, was identified with a score up to 0.98.

In nature, halogenated compounds are found frequently. More than 4,500 halogenated natural compounds have been identified so far, from the total number of which 2,300 are organochlorines, 2,100 are organobromines, the iodinated and fluorinated compounds are however rare (120 organoiodines and 30 organofluorines).[151, 152] Biological halogenations can occur on a wide range of organic scaffolds such as terpenes, polyketides or nonribosomal peptides. Organohalogen have been found in bacteria, fungi, plants and animals.[153-155] While the source of brominated compounds is commonly marine organisms, most organochlorines are produced by organisms of terrestrial origin.[156] Several chlorinated compounds have bioactivities ranging from the antibiotic vancomycin whose chlorine is responsible for its antibacterial activity to the antifungal antibiotic pyrrolnitrin.[153] Another example is the promising anticancer drug cryptophycin A. Halometabolites can play an important defensive role by inhibiting the growth of competing organisms.[153, 157] Halogenated compounds are thus undeniably some of the most medicinally useful natural products.

Table 37. Analysis of the complete sequence of pPD7E4.

Reading frame	ORF	Similarity		Identity	Position	Length (aa)
		Protein	Origin			
+2		C4-dicarboxylate transport protein	*Pseudomonas aeruginosa* 2192	366/436 (83%)	4943-6250	435
-1	TP	Transposase	*Pseudomonas putida*	25/53 (47%)	11167-11325	52
+1	PppA	Polyketide synthase	*Chondromyces crocatus*	237/514 (46%)	12948-15005	686
+1	PppB	Peptide synthetase protein	*Ralstonia solanacearum*	750/2192 (34%)	15002-21634	2,895
+1	PppC	Dioxygenase	*Paenibacillus* sp. JDR-2	72/231 (31%)	21663-22715	351
-3	PppD	Similar to serine/arginine repetitive matrix 2	*Rattus norvegicus*	41/130 (31%)	22690-23535	282
+2	PppE	Hybrid PKS-NRPS	*Chondromyces crocatus*	532/1561 (34%)	23771-32175	2,801
+1	PppG	Chlorinating enzyme	*Burkholderia vietnamiensis* G4	125/290 (43%)	32197-33147	317
+2	PppH	Thioesterase	*Bacillus pumilus* SAFR-032	68/243 (27%)	33144-33989	282
+1	PppI	GCN5-related *N*-acetyltransferase	*Herpetosiphon aurantiacus*	63/166 (37%)	34372-34905	178
+2	PppJ	Peptide synthetase ScpsB	*Saccharothrix mutabilis*	101/399 (25%)	34898-36161	421
+1	PppK	Transporter	*Pseudomonas aeruginosa*	184/361 (50%)	36276-37450	391

5.2 Results and discussion

The cosmid pPD7E4 was heterologously expressed using an expression system including the expression vector pJB861[158] and the host cell *P. putida* KT2440.

pJB861 (Figure 117) is a broad host range vector and a useful tool for expression in gram-negative bacteria.[158] *P. putida* KT2440 is a non-pathogenic gram-negative bacterium in nature. It is able to grow rapidly in different laboratory medium. *P. putida* KT2440 was certified as the host strain of the first host-vector biosafety system by the Recombinant DNA Advisory Committee (RAC).[159] In addition, *P. putida* KT2440 lacks a restriction system against DNA uptake, it is plasmid-free and its genome was sequenced completely in 2002.[160, 161] Owing to these important characteristics, *P. putida* KT2440 is an excellent host cell for heterologous gene expression, particularly genes being responsible for secondary metabolites.[162] Noticeably, its phosphopantetheinyl transferase (PPTase) is able to activate both ACPs and PCPs from different PKS or NRPS systems.[163]

To introduce the gene cluster into the vector, the cloning vector pWEB of pPD7E4 was replaced by the expression vector pJB861 (Figure 86b) by λ RED-mediated homologous recombination (http://streptomyces.org.uk/redirect/protocol). Homologous recombination is convenient for genetic modification in many organisms. In gene targeting technology, this method is applied widely to replace specific genes at specific positions (Figure 85). A short homology PCR fragment, ranging from 100 bp-1 kb, is sufficient for an efficient homologous recombination and is able to generate gene replacements. The recombinant construct is then transformed into a host cell, ready for production of target compound.[164, 165]

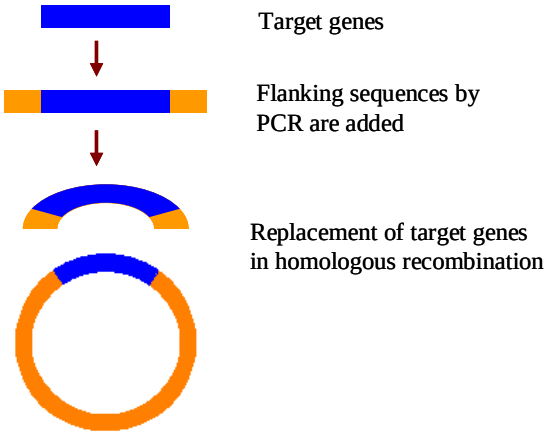
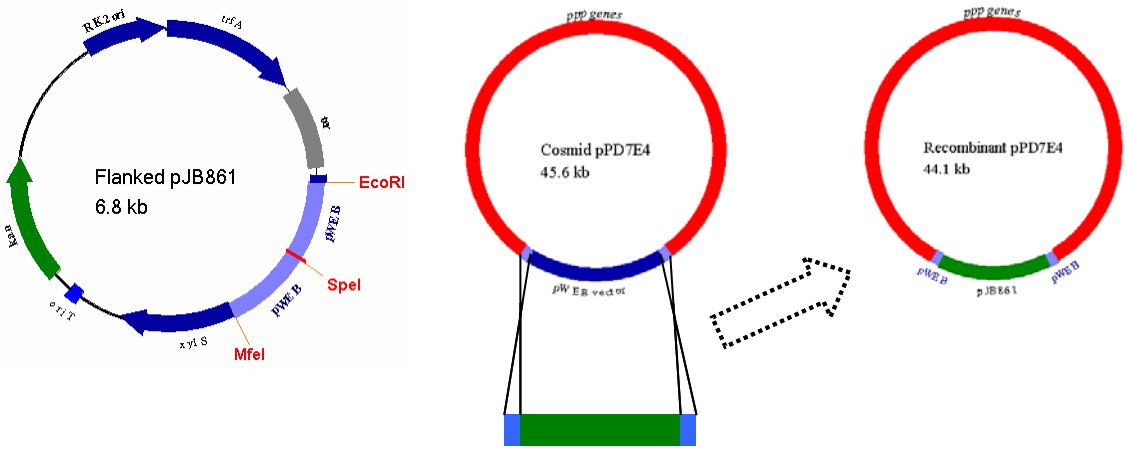


Figure 85. Flowchart of gene targeting by homologous recombination.

In preparation for homologous recombination, pJB861 had to be modified with two PCR fragments that were homologous to the two end regions of pWEB (Figure 86a).



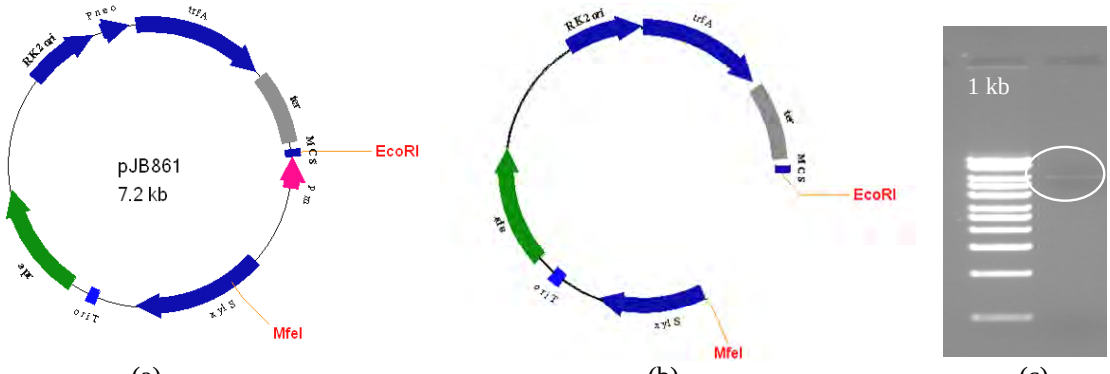
Flanked pJB861

(a) (b)

Figure 86. (a) Diagram of the flanked pJB861 and (b) diagram of a homologous recombination in which the vector pWEB was replaced by the vector pJB861.

5.2.1 Digestion of the expression vector pJB861

As analyzed in section 5.1, a natural promoter was predicted to be present inside the cosmid pPD7E4. Consequently, the promoter and *xylS* gene of the vector pJB861 (Figure 87a) were not required for this expression. Therefore, a fragment of 1,403 bp involving these components was deleted with two restriction enzymes, *Mfe*I and *Eco*RI (Figure 87b) (section 7.3.2.1). The digested pJB861 with a size 5,840 bp was recovered from the agarose gel, and 1 µl of DNA was checked on gel (Figure 87c).



(a) (b) (c)

Figure 87. Map of (a) the expression vector pJB861 (b) digested pJB861 with *Eco*RI and *Mfe*I and (c) agarose gel electrophoresis of purified digested pJB861.

5.2.2 Amplification of pWEB fragments

The vector pWEB had been employed to construct a cosmid library from the metagenomic DNA of the beetle *P. fuscipes*. The insert of the cosmid pPD7E4 was located at the position of the *Sma*I-restriction site CCC'GGG belonging to the MCS of pWEB (Figure 88a). For homologous recombination, the expression vector pJB861 had to be flanked by two short PCR fragments that were homologous to two regions of pWEB located on the two opposite sides of *Sma*I ((Figure 88b). These two amplified fragments were referred to as C side, G side fragments and named PTA24, PTA25, respectively. Resulting from this step, two similar regions between the cosmid pPD7E4 and the flanked pJB861 were introduced, ready for a homologous recombination.

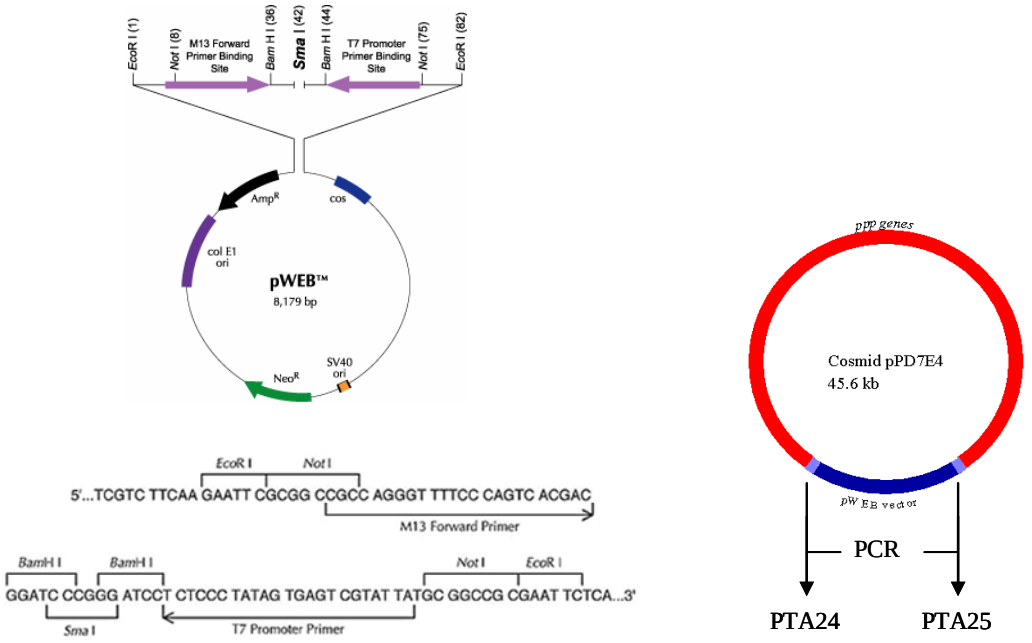


Figure 88. (a) Map of the vector pWEB (Epicentre) (b) The cosmid pPD7E4 including the *ppp* gene cluster and the vector pWEB. PTA24 and PTA25 were PCR fragments that were homologous to two end regions of pWEB.

After amplification by PCR with primers designed in section 7.3.2.2, PTA24 (395 bp) and PTA25 (384 bp) were yielded. In preparation for ligation with the expression vector pJB861, pTA24 and pTA25 were introduced sticky ends according to the methods described in section 7.3.2.2. Afterwards, these two fragments were purified and checked on agarose gel (Figure 89).

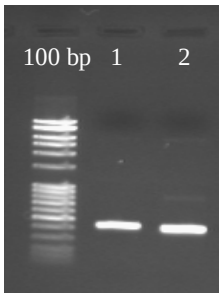


Figure 89. Agarose gel electrophoresis of purified plasmid (1) pTA24 and (2) pTA25.

5.2.3 Flanking of the digested pJB861

The two fragments pTA24 and pTA25 were then ligated with the digested vector pJB861 in section 5.2.1 to obtain the flanked pJB861 (Figure 86a), named pTA26 (section 7.3.2.2). Because homologous recombination only occurs with linear DNA, pTA26 was then linearized with *Spe*I and recovered from agarose gel (Figure 90) (section 7.3.2.3).

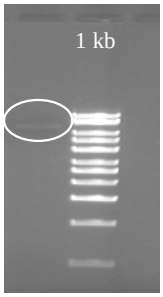


Figure 90. Agarose gel electrophoresis of the linearized pTA26.

5.2.4 Recombination of the cosmid pPD7E4

E. coli BW25113 containing the plasmid pIJ790 (Figure 91) was used as the host in the PCR-targeting system. A linear recombination cannot occur in wild-type bacteria due to the presence of the intracellular *recBCD* possessing 5' – 3' exonuclease activity that degrades the 5' ends of linear DNA molecules. The λ RED recombination plasmid pIJ790 was used, which involves the bacteriophage λ recombination system in order to inhibit any exonuclease activity (http://streptomyces.org.uk/redirect/protocol), so that linear DNA molecules are preserved. In

addition, this plasmid also has the promoter araBAD that can be regulated by arabinose and *rep*A*101ts*, a temperature - sensitive replication.

Figure 91. The λ RED recombination plasmid pIJ790. (http://streptomyces.org.uk/redirect/protocol).

In preparation for homologous recombination, the cosmid pPD7E4 was transformed into electrocompetent *E. coli* BW 25113 cells containing the plasmid pIJ790 (section 7.3.3.1). The transformed cells were selected on LB agar containing chloramphenicol (resistance gene on pIJ790) and ampicillin, kanamycin (resistance genes on the pWEB vector).

Subsequently, the linearized pTA26 in section 5.2.3 was transformed into the electrocompetent *E. coli* BW25113/pIJ790 cells containing pPD7E4 (section 7.3.3.4). In this step, a homologous recombination took place. The cloning vector pWEB in pPD7E4 was replaced by the expression vector pJB861. Upon this recombination, a modified genetic construct was obtained, namely pTA27 (Figure 86b). Recombinant cells were selected on LB plate with apramycin due to the presence of an apr^R gene inside the expression vector pJB861. Because the temperature-sensitive plasmid pIJ790 was lost during cultivation at 37 ^oC, cat^R gene was not present anymore and the amp^R and kan^R genes of pWEB were replaced by apr^R from pJB861.

5.2.5 Verification of the recombinant pPD7E4

The homologous recombination normally does not occur in all copies of the cosmid in one cell. Thus, there were transformed cells containing both the recombinant pPD7E4 and pPD7E4. For this reason, screening the cells possessing merely the recombinant construct had to be carried out.

Screening by antibiotic resistance

The cells containing only the recombinant construct pTA27 were resistant to one antibiotic, apramycin (resistance gene on the expression vector pJB861). In contrast, the cells containing both pTA27 and pPD7E4 were resistant to three antibiotics *i.e.*, apramycin (resistance gene on the expression vector pJB861 inside pTA27), ampicillin and kanamycin (resistance genes on the vector pWEB inside pPD7E4).

Screening by restriction analysis

This verification was based on the difference between the nucleotide sequence of the recombinant construct pTA27 and that of pPD7E4. pTA27 comprised 44,069 nucleotides, whereas pPD7E4 comprised 45,629 nucleotides. These two sequences were analyzed by the NEB cutter 2.0 program (Figure 92). Three restriction enzymes *Not*I, *Sca*I, and *Nde*I were chosen to cut these two sequences into distinguishable digestion pattern using electrophoretic separation (Table 38).

Table 38. Restriction enzymes used for checking the recombinant pPD7E4.

Enzyme	**Fragments of the recombinant pPD7E4 [kb]**	**Fragments of pPD7E4 (O) [kb]**
*Not*I	29.5, 10.8, 3.6	22.9, 10.9, **8.1**, 3.6
*Sca*I	20.6, 14.9, 8.4	20.5, 14.9, 8.4, **1.6**
*Nde*I	33.7, 7.1, **2.6**, 0.6	37.8, 7.1, 0.6

The screening experiments are described in the following sections.

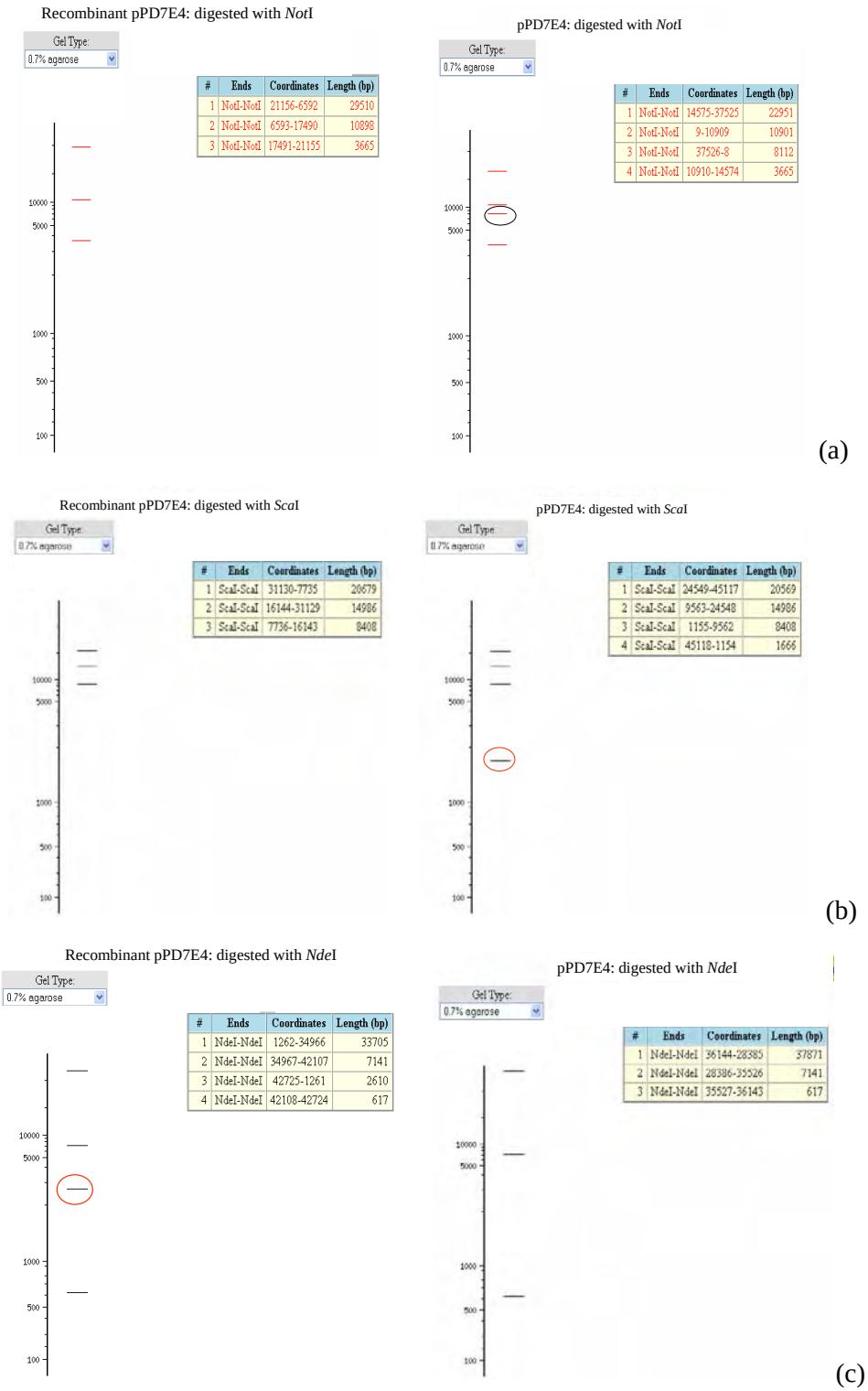


Figure 92. Theoretical digestion diagrams of the recombinant pPD7E4 and pPD7E4 obtained by applying the NEB cutter 2.0 program with (a) *Not*I (b) *Sca*I (c) *Nde*I. The distinct fragments between the two digestions are encircled.

5.2.5.1 Screening in *E. coli* BW25113

Forty single transformed *E. coli* BW25113 colonies were picked to examine antibiotic resistances as described in section 7.3.4.1. Four colonies grew on LB plate with apramycin (plate A) but did not grow on LB plate with apramycin, ampicillin and kanamycin (plate B), which indicated that they were the strongest candidates among the cells possessing only the recombinant construct pTA27. They were colonies No. 3, 4, 6 and 38 (Figure 93).

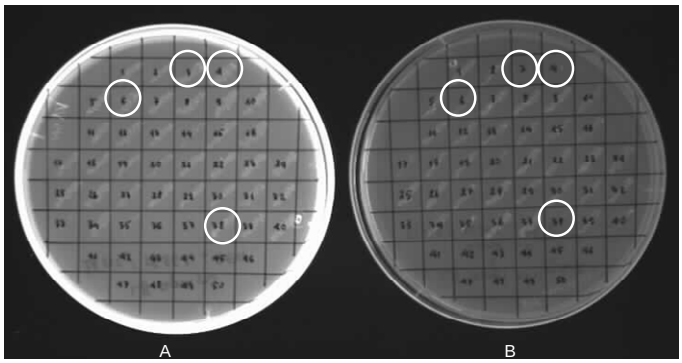


Figure 93. Checking the antibiotic resistances of the transformed *E. coli* BW 25113 containing pTA27 (A) on LB agar with apramycin and (B) on LB agar with apramycin, ampicillin and kanamycin.

DNA from the four above colonies was characterized by restriction analysis. The digestion reactions were performed and analyzed as described in section 7.3.4.2. Among the four colonies, colony 4 possessed a restriction pattern that was different from that of pPD7E4 (Figure 94). Therefore it was chosen for further experiments.

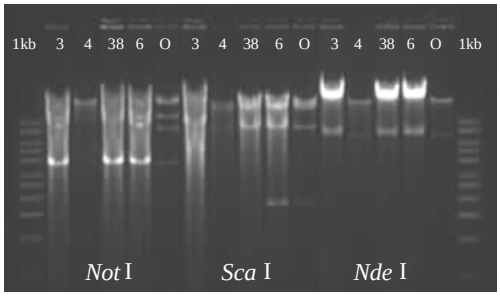


Figure 94. Agarose gel electrophoresis of checking pTA27 in the four colonies 3, 4, 6, and 38 of *E. coli* BW25113 and comparing with pPD7E4 (O) by digestion with three restriction enzymes *Not*I, *Sca*I, and *Nde*I.

5.2.5.2 Screening in *E. coli* XL1 blue

In order to ensure that a cell containing only the recombinant DNA was obtained, the verification was required to be conducted one step further. For this purpose, the isolated DNA from colony 4 of *E. coli* BW25113 was introduced into competent *E. coli* XL1 blue cells. Fifty transformed *E. coli* XL1 blue colonies were picked to verify them by antibiotic resistance screening and restriction analysis as described in section 7.3.4. All fifty colonies grew on LB plate with apramycin but did not grow on the other LB plate with apramycin, ampicillin and kanamycin. Five colonies No. 6, 17, 20, 40, and 50 were randomly selected to perform digestion with *Not*I, *Sca*I and *Nde*I. Their restriction maps were the same as those theoretically postulated and different from that of the cosmid pPD7E4 (Figure 95). This result confirmed that the pTA27 construct was successfully obtained. The pTA27 from colony 17 was chosen for expression in the host cell *P. putida* KT2440.

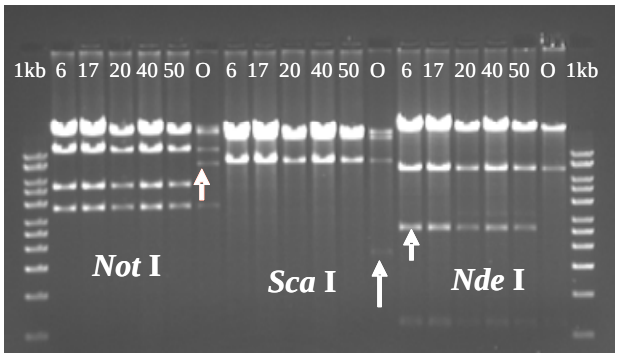


Figure 95. Agarose gel electrophoresis of checking pTA27 in five *E. coli* XL1-blue colonies. The arrows show the distinctive bands between the digestion patterns of pTA27 and pPD7E4 (O).

5.2.6 Verification of the heterologous expression in *P. putida*

According to the central dogma of biology, from the initial genetic material - DNA to the final genetic product - protein, this process includes three subsequent stages. The first is DNA replication for multiplying DNA molecule; the second is transcription from double-stranded DNA molecule to single-stranded RNA molecule, and the third is translation from mRNA to protein - a sequence of amino acids (Figure 96).[166] In order to evaluate the achievement of a heterologous expression procedure, each of the three above steps must be verified.

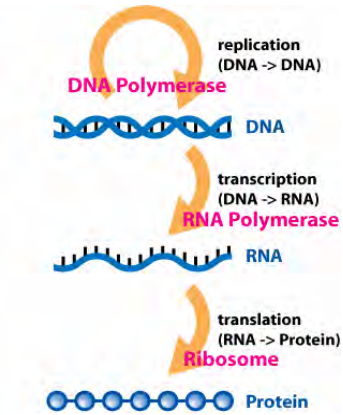


Figure 96. The central dogma of biology.
(http://elgolemrazonable.blogspot.com/2009/02/el-dogma-central-de-la-biologia.html)

Verification at the DNA level

The recombinant DNA pTA27 was transformed into *P. putida* KT2440 cells (section 7.3.5). Five *P. putida* TA27 colonies were randomly picked for harvesting of plasmid DNA that was checked by restriction analysis. Due to the small copy number of the plasmid, concentration of the obtained DNA was rather low. For this reason, the digestion reaction of plasmid DNA from *P. putida* was unsuccessful even though more DNA was digested. Therefore, the plasmid DNA from one transformed *P. putida* TA27 colony was selected to be verified indirectly as follows. This DNA was transformed into electrocompetent *E. coli* XL1 blue cells. The plasmid DNA from six transformed *E. coli* colonies were then isolated and digested with the three restriction enzymes *Not*I, *Sca*I, and *Nde*I (Figure 97). All the six examined *E. coli* colonies contained pTA27, which confirmed the presence of the correct recombinant DNA in the *P. putida* TA27 strain. This strain was then used for further verification.

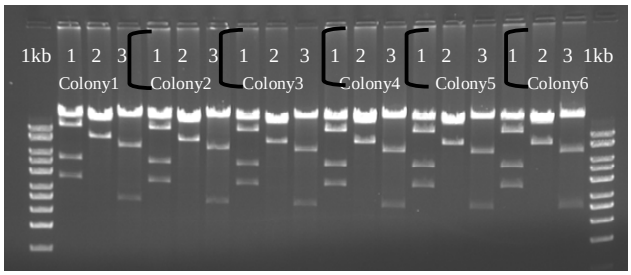


Figure 97. Agarose gel electrophoresis of checking pTA27 isolated from *P. putida* TA27 in six colonies of *E. coli* XL1-blue by digestion with (1) *Not*I (2) *Sca*I (3) *Nde*I.

Verification at the RNA level

A messenger RNA (mRNA) molecule is transcribed from DNA. Genes are only expressed when their mRNAs are synthesized. To check if transcription took place, RNA from *P. putida* TA27 had to be isolated and subjected to RT-PCR.

The growth rate of the host cell significantly affects the expression of a foreign protein. Metabolites can be lost due to overgrowth or even too rapid growth of a bacterium. The growth rate as well as the cell density therefore must be controlled.[143] Among the growth conditions required for cultivation of bacteria, temperature and duration of growth have an influence on the cell population. For this reason, in the primary survey, the transformed *P. putida* TA27 was incubated at two different temperatures, 16 ^{o}C and 30 ^{o}C.

5.2.6.1 Isolation of total RNA

There are several types of RNA inside the cell. Among them is the ribosomal RNA (rRNA) that constitutes up to 80% of the total cellular RNA. rRNA is the central component of the ribosome.[167] In prokaryotes, the ribosome can be broken down into two subunits: a small ribosomal subunit contains 16S rRNA and a large ribosomal subunit contains 23S rRNA. For this reason, two bands of 16S and 23S rRNA are detected on the agarose gel.[168] In most cases, the apparent size of rRNA does not exceed 2 - 3 kb.

The total RNA from *P. putida* was isolated using a modified version of the Qiagen kit procedure as described in section 7.3.6.2. 1 µl of RNA sample was checked on agarose gel (Figure 98).

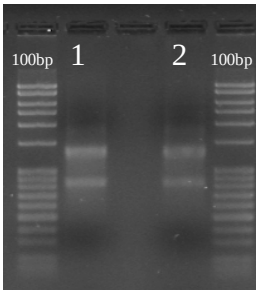


Figure 98. Agarose gel electrophoresis of RNA of the recombinant *P. putida* TA27 cultivated (1) at 16 ^{o}C and (2) at 30 ^{o}C.

This RNA was used for RT-PCR. The RNA sample was checked for genomic DNA contamination in a PCR with *Taq* polymerase (section 7.2.7.1). One of the four primer pairs that were designed for RT-PCR, the *pppD* specific primers were used, and DNA of the cosmid pPD7E4 was employed as a positive control to yield a fragment with a size of 520 bp. No amplification was observed, so the RNA samples were free of genomic DNA (Figure 99).

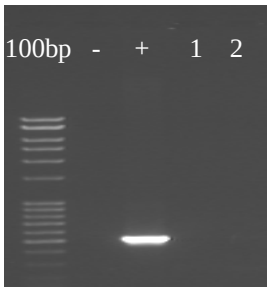


Figure 99. Agarose gel electrophoresis of checking DNA contamination in the isolated RNA (1) at 16 ^{o}C and (2) 30 ^{o}C.

5.2.6.2 RT-PCR of the isolated RNA

As mentioned previously, a remarkable feature of prokaryote is polycistronic mRNA that codes for more than one protein. A polycistronic mRNA can be transcribed from several functional genes of one operon.[169] Based on this characteristic, the ten genes on the cosmid pPD7E4 were localized to four putative operons (Table 39). *pppA*, *pppB*, *pppC* were in the same operon with *pppD*; *pppE*, *pppG* were in the same operon with *pppH*; *pppI* was in the same operon with *pppJ*, and the operon of *pppK* (Figure 100).

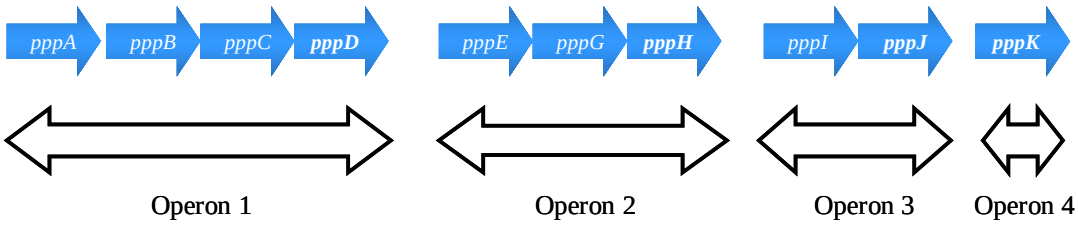


Figure 100. Four putative operons of pPD7E4.

Table 39. Positions of the genes on the cosmid pPD7E4.

Gene	Initial position	Terminal position
pppA	12,947	15,000
pppB	15,001	21,634
pppC	21,663	22,715
pppD	22,690	**23,535**
pppE	**23,771**	32,175
pppG	32,197	33,147
pppH	33,144	**33,989**
pppI	**34,372**	34,905
pppJ	34,898	**36,161**
pppK	**36,276**	37,450

The end region of each of the four operons must be present upon RT-PCR in order to ensure that all genes belonging to the corresponding operon are transcribed to mRNA. Therefore, the end regions of the four genes, *pppD*, *pppH*, *pppJ*, and *pppK*, were employed as templates to design primers for RT-PCR.

Table 40. RT-PCR products of transcription checking experiment.

Name of RT-PCR product	Length of product	Position of gene fragment on pPD7E4
PppD	520 bp	22,936 – 23,455
PppH	476 bp	33,450 – 33,925
PppJ	507 bp	35,583 – 36,089
PppK	475 bp	36,857 – 37,331

The RNA samples were used as templates for RT-PCR following the SuperScript^{TM}III One-Step RT-PCR System protocol. All RT-PCR reactions were positive, so the corresponding mRNAs were present (Figure 101). This implies that during the heterologous expression of pPD7E4, the DNA is transcribed into mRNA, which should be used for the translation step to protein. This strain was thus cultivated to obtain the target compound via heterologous expression.

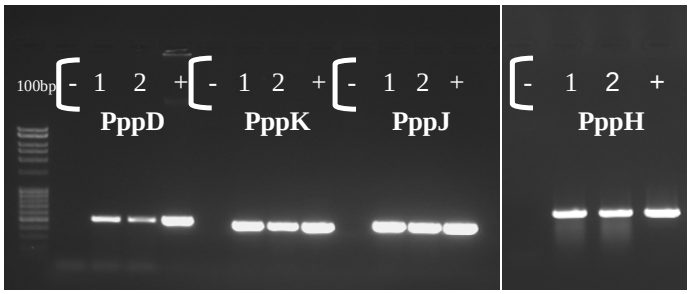


Figure 101. Agarose gels of electrophoresed RT-PCRs of the isolated RNA from the recombinant *P. putida* TA27 cultivated (1) at 16 ^{o}C and (2) at 30 ^{o}C with four primer pairs. Positive control (+) were PCR products of DNA from pPD7E4.

5.2.7 Detection of the expressed compound

5.2.7.1 Extraction of the expressed compound

To examine if the expected compound is produced, the recombinant *P. putida* TA27 and the negative control that was *P. putida* containing the flanked pJB861 were cultivated in 200 ml of LB medium with apramycin at 16 ^{o}C and 30 ^{o}C (Figure 102).

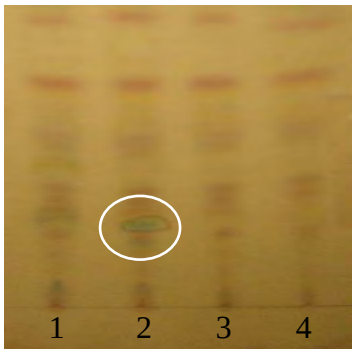
1 **2**

Figure 102. Culture of the recombinant *P. putida* TA27 at (1) at 16 ^{o}C and (2) at 30 ^{o}C.

Ethyl acetate was used to extract compounds from the bacteria culture (section 7.3.7). The obtained mixture in ethyl acetate was then analyzed in an attempt to detect the target compound encoded by the gene cluster of the cosmid pPD7E4.

5.2.7.2 TLC of the expressed compound

After developing TLC in eleven different mobile phases (Table 51) and using two different color reagents (ninhydrin and anisaldehyde), the mobile phase chloroform: methanol (6:1) at pH 9 and anisaldehyde were able to give a good separation (Figure 103).



Mobile phase: Chloroform: Methanol (6:1), pH 9
Color reagent: anisaldehyde

Figure 103. TLC of the extracted compounds from *P. putida* TA27 (2) at 16 °C (4) at 30 °C and (1) (3) negative control.

Discussion

Resulting from heterologous expression in *P. putida*, the mRNA transcripts of the *ppp* genes inside pPD7E4 could be detected by RT-PCR.

TLC showed that there was no difference between the extracted mixtures from the two cultures of *P. putida* cultivated at 30 °C. However, after staining, a green spot was detected at R_F = 0.15 in the transformed *P. putida* cultivated at 16 °C, which was not observed in the negative sample (Figure 103). It was therefore likely that the expressed compound would be present in this fraction. Only the extracted sample of the transformed *P. putida* cultivated at 16 °C would be used in the further steps.

The distinct green spot was detected in a normal phase TLC at a rather low R_F. This indicates that the analyzed compound is probably nonpolar. Furthermore, this spot is not fluorescent under UV light, so the corresponding compound is probably lacking double bonds (Figure 84).

Anisaldehyde is useful to detect different compounds in a mixture. It is able to detect several substances such as sugars, phenols, terpens, and glycosides, etc. in a wide range of color, green,

violet, blue, red, and grey. Nevertheless, it is rather difficult to deduce exactly the chemical scaffold of the target compound if based only on a green color of the spot.

Although the expressed compound was predicted to be an amine (Figure 84), the amino group could not be detected using the spray reagent ninhydrin. Perhaps, this primary amino group is transformed into derivatives. Further analysis has to be conducted in another PhD project, since the time frame of the present work did not allow for more detailed studies on the nature of the compound.

Chapter 6
Summary and outlook

6.1 Studies on the onnamide gene cluster from the sponge *T. swimhoei*

6.1.1 Isolation of the biosynthetic genes of onnamides from the sponge metagenome

Based on the previously sequenced genomic regions, by applying PCR screening of semi-liquid pools and the primer walking method, the remaining *onn* genes were successfully isolated from the 400,000 clone library constructed from the sponge metagenomic DNA. This fosmid library covering 32 Gbp of the sponge metagenome was sufficient for isolating the entire *onn* gene cluster. Six fosmids containing PKS genes were newly isolated. Among them, the fosmid pTSTA11 harboring an NRPS module that was predicted to be responsible for incorporation of an arginine residue into the onnamide structure was identified and completely sequenced. To obtain insights into the modular architecture as well as the biosynthetic pathway of the onnamide PKS, a corresponding ~ 99 kb sponge genomic region was sequenced and the *onn* cluster occupying ~ 64 kb was revealed. Both ends outside of this cluster were non-PKS genes including transposase genes, indicating that the entire onnamide PKS gene cluster was cloned. Upon analysis of this sequence, the genes required for onnamide synthesis were identified in ten ORFs, namely *onnA*-*onnJ*, and distributed on only one genomic region. The PKS domain organization is a striking counterpart of

that of the pederin PKS. Although the AT and some tailoring genes have not been found yet, all genes predicted to be involved in biosynthesis of the onnamide B core skeleton have been isolated.

In addition to the three fosmids containing the entire *onn* cluster, PKS genes were also found on other seven fosmids. This fact, as well as different termini that were identified in the isolated PKS regions was strong evidence of the presence of multiple variants of the *onn* gene cluster, which could be responsible for the production of individual onnamide-type compounds. This finding elucidates the fact that *T. swinhoei* is a rich source of diverse compounds belonging to the pederin family. This result showed that researches on distinct biosynthetic systems from rich natural product libraries from a diverse symbiont community can be reached. It will be interesting to explain the relationship between ecology and such diverse chemical libraries in the context of pathway evolution.

6.1.2 Insights into the biosynthetic pathway of onnamides

According to the isolated *onn* genes, the biosynthetic pathway of onnamide B was identified and is highly similar to that of pederin. The PKS architecture strikingly matches with onnamide B structure. For example, the GNAT should select the acetyl starter unit, two unusual CR domains seem to be responsible for an exomethylene bond branch, a tandem DH domain for tetrahydropyrane ring, MT domains for the formation of geminal dimethyl groups, two NRPS modules for incorporation of the two amino acid glycine and arginine residues, etc. In comparison with the *ped* cluster, as predicted, an oxygenase gene that is homologue to *pedG* is not present on the *onn* system. This corroborates the role of *pedG* located on the upstream region of *pedH* in order to catalyze a cleavage reaction, according to which the “sponge-type” intermediate compound bearing arginine residue might be transformed into pederin (Figure 62). The only discrepancy between the PKS architecture and the structure of onnamide B is the lack of an ER domain on module 11, which should be required for complete reduction of the intermediate. However, it has recently been shown that in *trans*-AT PKSs, the ER function is usually provided in *trans* as well (Figure 60).[138]

6.1.3 Identification of the onnamide producer

In addition to the *onn* genes, several non-PKS genes were also isolated from the sponge metagenome. Among them, *ftsZ* was a good candidate for investigation of onnamide producers due to its conservation among prokaryotes. *ftsZ* encoding a tubulin-like protein was subjected to phylogenetic analysis in order to identify its producer. When applying this method for a complex bacterial community such as sponge symbionts, horizontal gene transfer of genes between species occur frequently, resulting in taxonomic errors. However, it is usually assumed that proteins being a part of multicomponent protein systems *e.g.* FtsZ, are not often subjected to horizontal transfer. According to this analysis, the owner of *ftsZ*, also being the onnamide producer, belongs to the phylum Chloroflexi (Figure 27). In addition, the presence of multiple closely related *ftsZ* genes from different members of Chloroflexi inside the sponge metagenome is also further evidence of the existence of multiple closely related strains that could carry the variants of the onnamide PKS. Further analyses should be conducted to explore the preservation and evolution of the two similar PKS systems, namely the pederin and onnamide gene clusters, isolated from the two distinct bacterial phyla, Proteobacteria from the beetle *P. fuscipes* and Chloroflexi from the sponge *T. swinhoei*, respectively.

6.1.4 Outlook

Achievement of genetic studies on onnamide PKS from *T. swinhoei* starts a new era of researches on natural products isolated from diverse sources such as unculturable bacterial symbionts. Onnamide B was the first sponge-derived compound, for which biosynthetic genes were cloned by metagenomic approach. The results prove that genetic screening of metagenomic libraries is an effective method to target biosynthetic genes, even that of uncharacterized compounds as well as of multiple variants of complex systems in symbiont associations. The strategy applied for isolating the onnamide genes will be a good model for future works on natural products from diverse bacterial symbioses. With the possessing *onn* genes in hand, not only insights into the biosynthetic pathway of onnamides can be obtained but also fundamental data about renewable production by heterologous expression in culturable bacteria. This strategy could also be applied for scaled-up production of other rare marine drug candidates. Furthermore, comparisons between biosynthetic genes and chemical structures together with gene knock-out experiments or expression trials of target genes will be a useful way to elucidate unusual biochemical processes in nature.

6.2 Heterologous expression trials with a PKS-NRPS cluster on the cosmid pPD7E4 from the beetle *P. fuscipes*

Based on analysis of the complete sequence of the small PKS-NRPS cluster not belonging to pederin cluster that was isolated from the genome of the *P. fuscipes* symbiont, the structure of unknown polyketide was predicted. By λ RED-mediated homologous recombination, this PKS-NRPS gene cluster with a size of about 37.5 kb located on the cosmid pPD7E4 was introduced successfully into the expression vector pJB861 and then transformed into the host cell *P. putida* KT2440. Plasmid DNA from the recombinant strain *P. putida* TA27 was checked by restriction analysis. mRNAs from *P. putida* TA27 that were transcribed from the entire PKS-NRPS gene cluster were identified by RT-PCR. Up to this step, the microbial expression system including the expression vector pJB861 and the host cell *P. putida* KT2440 is recognized to be an effective tool in order to express PKS-NRPS genes. The extracted mixture of *P. putida* TA27 was analyzed by TLC with the spray agent anisaldehyde and a distinct green spot at R_F = 0.15 was detected. Therefore, purification and concentration of compounds in this fraction are future tasks in an attempt to obtain the target compound that will be structurally characterized and bioassayed.

Chapter 7
Materials and methods

7.1 Materials

7.1.1 *Theonella swinhoei*

The sponge *T. swinhoei* Y (yellow) specimens were collected by scuba diving at Hachijo-jima Island, Japan, at a depth of 15 m. Immediately after collection, they were preserved either in 95% ethanol at 4 ^{o}C or shock-frozen in liquid nitrogen, followed by storage at -80 ^{o}C.

7.1.2 *Paederus fuscipes*

Specimens of *P. fuscipes* were collected by J. Piel and by H. Başpinar, University of Aydin, Turkey and stored in 100% ethanol to preserve the DNA.

7.1.3 Chemicals

Table 41. Chemicals

Chemical	Source
Acetonitrile	Fisher scientific, Schwerte
Adenosine triphosphate (ATP)	Roth, Karlsruhe
Ammonium hydroxide solution	Riedel-de-Haën, Seelze
Acrylamide-bisacrylamide 37,5:1 (Rotiphoresegel 30)	Roth, Karlsruhe
Agar	Roth, Karlsruhe
Agarose NEEO	Roth, Karlsruhe
L-Arabinose	Fluka / Riedel-de-Haën, Seelze
Ammonium peroxodisulfate (APS)	Merck, Darmstadt
Bovine Serum Albumin (BSA)	NEB, Frankfurt/Main
BSA	Roth, Karlsruhe
Brilliant-Blue R250	Roth, Karlsruhe
5-Bromo-4-chloro-3-indoxyl-β-D-galactopyranoside (X-Gal)	AppliChem, Darmstadt
Bromophenol blue sodium salt	AppliChem, Darmstadt
Calcium chloride	KMF, Lohmar
Chloroform	Merck, Darmstadt
Desoxythymidine-5’-triphosphate (dTTP)	Invitrogen, Karlsruhe
p-Anisaldehyde	Sigma-Aldrich, Seelze
Desoxynucleotide (dNTPs)	Invitrogen, Karlsruhe
N, N-Dimethylformamide (DMF)	Fisher-Scientific, Schwerte
Disodium hydrogenphosphate dihydrate	Fluka / Riedel-de-Haën, Seelze
1,4-Dithiothreit (DTT)	Roth, Karlsruhe
100 bp DNA ladder extended	Roth, Karlsruhe
1 kb DNA ladder	Roth, Karlsruhe
1 kb DNA extension ladder	Invitrogen, Karlsruhe
Glacial acetic acid	Merck, Darmstadt
Ethanol	Merck, Darmstadt
Ethidiumbromide	Roth, Karlsruhe

Chemical	Source
Ethylenediaminetetraacetic acid (EDTA)	KMF, Lohmar
Ethyl acetate	Fisher scientific, Schwerte
D(+)-Glucose monohydrate	Fluka / Riedel-de-Haën, Seelze
Glycerin	Merck, Darmstadt
Glycine	Fisher-Scientific, Schwerte
Imidazole	Roth, Karlsruhe
Isopropanol	Fisher scientific, Schwerte
Isopropylthiogalactoside (IPTG)	Roth, Karlsruhe
Magnesium chloride hexahydrate	Fluka / Riedel-de-Haën, Seelze
Mercaptoethanol	Roth, Karlsruhe
Methanol	Merck, Darmstadt
Sodium acetate	Acros Organics, Geel (B)
Sodium chloride	Grüssing, Filsum
Sodium dihydrogenphosphate dihydrate	Roth, Karlsruhe
Sodium dodecylsulfate (SDS)	Roth, Karlsruhe
Sodium hydroxide	KMF, Lohmar
Ni-NTA Agarose	Qiagen, Hilden
Orange G	Sigma-Aldrich, Seelze
Perchloric acid	JT Barker
Potassium acetate	ABCR, Karlsruhe
Protein marker Broad-Range	NEB, Frankfurt/Main
Protein marker RotiMark Standard	Roth, Karlsruhe
Roti-Nanoquant	Roth, Karlsruhe
Seaprep agarose	Lonza, Switzeland
Sucrose	Riedel-de-Haën, Seelze
Sulfuric acid	Fluka/ Riedel-de-Haën, Seelze
N,N,N’,N’-Tetramethylethylenediamine (TEMED)	Merck, Darmstadt
Tryptone	Becton Dickinson, Heidelberg
Xylencyanol	Merck, Darmstadt

7.1.4 Chemical solutions

Table 42. Chemical solutions

Solution and buffer	**Composition**	**Store**
Alkaline lysis solution I (P1)	50 mM Tris-HCl pH 8 10 mM EDTA 100 µg/ml RNAse	4 ^{o}C
Alkaline lysis solution II (P2)	200 mM NaOH 1% SDS	Room temperature
Alkaline lysis solution III	3 M KOAc pH 5.5	Room temperature
Anisaldehyde	10 ml anisaldehyde 420 ml ethanol 15 ml sulfuric concentration acid 5 ml acetic acid	Room temperature
Bacteria cell lysis buffer pH8	25 mM Tris-HCl 0.5 M NaCl	Room temperature
Destaining protein solution	30% methanol 10% glacial acetic acid	Room temperature
Gel loading dye	0.05% bromophenol blue 0.05% xylene cyanol 0.05% orange G 40% sucrose	4 ^{o}C
IPTG	20 mg/ml H_2O	-20 ^{o}C
Lysis buffer for isolation of	8 M urea	No stock

Solution and buffer	Composition	Store
sponge DNA	2% sarkosyl 350 mM NaCl 50 mM EDTA 50 mM Tris-HCl, pH 7.5	No stock
Ninhydrin	0.2 g ninhydrin 100 ml methanol	Room temperature
Protein gel electrophoresis buffer 10X	250 mM Tris-HCl 2 M glycine 1% SDS	Room temperature
Protein gel loading dye 2X	0.09 M Tris-HCl pH6.8 20% glycerol 2% SDS 0.02% bromophenol blue 0.1 M DDT	4 ^{o}C
Staining protein dye	0.25% coomassie 30% methanol 10% glacia acetic acid	Room temperature
TAE 50X	242 g Tris-base 57.1 ml Glacial acetic acid 100 ml EDTA (0.5M) Ad 1 l H_2O	Room temperature
TE	10 mM Tris-HCl (pH 7.5) 1 mM EDTA (pH 8.0)	Room temperature
X-Gal	20 mg/ml DMF	-20 ^{o}C

7.1.5 Enzymes

Table 43. Enzymes

Enzyme	**Source**
Antarctic phosphatase	NEB, Frankfurt/Main
Hot Start Polymerase	Jena Bioscience, Jena
Lysozyme	Roth, Karlsruhe
Restriction enzymes	NEB, Frankfurt/Main
	Jena Bioscience, Jena
RNase	Roth, Karlsruhe
RNase-free DNase set	Qiagen
T4 DNA ligase	NEB, Frankfurt/Main
	Jena Bioscience, Jena
Taq DNA Polymerase	NEB, Frankfurt/Main

7.1.6 Antibiotics

Table 44. Antibiotics

Antibiotic	**Concentration**	**Source**
Ampicillin	100 µg/ml	Roth, Karlsruhe
Apramycin	50 µg/ml	Sigma, Steinheim
Chloramphenicol	12.5 µg/ml -25 µg/ml	Fluka / Riedel-de-Haën, Seelze
Kanamycin	50 µg/ml	Roth, Karlsruhe

7.1.7 Vectors and bacteria

Table 45. Vectors

Vector	**Source**
pBluescript SK II (-)	Stratagene (Amsterdam, Netherlands)
pCC1FOS	Epicentre (Hess. Oldendorf, Germany)
pHis8-3	Derivative of pET28a (Novagen, Merck KGaA, Darmstadt, Germany)
pJB861	Derivative of RK2 pJB861 (Blatny, J.M., et al, Plasmid, 38 (1), 35-51, 1997) NCB NCBI Accession No: U82000
pWEB	Epicentre (Hess. Oldendorf, Germany)

Table 46. Bacterial strains

Bacteria	**Genotype**
E. coli BL21 (DE3)	F^- *ompT* $hsdS_B(r_B^-m_B^-)$ *gal dcm* (DE3)
E. coli BW25113	$lacI^q rrnB_{T14} \Delta lacZ_{WJ16} hsdR514\ \Delta araBAD_{AH33} \Delta rhaBAD_{LD78}$
E. coli EPI300™-T1R	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1*
Phage T1-resistant	*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)]
P. putida KT2440	*rmo − mod+ hsdR*

7.1.8 Others

- CopyControl™ Induction Solution (Epicentre)
- QIAquick Gel Extraction Kit (Qiagen)or GFX PCR DNA and Gel Band Purification Kit (GE Healthcare)
- RNeasy Mini Isolation from Bacteria (Qiagen)
- Expand High Fidelity ^{plus} PCR system (Roche)
- SuperScript™ III One-Step RT-PCR System with Platinum® Taq DNA Polymerase (Invitrogen)

7.1.9 Bacterial cultivation media

LB: 10 g tryptone, 10 g NaCl, 5 g yeast extract. Added 1 l H_2O, pH 7.

LB agar: 10 g tryptone, 10 g NaCl, 5 g yeast extract, 15 g agar. Added 1 l H_2O, pH 7.

Semi-liquid SeaPrep medium: 10 g tryptone, 10 g NaCl, 5 g yeast extract, 800 ml H_2O, 6 g SeaPrep agarose. Put in microwave in 5 min. Added up to 1 l H_2O.

SOB: 20 g tryptone, 0.584 g NaCl, 5 g yeast extract, 0.186 g KCl. Added 1 l H_2O, pH 7.

All media were autoclaved at 121 ^oC for 20 min.

7.1.10 Primers

Table 47. Primers

Name	Sequence (5' – 3')	Annealing temperature
pTSSH3-1 primers	Forward: GAG AAT TCG GCA AGC AGA Reverse: TGT GAC CCA TAG CCC ATT C	54 ^oC
pTSSH3-2 primers	Forward: AAG TGA GTT CGG GCA GTG TCC Reverse: CCC GTT TGG CCT ACT TCC T	60 ^oC
pTSSH3-3 primers	Forward: GGC AGC CCG ATA CAC CA Reverse: TCG GGA GAG CAG GCA TCA	58 ^oC
pTSSH3-4 primers	Forward: AGC GGC TTC CAG ATG ACC Reverse: TTG CCG TGA ATT CCG CTT	54 ^oC
ftsZ primers	Forward: CGC TGC GAC GTA CCC AAC C Reverse: CGT CAT CGG ACA TAG GCG TAA C	56 ^oC
pTSTA4-KS primers	Forward: GGA AAT GTC ATC GTG AAA GC Reverse: CTG GAC AAA GGC AAG AAG	54 ^oC
pTSTA8-KS primers	Forward: CAC CGG CAC GAA GTT AG Reverse: ACC GCT CCT TGT TTC TGG	55 ^oC
DH primers	Forward: GCA TTA GTC TAC CGA CCT Reverse: CTC CAT GAA GAT CGA CCT	54 ^oC
A-OnnI2 primers	Forward: GAA TTC GCG GAC TAT TCC CTG CTG Reverse: AAG CTT TCA TCT TCA ATT GCA GAC TGA G	57 ^oC
A-OnnJ4 primers	Forward: GGA TCC TCA AAG ATC GCC TTC CTC AC Reverse: GTC GAC TCA TTC TCC AGC TTT GTG GAA G	61 ^oC

Name	Sequence (5' – 3')	Annealing temperature
Sponge8-primers	Forward: TGG ATC CAC AGC AAC GCC TCT TTT Reverse: GAC TAC AAT GCC CAC GCC TTC AGC	65 °C
Gap pTSTA2 primers	Forward: GAC ATT CTG TTG ACG GGA TTG Reverse: CTG GAT GAG ACG CCA GAC AT	52 °C
r1 primers	Forward: GAT GAC GGT GAA AAC CTC TG Reverse: CGG TAT CAT TGC AGC ACT G	51 °C
f1 primers	Forward: GCA ATT TTT GCG GTA GCC TTG GCA Reverse: GCA ATT TTT GCG GTA GCC TTG	52 °C
r2 primers	Forward: GTC TTG CGT CTG GAA CTG AC Reverse: CAC CTG GCT GAT TTT CAC G	51 °C
f2 primers	Forward: CAT CTA TTT TAG CCG CAC GTC Reverse: GAG CTG TAT CGT TCT CAG TCG	52 °C
f3 primers	Forward: GTA ATC ACA CGG ATG AGG AAG Reverse: GTG CCT AAT GAG TGA GCT AAC	52 °C
r3 primers	Forward: GTT GCA GTG AGC CGA GAT C Reverse: GAA GCT CCA TAG CCA CTA AAC	52 °C
G side of pWEB primers	Forward: ACT AGT CTC TCC CTA TAG TGA GTC GT Reverse: CAA TGG TCG CCA TGA TCG CGT AGT	55 °C
C side of pWEB primers	Forward: GAA TTC CTC TCA AGG ATC TTA CCG CT Reverse: ACT AGT CGC GAA TTC TTG AAG ACG A	54 °C
pppD primers	Forward: GAC AGT GGA CGT TTT GAA G Reverse: CTT CAT CGG CAA GTT CGG	55 °C
pppH primers	Forward: CTC GTG GGA CAC AGT ATG Reverse: CTC ACT TTA CGC ACG AAG	54 °C
pppJ primers	Forward: GAA AAG GGA CGC TTG TC Reverse: GAT GAA CCG ATC AAG CTG	52 °C
pppK primers	Forward: GTC CAG TTG TTC AGC CTC Reverse: CAC ATT ATC GAG CGT CTC	54 °C
pCC1/pEpiFOS primers	Forward: GGA TGT GCT GCA AGG CGA TTA AGT TGG Reverse: CTC GTA TGT TGT GTG GAA TTG TGA GC	

7.1.11 Equipment

Table 48. Equipment

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

Equipment	Manufactory
Pipette tip	Sarstedt
TLC silica-gel 60F254	Merck, Darmstadt
Ultraviolet crosslinker	UVP, Cambridge, UK
UV cuvettes	Sarstedt

7.2 Methods for chapter 3 and chapter 4

7.2.1 Isolation of sponge DNA

The metagenomic DNA (also termed as HMW DNA or total DNA) of the sponge *T. swinhoei* was isolated by using the following protocol: 1 g of sponge was cut and pulverized in a mortar under liquid nitrogen; afterwards, nitrogen was boiled off. The ground sponge was added into 10 ml of lysis buffer and mixed gently in order to resuspend the sponge particles well. The mixture was incubated for 10 min at 60 °C and extracted with 10 ml of phenol/chloroform/isoamyl alcohol (25:24:1). The solution was handled very gently without shaking or vortexing. The aqueous phase was separated by centrifuging at 11,000 rpm for 5 min and extracted with 10 ml of chloroform. CTAB was added to the upper phase to achieve 1% final concentration, mixed and incubated for 10 min at 60 °C. A whitish precipitate formed, which was removed by centrifuging at 11,000 rpm for 10 min. DNA was precipitated with two volumes of 100% ice-cold ethanol. The mixture was then mixed gently by slowly inverting the tube several times in order to mix the two layers and stored at -20 °C for 30 min to improve precipitation. The DNA solution was centrifuged for 30 min at 11,000 rpm and the supernatant was removed carefully without disturbing the DNA pellet. The DNA was washed with 70% ethanol, centrifuged for 10 min at 11,000 rpm and the ethanol was decanted. The DNA was air-dried for a few minutes and then resuspended with 500 µl of Tris-HCl (pH 8) buffer. The HMW DNA was stored at -20 °C.

7.2.2 Construction of a fosmid library from the metagenomic DNA of the sponge *T. swinhoei*

A fosmid library of the metagenomic DNA of *T. swinhoei* was constructed previously in our lab using CopyControl™ Fosmid Library Production Kit, in which semi-liquid pool method was applied for subsequent plating and screening.[132] The fosmid library comprising approximately

400,000 clones was distributed in 405 tubes referred to as pools. Correspondingly, every pool possessed about 1,000 clones. These pools were stored at -80 ^{o}C and labeled with three numbers: number of the box, of the row and of the column. For example, the pool 3A1 means the pool from box 3, row A and column 1.

7.2.3 The semi-liquid library method[132]

Identifying target genes from a total DNA, particularly from a highly complex DNA such as the sponge metagenome, is time consuming. In addition, the storage of a fosmid library of a metagenomic DNA on medium agar plates takes up a large space. The semi-liquid technique is therefore an effective solution. Semi-liquid medium is able to control the rapid growth of bacteria better than liquid medium, thereby minimizes a loss of target clones in the doubling period (Figure 104). Other advantages of semi-liquid medium are reducing risk of cross-contamination, fast and economic screening, and the library can be stored immediately at -80 °C.

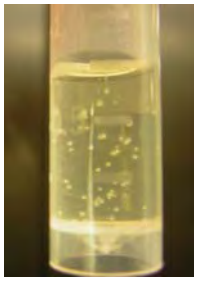


Figure 104. Bacteria colonies in semi-liquid medium.

Fixed standing 2 ml vials without cap were set up on wet ice. The appropriate colony-forming units (cfus) were added to semi-liquid SeaPrep medium and then mixed well. 1 ml aliquots of this suspension were pipetted in the above vials. These vials were kept on wet ice for one hour in order to complete the gelling process. The caps were put on the vials. The vials were then removed gently and incubated at 37 °C for about 18 hours. On the next day, each vial was vortexed for 5 sec until the suspension became homogenous. A 20 µl aliquot was taken for PCR. 500 µl of 50% glycerol was added into the remaining semi-liquid culture and stored at -80 °C.[132]

7.2.4 Creation of fosmid library superpools

Instead of PCR screening each pool individually, the pools in one row or in one column were combined and served as DNA template. Even in such complex mixtures (termed superpools), PCRs still give clear results in spite of low copy number of pCC1FOS.[132] Therefore, 45 row superpools and 45 column superpools were constructed from the 405 pools of the fosmid library of *T. swinhoei*. These superpools were labeled using two numbers: the number of the box and of the row or of the column from which they were derived. Each row superpool consisted of 30 µl x 9 of nine pools belonging to the same row, *ca*. 9,000 clones. Similar to a row superpool, a column superpool was formed by nine pools belonging to the same column (Figure 105). For example, the row superpool 3A comprised nine pools in the row A, A1-A9 of the box 3; the column superpool 31 included nine pools in the column 1, A1-I1 of the box 3.

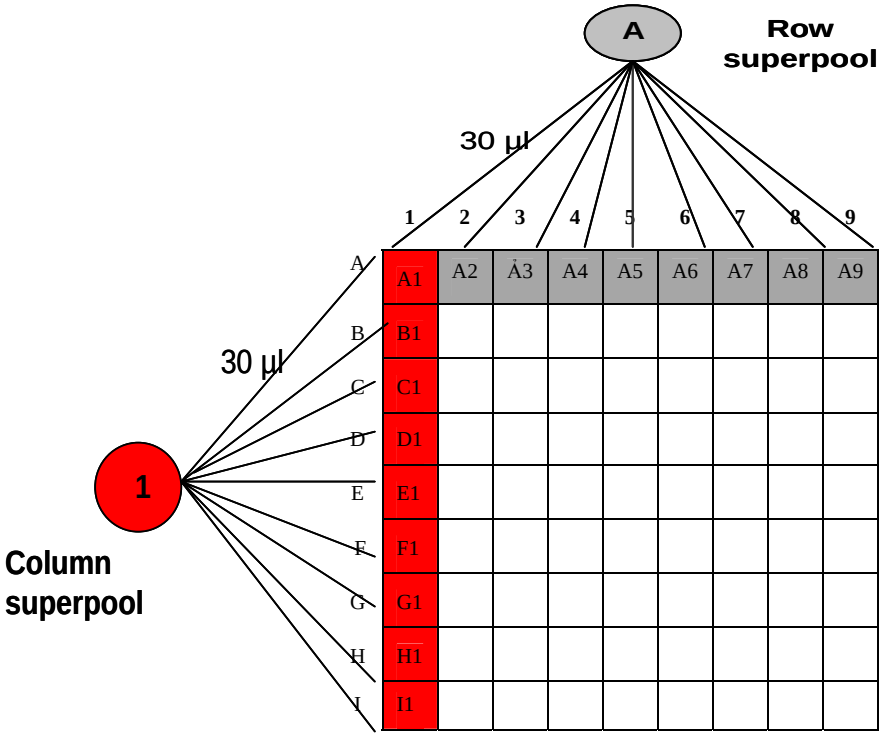


Figure 105. Creation of superpools.

7.2.5 Isolation of positive clones containing target genes from a fosmid library

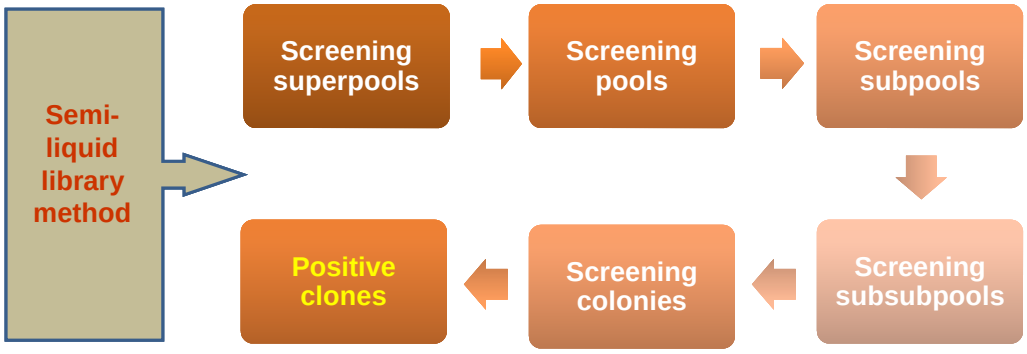


Figure 106. Flowchart for isolating positive clones from a complex fosmid library.

First, the superpools were PCR-screened with designated primers. From the identified positive superpool, PCRs were performed on the nine pools from which the superpool was derived. The positive pool was then diluted to subpools of ~100 cfus/tube and the semi-liquid library protocol was repeated.

15 µl bacteria culture of positive pool ($2x10^8$ cfus/ml) + 1,485 µl LB medium (10^{-2} dilution)

↓ $2x10^6$ cfus/ml

15 µl diluted bacteria culture + 1,485 µl LB medium (10^{-4} dilution)

↓ $2x10^4$ cfus/ml

150 µl diluted bacteria culture + 30 ml semi-liquid SeaPrep medium containing chloramphenicol 12.5µg/ml

↓ ~ 3,000 cfus

Distributing into 30 tubes, 37 ^oC, overnight (**subpools**)

After another round of PCR, the positive subpool was further diluted to subsubpool containing ~ 15 cfus and the final semi-liquid library procedure was carried out.

15 µl bacteria culture of positive subpool ($2x10^8$ cfus/ml) + 1,485 µl LB medium (10^{-2} dilution)

↓ $2x10^6$ cfus/ml

15 µl diluted bacteria culture + 1,485 µl LB medium (10^{-4} dilution)

↓ $2x10^4$ cfus/ml

150 µl diluted bacteria culture + 1,350 µl LB medium (10^{-5} dilution)

↓ $2x10^3$ cfus/ml

150µl diluted bacteria culture + 20ml semi-liquid SeaPrep medium containing chloramphenicol 12.5µg/ml

↓ ~ 300 cfus

Distributing into 20 tubes, 37 ^oC, overnight (**subsubpools**)

Once the positive subsubpool was recognized, the final dilution was prepared to generate ~ 200 separate colonies on LB agar plate.

15 µl bacteria culture of positive subsubpool ($2x10^8$ cfus/ml) + 1,485 µl LB medium (10^{-2} dilution)

↓ $2x10^6$ cfus/ml

15 µl diluted bacteria culture + 1,485 µl LB medium (10^{-4} dilution)

↓ $2x10^4$ cfus/ml

150 µl diluted bacteria culture + 1,350 µl LB medium (10^{-5} dilution)

↓ $2x10^3$ cfus/ml

Plated 100 µl diluted bacteria culture on LB agar plate containing chloramphenicol 12.5µg/ml

↓ ~ 200 cfus

37 ^oC, overnight

Individual colonies were resuspended with 30 µl LB medium and vortexed. At the end, the positive clone was identified by colony PCR.

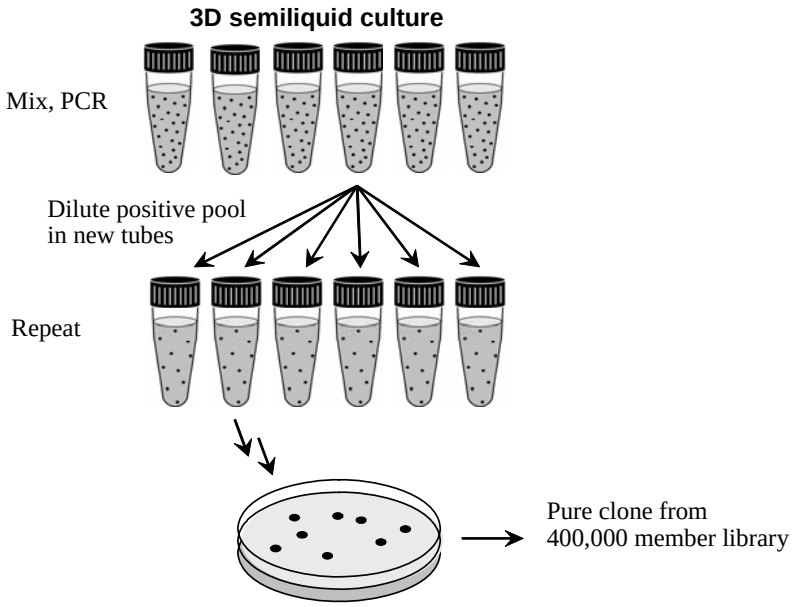


Figure 107. Isolation of positive clones from a fosmid library by PCR screening of semi-liquid pools.[132]

7.2.6 Primer walking

The purpose of this approach is to gradually reveal consecutive small fragments from a large DNA sequence which is too long to be completely isolated or sequenced in a single step. Particularly, in this study, the investigated biosynthetic genes were distributed in a giant genomic region, up to hundreds of kilobase pairs. The border of the previously sequenced region was used for primer design in order to isolate the contiguous portion of a gene cluster. This process was repeated until all target genes were isolated (Figure 108).[170]

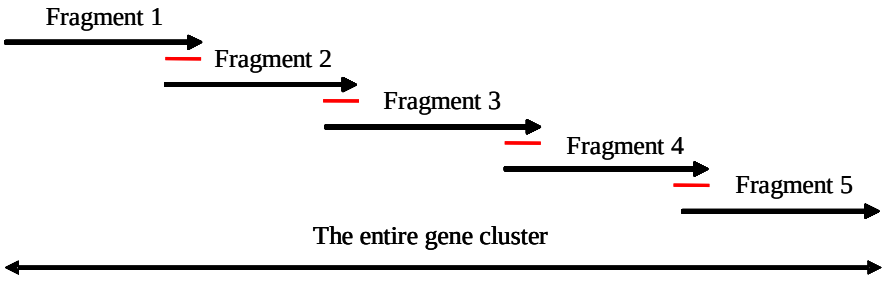


Figure 108. Diagram of the primer walking method. Primers derived from the sequenced fragments are shown in red.

7.2.7 DNA amplification by PCR

7.2.7.1 Standard PCR

The aim of PCR is to amplify a few copies of a target DNA to yield millions of identical copies. A PCR procedure commonly includes three major steps. In the denaturation step, the double-stranded DNA is denatured to single-stranded DNA at high temperature, 94 ^{o}C-98 ^{o}C. Next is the annealing step, when primers anneal to the single-stranded DNA at lower temperature, 50 ^{o}C-65 ^{o}C. And in the extension or elongation step, polymerase binds to the primer-DNA template hybrid for DNA synthesis in direction 5'-3'. The temperature in this step depends on the type of the used DNA polymerase; with *Taq* polymerase, the best working temperature is 72 ^{o}C. In addition, the final extension at 72 ^{o}C for 5-10 min following the last PCR step ensures that the single-stranded DNA is completely extended. A PCR typically consists of about 20-35 cycles, relying on the concentration of PCR compositions.[143]

PCR master mix		PCR program			
Buffer 10X ThermoPol	2.5 µl	Lid 106 ^{o}C, preheating on			
dNTPs 10 mM	0.5 µl	1. 95 ^{o}C	2 min		
BSA100X	0.25 µl	2. 95 ^{o}C	30 sec	2	34
Forward primer (50 µM)	0.25 µl	3. T ^{o}C	1 min		
Reverse primer (50 µM)	0.25 µl	4. 72 ^{o}C	1 min		
Taq polymerase	0.125 µl	5. 72 ^{o}C	5 min		
DNA template	0.5 µl	4 ^{o}C	pause		
dH_2O	20.625 µl	35 cycles			
Total	25 µl	T is the annealing temperature of primers			

In order to increase specificity of the PCR reaction, concentrations of the PCR components such as $MgCl_2$, DNA template, etc. as well as PCR conditions such as the primers' annealing temperature, the number of cycles had to be optimized.

7.2.7.2 Hot-start PCR

During preparation steps of a normal PCR, DNA polymerase can exhibit some unspecific activity at room temperature. This is the main cause of unspecific amplifications as well as primer dimmers,

resulting in a low yield of the target product. The Hot-start PCR inhibits polymerase activity at lower temperature by either manual technique or chemical modification polymerase. The Hot-start polymerase is bounded to an antibody that blocks polymerase activity during the initial denaturation step. It thus prevents non-specific amplification and the amplified DNA is cleaner and enriched.[143]

In Hot-start PCR, PCR master mix and PCR program were similar to those of a normal PCR, only one exception was that *Taq* polymerase was replaced by Hot-start polymerase of Jena Bioscience.

7.2.7.3 Touchdown PCR

It is also a modification of the normal PCR, in which annealing temperatures are varying. At the beginning of the PCR reaction, the annealing temperature T is 5-10 ^{o}C higher than the optimal primer melting temperature (Tm) and this temperature is decreased by 0.5-2 ^{o}C every cycle until the Tm of primers is reached. The PCR is continued at this temperature for the remaining cycles. In this stage, only very specific primer-DNA hybrid is amplified, resulting in reduction of non-specific PCR products and therefore, the target DNA yield is increased.[143]

PCR program

Lid 106 ^{o}C, preheating on

1. 95 ^{o}C	2 min		
2. 95 ^{o}C	45 sec		
3. T ^{o}C	45 sec		
4. 72 ^{o}C	1 min	2	14
5. 95 ^{o}C	45 sec		
6. Tm ^{o}C	45 sec		
7. 72 ^{o}C	1 min	5	29
8. 72 ^{o}C	5 min		
4 ^{o}C	pause		

35 cycles

T is the higher starting annealing temperature and Tm is the optimal final temperature of primers.

7.2.7.4 Colony PCR

This technique is convenient for identification of positive clones containing a target DNA inside a low copy plasmid DNA. Single colonies were picked up from the plate and resuspended in 30 µl LB medium. 1 µl aliquot was subjected to PCR and the remaining colony suspension can be stored at 4 ^{o}C for some days for cultivation. In a colony PCR, the initial denaturation step is prolonged for 5 min instead of 2 min as in a normal PCR, to promote cell wall and membrane disruption and DNA release.[171]

7.2.8 Induction of high copy number plasmid DNA

The fosmid library of the metagenomic DNA of *T. swinhoei* was constructed by insertion of ~ 40 kb fragments into the low-copy number fosmid vector pCC1FOS™. In order to ensure higher cloning efficiency of large insert DNAs and to maintain the insert stability, the clones were grown at single copy. Nevertheless, the vector pCC1FOS contains both the *E. coli* F-factor single-copy origin of replication and the inducible high-copy *oriV*. When a positive clone was isolated, it was then induced to high copy number, thereby resulting in increasing DNA yield. The induction procedure was carried out following the CopyControl™ Induction Solution protocol from Epicentre.

To be induced to high copy number, each clone was inoculated with 5 ml of LB medium containing 12.5 µg/ml chloramphenicol at 37 ^{o}C for overnight. In the next day, 500 µl of the overnight culture were added to 4.5 ml of fresh LB medium with chloramphenicol and 5 µl of 1,000X CopyControl Induction Solution. This clone induction culture was shaked vigorously at 37 ^{o}C for 5 hours. Afterwards, DNA was isolated following the protocol described in section 7.2.9.

7.2.9 Isolation of plasmid DNA

1.5 ml of overnight bacteria culture was centrifuged at 13,000 rpm for 3 min. Supernatant was poured off and pipetted off to drain the remaining liquid. 200 µl P1 (Table 42) was then added to the tube and vortexed until the pellet was evenly suspended. Afterwards, 200 µl P2 (Table 42) was added to the tube. The covered tubes were inverted carefully for a few times to mix the content. After a maximum of 2 min, 200 µl P3 (Table 42) was added to the tube, mixed again and centrifuged at 13,000 rpm for 3 min. The supernatant was transferred to a tube containing 500 µl

chloroform. The tubes were shaken vigorously and then centrifuged at 13,000 rpm for 3 min. The supernatant was transferred to a tube containing 350 µl isopropanol. The tubes were shaken vigorously and centrifuged at 13,000 rpm for 20 min at 4 ^{o}C. The supernatant was poured off and pipetted off to drain the remaining liquid. Finally, 500 µl ice-cold 70% ethanol was added to the tube and centrifuged at 13,000 rpm for 5 min at 4 ^{o}C. Ethanol was poured off and pipetted off to drain the remaining liquid. The plasmid DNA was then dried in a speed vacuum for 20 min and dissolved in 20-30 µl dH_2O.[143]

7.2.10 Analysis of nucleic acids by gel electrophoresis

Gel electrophoresis technique has been widely applied to separate electric charged molecules that differ in size, shape or charge through a polymer matrix *e.g.*, agarose, polyacrylamide. Nucleic acids including DNA and RNA molecular have negative charge due to the negatively-charged oxygen of the phosphate group of their backbone. Therefore, when placed in an electrical field, they move toward the anode (positive pole) of electrophoresis chamber.

An agarose gel was prepared by mixing agarose with 1X TAE and dissolving by boiling. The concentration of agarose gel depended on the size of analyzed DNA, normally ranged from 0.5 - 2%. Since DNA is colourless, it can only be visible under UV light by adding ethidium bromide into agarose gel to a final concentration of 0.1 µg/ml. In addition, DNA sample was mixed with loading dye (the ratio should be 10:1 of DNA: loading dye) in order to easily observe DNA migration through the gel and as well increase the sample's density.[143]

7.2.11 Subcloning of fosmids for end-sequencing

7.2.11.1 Digestion of fosmid DNA using restriction enzymes

a) Creation of a restriction map of fosmid DNA

Each positive clone was cultivated with the Induction Solution (section 7.2.8), then DNA was isolated and digested with different restriction enzymes in order to perform a restriction map. For the purpose of sequencing, DNA fragments of a fosmid were introduced into the cloning vector pBluescript SK II (-) following sticky-end ligation. Usually, ten restriction enzymes (Table 49) belonging to the multiple cloning sites (MCS) of pBluescript SK II (-) were chosen to digest fosmid

DNA. This resulted in introduction of cohesive end termini to the sheared DNA fragments. The most suitable enzyme was chosen which yielded separate fragments.

Table 49. Ten restriction enzymes belonging to the MCS of pBluescript SK II (-) were used to digest fosmid DNA.

Number	Restriction enzyme	Buffer	Number	Restriction enzyme	Buffer
1	*Bam*HI	*Bam*HI	6	*Xho*I	2
2	*Spe*I	2	7	*Kpn*I	1
3	*Eco*RI	*Eco*RI	8	*Sac*I	1
4	*Eco*RV	3	9	*Sal*I	3
5	*Hin*dIII	2	10	*Xba*I	2

Digestive formula of DNA

DNA template	25 µl
Buffer	7.5 µl
Restriction enzyme	5 µl
BSA (if required)	0.625 µl
dH_2O	11.875 µl
Total	50 µl

37 ^{o}C, 3 hours

b) Excision of DNA fragments from agarose gel

Following a digestion, DNA fragments needed to be extracted and purified. A 0.8-1% agarose gel in 1X TAE buffer without ethidium bromide was prepared. The sticky tape was used to create a large well for DNA loading lane. A 50 µl digestion reaction of fosmid DNA was normally performed. A 45 µl aliquot of the reaction was loaded into the large well, while the rest was loaded into the next well. 2 µl of 1 kb DNA marker were loaded on each side. The sample was resolved by gel electrophoresis at 110 V for 1 hour. After that, a slice of gel containing the marker well, the 5 µl well and 3 mm of the large well was removed. The gel slice was stained with ethidium bromide for 30 min. Afterwards positions of DNA fragments were located and marked under UV light. The

stained gel was then reassembled with the remaining portion of gel. Finally, the bands from the unstained portion were excised (Figure 109).

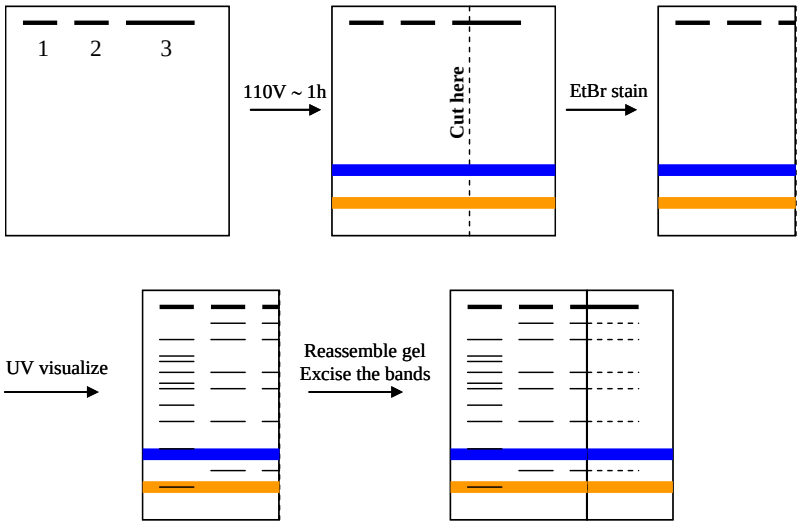


Figure 109. Excision of DNA fragment from agarose gel (1)1kb DNA marker (2) 5 µl digested DNA (3) 45 µl digested DNA.

c) Recovery of DNA fragments from agarose gel

The extraction and purification of DNA fragments from agarose gel were carried out following the QIAquick Gel Extraction Kit protocol or GFX PCR DNA and Gel Band Purification Kit protocol. The purified DNA was eluted normally in ~ 20 µl of dH_2O.

7.2.11.2 Linearization of the cloning vector pBluescript SK II (-)

The vector pBluescript is often used for sequencing purpose because of its convenient characteristics. Its MCS is flanked by T3 and T7 promoters; the selectable marker is Amp^R gene for antibiotic selection and the *LacZ* gene for blue/white color screening on IPTG and X-gal medium plate.[172]

In preparation for sticky-end ligation with a foreign DNA, the vector pBluescript SK II (-) had to be linearized by the same restriction enzyme as that used to digest the fosmid DNA.

Linearization formula of pBluescript SK II (-)

pBluescript DNA	20 µl
Buffer	5 µl
Restriction enzyme	2 µl
BSA (if required)	5 µl
dH_2O	18 µl
Total	50 µl

37 ^oC, 3 hours

7.2.11.3 Phosphatase treatment of the linear vector

In order to remove 5'phosphate group and prevent vector self-ligation, the linear vector was treated with Antarctic phosphatase.

Phosphatase treatment formula of the linearized vector

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

37 ^oC, 1 hour. Heat inactivate at 65 ^oC for 20 min

The treated vector was then purified using the GFX PCR DNA and Gel Band Purification Kit.

7.2.11.4 Sticky-end ligation of the DNA fragment and the linear vector

T4 DNA ligase was used to joint cohesive end termini of the DNA fragment and the linear vector. In addition, this enzyme also has the ability to catalyze the formation of a phosphodiester bond between 5'phosphate and 3' hydroxyl group in duplex DNA. A ratio of 4:1 of insert: vector was used for sticky-end ligation and for blunt-end ligation, the ratio should be 10:1.

Ligation formula

Insert : Vector	4 : 1
T4 buffer	2 µl
T4 ligase	1 µl
dH_2O	2 µl
Total	20 µl

16 °C, overnight. Heat inactivate at 65 ^{o}C for 20 min.

Finally, the modified genetic construct was transformed into competent *E. coli* XL1 blue cells by electroporation as described in section 7.2.14. The transformed colonies were then selected by blue/white screening on an LB agar plate with ampicillin (100 µg/ml) and also containing 60 µl IPTG (20 mg/ml) and 60 µl of X-Gal (20 mg/ml) which were plated freshly.

7.2.12 TA cloning

TA cloning is a useful method for cloning a PCR product. For preparation of TA cloning, in the PCR program, the final elongation step is often extended for 10 min, instead of 5 min as in a normal PCR. In this process, the single nucleotide A is naturally added to the 3'end of a PCR product resulting in introduction of an A' overhang terminus.

For cloning into pBluescript SK II (-), the cloning vector was linearized with the blunt cutter *Eco*RV and purified from an agarose gel. Afterwards, to the linear vector 3'-T overhangs were added in a reaction using dTTPs. Such vectors are called T-vectors.[143]

Formula of introduction of 3' T overhangs into pBluescript SK II (-)

Linearized vector by *Eco*RV	15 µl
Buffer 10 X ThermoPol	2 µl
dTTP (100 mM)	0.5 µl
Taq polymerase	2 µl
Total	20µl

72 ^{o}C, 2 hours

The T-vector was purified using the GFX PCR DNA and Gel Band Purification Kit. The 3'A overhang PCR product could thus be introduced into such a T-vector using T4 ligase following the formula in section 7.2.11.4. Finally, the modified genetic construct was transformed into competent *E. coli* XL1 blue cells by electroporation as described in section 7.2.14 and selected by blue/white screening and antibiotic selection.

7.2.13 Preparation of electrocompetent *E. coli* XL1 blue cells

Salt content in bacteria culture causes arching in electroporation process and as a result, high salt concentration reduces the transformation efficiency. For this reason, *E. coli* or other bacteria used as electrocompetent cells must be washed thoroughly to be "salt-free" by cold 10% glycerol and can be stored at -80 ^{o}C.[173]

Usually, cells were harvested in the 'log-phase growth' when the culture reached an OD_{600} of 0.4-0.8. This was done because an excessive number of living cells in an electrical field might be the cause of apoptosis and necrosis phenomenon in which the cell morphology was changed.[174]

5 ml LB medium with *E. coli* XL1 blue was inoculated via toothpick. The culture was incubated overnight on a shaker (200 rpm) at 37 °C. On the next day, the 5 ml overnight culture was poured to 200 ml LB medium in EM-flask and incubated on a shaker (200 rpm) at 37 °C till OD_{600} was around 0.4-0.8. The culture was placed immediately on ice and the cells should be kept on ice for the subsequent steps. The 200 ml culture was centrifuged at 5,000 rpm, 4 °C for 5 min and supernatant was discarded. 100 ml of cold 10% glycerol was added and then centrifuged at 5,000 rpm, 4 °C for 5 min and supernatant was discarded again. 50 ml of cold 10% glycerol was added and centrifuged at 5,000 rpm, 4 °C for 5 min; supernatant was discarded. 10 ml of cold 10% glycerol was added and then centrifuged at 5,000 rpm, 4 °C for 5 min and supernatant was discarded for the last time. 2 ml of cold 10% glycerol was added into the pellet. 70 µl of cells was pipetted to pre-chilled 1.5 ml eppendorf tubes. The aliquots were stored at -80 °C for 2-3 months after shock freezing with liquid nitrogen.[143]

7.2.14 Transformation by electroporation

Electroporation is an effective method used to transfer a foreign DNA into a host cell. In this process, an electrical pulse higher than dielectric strength in few milliseconds temporally disrupts the phospholipids bilayer of the cell plasma membrane. This causes formation of the pores that allow charged molecular DNA to cross. Once DNA passes, the pores are rapidly resealed and the phospholipids layer is spontaneously reassembled by membrane-repair mechanism.[175]

2.5 µl plasmid DNA was mixed with 50-70 µl of electrocompetent cells on ice. The mixture was transferred to an ice-cold 2 mm electroporation cuvette and an electroporation was performed using a Bio-Rad electroporator set at 2.5 kV. The expected time constant should be close to 5 milliseconds. Immediately after electroporation, 1 ml of LB medium was added into the cuvette and then transferred into a 1.5 ml tube. The electroporated cells were incubated on a shaker (200 rpm) at 37 °C for 1 hour in order to ensure the expression of the antibiotic resistance gene. Afterwards, 500 µl of transformed cells were plated on LB agar plate containing appropriate antibiotics and IPTG/X-Gal if required. The plate was then incubated at suitable temperature for 16-18 hours.[143, 172]

7.2.15 Overexpression of proteins encoded by the A domains

Proteins encoded by the A domains presented in the *onn* and the *ped* gene cluster were overexpressed following the procedure depicted below (Figure 110).

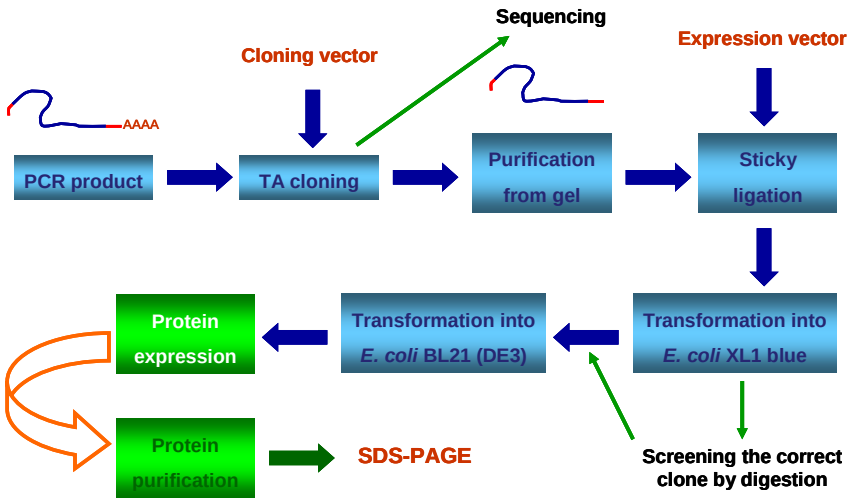


Figure 110. Protein expression diagram.

7.2.15.1 Amplification of fragments carrying the A domains of the *onn* gene cluster

The A domains were expressed using the expression vector pHis8-3 (Figure 111). Therefore, in preparation for sticky-end ligation, the sequences of restriction enzymes belonging to the MCS of the vector pHis8-3 had to be introduced into the sequences of these two A domains. Obviously, the nucleotide sequences of the chosen restriction enzymes should not be present in the A domains to avoid cleaving them during the genetic engineering process. As a further requirement, for preserving ORFs of the target proteins, the nucleotide sequences of these restriction enzymes were obligated to be present in the same triplet codon frame with the start codon of pHis8-3 (Figure 112). Fulfilling all three above conditions, the sequences of *Eco*RI, GAATTC and *Hin*dIII, AAGCTT were chosen to be introduced into the A fragment of OnnI2 (named PTA14); while the sequences of *Bam*HI, GGATCC and *Sal*I, GTCGAC were chosen for the A fragment of OnnJ4 (named PTA15).

In addition, to dominate the target protein synthesis, the presence of the start and stop codon in the expressed A domains had to be controlled. The start codon ATG was already present in the expression vector pHis8-3, so only the stop codon TGA was required to be added on the reverse primers.

A-OnnI2	Forward primer: GAA TTC GCG GAC TAT TCC CTG CTG
	Reverse primer: AAG CTT TCA TCT TCA ATT GCA GAC TGA G
A-OnnJ4	Forward primer: GGA TCC TCA AAG ATC GCC TTC CTC AC
	Reverse primer: GTC GAC TCA TTC TCC AGC TTT GTG GAA G

To ensure that all sequences were free of any modification, the Expand High Fidelity^{Plus} PCR System with HiFi polymerase was used.

After PCR reaction, in order to sequence, the amplified PTA14 and PTA15 were subjected to TA cloning as described in section 7.2.12.

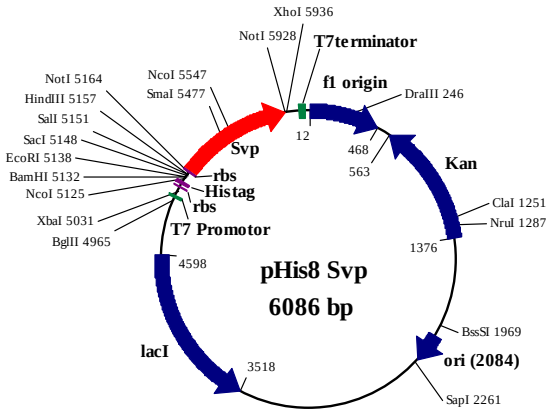


Figure 111. The expression vector pHis8-3.

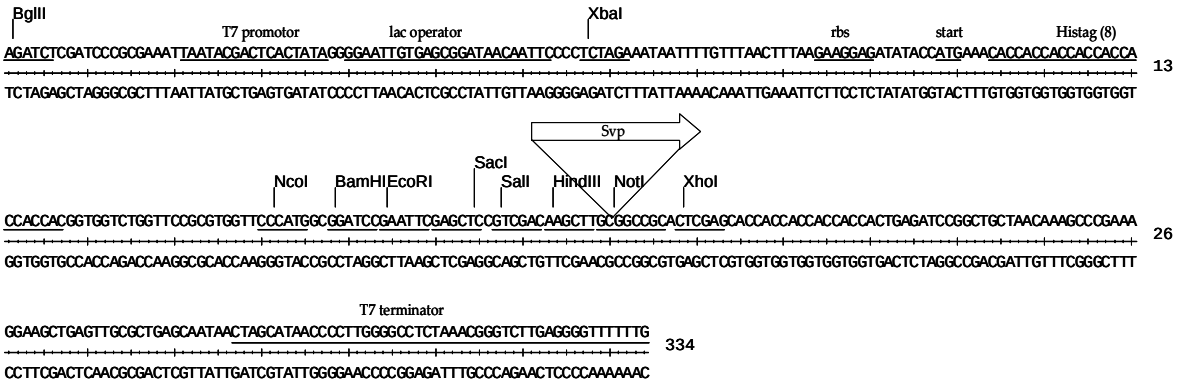


Figure 112. The MCS of the expression vector pHis8-3.

7.2.15.2 Ligation of the amplified fragments with the expression vector pHis8-3

In preparation for sticky-end ligation with pHis8-3, pTA14 was digested with *Eco*RI and *Hin*dIII, whereas pTA15 was digested with *Bam*HI and *Sal*I. Meanwhile the vector pHis8-3 was also digested with the corresponding enzymes. After purification from agarose gel using the Purification Gel Extraction kit of Qiagen, the ligation between the digested pTA14, pTA15 and vector was carried out to form pTA16 and pTA17, respectively.

Digestive formula of DNA		Digestive formula of vector	
Buffer	7.5 µl	Buffer	6 µl
Restriction enzymes	5 µl	Restriction enzymes	4 µl
DNA	25 µl	DNA of pHis8-3 vector	20 µl
BSA	0.625 µl	BSA	5 µl
dH_2O	11.875 µl	dH_2O	15 µl
Total	50 µl	Total	50 µl

37 ^oC, 3 hours

Ligation formula

DNA	7 µl
pHis8-3 vector	1 µl
T4 buffer	1 µl
T4 ligase	1 µl
Total	10 µl

16 ^oC, overnight. Heat inactivate at 65 ^oC for 20 min

7.2.15.3 Transformation of the recombinant DNA into *E. coli* XL1 blue cells

pTA16 and pTA17 were transformed into competent *E. coli* XL1 blue cells by electroporation (section 7.2.14). The transformed cells were selected on LB agar plates containing kanamycin (resistance gene on pHis8-3).

7.2.15.4 Transformation of the recombinant DNA from *E. coli* XL1 blue into *E. coli* BL21 (DE3) cells

pTA16 and pTA17 from the transformed *E. coli* XL1 blue were introduced into electrocompetent *E. coli* BL21 (DE3) cells (section 7.2.14) and selected on LB agar plates containing kanamycin.

Coexpression of the A domains of the ped *gene cluster with the chaperone plasmid pKJE7*

pKZ183-1 comprised the A domain of pedF2 and pHis8-3; and pKZ174-10 comprised the A domain of pedH6 and pHis8-3. They were transformed into the competent *E. coli* BL21 (DE3) cells containing the chaperone plasmid pKJE7 of Takara[144] by electroporation (section 7.2.14). The

transformed cells were selected on LB agar plates containing kanamycin (resistance gene on pHis8-3) and chloramphenicol (resistance gene on pKJE7) (Figure 113).

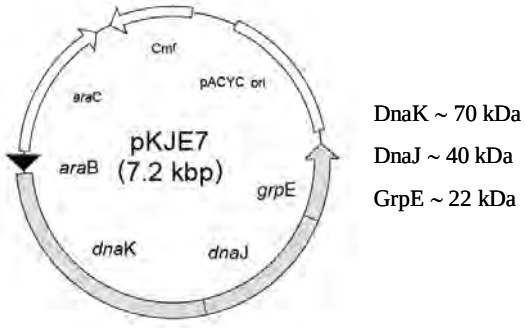


Figure 113. The chaperone plasmid pKJE7.

7.2.15.5 Overexpression of the proteins encoded by the A domains

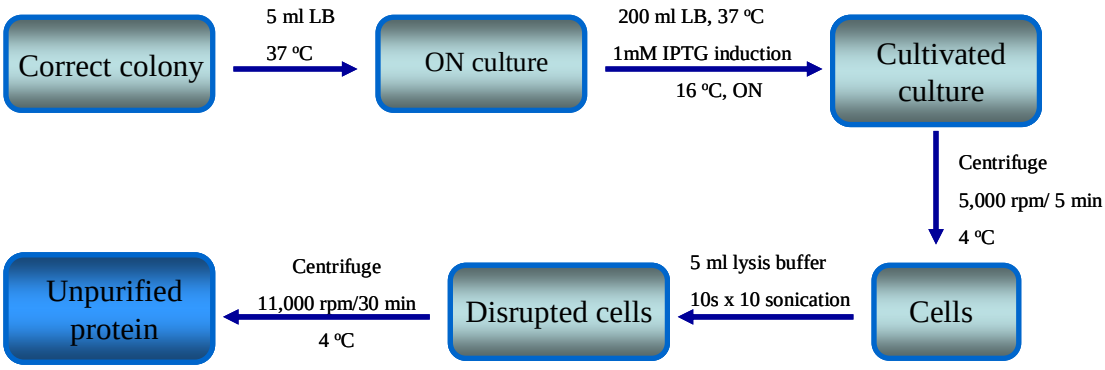


Figure 114. Protein overexpression flowchart.

The correct transformed *E. coli* BL21 (DE3) colony containing the A domain was inoculated overnight with 5 ml LB culture at 37 °C. In the following day, 5 ml overnight culture was poured to 200 ml LB medium in EM-flask and shaken at 200 rpm at 37 °C till OD_{600} reached 0.5-1. 200 µl of 1M IPTG stock was then added to a final concentration of 1 mM as the inducer of pHis8-3 and continued to be cultivated overnight at 16 °C.[176, 177] In the case of coexpression with the chaperone plasmid pKJE7, in addition to IPTG, arabinose was used as the inducer of pKJE7 with a concentration of 0.5 mg/ml[144]. The 200 ml bacteria culture was centrifuged at 5,000 rpm, 4 °C for

5 min and the supernatant was discarded. The cell pellet was resuspended in 5 ml lysis buffer (Table 42). Sonication was then carried out in order to disrupt the cell membrane and release protein using high-frequency sound waves.[178] During the sonication process, the cells were kept on ice. The cells were sonicated for 10 times x 10 sec with 10 sec pauses, at 90% amplitude. After the cell disruption, the homogenate was centrifuged at 11,000 rpm, 4 °C for 30 min and the supernatant was transferred into a new tube. The pellet was kept to check on gel.

7.2.15.6 Purification of the expressed proteins

Octahistidine-tagged recombinant protein is purified by immobilization on Ni-NTA matrices based on the tight bond between imidazole ring of histidine residue and nickel ion.[179, 180] In our experiment, Ni-NTA agarose was used to improve binding capacity and inhibiting nonspecific binding.

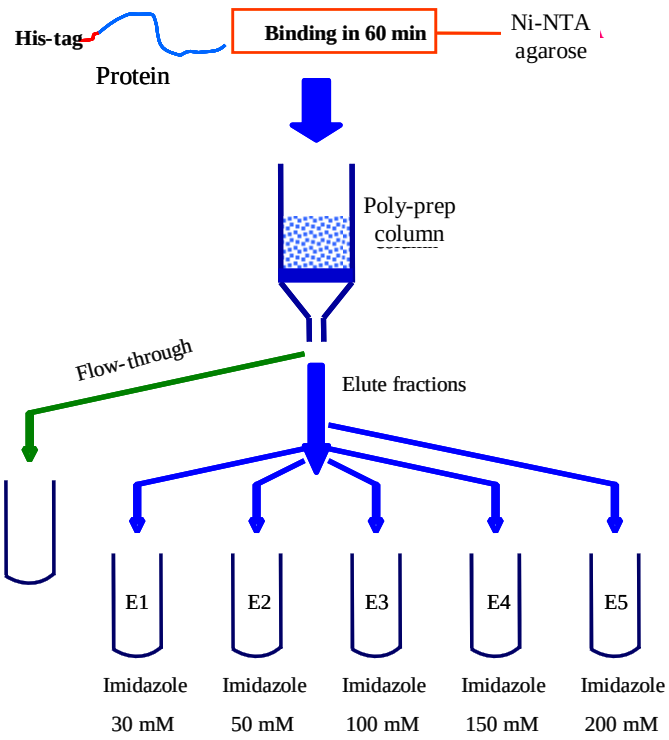


Figure 115. Purification of the His-tagged protein applying Ni-NTA technology.

200 µl Ni-NTA agarose was added to the lysed cells, incubated on ice for 1 hour and shaked at 100 rpm. After that, this mixture was loaded into a poly-prep column and the whole flow-through fraction was kept to check on SDS-PAGE. In the elution step, imidazole was used because of its ability to bind to the nickel ion. In particular, at high imidazole concentration, this affinity is stronger than that of histidine. The His-tagged protein is therefore released as it is not able to compete with imidazole in binding sites on Ni-NTA resin. A gradient of imidazole concentrations, 30 mM, 50 mM, 100 mM, 150 mM, and 200 mM in lysis buffer were prepared and 1 ml of each was loaded into the column for elute fractions (Figure 115).[142]

7.2.15.7 Analysis of the expressed proteins by SDS-PAGE

a) Separating proteins by SDS-PAGE

SDS-PAGE is a powerful technique to resolve proteins according to their sizes. However, the movement speed of a protein in a matrix depends on not only its size but also its secondary, tertiary, or quaternary structure. Therefore, in this method, proteins with different levels of structure must be linearized by incubating with SDS. SDS is an anionic detergent that is able to denature levels of protein structures and cover negative charges outside the heat denatured proteins (Figure 116).[181]

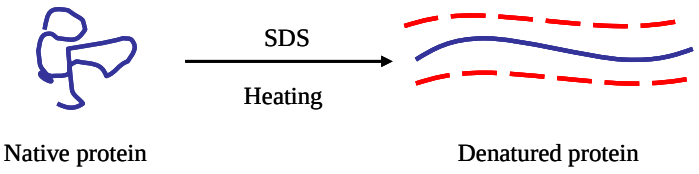


Figure 116. Denaturation of protein by SDS.

Afterwards, the denatured proteins are separated by loading them into a discontinuous polyacrylamide gel. This gel is placed in an electric field in which the denatured proteins migrate toward the positive pole. This entire process is referred to as PAGE. The polyacrylamide gel system consists of two sequential layers of gel. They are the top gel, also termed as the stacking gel and the lower gel, known as the separating or resolving gel. The stacking gel is slightly acidic (pH 6.8) and has a lower acrylamide concentration. Under these conditions, the SDS-coated proteins are resolved weakly but concentrated to several folds and a sharp starting band is formed. In contrast, the resolving gel is more basic (pH 8.8) and has a higher acrylamide concentration, normally in a range

of 5-15% depending on protein sizes.[182] In this gel, proteins are separated based on their sizes, smaller proteins travel through the gel more easily and rapidly than larger proteins. In addition, APS and TEMED are added into both stacking and resolving gel because their presences affect the rate of polymerization and the properties of the resulting gels. Their sufficient volumes in gels are able to improve the separation of adjacent protein bands.[183]

In our experiments, 50 µl purified protein was mixed with 50 µl protein gel loading dye 2X. This mixture was heated at 99 °C for 5 min. The samples were then loaded into 5% stacking gel and 12% resolving gel (Table 50) because the sizes of the target proteins were in a range of 61-64 kDa. 8 µl of protein marker of NEB and 3 µl of that of Roti were loaded into the two outside lanes of the gel.

Table 50. Composition of the stacking gel and the resolving gel.

Composition	**5% stacking gel (3 ml)**	**12% resolving gel (10ml)**
dH_2O	2.1 ml	3.3 ml
30% acrylamide mix	0.5 ml	4 ml
1 M Tris-HCl pH 6.8	0.38 ml	-
1.5 M Tris-HCl pH 8.8	-	2.5 ml
10% SDS	0.03 ml	0.1 ml
10% APS	0.03 ml	0.1 ml
TEMED	0.003 ml	0.004 ml

After that, electrophoresis was carried out through two stages. In the first stage at 80 mV for about 30 min, the proteins moved in the stacking gel till they concentrated at the same level. In the following step at 120 mV for about 90 min, the proteins were separated in the resolving gel.

b) Staining proteins

Since almost proteins are colourless, their bands on acrylamide gel are detected by staining with coomassie, the most popular protein gel stain. This dye was dissolved in a mixture of methanol and acetic acid (Table 42) which fixed proteins on the gel. The gel was stained in about 200 ml dye solution for 20 min and shaked at 150 rpm.[184]

c) Destaining proteins

The dye was then removed by destaining with a solution comprising methanol and acetic acid (Table 42) till the blue bands of proteins appeared on a clear background.

7.3 Methods for chapter 5

7.3.1 Heterologous expression

Together with the metagenomic approach, heterologous expression is an effective strategy in studying natural products. After biosynthetic genes of a promising compound are isolated from a metagenomic DNA, they can be heterologously expressed by recombination with a suitable expression vector and then transferred into culturable bacteria, ready for obtaining target compounds. Requirements of a successful heterologous expression are achievements of the replication, transcription, and translation of the genes of interest.

7.3.2 The expression vector pJB861[158]

7.3.2.1 Deletion of the promoter of pJB861

The promoter of the expression vector pJB861 was removed by the two restriction enzymes *Eco*RI belonging to the MCS position and *Mfe*I belonging to the *xylS* gene (Figure 117). Then, pJB861 without promoter was recovered from agarose gel using the QIAquick Gel Extraction Kit.

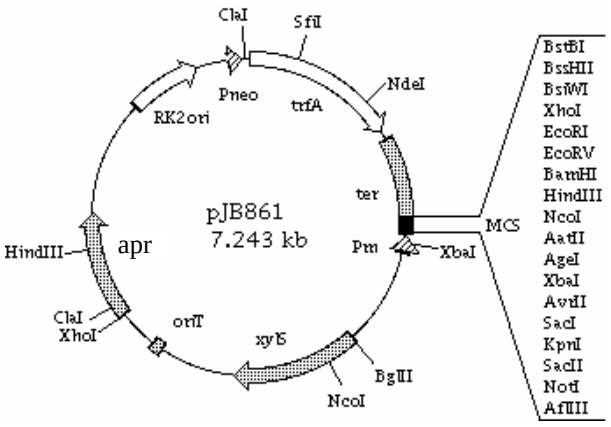


Figure 117. The expression vector pJB861.[158]

Digestive formula of pJB861

DNA pJB861	20 µl
Buffer 4	5 µl
BSA	5 µl
*Eco*RI	1 µl
*Mfe*I	1 µl
dH_2O	18 µl
Total	50 µl

37 ^{o}C, 3 hours

7.3.2.2 Flanking the digested vector pJB861

The expression vector pJB861 was flanked by two short PCR fragments PTA24, PTA25 (explained in section 5.2.2). As described in section 7.3.2.1, pJB861 was digested with *Eco*RI and *Mfe*I; hence the sequence GAATTC of *Eco*RI or CAATGG of *Mfe*I were required to be present inside two PCR products PTA24 and PTA25 for compatible termini. In addition, the flanked vector pJB861 would be linearized with *Spe*I; therefore, it was also necessary to add the *Spe*I sequence (ACTAGT) into PTA24 and PTA25. Based on the pWEB sequence, two pairs of primers were designed to amplify PTA24 and PTA25 (Figure 118).

Primers of PTA24

Forward primer: **GAA TTC** CTC TCA AGG ATC TTA CCG CT
Reverse primer: **ACT AGT** CGC GAA TTC TTG AAG ACG A

Primers of PTA25

Forward primer: **ACT AGT** CTC TCC CTA TAG TGA GTC GT
Reverse primer: **CAA TGG** TCG CCA TGA TCG CGT AGT

Figure 118. Primers for amplification of PTA24 and PTA25. In diagram of pPD7E4, red: the *ppp* gene cluster that was inserted into the position CCC'GGG of *Sma*I inside pWEB, blue: two pWEB fragments needed to be amplified.

The amplified PTA24 and PTA25 were then purified from agarose gel using the QIAquick Gel Extraction Kit. In order to have cohesive termini, PTA24 and PTA25 were subjected to TA-cloned (section 7.2.12). After that, the plasmid DNA containing the pTA24 was digested with *Mfe*I and *Spe*I, whereas the plasmid containing pTA25 was digested with *Eco*RI and *Spe*I. These digested fragments were recovered from agarose gel and cloned with the digested pJB861 in section 7.3.2.1.

Ligation formula of the digested pJB861 and pTA24, pTA25

Digested pJB861	14 µl
DNA of pTA24	1 µl
DNA of pTA25	1 µl
T4 buffer	2 µl
T4 ligase	2 µl
Total	20 µl

16 °C, overnight. Heat inactivate at 65 ^{o}C for 20 min

7.3.2.3 Linearization of the flanked vector pJB861

The flanked vector pJB861 was introduced into electrocompetent *E. coli* XL1 blue cells, after which isolated DNA was linearized with the restriction enzyme *Spe*I. The linear construct was purified from agarose gel using the QIAquick Gel Extraction Kit.

Flanked pJB861	16 µl
Buffer 2	2 µl
*Spe*I	1 µl
BSA	0.2 µl
dH_2O	0.8 µl
Total	20 µl

37 °C, 3 hours

7.3.3 Homologous recombination of the cosmid pPD7E4

The procedure of gene targeting by λ RED-mediated homologous recombination was carried out following the "PCR-targeting system in *Streptomyces coelicolor*" protocol of Gust B, Kieser T and Chater K F (2002).

7.3.3.1 Preparation of electrocompetent *E. coli* BW25113/pIJ790 cells

Electrocompetent *E. coli* BW25113/pIJ790 cells were always prepared freshly before use. 20 µl of *E. coli* BW25113/pIJ790 were used to inoculate with 3 ml LB containing chloramphenicol overnight at 30 ^{o}C. In the next day, 700 µl of *E. coli* BW25113/pIJ790 from the overnight culture were inoculated in 30 ml SOB medium containing 20 mM $MgSO_4$ and chloramphenicol. Since a large foreign DNA (~ 40 kb) was introduced into competent cells, a rich nutrient medium such as SOB was required to ensure the stable growth of the transformed cells.[173] After growing for 3-4 hours at 30 °C with shaking at 200 rpm to an OD_{600} of ~ 0.4, the cells were recovered by centrifugation at 4,000 rpm for 5 min at 4 °C. The medium was discarded and the pellet was resuspended by gentle mixing in 30 ml of cold 10% glycerol. Then it was centrifuged and the pellet was resuspended in 15 ml of cold 10% glycerol, centrifuged and decanted. Then finally, the cell pellet was resuspended in 100 µl of cold 10% glycerol.

7.3.3.2 Transformation of pPD7E4

Electrocompetent *E. coli* BW 25113/pIJ790 cells were mixed with 2.5 µl DNA of pPD7E4. Immediately, electroporation was carried out in an ice-cold electroporation cuvette using a Bio-Rad electroporator set at 2.5 kV. 1 ml cold LB was added to the shocked cells and incubated shaking for 1hour at 30 °C. The cells were spread onto LB agar plate containing chloramphenicol, ampicillin, and kanamycin. The plate was incubated overnight at 30 °C to maintain plasmid pIJ790 inside the strain.

7.3.3.3 Preparation of competent *E. coli* BW25113/pIJ790 cells containing pPD7E4

Electrocompetent *E. coli* BW25113/pIJ790 cells containing pPD7E4 were prepared freshly before use. One isolated colony of *E. coli* BW25113/pIJ790 including pPD7E4 was picked, cultured overnight at 30 ^{o}C in 3 ml LB with ampicillin, kanamycin and chloramphenicol. 30 ml of SOB medium containing 20 mM $MgSO_4$, the three above antibiotics, and 1% of the overnight culture of *E. coli* BW25113/pIJ790 including pPD7E4 was used for inoculation. 300 µl of 1 M L-arabinose stock solution was added for induction of the λ *red* genes on pIJ790. It was grown for 3-4 hours at 30 °C with shaking at 200 rpm to an OD_{600} of ~ 0.4. The cells were recovered by centrifugation at 4,000 rpm for 5 min at 4 °C. The medium was discarded and the pellet was resuspended by gentle mixing in 30 ml of cold 10% glycerol. Then it was centrifuged and the pellet was resuspended in 15 ml of cold 10% glycerol, centrifuged and decanted. The cell pellet was resuspended in remaining ~ 100 µl of cold 10% glycerol.

7.3.3.4 Transformation of the linear flanked pJB861

Electrocompetent *E. coli* BW25113/pIJ790 cells containing pPD7E4 were mixed with 2.5 µl DNA of the linear flanked vector pJB861. Electroporation was carried out in ice-cold electroporation cuvette using a Bio-Rad electroporator set at 2.5 kV. Immediately, 1 ml cold LB was added to the shocked cells and incubated shaking for 1hour at 37 °C. The transformed cells were spread onto LB agar plate containing apramycin. Because no further gene replacement continued to be carried out, the cells were incubated overnight at 37 °C to promote the loss of pIJ790.

7.3.4 Verification of the recombinant pPD7E4

7.3.4.1 By antibiotic resistance

Individual transformed *E. coli* cells were streaked out on both LB plate containing apramycin and LB plate containing apramycin, ampicillin and kanamycin. These two plates were then incubated at 37 ^{o}C for about 16 hours and the growths of cells were observed.

7.3.4.2 By restriction analysis

According to the antibiotic resistance screening, DNA from the transformed cells that could only grow on the LB plate with apramycin was isolated and digested with the three restriction enzymes *Not*I, *Sca*I and *Nde*I as mentioned in section 5.2.5.

Digestive formula of DNA

DNA	3µl
Buffer	1 µl
Restriction enzyme	1 µl
BSA (if required)	0.1 µl
dH_2O	3.9 µl

37 ^{o}C, 3 hours.

Afterwards, the digestion reactions were analyzed using 0.7% agarose gel.

7.3.5 Transformation of the recombinant pPD7E4 into the host cell *P. putida*

4 µl DNA of the recombinant pPD7E4 was mixed with 70 µl competent *P. putida* KT2440 cells. Electroporation was carried out in an ice-cold electroporation cuvette using a Bio-Rad electroporator set at 2.5 kV. Immediately 1 ml cold LB was added to the shocked cells and incubated shaking at 30 °C for 3 hours. The cells were plated onto LB agar plate containing apramycin and incubated overnight at 30 °C.

7.3.6 Verification of heterologous expression of pPD7E4 in *P. putida*

7.3.6.1 At the DNA level

The multiplied DNA from the transformed *P. putida* colony was isolated and checked by restriction analyzed as mentioned in section 5.2.6 and section 7.3.4.2.

7.3.6.2 At the RNA level

a) Isolation of total RNA

Total RNA from the recombinant *P. putida* TA27 was isolated following the RNeasy Mini kit protocol of Qiagen with some modifications. This isolated RNA was used for reverse transcription PCR (RT-PCR). Genomic DNA contamination in the RNA sample had to be removed by DNase treatment. Thus, even though RNeasy silica-membrane technology is able to eliminate most DNA, a modified process with three times of DNase treatment was applied to ensure complete removal of contaminated DNA.

Bacteria were harvested by centrifuging 500 µl overnight culture at 6,500 rpm for 5 min at 4 °C. The supernatant was decanted and all remaining media was removed. After that, bacteria were resuspended thoroughly in 100 µl lysozyme-containing TE buffer by vortexing, incubated at room temperature for 5 min. 350 µl buffer RLT in β-mercaptoethanol were added to the sample, mixed by vortexing vigorously. And then, 250 µl ethanol 100% were added, mixed thoroughly by pipetting. The sample was applied to an RNeasy mini column placed in 2 ml collection tube. The tube was closed gently, centrifuged for 15 sec at 10,000 rpm. The flow-through was discarded. 350 µl buffer RW1 were pipetted into RNeasy mini column, centrifuged for 15 sec at 10,000 rpm to wash. The flow-through was discarded.

The first DNase treatment: 10 µl DNase I stock solution were added to 70 µl buffer RDD, mixed by gently inverting the tube. 80 µl DNase I incubation mix were pipetted directly onto the RNeasy silica-gel membrane, placed on the benchtop for 30 min instead of 10 min as protocol. 350 µl buffer RW1 were pipetted into the RNeasy mini column, centrifuged for 15 sec at 10,000 rpm. The flow-through was discarded. The RNeasy column was transferred into a new 2 ml collection tube. 500 µl buffer RPE were pipetted onto the RNeasy column. The tube was closed gently, centrifuged for 15 sec at 10,000 rpm. The flow-through was discarded. Another 500 µl buffer RPE were added to the RNeasy column. The tube was closed gently, centrifuged for 2 min at 10,000 rpm. The RNeasy column was placed into a new 2 ml collection tube and centrifuged at 13,000 rpm for 1 min. To elute, the RNeasy column was transferred into a new 1.5 ml collection tube. 20 µl of RNase-free water were pipetted directly onto the RNeasy silica-gel membrane. The tube was closed gently, centrifuged for 1 min at 10,000 rpm.

The second DNase treatment: a new RNeasy mini column was used. 20 µl of RNA sample, 70 µl of buffer RDD and 10 µl DNase I stock solution were mixed and incubated on the benchtop for 45 min. Then, 350 µl of buffer RLT were added and mixed well. After that, 250 µl of 100% ethanol were added and mixed well by pipetting. 700 µl of sample were transferred to a new RNeasy mini column. The tube was closed gently and then centrifuged at 10,000 rpm for 15 sec. The flow-

through was discarded. 350 µl of buffer RW1 were added to the RNeasy column. The tube was closed gently, centrifuged at 10,000 rpm for 15 sec. The flow-through was discarded.
The third DNase treatment was done according to the first treatment step.
RNA was then stored at -80 ^{o}C.

b) Checking the isolated RNA

The isolated total RNA was checked on 1.8% agarose gel. In preparation for this step, electrophoresis equipment had to be cleaned with 70% ethanol in order to protect RNA from degradation by ribonucleases. Running voltage was applied up to 10 V/cm because higher voltage could damage RNA.[185]

c) Reverse-transcriptase PCR (RT-PCR)

RT-PCR is the most sensitive technique for detection of mRNA, which is a two-step process. Firstly, complementary DNA (cDNA) is synthesized from mRNA by reverse transcriptase. Then, the cDNA is amplified by a normal PCR. Currently, RT-PCR is often carried out in one-step RT-PCR hence all reaction components are mixed in preparation of the master mix.

RT-PCR was prepared following the SuperScript III One-Step RT-PCR System protocol with *Platinum Taq* DNA polymerase (Invitrogen). The RT-PCR products were then analyzed on a 1% agarose gel.

RT-PCR master mix

2X reaction mix	5 µl
RNA	0.25 µl
Forward primer (10 µM)	0.5 µl
Reverse primer (10 µM)	0.5 µl
Enzyme	0.25 µl
dH_2O	3.5 µl
Total	10 µl

RT-PCR program

Lid 106 ^{o}C, preheating on			
55 ^{o}C	30 min		
94 ^{o}C	2 min		
94 ^{o}C	15 sec		
T ^{o}C	30 sec		
68 ^{o}C	1 min	3	39
68 ^{o}C	5 min		
4 ^{o}C	pause		

40 cycles
T is annealing temperature of primers

7.3.7 Extraction of the expressed compound

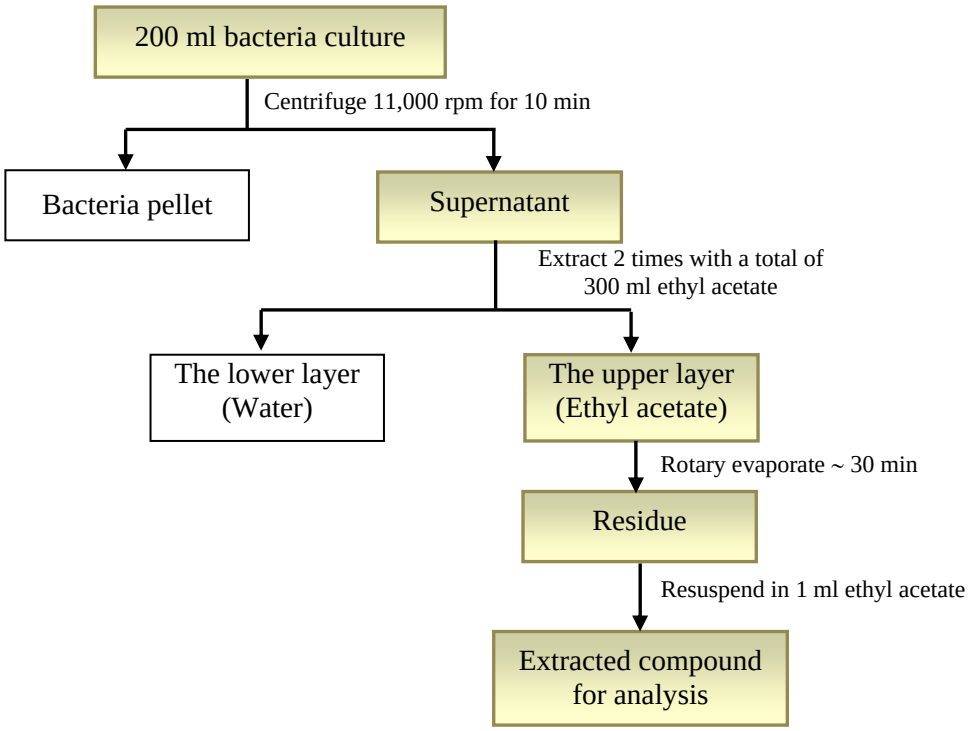


Figure 119. Flowchart of extraction of compounds from *P. putida* TA27.

A negative control was prepared by transforming the flanked pJB861 (section 7.3.2.2) into *P. putida* KT2440. This enabled a comparison between the extracted compound of *P. putida* TA27 containing the recombinant pPD7E4 and that of *P. putida* containing merely the pJB861 without pPD7E4. The recombinant *P. putida* TA27 and the negative control were cultivated in LB medium with apramycin under the same conditions at 16 ^{o}C and 30 ^{o}C in four days. Ethyl acetate, a non-polar solvent, was used in an attempt to isolate the target compound from the recombinant *P. putida* TA27 (Figure 119).

7.3.8 Thin layer chromatography of the expressed compound

Thin layer chromatography (TLC) is often the first technique that is applied in order to analyze chemical characteristics of certain compounds, particularly in the case of mixture of unknown compounds.[186, 187]

The mobile phase is the most important factor that affects the movement speed of compounds on thin layer, resulting in separation of compounds in a mixture. For this reason, eleven mobile phases that were different in polarity were developed to detect the target compound from the extracted sample (Table 51).

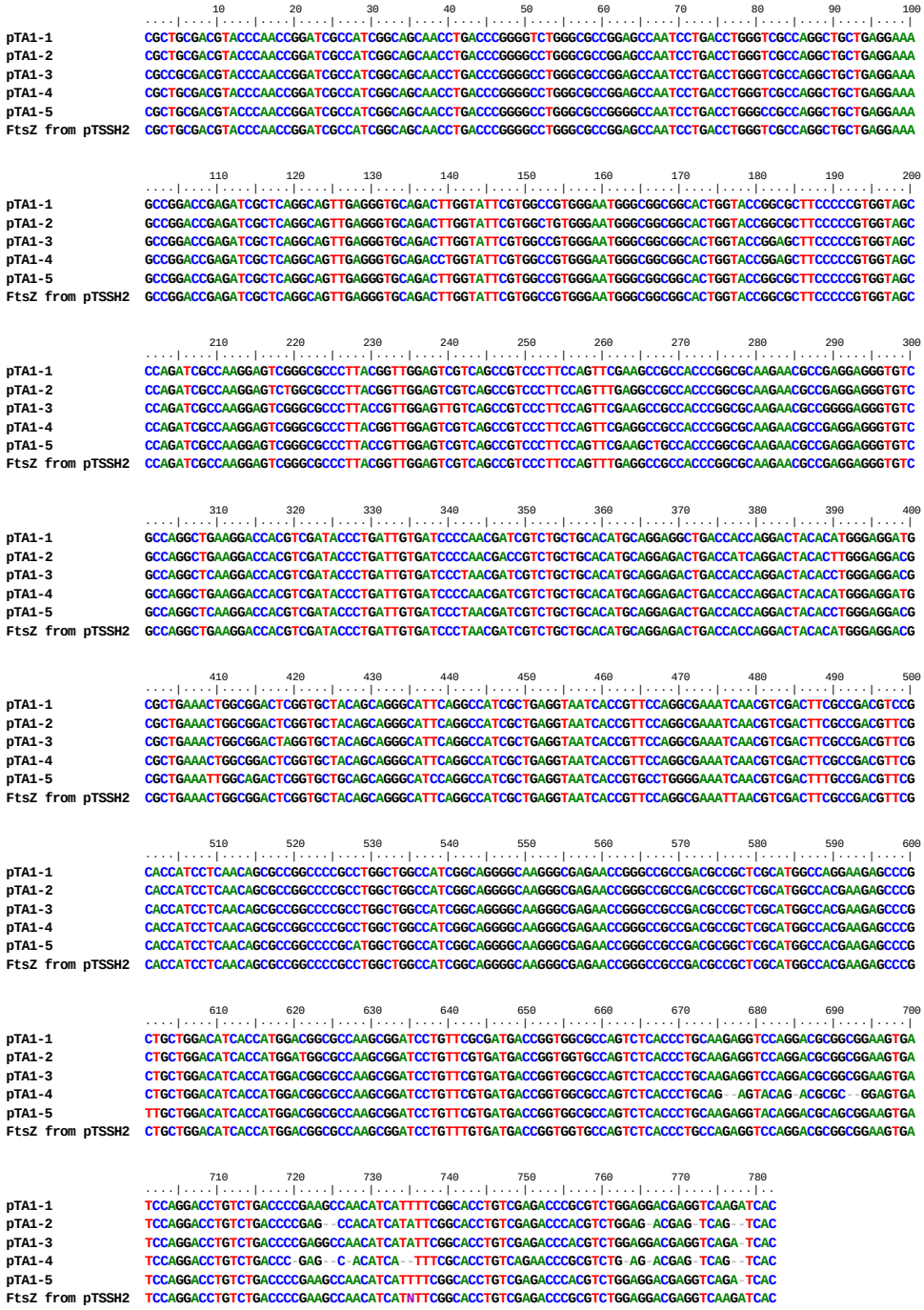
Two universal colour reagents for staining TLC plate, anisaldehyde and ninhydrin, were used. Anisaldehyde is sensitive to many functional groups, thus several spots can be detected based on various colours. Ninhydrin is specific for detection of amino acids, amines or sugar amines based on pink spots.

Table 51. TLC mobile phases

	Mobile phase		**Mobile phase**
1	Methanol	6	Chloroform: Methanol (6:1), pH 9
2	Acetonitrile	7	Ethyl acetate: Methanol: Water (80:15:5)
3	Methanol: Ammonium (100:1.5)	8	Acetonitrile: Ethylacetate (9:1)
4	Methanol: Water (50:50)	9	Methanol: Water (75:25)
5	Acetonitrile: Water (50:50)	10	Methanol: Water (90:10)
		11	Chloroform: Methanol (9:1), pH 9 -10

Appendices

Alignment of FtsZ sequences



The protein sequence of the onnamide B gene cluster

Sequences with colored background are PKS conserved motifs. Background in (yellow) KS domain, (pink) ACP domain, (bright green) DH domain, (turquoise) KR domain, (gray) MT domain, (red) NRPS domains, (dark yellow) ER domain

OnnA HMGS 420 aa

MITAGIEAMNVFGGTAFLDVRKLAEHRNLDMTRFDNLLMEEKAVALPYEDPITFAVNVAKPLIDELSPRQKARIEMVVTC
SESGIDFGKSMSTYIHDLLELNRNCRLFELKQACYSGTAGFQTALNFILAQTSPGAKALVIATDISRFFVAEGGEALTHD
WSYLEPSGGAGAVAMLVGESPHVFQVDVGASGYYGYEVMDTCRPALDSEAGDADLSLLSYLDCCQQSYLEYERRVSDVNY
RDSFQYLAFHTPFGGMVKGAHRKMMREMVKAKPPEIETDFRDRVMPGLTYCQRVGNIMGGTVFMSLASTIDKGHFPTPQR
IGCFSYGSGCCSEFYSGIVTGEGQAMQRRFAIDQSLDRRYELSMEDYEKLLRESIKVRFGTRNVSLDFDIIPGAIEGGGT
GQPRLYLERIKEFYREYVWR

OnnB ACP-KS1-KR-MT-ACP-KS2-ACP-ACP-KS3-ACP 4,376 aa

MENRLKHQTRNQTQKAAVEQYILTHWAEALWSQSNADLLPAADPILVDAELFLVSAAKLGFFAKTQFLSHPGHEPIDQKA
LHQLEKRDYQVRQARAEDLGVLQSLEALCWEEGLQTSAEGLAKRLQDYPEGQFVLVFEEQVVGVIYSQRIADTRVLYGKN
AANVAVLRRNDGPIVQLLAINISPQVQQRQWGDQLLEFMLQRCAVMHGVEAVAAVTLARDYHKQKQMPMGAYIHHRTEQG
WLADPILHFHELHGADIIDLIPQYRPKDVKNEGFGVLVSYDIHHRKRRNIESRTSRNEAGGEAPSLPQREHTLEAISDYL
EEAVKTILGPEQECLFSKETPLMEMGLDSADFLGLREQIAVRFQTALEPTFFFQSKTPQDIISYFAKEMYPHPEKEAKAT
PLTATSVQAAPSHSRQGHIAIVGLACRLPGGIETPQALWQLLKNGESAVGSLPSGRWNWPADIDPDNRHRGIDQGGFLDD
IAGFDAAFFRLSTTEVESMDPQQRMLLELSWQVLEDAGYAPKDLKKSQTGVFIGASGSDYSYLLNQSPVSVEAHFGTGSA
MAVLANRISYFYDFYGPSLVVDTACSSSLVAVHKAVQSLRVGECDQVLVGGVHVMCHPANSLAYYQAGMLAKDGKCKTFD
QQANGYVRAEGAVMLLLKPLEAAVADQDQVFAVIRGTSCNHGGLASGLTVPNPEQQAALLQQAWRDARISPLELSYLEAH
GTGTALGDPIEIQGMKDAFAGYVKARSLPVPEIRSCGLGSIKTNLGHLEAAAGIAGLLKVVLAFRHRELPPLLHFKQLND
HIDLANTPFYPVDQLRSWDVPEGAIRKAGVSSFGSGGTNSHVVLEEYPQRPARPITPKGPLIFVLSARNQENLRAYAQKF
LDWLRGEKAATPPLREMIYTLQVGRQAMEQRLALAVSDIDDLCAQLTRFCGGHDMNDELPEMAVWWKSGEQVDWASLYKE
ETRPRRISLPTYPFAHEAYWPAIVQPTREVRELLLAPVWSPVPFPLTSGITPPPSKKTVVVGANPQQQQWIQQVYPQALC
LEDSLRWSVQAWEDKWRELEGVDHIIWAGSDQFPESIADEQLVQEQEAGILQVFRMVKSLLALDQDRVAISWTLITTQTQ
PVHFQERINPTHAAIHGFAGCLAKEYPRWHVRLFDVAAGTAWPVQEMFACEPQQGKALAYRGSKEKGGQWFQQELLPLQT
SVQGGQERLPYRQNGVYVVIGGAGGLGEIWSRWMVETYQARIIWIGRREKDDEIQGKMDALSPAPHYIAADASDREALQR
AYAEIRQCHPRIHGVVHSAVGSFDLSLAEISEERFRQILAVKVDLSLRIAQVFAQERLDFALFFSSVLSFLKSSGVSGYT
AGCCFKDAFARQLAHAWSCPVKVMNWGYWRVGTGARISKAVKIRATQSGFETIDASEAMQALNRLLHGPIDQLLFLKTNN
RRPLDMIQPEQTISCFESSHLSSIEAMMQRLDESPSDPQSRIKSMSAASIFKNEDMEPLLLKFLAGIVASKPVIRVPFYQ
NWLLQSERILREKGIANAADFSLETCWQEWESARKDWLQDANKTAALTLVETCLRAFSDILAGKRKATDVMFPHSSMTLV
EGIYKNNLVSDYFNQVLGDVLVAFMTTRGHEEPVRILEIGAGTGGTTATLLEKLRPFQEQIAAYCYTDVSKAFLFHAEEH
FAPEHPFIGTAIFDVEQPLANGVIKPADYDIVIATNVLHATKNICETLRNAKAALKQHGLMLLNELSDQSLFAHLTFGLL
EGWWRHEDASIRIPGSPGLFPEAWQSVLEREGFTSVCFPARAAHQLGQQIIVAESNGIVRQQPLAPGHNAPSIRQPTAGS
EPRRLPGVKTQKDRSQGMLREACTHYFQTLIGKVLRIESQRIDPAESLEAYGIDSILIVQIIEALQEIFEDVSSTLFFEV

QTVDALVTHFLQHQREALIALMGIDEFTDSEPSEEVVAPAASVRGAPSNRLEREPIAIIGMSGHYPMADTPDAFWEMLKN
GQDCIREIPPDRWPLEGFYLADQDQAVASGKSYSKWGGFLENPFDFDARFFNISPKEAKDMDPQERIFLQAAWEALEDAG
YDKKSLATRYRQRVGVFVGITRTGFDLYGPQLWEQGNTAYPHTSFSSVANRISYLLDLRGPSMPIDTMCSSSLTAIHEAC
QRIQFGECEMAFAGGVNLYVHASSYVGLCASRMLSKDGRCKSFGTGSNGFVPGEGVGVVLLKPLSQAIKDHDPIHAVVRG
TYVNHGGKTHGYTVPNPNAQGELIREALNRAGVHARTVSYVEAHGTGTELGDPIEVTGLTQAFRQETQDSGFCALGSVKS
NIGHLESAAAMAGLTKIILQMKHGMLAPSLHARELNPKIPFEKTPFVIQQELAPWQRPTVSLDGDIKEYPRIAGISSFGA
GGSNAHVILEEYIPPEHQQTAPSGRSNPSYFVPLSASNETQLRLYAQKLCDWLEQNRSRSLSCADLAYTLQVGREAMEAR
LGLIVRSVEELEEGLKAFLNPGTTPGHSSVAWVRGEVRKHKETIRLFTLDEDMAKAVDAWIHKRRFAKLAEFWVKGYPID
WHRLHLQQTPFRISLPTYPFIKTTYWWERTSATGQATDIKAAETTIAEAGSGVKTPSKRELQPLSSARRQLPEIEIAKPR
ITLPDLIGRSVKTDREMQAVGQQSQKAKTAKPHVSLSDLDRKLNGEAQPLRQTSHVVLHDDGQGIFSIQMASPNHEGVPF
EAVMADLKVCFQSIQSLTEAANHNAKVVLFQGEEQFFFSGRSESRTAFMEQEIHRLPLDCRLPVIAVVRGQGKGLGCLLG
SFCDVMICSRGAVLGYGRPQWPLSEEERAFFRDRFGRHLGEAILRAQVSGTTFQEKGQGMTVLPETELDDYVAALARRLS
GFSQQSLMVLKKQLARDSVEHAMKLVHASTPIDGDGQGEACTTSAQPSVNLPKTQNNPQAQTPQRIRLQSQVITVDAWPN
GVVLVSLCDRENKNMFSKAFLQGFDEAFEHIRGNAAYKVVIITGYDSYFACGGTKESLQAIQQGAEKFTDTRIYGRPLEC
EIPVIAAIQGHALGGGWSMGMYCDQVIFSLESLYQSPYMQFGFTPGAGSTLIFPHRFGKDFAGEILLSASAYRGSDFERQ
GIQMPVLPRRRVLAHAMEMANEMALSSRDELVETKVVRCQHLRHRLEETYLQELAMHEKTYVGNPDVFAKIQMHFNDGIQ
VEDRLDGIQEQPCRDRESEKAPVVSMQQVRQELRRFLAEELHMTPDMVEEDIEFVKMGLDSIIAVSWVQKINQAFGLALG
ATIVYTYTNLLDLLQHIFPEIAKAPSSTIPEPELAVSSSDSDIYPGFKPIALQPTVAAKKTGNANGSSLHAEKQKPTPAG
RPDGAIGQGDLKDVHVKLRQLLAEELHMTPEAVEDDVSFVEMGLDSIIAVSWIQKINQAYGLSLEATVVYTYTTLLDLAQ
HIFPETATTVFTSTTRQPSANAVGEPEEKAVAMMAELRQLLAEELHMAAEAIEDDVNFVEMGLDYVMAGSWVQKLNQAYG
LSLEATVIYTYTNLLDLAGHLASEMSHRLATARPLAKLERERPRRTPRADAKPVVSPPASQPLTPVSKRWLPEQKTGRIA
IIGISGKFPKASTLDQFWENIAEGRNCVSEVPESRWSVDEFYDADGKVPGKTMSKWMGILEEVDQFDPLFFAISPRDAEL
MDPQQRLFLQACWSCIEDAGYNPKTLSGSSCGVFVGCDMGDYGRSVQYQELDAQSLLGGVVSILPARISYFLNLQGPCLA
VDTACSSSLTAIANACDSLLLGHSDCAVAGGVCVMTGPEIHIMMSKAGMLSPNGTCFTFDQRANGFVPGEGVGAMFLKRY
EDAVADGDPIYAVLRGWGINQDGKTNGITAPNARSQTRLEKRVYEQCGIHPEDIQLIEAHGTGTKLGDPIEVEGLRDAFA
HFTEKQHYCALGSVKSNIGHLATAAGVSGMIKLVLALQHQKLPPTVNHEKLNEHIRLEGSPFYINTACRDWVVPEGKTRC
AAISSFGFSGTNVHMVVEEHQPTENRRLTDAAPYLIVLSAKSDPQLEASLQQLLGFTRKHTEHAETTLHIADLAYTLQTG
REAMTERLGLLVSSLDELADKLDMLLQQKSCDQIADCYRGTVIKQNIETFSSLADEDDIASWITKRQYAKVLNLWVHGAE
IPWNSLYETVACMPSRIHMPTYPFAKERYWVARRKETGMLTEQNHEQREPTSEKADHEFFWKIEELIEEWERQQQYGGHL
TPETLTKVFAYEVVKNLGPEQIHRVDQLLAATPKPATAPPPPVQQAAAVVPEPKLVRSEDIPQAVCEAISEVLKLQTIGE
HDRFQDYGLDSISAMKLSVRLEEKLGRKVRPQWIHDFPSVGTLSRRLMEQDELVDA

OnnC OXY 386 aa

MSATIQNLFDMIVPHIEEKKPDVSAATMAFSFLFFSDVHKNISDAEKYDFLRHITRFADQEGFTAIYLPERHFYEFGSIY
ANPAIVASYLIPQTQRIRFRTAGVSLPLHHPAEVVEWWSINDILSNGRVDLGFGSGWNKSDFIYAPEAYENRRAICSERI
GMVQKLWRGETLFFPGPGGETVSINTYPRPIQKELNVWLLITANEGAFFHAGREGYNVFTMLYGYDLEALSKKIAIYRKG
RAEGGHDPETGVVTLMLHTMVYKDRETVRQVVEVPFKQYIKSALDAHVQAGVGKRQGFAEMSADEKEKVLEYAYNRYFKA
CAIFGTVEDGRQMVDSAFDAGVNEIACVVDFGVDYALVKDALSYLKCLVSAYIADGSDGPEPSHAV

OnnD MT 317aa

MTVTQYEFKNLVKERLFNCIARQIDMIGHVPGDLSRMVGSSLDAADDHSEGHWERPTELTELINAHYDRFFYEQHGVERL
IREETDFKNLGYWDDTTLDLNAAAERLFKTLMAMIPKKSGRILDAGCGTGGATRRLLESYPPENVWAINISAKQIETTKQ
NVKGCHAIVMNAVDMTFEDNFFDTVLSIEAAMHFETRRKFLEESFRVLKQDGCLVLSDILFTSQERLEQNDYFGGVSNHI
ETIEDYQQLMEEIGFRNVVVKDVSKAVWGSNFLYNINKLHKEFYHGRLDLIQLTETLWSMYYVNSIFSVCLFTFGQK

OnnE 356 aa

MHTILSEEYLNEYYTDGYINTDIRLTDHLVQEMRQHYNSLEEMRNDYPQYFTKNEHQAYLGGKLTGILYNFIPGYAAKKL
KKMYSDAYCKAVHAEQIFIEKIFEQLLEKNFHKFFKTKYIVASYDIYLNNDYKHLSFTDIHSDIPNFHHFYETENDLTIY
LPLVDLTEENGGRLSILPEAKSNLKIPGNVLLKLYEEAFLGNKNNLDENGYIVPEKINENDYKAFTKNKTYWELLENYNI
NTELVAKYYMNDFVKHDWKAGNIVLFNNKTFHAAETWRNQQYNREIYMIRLLPIYDAKVRLKGIIHGKPFNRFLIDTEKG
EFLRFDEPVDFSRIPEEDKLKLSGCYQPSVAARANV

OnnF 340 aa

MYNIFKDKYFKEYYSNGFVETDTVLSHDLVDEIRAHYQAKAEGHNDFPKFFVNNEHQIYLEGEESGRAFSKLPKSVAHQK
IRELYDTSYSKAVYCEQVFMERVMTELLEKDFLRFFKTPYLIVSYDIYLTNDCNRPGAGIHTDPPNFHHFYETENDVTIY
IPLVDLNDKNGGRISVLPESKLKLSGNVLLKMMEEAFGGEQQWLDENGYIDPDKIDDKALSDFIKSRPYQRLIEHQRNVI
SMARQYYTDEFTYINESKGRVLLWNNKNFHAAEAWKEKSFNREVYIIRCMPIYNTKIKLKNKLHGKLFNNLLVHVKTGTL
EKFDREVDVSQIPEMDKVAI

OnnG MT 321 aa

MNNQLADALKERLFNSISANLDMIGYVEDDLRRLAGPEFDPEKRVSLLQDDGVDDIARARIPLTDEINSFYDHQFYSQDS
IFGLLLGDTKFRNIGYWDETTPDQNAAAEKLQDMLLEMIPEKTGRILDVACGMGASTRRLAELYSPENVWAINISEKQIE
STRENAKGCHVQVMSAVEMTFDNDFFDTIMCIEAAFHFETRRKFFDDSLRVLKQGGRLVLSDTLFTSKERLEQSSIFPSP
ENHIDTLEEYRQVMEEAGFRNIVVKDVSKNVWEAHFLYVINKIHEGFYHGRLSLIQLTEILWSYYYLKAITGICVFASGQ
K

OnnH MT 268 aa

MSRSHLEEIAELYDSAEGHVGNLIFDGQVHWGYWDERNADASLAEGADRLTQIMIDKTTIEKGQKFCDLGCGWGGPAVAL
AKAKGCYIDGITCSGQQQQNAVKKAQELGMDDLLNFIHGDALNMPCKDQTYDGGWFFESIFHMGHREALLEANRILKLGA
TLLITDAYLLSTASEDFKEHTSRRVHSRFMPKDIYPGVLEETGFEAVEVLDVTQYVMRPLAQKLKDACVAYREEILKLVP
EEAIDDWLWGFEDFCANLGYLLVTARKK

OnnI (KS4)-ACP-CAT-KS5-KR-ACP-KS6-KR-MT-ACP-ACP-KS7-DH-DH-KR-ACP-KS8-KR-ACP-KS9-DH-KR-ACP-KS10 9,966 aa

MSRHQELERIKDQIFSAVLEDTPNMSTPLAGGRENLYREREESVASSLANIQRKLFDEAGDDACEFKENKENRRLEPIAI
IGLSGSLPKSQTIAEFWRSLDQDLSLIEEIPRSRFNWEEVYDPDGKDVDKMRTKWGGFLRDIYGFDPHFFKILPRDAAVM
DPRQRLLLMSVYQTLADAGYAPETFKKSKTGVFFSIQDNEYLQLLREGGVDRGEGFGHASMIANRIAYFFDFRGPSEFVD
AQCAGAAVALYRAVSTLRSGDITYAVVGAANLLLRAEPFAVLTRANQLSPTNCVNSFGKDAQGHLRAEGVVSLLLKPLSK
AEADGDPIYALIKNTACNYNGQGGMSIAAPNVDSHAELIETCYEQVQVDPGEIRYIEAQGMGNPLSDLGEWHAYNQALQS

MAKKRGVVLPQGQCAISTLKPMMGHMESVSSLGAIMKVIRSFKTNTIHKILNVQEISPDLDPQGMPCRLLTETEPWPEQA
RPRLAGLHSFGIGGNNVHILLEEYKAEPDCRAERGEIRLDEPRLVVLSAKTERSLLGIAGELRTFLLDATGTSQPSLFLG
DLAYTLHMGRDAMACRVAFVVRTLSELINGLGHYLSRQETQDGDVPIYVGNTEDGAEHIQPLLLRENAARMQEALFAERD
LEKMALYWTLGGKLSWERLYQGERVRRISLPSYHFNKKVYVFSSQTQDVDLKSQSIETPVETSAPQPVLQQDNVLEKKPL
PETRSTETSSCGSPREQQVLAEEMKDYLVAELSKELELPANEIKVDCHLQDYGIDSMVGMRLCRGLTERFGVEVLGREMF
RHPTIDSLSAYLAQKVTGQQTSEATLPKTEPVVAAEIDHTPYDLSEGQKGLWVLQKMVPGMSAYNVPLGFRTQCALDTER
FRQTCQFFLDQYPILNTVIKETENGLCQQISRDKQLSIQHEDISSLDSWQIPAYLQEKNNIPFDLANGPLVRFHLLKRSQ
QEWFVLISVHHIVFDGSSFLPAINTILKGYQTLSQGGTPLPTSFEASYRDFVRWEQQMLAGADGEKHRAYWKQQLSGKMP
ILNLPTDRPHSAVQSFRGAVHSMTLPGEASRHINAFSRELRINPSTFLLGVFKLFLYLYTKQDDVIVGMPNMGRPQKRFG
DLMGYFINMLPIRSRKLAELSFRDFVEELQLTMLDAFDHASYPFSSMIQDSKITLSNEHSPVFQVAFEYQNVLSPTDVQD
FQAQYQDSLPMAFVQGVHQMGEYEIVLEVLEEENAFILNLKYNPDLFDATTMTRMLEQYEVLCDKVIGDPHQKNADYSLL
SSKEERLLLSDWNDTSSAYPSDCVHELFERQAETNPAAIAVMCEGKSLTYGELDQRSSVLAKQLQDLGIGPDQLVAICVT
RSLDMIVGLLGILKAGGAYVPLDPEYPTERLAYMLEDSQAEVVLTQTALRHQVTALSSGAARRPHILAVNEVLEAWNGEA
ISKGTLRREVQNTHLAYVIYTSGSTGKPKGVMIPHNALINFLTSMAKEPGLNSNDRLLAVTTYCFDIAGLEFYLPLIMGA
QCVICSSETIVDADRLKQEISHCRPTVMQATPSTWSMLFHCGWKNEEKVKVLCGGEALPMALQHQFMACNCEVWNLYGPT
ETTIWSAVQRIRENEAIVIGKPIQNTGIYILDHNRALVPIGVPGELCIDGDGLARGYLHKPELTAEKFMDHPFHRGKKIY
HTGDLARWLPNGTIEYLGRLDLQVKIRGFRIELGEIEFQLKQLPGIRDAVVVAKDMAKDMEGLHQLVAYFIPGGGPESVD
TEVLRAGLKQKLADYMVPAFFIPLEEIPLTPNGKVDRKTLMERDILVNRSQEHTQPQSAIEESLLEIWQDILKIDNISPT
DGFFALGGNSVLAVRLAERIAQTLGMPFNPTALFKHASIREIGHYYRSANQQTSSVSHGGAEPQKTPDATQSGDDYPDYY
QDSVAIVGISCHFPGAADHHTFWRNLRDGKESASFFPEEELRAAHVPEARIYNPNYVPLKLTIEGKDLFDAEFFNISPAN
AVYMDPQLRLLLTHSWQAMEDAGYCARDIPDTAVFMSACNSFYKTLLHRANAIGEADEYAAWIASQSGTIPTMISYQLGL
KGPSAFVHTNCSSSLSGLYFAVQSLQTGQAKAALVGAATVFPLPGIGYVHQPGLNVSSDGHIRTFDAAADGLTGGEGVGV
IMVKKAQDAIADGDHIYALLRGISLNNDGSDKTGFYAPSVKGQSEVIGKVLRATNVDPSSISYMEAHGTGTRLGDPIEVM
ALSDAYRQFTSQTQFCGIGSVKPNIGHLDTAAGLAGCIKVALSLSHGEIPPSINYRQPNPEIDFKTSPFYVVDQLQAWKT
DAVPRRAALSAFGIGGTNVHAIMEEFVVPASAENRRLSDGPYLVALAAKNPERLRAYAQKLLAFVKADAGKRDDQPLNLA
DLTFTLQTGREAMEERVVFVVCEQHELVAMLDAFIAGKERISGCYTGSARHAGNVLRILGDEDAHELIQRWTETGKHAKL
AELWVNGIAIDWKALYRGHSPRRISAPTYPFAGVSYWVEEETAALSRTSPSQHLASRGVLMRLPQWQDKAISHTDHGVSA
AHHLILWAGNGMAPVAQRENTSVIHLQSQGPTAAERFLDYSVQAFESIKQIREQHRGPVFIQLLFPSHGEKQLVAGLGGL
LKTAHRENPKTTYQLIEWDVEAPNREALLEILQENSLCAGDAHVRYCQGQRQVLVWQEVTHQSQPAIPWKDRGVYLITGG
AGRLGILFAREIAQQTTGATLVLVGRSELGGQSLGQIQQLESLGASVVYRQVDVSRREAVTALIYDVGENVGEINGILHC
AGVIQDNFMVKKQVEEWRSVLAPKVTGAVNLDLATRHCPLDFFVLFSSAAGAIGSPGQADYATANAFLDAYAGYRNRLVT
AHRDQADGPRGFTLAMGWPLWQEGGMRADAVSEKITRETTGMLPMQTANGIRALYQCLALGQSQVMVMEGHLPTLRAFLG
AGQLANPEGAEKPERVVAEIDDQELATRTLQEMKTLLGSVIGLVPDEIDAQKPLENYGLDSIAIIQLNEKLDGVFADLSK
TLFYEYQTLDELVDSLLADCPHACRTWVGRFEAPPVREPEPETETVNGLCEPASQVPQPHARSAKKSCGNSVQQEPIAII
GISGRYAQADTLDDFWHNLKAGKDCVTEIPQERWSLDDFYLADREAAVAQGKSYSKWGSFLDGFADFDARFFAIAPREVR
SLDPQERLFLQCSWEALEDAGYTRESLQRHHQRNVGVFVGITKTGFDFYEPELLAQGETLHPHTSFGSVANRVSYFLNLQ
GPSMPIDTMCSSALTAIHEACEHLRHGACELAIAGGVNLYVHPLSYVRLCSARMLSGDGRCKSFGEGGNGFVPGEGVGAV
LLKPLSAAIRDRDHIWAVIRGSGVNHDGKTNGYTVPNPNAQGQLIADTLKKAGVPARAISYVEAHGTGTELGDPIEVTGL
TQAFSQDTNDRGFCALGSVKSSLGHLEAAAGMAGLTKIVLQMKHGQIVPSLHARVLNPNINFDKTPFVVQQELGIWKRPQ
IEINGMVQEIPRMAGLSSFGAGGANAHLIIEEYVANRQEMASLTERPGPYAILLSAKNEIRLEEVVENLRQFLIAQSRDG

QGSDLERVAYTLQVGREAMEERLGLVVDSVDELQDKLERFLAGEMHVEGLYRGRVQRGAQAEIGQEKVDAWLEQRLFSKL
LDLWTKGMSLDWSKCYGAVGSHEMPYRMSLPTYPFAKERYWIQPQSRELRDNEAAEGPQVEPLTLAPLWHTVSNLESQPI
FPAEDTRIVIMGGTHEQHQTIRATFPNARSLQLDPEASCEQIAAQLGAMTSLEHIVWITSDRPWQSVASESIPEEQNHGV
LLLFRLVKALLALKFGDRELGWTLITTHAQSVWQQGALNPNQASIHGFAGSMAKEYPHWQIRLVDLEASAAWPVQQMWRV
PAHAQGEALVYRGHEWFRRTLAPVHTQVEKASLYRKNGVYVVIGGAGGLGQIWSKAMIEQYQAHIIWIGRRKKDEAIQAA
LDHLASNIGPAPSYIAADARNYHELQAAYEQIRQEHPHIHGVVHAALGPYDQRLFDMGEDQFRAMFSVKVDVSVRLAQVF
KREPLDFILFFSSMIAFGKSAGMAAYSAGCAFKDAFALQLSREMPGAVAVKVLNWGYWNVGGGTRISASLKRRVAQSGVQ
SITADEGMEAVDRLLSGPVKQLAFTKTAKPGLIETLDATERIIHYRERGRSCAPTLQQYRPHGEAPVISPRADELNAWMV
KLLFVQLHDLGIFQSGHAAHQETVRKHMQILDKYERWWLESLKLLENYGFIRLEDGRIQAREEAAVEDTQTVWAAWEASR
QAFLQTSETRTLAVLVHDCLLRLPDILRGTLLVTDVLFPNGSMEKIEGLYKNNHVCDYFNQVVAEVVQTYIHQRLAVNPK
ATIRILEIGAGTGGTTSMVLPALRPFQDHIDTYSYTDLSKSFFIHAKERYGTAYPFVEYKILNIEKPLAKQDVTLLGSYD
IAIATNVLHATKSMRNTIRNVKAALARNGIAIINEMTTKTVFATVLFGLIDGWSLSEDTVLRIPGSPGLYAETWHQLLEE
EGFRSILFPAHPARELGQQVIVSESNGVVRQKTGNPLSQAMEVVEEEVGAPGDEDPCREQVTQXVLSQLAQTLEVPENCI
DLDVPFSDYGVDSILGVNFITQINDHMGIEMNTTVIFDHTTVHDLTAHITRTYSDRIGLSNAPQLEVPDHGELVHRLVVS
HLAKVLDVAESTIEGDVPFSDYGLDSILGVNFITQINDDLGLEMNTTVIFDHTSVNALADHISKTFQGDLNIREKVKETG
DLKKGITDEPKWEPDPPADMSVEFGEKDFGSSSPRIAIIGMSGQFPGAKDMESFWQNLISGDDPIGELPPHYLPPESFSP
DPLPGKSYCKWGGILEDRDCFDPLFFKLSPREAESMNPHQRLILQESWKALEVAGYAPKSLNGSQTGVFVGSEPTGYVHE
TFTGSSEAIIASRLSYFLNLRGPAFVVNTGCSSSAVAIHLACESLRHNESDMALAGGVFAAMNPTVLISLSQAGMLSASG
RCSPFDATGDGMILSEGVGMVALKRLRDAVEDGDPIYGVIRGSGINQDGTSNGITAPSGAAQEALISEVYRRYQINPEEI
SYVETHGTGTKLGDPIEANALVRAFKQFTDKQHYCRIGSAKAHIGHTAASAGVIGLIKVLLCLKHREIPGLPHFNQLNPR
IEFANAAFYVDAHLSAWESPPGKPLMAALNSFGHGGTNAHIVVEEYVGSRPENPRDHSFNTSHLVVLSAKNDERLRVYAR
DIYNHQKAHLDALENLAYTLQVGREAMDERVAFLVHDRDTLLERLQAFAAGEQQIANCWRGGIEDRQTREAAGEAEIASW
IASGELSRVSDYWAKGGVIAWSQLHQVTQPRRIHLPPYPFAWERYWRAEPEAGTTAVPVAASQLHPLLHQNTSTLFQQRF
GSSFSGEEFFLADHQVNGRRTLPGVAYLEMACAAATYASEAADQTGLQLQNITWLRPLHVDDRQADVSIRLLAEAPGRIR
YDIYTQGGADETALCIHNQGTAILSGEDHQPVLDLAGLRTASQLELSVEQCYAAFREKGIIYGPGHRGIQAVYAGHRQVL
AKLAIPAAVADTAGQFMLHPSLMDSALQASIGLTLGNRQTSAEDRRTLLPFALERLKITGHCATEMWAWIRFSKDSVPGD
SIQKLDIDLCDKQGVVAVSLRGFSSRVVEKIETFRHSLLQTLHADGMGKQGTATFTGSEFFLRDHGHVTPGALYLEMAYA
AGTFGQEKQVTGLRNVIWSKPMIVQGKLDVTLRLEPEGDHYRYVISVPNRDGQIVQPETCCQGHIITAEQEPNVAERLDL
DAIQARCGSAKNTDACNQIMQSTHGPSLLSITQLHYNGHEALAELVLPTVLGGDLQGYLLHPSMLNGAILSSVIFSLIQE
PASPLPMPFSLDQLVIHQTLPQRAYVHVKASPPGHNPQSHAHMKRYDMDICDANGNLAVSFKGFTTIFPTDRPYRRLTAT
PQWQNQPLSQPQTPAPVAAPVFFLAGEQPALRQMLAERWPQSQVETLLPALENQAAGIQTNFLRVFHMMQALIRQAPRQT
QPILLLAPETHDAEQYAALVGLLKTAHLENTRIAVKTISYVPEEPRDQRLIALLEQELCTPFGEVEIRYDCNGAREVKLL
EEVTLPVASQEPIAGLQKGDVVWITGGLGGLGKTFARYLSAVKGVKVILTGRSPLDETRRQILGELQTENAQVTYLQADI
SNSDAAHAAVQTIIDTHGTLNGIIHSAGVIKDAYLLEKTDEAVCQVLSPKVAGVLAIEQATRHIRLDFMVLFSSIAGALG
NPGQADYAGGNAFLDAFAIRRNRRVQQGLAFGRTLAFNWPLWAEGGMQMDRSNEILMKQATGMVAMDSQSGLEVLEQALR
SPYGHLLVAYGDVERIRTGLLQFKHVPVGESLHETKQSDPQDKAKLTAALRQDLVKMVADLQRIALERIQVDRELSDYGF
DSISFTQLANELNAFFGLTLMPTLFFELPNLAALGSYLLEHHGQAVQKKYAGQQQTTAAEAAHEGGAVRPTPWPASQPRF
MEPQPQPLPATPVRQTASTAIAVIGISGRFPGADSPEALWEQLVANQDLITEVPEDRWDWRAIYGDPHQEPGKTKIKWGG
FLSDIDRFDPLFFGISPREAEVLDPQFRLFLETVWAAVEDAGYRASTLSEGKTGVFAGAATADYKDLWQQARAEFGNQVS
AMEPFPFIIANRVSYLLDLHGPSEVIDTACSSSLIAVHRAVESIRQGSCDRAIAGGVSIIAHPRITISSSQVGILSQDGR

CKTFDASANGYVRSEGVAAIFLKPLEQAIKDGDRIYGLIRGSSENHGGKATSPTAPNPLAQKELLVSAYRDAGIDPRSIG
YIEAHGTGTELGDPIELNGLKGTFDALYQEQGHARSEQPYCGLGSLKTSIGHLEAAAGVSGIVKVLLMLKHKLMPGNIHL
NQPNPYLELEGSPFYLVKENRAWDACPDEQGQPLPRRAGVSGFGVGGSNAHVIIEEYVAAQDTSASPVAGAERPCLIVLS
AKTEARLQDMVRNLGDYLRKNRQRISNQLESLAYTLQVGREPMDERLGFVVSSLEELEQTCRDFLAGQLDPNRGLRGNLK
RDQERVAAFSEDEDMADTIDAWIRQHKLFELLDAWVKGAAFDWNGLYPNGKPGRISLPTYPFARDRYWLPQSRGTATPKL
ENRSSQAVAAHETQPPSTDASPFETMTFEEVWQERALEKGSPIAIRTLVCFLSEPENQIAFAEAMTRLSPQTQVVFIEQG
TRYQDLGGDRYRVASSDVATYGKAFKHIVKTHGGAQAVVYLWPLEDAGCIRDATFVVSILQAMVAEKLGVQRFLPAGQFA
PDTPEGSYLESWIGFEPSLGLTLPRMQVAVVLRELAAQPGKDADMETWAQDLWGELQSVKARSVLYRGHRRHELRILPTT
LRDVTHSPFRSGATYLITGGCGGLGYLCAQHLAETCGARLVLTGRSPMDVQKQAKLDQLERLGAQAIYVQADVADLEAMK
QVADRAKAAFGGIHGVVHAAGMTGSPSVLTKDIEDFHRVLAPKVEGTLMLDQVLGHEPLDFVCYFSSVSAILGDFGSCDY
AIGNRFLMTHAKLRNVLQVKGKRSGKTVVINWPLWEDGGMQLGEEEQTHFYLESSGQRAVQAQEGLALWEALMSQDKTQQ
VVVIGKPERVQRFLGIVETPVQAAPSEIMPSQGTVKRTELKGLDLGQCVTWDLTELISGQLKVERDDLDDESNLADFGFD
SILLAEFSRVLGDFYDLDITPSVFFGHSTIKQLTRYFLTEFPDAMEGFYLVSQPVVPERPRAVAQVTAASSAPPVVQPGA
TEPIAIIGMSGRFPKARDVASMWDILAQGINAVDEVPANRFDWKTLYGGPQRADNKTNSKWCGSIPGIEEFDPLFFEIAP
LEAELMSPGQRHLLQEAWNALEDAGYGPRHLEEQTIGMFVGVEEGEGYAKRVKEISLTAGSNGILAARLAYFLNLNGPTM
AINTACSSGLVAAHQACLSLRQGECDTALAAGVNLMVSPEAYVGMSQAGMLSADGACFTFDKRANGMVPGEAVAVVVLKR
LSRAVADGDPVHAVILGSGINYDGKTNGITAPSSVAQTELIRRVYRDYQINPEQIEYIVTHGTGTQLGDPVEINALNDAF
KDTARTHFCALTSTKTNFGHTFAASGLVSLISLVMAFRHETIPMNLHFNEGNPYIQWENSPFYVNCANKPWPAHPEGERS
GAVSAFGMSGTNAHMVLQSYMPDQLVKLSPDGLPPYFLLALSAKTRDALNERVRLMVEAFQQGPVSEPELHATSYTLLAG
RHHFQHRLAIVISDRENALHVWSQVDQVEKQANLFLGEVPQKFTVQKAILQYSKDLLQQSRLCHDNPEQYHDNLCALADL
YCQGYELPWPTLFGDHSPQLLHLPTYPFAREPYWISETSANPETQLAASPVLHPLLHQNTSDLKGQRFSSTFSGQEFFLA
DHVVKGQKVLPGVAYLEMARAAVLRAGGTPDAQTAMRLRNIVWTQPVLIEEQAKEVHIRLFAQEQGRIQYDIYTMPDATS
DGSATFFDTPDVERLDIPELQARLSWQSFGSAACYAAFESIGIHYGPGHQGLETVYLGHGEDDEAQILAKLSLPASVAQTL
NQLEAHPSLLDAALQAFIGFALGLGDQGTQNLSLPFAVEEVDIIAPCEANMWAWVRRSDGQSRETSASVEKYDIDICDEQG
HICIRLKRFSSRALNDPVQTGENGQTLCYKPLWQEKAVTVTARRLEDAQHLVIVAGMKPFASDLLAAPQGHITYLDLRSDT
AYPLAYQELALQVFKTVRKRLAARDRSRVLIQLLVPRESRHQLAAGLNALLKTAHLENPAIAGQLIEVSPDETAAELFLKL
EENRRCLEDTRIRYNGNQRLVACWEEVEPQDQHLPTLPWKERGIYLITGGMGALGAVFAREIASRVEAPTLILTGRSPLHD
VQPSLIDDLENLGAKVVYRAADMSRQQAVDAIIQSTARDFGSINGILHCAGMTRDNFIVKKTDDEFKTVLGPKVSGLVYLD
EATRDLKLDFFACFSSLSGVFGNVGQADYATANAFMNAYAHYRGELVLSNQRHGRTVSINWPLWLDGGMDVDEHIKENMLQ
KAGMVPMDTESGIWAFXXXLAGEQHQVAVLHGNGARLRSHMEATSAPRETGTTSKPVPVLDDGFRTQTVNFVKNALSQSIN
LPVERIQTETPFEKYGIDSILQVKVIQDLEAVVGELPRTILFECENIQELVSYLIDNHADALLSSLGPKQTPEQETAKKET
AESAENLSAENLPVQAPLKRQVISTPQRASQPQSQPQHQREEDDIAIIGISGRYPLSKTLDALWENLKAERNCITEADASR
WRQALDGIVAQGGPVPPCRYYGGFLQDVHGFDHALFDIAPEQVAGLSPELRLFLEITWETFEDAGYAKHALQALQAREQQG
VGIFVGTMYSQHSFTAPNLTEAAYLSNGTDWQIANRTSHFFDLTGPSIAVNSACSSSLTAIHLACESLKQRSCSMAIAGGI
NLTLLPSKYDALSRSKMLGSGHESKSLGVGDGYIPGEGVGAVLLKPLGAAKRDHDRILGVIKSSFINHSGGRQMYTAPDVK
RQAELIINSIERSGIDPETIGYVESAVNGSELGDPIEISALQKAFATFTNKRQFCALGSVKSNLGHLEAASGVSQLSKVLL
QHQHQMLVPSINANPMNPHVKLQKTAFYLQQACSPWEPLHHPETGERILRRSMINSFGAGGSYANLIVEEYGEKAPQDLAD
ARVEVQEHLLVFSAKTEKSLVDYLGKMVAFLEDNGASVTLGEVSRTSCRCNASPSSPRPCLGFPSSWQGCVYYALTSNLDE
QTPRREFIQQALKDYDLKRLAAFWLEGEVIDFRQLPGQSQPGIALPKYAFEHADPQTEHDASKTGLKAEAEEIDEAFFLDI
SQKIANDHMSEDEFEKLMLARIHQENEHV

OnnJ ACP-KS11-DH-ACP-KS12-KR-ACP-CAT 4,148 aa

MYDQNQIKLLYRQLKEQKIRPEEAAGKLRELQGQRTKAEDGAPEIYRYDESFLKDHTVNGEQVLIGMTHASLAMRTFFEW
FPEQACVRLQRLNFVTPVAVERNGQVEVRIKPGVSEGKDTPLDFQAVYRENPSATWNLTATGKLQAAEFHHQRIDIQQVK
SGLALFPDLTQLYADGTLIQFGESFKTVTQLHTGEDDVLARVALTRHAQEEPHSYEIHPLIAHSAFLAIVPILENAKGGG
SFLPFGIKEFTYRKTDGLENCWIWAHLVNDAGEMILFDANVMTDDGTVVAHLAGCSLQRLRSDVQQPRTVVKPLAAQTPV
SPQAALTDLAGSIQNYLVGKLVRLTGVDEGAYADTSTNLMELGLDSIQLVTLSDEIKQETDIDLDPTLFFEYPNLKELTR
FFFEEHADAFRSMEGFHAAQSEPAEELAQAEAVGQPAIPHSEQPQQPQWQSSTIPHQTAILPENSGKETVRREDIAVIGM
HGIFPDGADVDQFWRDIRDAKDLIREIPLDHWDVGPWYDENPEAKDRTYSKWGSFIADVGGFDPGFFSISPREAEWMDPQ
VRLTLQSIYATAEDAGVINRLRGSDTGVFIGICFNDYADKIADLRLPVDPYSGTGSSGIAANRASFIFDLTGPSLVINTA
CSSSLFALHAACHALRNGECGMAFVGGANLLLSSFHYRYFSAIKALSPTGRCHAFDAAADGYVPGEFVGSILLKPLSRAQ
ADGDHIYAVVKGSAALHGGHSPSLTAPSVAGEWEDAGINPETISYIEAHGTGTKLGDPVELNSLQKAFRRYTQKEGFCAVG
SVKANIGHTEGAAGMAGIMKVILQMQHREIPPLALFENLNPYIRLDQSALYINRESQAWDVSDAPRRAGINSFGFSGSYAH
VVLEEYLPQRAHEAMPDTAPVIIALSARNKERLRGYAEKLLAFVQAPTAENINLQDLSYTLLVGREAMEERLGFIVGSMQE
LVHKLQQVVEGKPGGDFYQGNVKRAKPSDDDFSADEDMARTLELWLEKGKLAKLLNLWVMGLIGDWSGLFNTTYGAHDGAH
PRPRRISLPTYAFAKERYWVPESALTSAPAVGHAAAKVIHPLLHENTSDLTQQCFTSRFTGQEFFLADHVVEGQKVLPGVA
YLEMANIAVAKAAASLMDGPPEVHLKNLVWVRPIVVNDAQDVHIQLFQERNGEIAFHIYVPAEEALGQPLTCCQGTAIIKT
SDHPAADHGMYLDLDALRQKINRNRILPQDCYQAFEAMGIEYGPGHRAIEAVHVGDQEVLAALKLPASVADTQGQFTLHPS
LLDSALQASIGLSFGSAERDNRASLPFALDSLEIFRPCSESMWAWIRLCPSAKTAITKLDIDLCDGEGRICLKMSGYSARL
LDKTQAEKSVLASSVAVAESGQEKPQTEPPRQIETEATGQFAPVKLEPEVFEKKAVAYFREQLANVLKLPADRVQPEEPLE
TYGIDSVLVMELTQELEKTFGSLSKTLFFEYQTIEEVSQYFVERHGDQLARLLQPEQMEQKPSPQETKAQPSTPETKTQTL
PEQAMPRHGQSLTGLRQSRFSAVGQSPSTAQPSPASESGRRADAPLDIAIIGLSGKYPQSSDLDAYWRNLRDGRDCITEVP
EDRWDWRHYYTEDRSLPGHHYSKWGGFIEDMDTFDPLFFQISPREAPYMDPQERLFLAYAWMALEDAGYQRSDFNRKDGVN
QGDQVGVYAGVMYGHYQLYGAEETLRGNPIAVGGSYASIANRVSYILNLHGPSMTLDTMCSGSLTTLHLACQDLKHGHTDL
AIAGGVNVTVHPNKYLMLSAGQFISSHGHCASFGEGGDGYVPSEGVGVAILKRLEDAERDGDHIYGVIKGSALNHGGKTNG
YSVPNPNSQQAAIEKALKESGVDPRTISYIEAHGTGTKLGDPIEITGLSKAFGPNTEKQYCRIGSAKSNIGHCESAAGIAG
VTKVLLQMKHGQIAPSLHSQVLNPNIDFSTTPFMVNKELRTWEQPVVDSVPYPRIAGVSSYGAGGSNAHMIIEEYQQAQKP
VVDSSMPVVIPLSARNEERLKAYAERLLRFIREDTGSTARINLADLAYTLQVGREAMEERFACLVQSMPELEQRLRDFIAG
HENEDLHLGNIHTHKDALAAFVSDEDLPTMIDLWISQGKMHRLADLWVKGLNVDWHKLYEGHDMQRISAPTYPFAQTRYWI
KISGSQVPNSSVPEKHEAAFGEKTEAAFGTLIVKPAWKPQNSRLGGLSINDRDHLVLLCDLHERSQAMVNRAPRISFADLE
ATQQNGEQRFQDYALQLFRKIQTILQNKPKDPVLLQVLVPNDGPEQVFTALAGLLQTARMENPNILGQLVAVPAGISGQDL
MDIVQDNSQCPDDHLIRYDQDQRFVASFAEVTHFVRESKLEQLPWKNGGVYLITGGAGGLGLMFAKEIARKVNRPQLILTG
RSPLNRERREKISEIERLGAQVTYATVDVCDKSAVDALIQDIQESAGGLNGVIHSAGVIRDNFILKKTETEFQQVLDPKVT
GLMNLDLATQQLELDFFVLFSSIAAAGNVGQADYAVANAFMDAYCHNHNSEINYVSANDRRPRHRMLSINWPLWREGGMGL
DASLEKSMKGLGIVPLETPSGIHAFYQALACGESQVMVMEGELNKLRGLFGLDQQQVAPEHIESVQQREEREQVVPASQDM
LEEKALAYFTEILSTIIELPADQINADTPVESYGVDSVMMLEMINVLEETFGSLPKTLFFEYSDVRALSEYFLATYRGPLQ
RVLKWEQEPAGAVGSAAVDANEAPPAISPLSFGQKTLFFLNNLNPESAAYNIGGAIRIRAALDVDVLKDSLQSLIDRHAIL
RASFFLEDDQFFQKAQPHQEVAFTLEDIPDLPLDELRSKMEKDQKLPFELTQGPLVRTNLYKRNAEDFILHISIPHIIFDA
WSFPLFFSELWQIYFGRVSGKPLALPELQFDYSDFVSWQASMLESDEGQIHKRYWLKQLSGELPVLDLKTDRPRPPIQTHE
GATHRFELDTETVNGLDDLAKQTRCTRYMILLAAFQIMLQRYSGQSQVLVGSPTSGRTQRKFSGILGHFVNTVVLRADFSD

DLSVLDFLTDQVRTTVLGALDHQDYPLWSIIEQLQPERDPSRSPLFQVLFNYNRVSKTSPFPQADFTIELFNITQQEGQFD
MTLSLEDVRSGITAGIDFNTDLFDESTIIRMGRHYQNILAAMVNQPQQTISKIAFLTQPETQLLAQWQAIPRNHGDRSCIH
QLFEAQAEARPDATAVVLYQAEDRLEAQSQQVATYGELNEKANQLARYLRKLGVGREAIVGICVERSLEMIIAVMAVLKAG
GAFLPIDPILPRKRREFMLNDASVSVLLTKEALLGKGIDDQIPVLFLDQNWDSIAQKSSSNLHCDVSSDSLAYVIYTSGST
GEPKGTMIEQGSWADLYIALEEIYQFRSIESFLQMADFSFDVFFLDFIRAFCSGGKLVLCPKEFLFNPEMLYGLICKEKIE
FADFVPGVMSHLMDYLESHGRNLETMKTIIVGSDVFYAKDCHRLKKLIGEQTRLFNAYGLTETTIDNLYFETNALAITDAM
IVPIGRPLPNTSVYILDQHLHPVPIGVPGELYIGGSNVARGYLNLPEKTQEKFIRHPFSHADDERLCRTQDLVYYQQDGTI
HYMGRADSQVKIRGHRIEVSEIETVLLQIPQVMEAVVLVKEREAGGRFLVGYIVPEAGSPIDSHQLRVTLKEKLPDYMIPS
QHVFLDKMPLLVSGKVDRQALPEPDFDEFRAHPYVQPSTKLEKLIAYIWAETLGLKRVGKYDGFFEIGGDSLLCIKIIAKM
KEKQITITPRLLLLNSTVADLSRAIESSDAGGKAWHVDHDPKPEPDHGVEPLHQIFFVPGLDQNIFATVDLEKNLGADYHI
HTLLSFNAEENGYAYTSLREMAKGYIQEIKAAQPQGPYHLIAYSWGGSVAVEMVRQLQASQEKTGFVGLIDAYYNEFHCLP
LDSKKVYIAASHYFKVASLLSFLQRRCGVTLSSEETHTHLKDMKDMSNEAMIDFLKAAISWDEKAGIDVGQSLLTFLNMVD
ISEKISRDELSLDFDPPRISFFSATKTDAESHMIRNLLNDTVPGAKVYFDFADKSALDSWSSFKTHFEEIKIAADHQTILF
EPYVKQITNHVKEKLTVLS

The protein sequence of the second PKS-NRPS cluster from *T. swinhoei*

Only sequences of PKS and NRPS protein are shown, additional genes shown in parenthesis.

PksA Acyl-CoA dehydrogenase 176 aa

MDERRSIPPHIILDLGNHGFLGLQVPERYGGMNLNYRDALWVYEQVASIDLTLASFLAVHNALGIRPILKYALMEQRNAL
IPKLATGRELASFAITEPGAGSNPRAIAATATPIGENKWQLNGQKSWIGNGSWAGVTNLFVQLLDNEQTPMGITGFVLSQ
EITPLSQGPEALTLGL

PksB Acyl-CoA dehydrogenase-ACP 524 aa

MVQNTIYLDGVSVESNQLLGKPGEGMVVAQDTMMFTRLVIAAMSVGAMKRCAQLMLRYAQRRTVTTGRLLENPVTITRLS
ELTAAVVVVETYVKRIAEFLDTGIKVPEEAYIVCKTGGPEFLGDATDALIQLLGGRGYIETNIAPQLWRDARILRIFEGP
TETLTMYLGSRACNKSEIIHQFMTKSLGASAIANELREAVAQIKTDFLDHQQTECGLPFASPFSDTVFTDRWIYTVVGEL
VTLGMLWAALEGAVRHLDTEKMDQEEMDHGLRWIRTQFTQKREQIHSHSAIETIRSKSELECWIDGYAKKIGDLEQRLPA
EEHELDPLLRQTQVMAPQTQSQLALEVEEEISDLHVPRPCSKQDETQMVTLDVTLGSSQKRPSFATPIYTVDAIQTWLIQ
WLATHLQTVVDKIDVTTALADYGLDSVMALELLDALETWLQEPLEMTPTLFWNYPTISELAAYLANRTLQADELTIREKM
QSQPQLDVSAGLISEKMNLSSLSDSEMAQLLAQEIAASRTRRAE

PksC KS1 238 aa

MSSTPEITDYRTLLQDAFVELQELRAEYDRLKQKTSAPEQLESIAVVGASCRFPGGANTLSSFWHLLCNGIDAITEVPAD
RWDATACSRHTSNQVVPPYGGFLDRVDHFDAPFFNISPLEAEMMDPQQRLLLETHWEALENAGIAPTTLRGSKTGIYVGI
ATNDYMLLQAKAQVDPDAYFGTGTANSVAAGRIAYALGLQGPTLSRDTACSSSLVAVHLACQSLQRDECTLSFLAILV

PksD (KR)-ACP-(KS2) 882 aa

MGREALPPRAAWSEYLLNPAGDTTLRQKIADIQDLEEAGAEVLVLTGSLTNRTAMLGYKERVKAKMGPIFGVIHCAGLVS

QENPAFIRKPLDEIRAVCAPKVEGLDVLHELFATETLHFFVLFSSVSAIVPTLGAGQSDYAMANAYMDYFATHQQGQGRT
HYVSIQWPNWQETGMGLVQSDAYRQSGLCNHTDEEGLTLLDAVLALKPGPVVLPAVARTAQFHPEHLLMPHLVWQEAAAS
LSLDGSSVLYTKIAEWLGSLLSQELKLSADRLEMTLPFQAYGVDSILLAQLVARIDRELPGVKLDPSLILMYPTLQSLSE
YLVKTYPEALAALFKEPAMTAEKGKPLHRDLENILNQNRNSLTESQWPKSGGLLAAVEPSAKATAKIASQTETQIAVVGM
ACHFPDAPTLQHYWDNLKHGRDSICEVPDSRWTVDDYYDPNSNQRGKSISKWGAFLADIEAFDPHYFGISEALASQIDPL
ERQWLEVSVEALADAGYEKEDLWGQAVGIFVGSRVSNFADKLETLQKDALVGLGQNFIAAHLAHLYNFKGPNMVVDTACA
SSLTAIHLAMQSISRGESTVALAGGVDILLDEAPYIGLSAAQVLSPDGRCKTFDENANGIGIGEGCGVLVLKSLDQAIAD
GNKIYGEIEGSAINQDGYTMGVTTPNPEAQQALIETVLREAQVDPRTITYVETHGTGTLIGDPIELKGLTDAFGQAPNDK
QFCGVGSVKSNIGHLLSAAGAAGIIKVLLSITHKALPPTLHCEHPNPRFNFAESPFYPVRQLTDWAGEAGIHRAGVSAFG
LGGHNAHLIVSGAGIPSTQRASLEPRGVPVVFNRRRYWPTGWTLPPANGISAAAIDDERECHEELIRFFEEDLEIIIQEE
TV

PksE DH-ACP 536 aa

MPGVTLLDMTYRLGKKALGHSTFRLNNVLFQQPIATHETFDQTLRVTIEMQSANEGWVQVQSQKVKARQALEQGWSDIMN
CGLSLQATQTQELLRATFDIPGFIETASRQWDMDQSYEAARAVDIWHGPFMKTWGTVYQKQNERLVALHLGELAEDYRDK
FYAHPAFLDGATLVGFPEGRNGVDLVDDDVPYIPFSIRRFAIYQPFPQTIYVYTNPPEATRDIVKLPDLVASDITIYDGE
GCVIAEFNGLAAKRIREAHLISQLVSQVLLPEEKVQAAPRVACLQMDKGPAAWDGENEQSNLSRQPDIPTHPKATAHDDE
TKATTTRYLQQEVAKILNKPAAEISPHTPFYELGLESVNLLVLTKELEQELGLPLYPTLLFEYTTIAELTAYLFRAGATY
EPPPEPMPAKPITVPKNPVPEEARQSAEASFYQPRWEEQTLPVDPQFDAASPLILLYSPDKNTIHKLKEQGIRGNGRRSV
LDFRQFWPEVSENWRQSLRSESGSGRGMVTPYRDPARKQPHTKKRRFLEYARTWRF

PksF KR-KS3-ACP-(KS4) 1,648 aa

MLHHHLFCSIVRIKTQFINSKSRASEEMGVGLSWIFVSFGQRYQKIGDNHYVLNLAQEGAWSHLIATLQENNHIPKSVVF
LSTLEHGDFNPDEEMIEELEVYLHHHLYPLFQLAGSLIKAKMPQDVQLLYLYQAQDSLRACFEQAMTGFGKTLEQENPKF
RFKTIGMENSSDLSLSTLISQEVVANVGDIVPQKSFVPRLPEPRRDKTFLGRYSTNIRYQAGTRQVQTFAPCTHLLNEAA
IPQLCKQGTYIITGGTGGIGLIVAEGLLRNYAANLVLVGRSPLNAEKTDRLEALKGLGTGTVTYLQADITSLEETTSLVK
RTQSQFGKIHGVIHAAGVIQDSLIPQKTLPQLRTVVVPKIMGALHLDRLLADEMLDLFVLFSSTTAILGNVGQADYGYAN
RFLDAFASYRQCQVAQGQRHGHTVAINWPLWENGGMQTTPETRQTLAKAGLQPLDDQTGWRAFDTAVNQQHFHQLLFLWQ
RAHQADETAKRTDTQSARPTVALPQAIDAYLSVPSIPEVEKKGSTDPNATDIAIVGLSGRYPMAESLAEFWENLKAGRDC
ISEIPTERWHHERDFNRGKHKASKSYSKWGGFMADVDKFDPLFFNISPREAEMMDPQERLFLEIAWQTLEDAGYTRETLA
TQTQGQVGVFVGAMWSEYQLYAQEQMGLSSSQASIANRISYALNLHGPSLAVDTMCSSSVTSIHLACESIRRGECVAALA
GGVNVSIHPNKYRFLAQSNFVSSDGHCRSFGEGGDGYVPGEGVGAVLLKPLSLSEADGDTIYGVIKGSSVNHGGKTSGYT
VPNPVAQAQLIAATLEKSGIAPESISYVEAHGTGTSLGDPIEIRGLSQAFGEQSETYSCAIGSVKSNIGHLESAAGIAGL
TKVLLQMKHQQLMPSIHSETLNPHIDFDSTPFAVQRELAEWKQPILEVAGEQQIVPRRAGLSSFGAGGANAHLIIEEYRK
EGKETGNKGLGIKKEDWVDSEPQVIVLSAKNEERLREYAHKLLVYVEQTSVPNPAIRARDESNHQTSPQVGPSLANIAYT
LQVGREAMGERLAFVVSNLEELIVRLRQYQLKENATGQVYQGNTNNSKQTVNLLSGGIIGSAIIKTVVDNREYDKLAQLW
VSGVEIEWRLLYGDETPRRIPLPTYPFAKKRYWFDEHMAKSKEEQLATPASMAPLPTTVLPDPPNASAENVPLKHGKQKV
VLRMPDTHILPEEIGSAQDNSMRLALASLTEEVENVPGKHQSLEHMPVEPPVQDTAQVRAMIAQQLRQQLVDLLYVELEE
LDDQAKFLDLGLDSIAGVEWVKRINDTFGLEVNATRFYDYPTIVHLASYLAGMIAQHTTEPVVIDVPVPAQLPHLHLIET
QGLHEQADPKEENIQTTEQETSTIDVSLIESRSGESAQQSMQTQVDKVPTGFSDSLTEGIAIIGMSGRFPGANDVEAFWQ

NLANGVDSIAEVSPERWDVDHYYDANRQAPGKSYSKWLGALEDIDKFDPLFFEISPREAELMDPQQRLFLEEAWKALEDA
GYTGHTLSERKCGVFVGASTGDYHASLNQGDSEPDAYTLMGSSPSMLAARLSYILNLKGASLAIDTACSSSLVAIHEGCE
SLRTGRNEMVLAGGVCVLTGPQMHIMTSTLFLRNESVILFVHATRGGF

(Transposase)

PksG (KS4) 206 aa

MQLEGVSNRPLDRISSKSEGHENKDVNLAGQSVKVSRSEKRENQVFVIDLEQFVSAGKVQI
MLSSDGRCKTFDKQADGFGIAEGVGVVVLKRLADAVRDGDCIYGVIKGSGINQDGSTNGITAPSANSQTALELDVYRQAA
IAPATIQLVEAHGTGTKLGDPIEIEALTTALQAYTDQTGYCAIGSVKTNIGHTLTAAGVAGVIKV

(Transposase)

PksH (KS4)-DH-ACP-(KS5) 1,155 aa

VNSEQLAVNSEPQVIVLSAKDEERLREYAHKLLVYLSASAPNAVTQELDATVGTQTVQYTLREMVADILGVDMTEIESEQ
AFAECGLDAVQLSRLQTMVEERVGCDLPSTLLTAGASVASIAQHIASLGLVAPKGKYAPHPQTVPSLASIAYTLQVGREA
MAQRLAFVTSSVTDMKQKLTAYVDNKQAGIYQGHRGANRERFDLLTEGEEAKVFIETLMYNGKLERVAQAWVSGVEIDWS
LLYGEGRPRRISLPTYPFARESYWFRRHQKPESPNGAVHSVLQQAVRPVIRSGSSNKLKLKSLQPQASSLFFASRENQPG
EFTTRMTLAENCLLREHIVFGHYVLPTDSLLELVYEAAIVYLQTERLDFEQLYLHKPLIGLPNTETTARVTFGEKGSVAT
FELSSQVAGLSQSLAKNMSGTLRIHQETLTSEHSYKQVLTDFEKKGKMTDLFSADYPLQVGDFYQSLQEVYLAGNRAVGV
LKLAAAAQPTKSQFLLSPSILDSLLASVVCLAHHCATAVPELNTPDYAFIPIYIDKVHIAKPLTDEVYTSYVQLVKQEDE
FMRFDVALIDHQDQVVLCLDGLDEKRITGADIEQSLAEMIQLDDLSKESKTSLHSEGIPPMKRTVEGYSVRGDGVETRTY
ILQTVRNFLSQTLLLPDEKIKENANFLEMGFDSILGMEFIQKVANHFEIQLKGTALYEYNTLNKLTDHITQTYPQISQLA
INPNILPFRQVDIPDHEAADVRALKLKSRTARMGAYSDEPAVAAGSVEESVQILETPQPQSEISSSPKDIAIVGLACRFP
EAENAASFWQQIAASRYLITDIPESRRAAWRDVADVYCQKGSFINDVDQFDPLFFNISPREAASMDPQLRKIFEVAWEAI
EDSGQVIRGSHTGVYLGNCFNDYAQLLKEQGQLDDQYVGMANANSALANRISYTLDLTGPSLTVDTACSSSLVALHLGCE
ALCNGDIEQALVGGVNLSLHPEKYKTFCAMGALSTGGVLRPFDERADGYIPGEGIAMLVLKPLTQALVDSHHIYGIIKGT
AVRHGGYSGGPTVPNVEEQTEVMRAAWHAGNIPAHTLSYYEAHGTGTRLGDPLEIQAIKKALAQGGEQTGACAHWDSQGQ
HRPYRGYGWVSGCDQSVVDDEASAVTSPAQTGKFK

PksI (KS5)-DH-KR-MT-ACP-ER-(KS6) 2,586 aa

MAPVLGWVTRWRSRLSKRLWPRAANRQVLVPIGTVKANIGHTEATAGLAGVIKVLLMMKHQLLPALPKLENLNEMIELEG
SGLYINDVTQPWSSENNLPRRAGVSSFGMGGTLAHAVIEEYQGSNLQVGSFVDDSKPQVIVLSAKNEERLREYARKLLVY
LQQTSALNAGTQEPDATIDAQSLQQTLCEMVAELLGVDVADIEREQPFAEGGLDVVQLSRLQTMVEGRYGCELSSTLLVA
DASVASMAQHLASVDLMVPIGRHRHRPQPDPSLASVAYTLQVGREAMEARLAFVVSSLEELIARLNQYEQQESTTGHIYQ
GNIKESEEMTSLLVDGAAGEAYIKVVVDNREYDKLARLWVTGVEIDWSLLHGESKPVRISLPTYPFARERHWIPESEPLE
MGIKGDSAAKLDSIQHLHPLVHQNVSTLTEQRFSSSFTGNEFFLRDHQVQGERVLPGVAYLEMARVAAEMAVEEGRVTHL
QDVVWIRPLVVQDAAQEVHLGLYPDAQGTLRYEVRTLDRSGTAIVHSQGGVLFGGERAEAEVIDLEAIRARCPEKKSSQE
CYQQFEQQGLGYGPTFQALEEVVCNESEALAQLRLPNEVVQESYFLHPSLLDGALQSIIGLLDDDSEATPTTYLPFAVEL
VEVSAPTPTMGYAYVTRLASGTDTARYADMARFNVALLDESGRVCLKMHTLTVKAALPVEAGWYYQPCWREEVLPDTDRE

DGPSEDNIVLVMPKGLQELGDGLAGRYPQQRVIRLLLSDTCLSLNETTWEINLANDAGMAACLNVLPEIDTLYFLGGLSN
GQFAPAELDELAGHQEEGVISLFRLVKLLEERGTMNGLSALKVVTNHTLPLDARDESYPWDAALSGFTMSLAKEYPAIDV
SNLDIVLKRDEAGQVVLDEQDLTVMVAEHGCHGQPILLRRGSRYIRQVVRLSLPDVEQVSRWASGQVESETKSTENDTAL
QSCDLQPASLLPYRQGGVYLILGGTGGIGLETAVHLAEQVQAKVVLVGRSALTTAKEAELKRIQTAGGEYLYCQADGTDL
DQMQAVVRQAKAAFGTIHGVIHSALVLKDSTIRTMDEATLLAVLAPKVTGSVVLAEVVKEESLDFMLFFSSTQSFWGNGG
QSNYAAGSTFQDAYARYLNQLRPYPVKVINWGYWGSVGVVATEEYRQQLTQQGIQSIEVKEGMAAIDQILAAPLPQVAAI
KADKSVLAQMGIAFEHKVEVIAPDYASIAQDTVAALTATISALTPDLSAMEQESIGFEALESLARLALLQAFQWMGVLQN
PGEYASRYDLSQKLNMLPKYQRLFDACLHILEQGGYINLHGATVEATDSVAQAQEETLPQQQEEMVYSHPMLQPHLFLLT
TCLQALPEILQGRIPATDVMFPNSSLDLAQGIYQGTPVTNYYNQLVALAVKSTVEQRLPTLQMDEKIRIIEIGAGTGGTT
GFVLEVLAPYQDHIDYLYSDISQSFLQHGQRHFGHHYPCMHIGLLNIESTLTRQGLDLNQADIVLGANVVHATKHIQQTL
KNLKGLLKANGLLILNEATRFSAFATLTFGLLDGWWAFEDGEQRIPYTPLLGSEQWQRALDAAGYKQIQIGGEATKKTTD
QIVILAESDGVCTQWIKANPVGRNIEPQPRKSQVGRAFMPEETKQRALSRSPAKRPSALGPLSAKRQPDGLRTKSDQVDE
LATALEQTILEQVARATGLNPDQVDLDKSFVEYGVDSIVGIELINGISQAVGQQLRTTTLFDYPTVHTLANYINTEYGPA
LATTKAAADLQVPSVAEPTTSVDQLFNSQEINPGQIPFAVHSIQVDEGVGESLFSLSGAAAKAQAVVIEKPGNLEALQIQ
SVSVVPPNTGEVQIAVRAFALNFGDLLCLRGLYPTMPPYPFIPGFEVAGIVQQVGQAVSDFVPGDEVIALMGDPMGGHAT
MVNTRAELVLKKPANIDFETACAAPVTFLTAYDALEKAALMAGEKVLIQTAAGGVGLMAVQLAQRQGAEIFATAGSTEKC
NYLKKLGIHHVINYRQEDFAAAIQRLTNGIGVDVVLNTLSGENIQKGLNSLAPGGRYVEIAMTGLKAATSVDLSNLVHNQ
SLHSIDLRRLLKSAPEVAGRYLQTLKEMLHEGEITLTIDRVFPFAQLRQAYHYLDAAQNIGKVVVSTADVTTLSSVESAV
ADTQRQPQLERNHHDIAVVGMAGRFPGANDVTEFWENLAQGRSTITEVPPGRWSMAEHYDPNPDSPNKTYSKWGGFLDDI
DCFDPEFFNLSGREAELMDPQQRLFLEACWSALEDAGYCCNQLDSRPFGVFLGGGRGDYQQKMLEAGVEAHAYSFMGNDA
SISAARIAYYLNLKGPVLAINTACSSSLIAIHLACQSIQQGESEVALAGSVFINTGPQFHILSSKAGMLSPDGKCKTFDN
RADGFVPGEGVGVIVLKSLAQAIADRDPILGVIRGSAINQDGKTNGITAPSTLSQTELEKAVYAQSQIHPETISYVETHG
TGTKLGDPVEIEALTNAFRHYTDKQQFCAIGSVKTNIGHLATAAGVAGVIKVLLAFKHQKLPPSLHFEQPNEHIDFENSP
LLCKHRVAGLENRVWSSQSGRQFLWL

PksJ (KS6)-ACP-KS7-ACP-ER-KS8-KR-ACP 3,327 aa

MNISILRTVPYYVNTELRDWKTESGPRRAAVSSFGFSGTNAHLVLEEYQAAWQGDKVAAPRGYAARTSAHCAIGQE
MQNDLEEASAPNPVSQESDATIDAQSIHQTLCEMVAGLHGVDVAEIESEQTFAECGLDAVQLSHLQTMVEKWVGCELLVT
PADSAGSLVQQVLLDLKDTAIDNLQPANLQPTLSLASVAYTLQVGREAMAERLALVVSSIEELMEQLKQYDHKESDTGQV
YQGNIKTSQQTANLLIEGATGEAFVKMVVDNREYAKLAQLWVLGVEIDWRLLYGEEKPRRISLPTYPFAKKRYWFNGHLN
RNATTIQKRKDTKLSLPDKIDSGTIDNIAAETNLLMSNNGPDRRPKLLLQSLDEVSIPKTVGDAQEQTGRLMQDSIAFSK
KPAINPAVDQSAKSSKRDVATIAKQLQQQLAALLFVDLEDISNYERFVDLGLDSITGVEWVRQINAEFGADIHAARLYDH
PTVERLAKYLTEITPQNSPESMVPHAPARTQSAQIQLKHLTAIQRAEQPLHGTTKGASTSEKESIQQQTRDEHLQSIPIQ
VDKYSNRINDSSTESIAIIGMSGRFPGANDVDELWQNIATGVDSITEVSPERWDVERYYDPHKEAPGKSYSKWLGALDDI
DKFDPLFFEISPREAELMDPQQRLFLEEAWKAFEDAGYAGCSLAEGRCGVFVGASAGDYHTRLNQSGSEPDAYMLMGSSP
SILAARLSYILDLKGASLTIDTACSSSLVAIHEGCQSLLSGQNEMVLAGGVCVLTGPQMHIMTSKAQMLAADGRCKTFDN
QADGFGIAEGVGVVVLKRLADAERDGDQIYGVIKGSAINQDGATNGITAPSADAQTALALEVYRQAAIDPETIQFVEAHG
TGTKLGDPIEISALTDAYRKWTDKRNYCAIGSVKTNIGHTLTAAGVAGVMKVLLALKYKMLPPSLHFENANEHIDFENSP
FYVNTQLRDWSIASGPRRAAVSSFGFSGTNAHLVVEEYQGAREQSGKVTGDESAQIIVLSARNEERLQAYAHKLLAYLSA
SAMLVPNDLEVASTPNLATQAIAETPDIQAIRQSLRELVADILGVDIAEIEDDQTFTDCGLDTIQLSRLQTMVEERVGCE

LPSTFLTTEASVASVAKEWASQGSVVPNSEYTPQASLALPSSASLVNIAYTLQVGRESMEERLAFVVSSIEELIEQLNQY
EQKESVTGHFYQGNTKDSKETTSLLIDGEKGEAFIKAIINDEEYDKVAQLWVAGVEIDWRLLHGDETPRRISLPTYPFAK
KRYWFDEHMAKNNQQQLATPASMAPLPTSVLSDPQNATAESVALKHGKQKVVLKVLDAHILPDENRTAQENSMRLSLESL
TEEAGKTARKHPSLPAHAPVETPVQNTAQVQVMIEQQLRQQLVELLYVEPDDVETQKKFLDLGLDSITGVEWVKRINGEF
GLNIPATTLYDYPTLQRLAQYLAEMIPHESSRVHVTKSGATRATLPQGTKAETRLPIVLNPLSAELAQQVGLVVSGVQAV
DDTHLTTWVVPSPQANEIQIQVQASAINFPDIMCIKGLYPTMPSYPFVPGFEVSGIVTAVGRDVQQFAEGDEVIALTGQQ
LGGHAHYVNVSPFAAVKKPRNVTFEEACSLPVIFLTIHHALQVANLKEGEHILIQTAAGGCGLMAVQLANLRHAVLYGTS
SRDEKLAFLQQIGVDYRLNYTTDFDESILAITEERGVDVVLNMLGGLAIQKGLNVLAPGGRYVELAVQGLKASPTLDLSG
LVHNQTFYSIDGRRGGFGDPERLSEYLHEMVQLVEHEEIYPIVHRIYPIAQIQEALRYVESGAHIGKVVISHTQAEAVDL
TQKCRSDLIQQKARSRQKPAPVHPGAQTIQNTPVISSMRHEEIAIVGMSGRFPGANNVEEFWQNIAGGVDSITEVSPARW
DVNCYYDPNRQAPGKSYSKWLGALEDIDKFDPLFFEISPREAELMDPQQRLFLEEAWKALEDAGYAGHTLSERRCGVFVG
AGTGDYHASLNQNKPELDAYTLMGSSPSILAARLAYILNLKGASLAIDTACSSSLVAIHEGCQSLLTGQNEMVLAGGVCV
LTGPHMHIMTSKAQMLAADGRCKTFANQADGFGIAEGVGIVVLKRLADAERDGDHIYGVIKGSGINQNGATNGITAPSAH
SQTALELEVYEQAGIDPATIQLIEAHGTGTKLGDPIEVQALTDAFRTYTDRTGYCAIGSVKTNIGHTLTTAGVAGVLKAL
LSLQHKQLPPSLHCEEPNEHIDFDNSPFYVNTMLRDWSTESGPRRAAVSSFGFSGTNAHLVLEEYEPLAVKPLTLMRQEA
GTLWVRPRPEGSRGNNVSAHLAMKSGPQVIVLSAKNEERLREYAHKLLVYFAGFATSNPTPQALDETPDIQTIRQSLREI
VANVLGVEVAEVAHDQTFDECGLDAVQLSRLQTTVEEQYSFELAITPSDSVESLVQQLFSLWGDGTSSRQLVVNSKHLAV
GSSQPPALADIAYTLQVGREAMEERLALVVSSTEELAELLKQYQQRKEVPDQVFQGNIKNAKQTINLLVDSASGETYIKT
VFKNQEYATLAQLWVLGVEIDWHLLYEGEPRRHISLPTYPFEKKRYWFPESSSTDQPPRAAHETWQSRQQNVMTKTAQIL
NFEAAWIEQALPVHDVDWVERISAEGDKRILVMYEDPTDYAAMDKLLGGVDVALQESGQARSLIVQFVGINSGLSAVEGN
TSFMLDKTSQEDIAQWIKDVKHNGQLPDVILFLGHPPETDMLADGYPLSSAYKDEIRFLFHLIQGFMQHAWDNSIRGYYL
FASTEAIPRLALEALSGLARATVLENPQQIYKTICVDPTLTADQRAVVLLQEWLLERSAATNTNHLAMVRYQGLRRFVSS
LQERASTSPEATQVFRPGGTYLITGGLGAIGQHLCAYLSQHYQSTLIILARSKRTEETEGQIQTIEQYGGKVWYYPVDIT
DRQSLQETVARAKTKVDQIDGIIHLARTVEDDLIVNKHFDSFARVMAPKVEGTLYLDQATQDEPLDFFMIFSSAASYGIK
GSPDYAYATAFQNAFAEWRQGLVHKGERSGQSLALCWGQWAMDPYSNEQRNALLKQMGFDFLTIESGMQILKQSLAQDNA
IIGAIAVSEPQKVRRFLGIRPASREQNASDLTIEQLLMDLQDKTKNPEDLFRQIAKFDTRQLSDRQIKTLHAHLTRHSVS
RAKPEIDGLNVEEVLEYEPTTAFLPDIQPADIQPADMQTVILAMLKKTLKLEEDRFDPQMTFLEYGMDSITGMQMATNLE
QSLGIEVPPRWLIEYPTLEQFAGKLQTIKRQGTKQQADIELKGKENEKRRA

PksK Oxygenase 323aa

MVFDIATFADQQGFHAVWMPERHFHSFGGIYANPSVLASALAVKTGQIRLRSGSVVLPLHHPVEVVEAWAMIDQLSGGRVD
LGFASGWNPNDFILSPTTYDNLREVWHARIPLVQRLWSGESMPFRNGQGEEVAIQIYPKPLQAKLNVWLTISKSTASFRYA
GSQGYNVLTMLQGIDLNELGKKIAVYRLGRQEGGFDPKTGSVTLMLHTLVHKDLKTVEEAVRDPFFNYIKSALTGHIQQVA
ADQRPNETELNKMVDYSYERYFKTGALFGSVAETEQIVAKAIAVGVNEIACLMDFGVDASVVMDSLGYLNALKNKVMKGI

PksL KS9-ACP-C-A-T-E 2,367 aa

MAGRFPRSPDLASFWDNLQAGSNLIHEIPQDRWDWPADQGDFEADDHHPMVQHIGAMDDIFSFDAGLFGISPREAEAMDP
HQRVLLEVVWETLEHAGYAPEQTSGTKIGVFVAMYNTDFLIRSQKVRWDKESQPYLATGNGLLISNRISHEFNWHGPSEI
VATACSSALVALHKARQAIRGGDCDMALVGGVSLLLSPERIGILDQLGILSQLGYCAPFDLNSPGEVLGEGVGAVILKPL
SHAKRDGDAIYAILKATGTNHHGNVSGSITMPSVQAQRDLMIETYQNAQIDPRTISFIEAHGSGNIGSDKVEILAFQQAF

AALAHEQQVVLEKQVCGIGSGNTGFLEAAGGMAQLFKVILSMQHKTLPATRNFNAVDESLNLSESPFYIVDQTKAWDAFR
DEDGQILPRRGAINAYGLGGTNAHIVIEECVQSATEESSARFEAEAPQLILLSAQSEAQLKQSAQNLCHHLTGQPQLSYC
DLQSIAYTLQVGRAVMEHRVTFIVTDIDTLQNKLSAYLSGAEVIENFTAGHAKREQALTNIFMQDEDLQALINQWVGRRK
LDRLARLWVSGVEIEWSLLYPDSTPPRVPLPTYPFARNRYWVTDSDTIPTASGNNGVQKGSSSWLESGPEMLLVEADATK
TKQKNLEAYLLRLLSQLLKVPPHQIALDKTLYEYGVDSIIGMQVRRRLEETLDVQITGREMLEHGTIHSLAALLDRKVNH
FPQAQQPEALETSPSEPADGPEEHLRQSPLHQFPLSEGQKGLWLLQKLTPEMSAYNVPIALRFQQPLDAERWQQACAHLV
THYPILQTVFYDEDGQLRQVIDPTQPLFFEIEDIGHLPETDLILYLQTRVQEPFDLAIGPLLRVQLFACAEEEYILLLTI
HHILVDGTSILPVLRTLLQIYQALAEGKDPNLAADDGTYHDFVTWEQEMLAGEEGCTYRAYWQQQLAGELPILSLPLDKP
RPPMQRFRGDTITTQLDPDVTQCLRDLAKTSQVNMASLMLAIYKVLLYRYTGQEDILVGMPTAGRPQRRFEETIGYFINM
VVVRSQLDGTIPFTEYLKQLQLIVVDALDHGAYPFPRLVAELNVPPNQTISPLFQVSFAFQNFIQPNSLAAFERQYQETL
PFEVVDGIHQVGEFDLDLEIVEVEDTCLLTMKYSPDLFDPSTIERMLGHYTKLATEVVTAPDKAISAHELLLEDEKQKIL
VEWNAPRGLRQADSLSDRFVPHFRESERGIVITRPDPRKWDAQLADKCIHQLFEAQVERTPEAIAVVFGDEKLTYRELNG
RANQLAHHLIAQGVAPKGGAGEQSSQPQHDAPQGADVLVGICVERSIEMAIGLLGILKAGGAYVPLDPVYPQERLAYMLS
DADVRVLVTQKSLVNGNIVPINNEQMTVVCLDTGWENISKQSTANPITSVQPDDLAYVIYTSGSTGMPKGVAIAHRNLVN
AYRAWEDAYQLRARASSHLQMASFSFDVFSGDFVRALCSGAKLVLCPQELLLIPEQLYELMQDEAIDCAEFVPVVLKNLI
RYLEETKQNLHFMRLLVVGSDSWDMQDYRDVHRFCGPETRLINSYGVTEATIDSCYFEGTCIDLDNARPVPIGRPFNGVQ
IYTLDSYGQPVPTGVAGELHIGGDGVAQGYLNRPELTAEKFITISLDPQANPVRLYKSGDMARWLPDGQIELLGRMDHQI
KIRGFRVEPSEVEAVLDRHPAVEEAVMVARTAQNGQLGLVAYIIPMGQRGNGSNAKADAATLRDYLKRQVPEYMVPSAFI
LLDELPLTPNGKVDRKALAAMPTDDIGHLLKQAEFVEPRTPIETALADIWRQLLHLERVSIHDNFFTVGGDSILSLQLVP
RARQVGIHLTPQQIFQHQTLAELSTVVDCEPNLQAATTGQPPVYQPWTLENQALVTGEAPLTPIQQAWDALNLPNPHIYT
QSILIEIDSQADSALLRQSLHILLRHHDALRLRYQETSSGWRQTFSPPIDDVLLREVDVTGLSESALQTTVLAEMQRVEC
TLNLTDGPLMGLALLRRIDGPSKLLWITHHLVVDWVSWRILLADLETVYQQLAHEQPVQLPTKTTSFQQWARWLLERGPS
AFADEREYWVSRFEGARPLSQDNPNGAPTWASMRVVSAELTLEETQALLQRAPAAYATQINDLLLTALLQTLGKWTQEKT
LLIELESHGREPIDHLDPDDDPVIDLSRTVGWFTAAFPVRLNHRSGDLSDQILSVKEQLGQIPNHGLGYGILHYSRETPE
LRQVPMPPICFNYLGQFDQPAENQGYTPSNDNEPTSLLRGTGMEGLTAFDALQNMRQSEFIQQNGTAQLWNLFDLELMVV
AGQLRVLWLYSENVHDRSTVQQLTDQFMTHLRQLIAHCQMRVGLELT

(Heavy metal translocating ATPase)

PksM CR 192 aa

YDNYFCCGGTQEELLNIFEGKIKFNVLPFYRILLDCEIPTISAMQGHGIGGGLAFGCYADLIVMAEECLYSANFMKYGFT
PGLGATYIIPRKFGELLGAELLYSANNFHGGKLRERGIPAKVVKKQEVIATALTLARDLADKPRTSLKLLKQHLTQTIKA
ELPTIIEQELAMHEISFSQPEVRERIETLFGG

(Aldehyde oxidase)

PksN (KS10)

MRMNDMSSSLNKPSRLVNSKSSAQTIAREAIAIVGLAGRYPQAENLREFWENLRQGRDCITEIPPERWDYRHYFDKDKNA
AGKSYSKWAGFMADVDKFDPLFFHISPHEAAMLDPQERLFLETVWAAVEDAGYSRSTLAKVRQIGVFVGVMNVDYSRIST
EAWANGSDHDERAYLWSIANRVSYIFDWRGPSLVVDTACAASLTAIHLACASIKRGECKLAVAGGVNLILHPQQHVIESK

MKMHSTQGRCRPFAQDADGIVDGEGVGAVLLRPLADAVEAGDHIYGIIKGSGINTDGRTASYMVPSFEAQAALIRTVLTA
ADIDARTISYVEAHGTGTVIGDPVEIRGLTEAFGADQIESQTCAIGSVKSNIGHLESAAGIASLTKVLLQMKYQQLVPSL
HAENLNPRIDFGATPFKVQGTLAEWKRPVVEIDGEVKEYPRRAGISSFGAGGANAHLVVEEYEGVERQSGPADPSVAKLQ
GNSDSEQPQVIVLSAKNEERLREYAHRLLVYLEQASAPNPGRQESDATVGSPTIQHTLCEMVAELLSIDIIELEVEQSFA
ECGLDAVQLSRLQTMVEERYGCELPRMLFAAGTMVKDVAQRISALTSAVADGEHRQQAPLQKGLCLASMAYTLQVGREAM
EERLAIVVSRIDELIELLRKYKQKANAIGPVYRGNIKASQETIGMLVEGTAGEAYIKVVVDNREYDKVAQLWVSGVEIDW
QLGS

Abbreviations

A domain	Adenylation domain
ACP	Acyl carrier protein
APS	Ammonium persulfate
AT	Acyltransferase
bp	Base pair
C domain	Condensation domain
CoA	Coenzyme A
CR	Crotonase
CTAB	Cetyl trimethylammonium bromide
DH	Dehydratase
dH_2O	Distilled water
DNA	Deoxyribose nucleic acid
dNTPs	Deoxynucleoside triphosphates
dTTP	2'-deoxythymidine 5'-triphosphate
GNAT	GCN5-related *N*-acetyltransferase
His-tag	Histidine-tag
HMGS	3-hydroxy-3-methylglutaryl Coenzyme A synthase
HMW-DNA	High molecular weight DNA
IC_{50}	Half maximal inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobase pair
kDa	Kilo dalton
KR	Ketoreductase
KS	Ketosynthase
min	Minute
MT	Methyltransferase
Ni-NTA	Nickel-Nitrilotriacetic acid
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthase

onn	Onnamide gene
ORF	Open reading frame
OXY	Oxygenase
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
ped	Pederin gene
PK	Polyketide
PKS	Polyketide synthase
RNA	Ribonucleic acid
RNAse	Ribonuclease
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second
T domain	Thiolation domain
TE	Thioesterase
TEMED	Tetramethylethylenediamine
TLC	Thin layer chromatography
TP	Transposase
Tris-HCl	2-Amino-2-(hydroxymethyl)-1,3-propanediol, hydrochloride
X-Gal	5-Bromo-4-chloro-3-indolyl beta-D-galactopyranoside

References

1. Faulkner, D. J., *Highlights of marine natural products chemistry (1972-1999).* Nat Prod Rep, 2000. **17**(1): p. 1-6.
2. Newman, D. J., Cragg, G. M. and Snader, K. M, *The influence of natural products upon drug discovery.* Nat Prod Rep, 2000. **17**: p. 215-234.
3. Koehn, F. E and Carter, G. T., *The evolving role of natural products in drug discovery.* Nat Rev Drug Discov, 2005. **4**: p. 206-220.
4. Van Lanen, S. G and Shen, B., *Microbial genomics for the improvement of natural product discovery.* Curr Opin Microbiol, 2006. **9**: p. 252-260.
5. Proksch, P., Edrada-Ebel, R and Ebel, R., *Drugs from the Sea - Opportunities and Obstacles.* Mar Drugs, 2003. **1**: p. 5-17.
6. Cardellina, J. H., *Marine natural products as leads to new pharmaceutical and agrochemical agents.* Pure Appl Chem, 1986. **58**: p. 365-374.
7. Simmons, T. L., Andrianasolo, E., McPhail, K., Flatt, P and Gerwick, W. H., *Marine natural products as anticancer drugs.* Mol Cancer Ther, 2005. **4**(2): p. 333-343.
8. Harvey, A., *Strategies for discovering drugs from previously unexplored natural products.* Drug Discov Today, 2000. **5**(7): p. 294-300.
9. Rajeev, K. J and Xu, Z., *Biomedical Compounds from Marine organisms.* Mar Drugs, 2004. **2**: p. 123-146.
10. Kerr, R. G and Kerr, S. S., *Marine natural products as therapeutic agents.* Expert Opinion on Therapeutic Pattern, 1999. **9**: p. 1207-1222.
11. Burres, N. S and Clement, J. J., *Antitumor Activity and Mechanism of Action of the Novel Marine Natural Products Mycalamide-A and -B and Onnamide.* Cancer Res, 1989. **49**: p. 2935-2940.
12. Donia, M and Hamann, M. T., *Marine natural products and their potential applications as anti-infective agents.* Lancet Infect Dis, 2003. **3**(6): p. 338-348.
13. Haefner, B., *Drugs from the deep: marine natural products as drug candidates.* Drug Discov Today, 2003. **8**(12): p. 536-544.
14. Faulkner, D. J., *Biomedical uses for natural marine chemicals.* Oceanus, 1992. **35**: p. 29-35.

15. König, G. M and Wright, A. D., *Marine natural products research: Current directions and future potential.* Planta Med, 1996. **62**(3): p. 193-211.
16. Thakur, N. L and Müller, W. E. G., *Biotechnological potential of marine sponges.* Current Science, 2004. **86**(11): p. 1506-1512.
17. Blunt, J. W., Copp, B. R., Hu, W. P., Munro M. H. G, Northcote, P. T and Prinsep, M. R., *Marine natural products.* Nat Prod Rep, 2009. **26**(2): p. 170-244.
18. Brusca, R. C and Brusca G. J., *Invertebrates*. 2nd ed. 2003.
19. Wulff, J. L., *Ecological interactions of marine sponges.* Can J Zool, 2006. **84**: p. 146-166.
20. Linington, R. G., Robertson, M., Gauthier, A., Finlay, B. B., van Soest, R and Anderson, R. J., *An antimicrobial glycolipid isolated from the marine sponge Caminus sphaeroconia.* Org Lett, 2002. **4**: p. 4089-4092.
21. De Silva, S. D., Williams, D. E., Andersen, R. J., Klix, H., Holmes, C. F. B and Allen, T. M., *A potent protein phosphatase inhibitor isolated from the Papau New Guinea sponge Theonella swinhoie Gray.* Tetrahedron Lett, 1992. **33**: p. 1561-1564.
22. Mayer, A. M. S and Hamann, M. T., *Marine pharmacology in 1999: compounds with antibacterial, anticoagulant, antifungal, anthelmintic, anti-inflammatory, antiplatelet, antiprotozoal and antiviral activities affecting the cardiovascular, endocrine, immune and nervous systems, and other miscellaneous mechanisms of action.* Comp Biochem Physiol C, 2002. **132**: p. 315-339.
23. Newman, D. J., Cragg, G. M and Kingston, D. G., *Natural products as sources of new drugs over the period 1981-2002.* J Nat Prod, 2003. **66**: p. 1022-1037.
24. Potts, B. C. M., Faulkner, D. J and Jacobs, R. S., *Phospholipase A2 inhibitors from marine organisms.* J Nat Prod, 1992. **55**: p. 1701-1717.
25. Shoji, N., Umeyama, A., Shin, K., Takedo, K and Ashihara, S., *Two unique pentacyclic steroids with Cis C/D ring junction from Xestospongia bergquistia Fromont, powerful inhibitors of histamine release.* J Org Chem, 1992. **57**: p. 2996-2997.
26. Burgoyne, D. L., Andersen, R. J and Allen, T. M., *Contignasterol, a highly oxygenated steroid with the unnatural 14beta configuration from the marine sponge Petrosia contignata Thiele.* J Org Chem, 1992. **57**: p. 525-528.
27. Wijffels, R. H., *Potential of sponges and microalgae for marine biotechnology.* Trends Biotechnol, 2008. **26**(1): p. 26-31.

28. Pham, N. B., Butler, M. S and Quinn, R. J., *Isolation of Psammaplin A 11'-Sulfate and Bisaprasin 11'-Sulfate from the Marine Sponge Aplysinella rhax.* J Nat Prod, 2000. **63**(3): p. 393-395.
29. Fernández, R., Dherbomez, M., Letourneux, Y., Nabil, M., Verbist, J. F and Biard, J. F., *Antifungal Metabolites from the Marine Sponge Pachastrissa sp.: New Bengamide and Bengazole Derivatives.* J Nat Prod, 1999. **62**(5): p. 678-680.
30. Dewick, P. M., *Medicinal Natural Products: A Biosynthetic Approach.* 3rd ed. 2009: John Wiley and Sons.
31. Cseke, L. J., Kirakosyan, A., Kaufman, P. B., Warber, S. L., Duke, J. A and Brielmann, H. L., *Natural products from plants*. 2nd ed. 2006: CRC Press.
32. Leeper, F. J., Vederas, J. C and Croteau, R., *Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids*. 2000: Springer.
33. Van Broekhoven, A., Shapiro, F and Anné, J., *Novel frontiers in the production of compounds for biomedical use*. 2001: Springer.
34. Staunton, J and Weissman, K. J., *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**(4): p. 380-416.
35. Piel, J., Hui, D., Fusetan, N and Matsunaga, S., *Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia.* Environ Microbiol, 2004. **6**: p. 921-927.
36. Schwarzer, D., Finking, R and Marahiel, M. A., *Nonribosomal peptides: from genes to products.* Nat Prod Rep, 2003. **20**(3): p. 275-287.
37. Toone, E. J and Purich, D. L., *Advances in Enzymology and Related Areas of Molecular Biology*. 2008: John Wiley and Sons.
38. Rawlings, B. J., *Type I polyketide biosynthesis in bacteria (Part B).* Nat Prod Rep, 2001. **18**: p. 231-281.
39. Frey, P. A and Hegeman, A. D., *Enzymatic reaction mechanisms*. 2007: Oxford University Press US.
40. Setlow, J. K., *Genetic engineering principles and methods.* Vol. 25. 2003: Springer.
41. Atta-ur-Rahman., *Bioactive natural products*. Vol. 22. 2000: Elsevier.
42. Romeo, J. T and Dixon, R. A., *Phytochemistry in the genomics and post-genomics eras*. 2002: Elsevier.

43. Piel, J., *A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles.* Proc Natl Acad Sci U S A, 2002. **99**: p. 14002-14007.
44. Cheng, Y. Q., Tang, G. L and Shen, B., *Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3149-3154.
45. Reeves, C. D., *The enzymology of combinatorial biosynthesis.* Crit Rev Biotechnol, 2003. **23**(2): p. 95-147.
46. Piel, J., Wen, G., Platzer, M and Hui, D., *Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase.* ChemBioChem, 2004. **5**: p. 93-98.
47. Nguyen, T. A., *et al.*, *Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection.* Nat Biotechnol, 2008. **26**(2): p. 225-233.
48. Finking, R and Marahiel, M. A., *Biosynthesis of nonribosomal peptides.* Annu Rev Microbiol, 2004. **58**: p. 453-488.
49. Marahiel, M. A., *Multimodular enzymes in biosynthesis of nonribosomal peptide antibiotics.* Faseb J, 2000. **14**(8): p. 550.
50. Matsunaga, S and Fusetani, N., *Nonribosomal peptides from marine sponges.* Curr Org Chem, 2003. **7**(10): p. 945-966.
51. Eppelmann, K., Stachelhaus, T and Marahiel, M. A., *Exploitation of the Selectivity-Conferring Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics.* Biochemistry, 2002. **41**: p. 9718-9726.
52. Marahiel, M. A., Stachelhaus, T and Mootz, H. D., *Modular peptide synthetases involved in nonribosomal peptide synthesis.* Chem Rev, 1997. **97**(7): p. 2651-2673.
53. Döhren, H., Keller, U., Vater, J and Zocher, R., *Multifunctional peptide synthetases.* Chem Rev, 1997. **97**: p. 2675-2705.
54. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C and Walsh, C. T., *The EntF and EntE adenylation domains of Escherichia coli enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to hiolation domain cosubstrates.* Proc Natl Acad Sci U S A, 2000. **97**(6): p. 2509-2514.
55. Arnez, J. G and Moras, D., *Structural and functional considerations of the aminoacylation reaction.* Trends Biochem Sci, 1997. **22**: p. 211-216.

56. Conti, E., Stachelhaus, T., Marahiel, M. A and Brick, P., *Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin.* EMBO J, 1997. **16**(14): p. 4174-4183.
57. Stachelhaus, T., Mootz, H. D and Marahiel, M. A., *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chem Biol, 1999. **6**(8): p. 493-505.
58. Turgay, K., Krause, M and Marahiel, M. A., *Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes.* Mol Microbiol, 1992. **6**: p. 2743-2744.
59. Paracer, S and Ahmadjian, V., *Symbiosis: an introduction to biological associations*. 2nd ed. 2000: Oxford University Press US.
60. Flachshaar, D and Piel, J., *Symbiosis*, in *Wiley Encyclopedia of Chemical Biology*. 2007, Wiley VCh.
61. Kobayashi, J and Ishibashi, M., *Bioactive Metabolites of Symbiotic Marine Microorganisms.* Chem Rev, 1993. **93**(5): p. 1753-1769.
62. Dudler, R and Eberl, L., *Interactions between bacteria and eukaryotes via small molecules.* Curr Opin Biotechnol, 2006. **17**: p. 268-273.
63. Tsuruo, T., *et al., Rhizoxin, a Macrocyclic Lactone Antibiotic, as a New Antitumor Agent against Human and Murine Tumor Cells and Their Vincristine-resistant Sublines.* Cancer Res, 1986. **46**: p. 381-385.
64. Piel, J., *Metabolites from symbiotic bacteria.* Nat Prod Rep, 2009. **26**: p. 338-362.
65. Fu, X., Do, T., Schmitz, F. J., Andrusevich, V and Engel, M. H., *New cyclic peptides from the ascidian Lissoclinum patella.* J Nat Prod, 1998. **61**(12): p. 1547-1551.
66. Salomon, C. E and Faulkner, D. J., *Localization studies of bioactive cyclic peptides in the ascidian Lissoclinum patella.* J Nat Prod, 2002. **65**(5): p. 689-692.
67. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A and Haygood, M. G., *Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella.* Proc Natl Acad Sci U S A, 2005. **102**: p. 7315-7320.
68. Schmidt, E. W., *From chemical structure to environmental biosynthetic pathways : navigating marine invertebrate-bacteria associations.* Trends Biotechnol, 2005. **23**(9): p. 437-440.

69. Long, P. F., Dunlap, W. C., Battershill, C. N and Jaspars, M., *Shotgun Cloning and Heterologous Expression of the Patellamide Gene Cluster as a Strategy to Achieving Sustained Metabolite Production.* ChemBioChem, 2005. **6**(10): p. 1760-1765.

70. Pettit, G. R., Herald, C. L., Doubek, D. L., Herald, D. L., Arnold, E and Clardy, J., *Isolation and structure of bryostatin 1.* J Am Chem Soc, 1982. **104**: p. 6846-6848.

71. Kerr, R. G., Lawry, J and Gush, K. A., *In vitro biosynthetic studies of the bryostatins, anti-cancer agents from the marine bryozoan Bugula neritina.* Tetrahedron Lett, 1996. **37**(46): p. 8305-8308.

72. Davidson, S. K., Allen, S. W., Lim, G. E., Anderson, C. M and Haygood, M. G., *Evidence for the biosynthesis of bryostatins by the bacterial symbiont "Candidatus Endobugula sertula" of the bryozoan Bugula neritina.* Appl Environ Microbiol, 2001. **67**(10): p. 4531-4537.

73. Hildebrand, M., *et al., BryA: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan Bugula neritina.* Chem Biol, 2004. **11**: p. 1543-1552.

74. Sudek, S., *et al., Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from "Candidatus Endobugula sertula", the Uncultivated Microbial Symbiont of the Marine Bryozoan Bugula neritina.* J Nat Prod, 2007. **70**(1): p. 67-74.

75. Morgan, E. D., *Biosynthesis in insects*. 2004: Royal Society of Chemistry.

76. Kellner, R. L. L and Dettner, K., *Differential efficacy of toxic pederin in deterring potential arthropod predators of Paederus (Coleoptera: Staphylinidae) offspring.* Oecologia, 1996. **107**(3): p. 293-300.

77. Clardy, J., *Using genomics to deliver natural products from symbiotic bacteria.* Genome Biol, 2005. **6**(9): p. 232.1-232.4.

78. Piel, J., Butzke, D., Fusetani, N., Hui, D., Platzer, M., Wen, G and Matsunaga, S., *Exploring the chemistry of uncultivated symbionts: Antitumor polyketides of the pederin family.* J Nat Prod, 2005. **68**: p. 472-479.

79. Lee, Y. K., Lee, J. H and Lee, H. K., *Microbial symbiosis in marine sponges.* J Microbiol, 2001. **39**(4): p. 254-264.

80. Haygood, M. G., Schmidt, E. W., Davidson, S. K and Faulkner, D. J., *Microbial Symbionts of Marine Invertebrates: Opportunities for Microbial Biotechnology.* J Mol Microbiol Biotechnol, 1999. **1**(1): p. 33-43.

81. Taylor, M. W., Radax, R., Steger, D and Wagner, M., *Sponge-Associated Microorganisms: Evolution, Ecology, and Biotechnological Potential.* Microbiol Mol Biol Rev, 2007. **71**(2): p. 295-347.

82. Vogel, G., *The Inner Lives of Sponges*, in *Science*. 2008. p. 1028-1030.

83. Fieseler, L., *et al.*, *Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges.* Appl Environ Microbiol, 2007. **73**(7): p. 2144-2155.

84. Taylor, M. W., Hill, R. T., Piel, J., Thacker, R. W and Hentschel, U., *Soaking it up: the complex lives of marine sponges and their microbial associates.* ISME J, 2007. **32**: p. 187-190.

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

86. Proksch, P and Muller, W. E. G., *Frontier in marine biotechnology*. 2006: Horizonbioscience.

87. Imhoff, J. F. *New natural porducts from marine microorganisms*. in *IFM-GEOMAR report 2002-2004*. 2004.

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

89. Beer, S and Ilan, M., *In situ measurements of photosynthetic irradiance responses of two Red Sea sponges growing under dim light conditions.* Marine Biology, 1998. **131**: p. 613-617.

90. Unson, M. D., Holland, N. D and Faulkner, D. J., *A Brominated Secondary Metabolite Synthesized by the Cyanobacterial Symbiont of a Marine Sponge and Accumulation of the Crystalline Metabolite in the Sponge Tissue.* Marine Biology, 1994. **119**(1): p. 1-11.

91. Bewley, C. A., Holland, N. D and Faulkner, D. J., *Two classes of metabolites from Theonella swinhoei are localized in distinct populations of bacterial symbionts.* Experientia, 1996. **52**(7): p. 716-722.

92. Magnino, G., Sarà, A., Lancioni, T and Gaino, E., *Endobionts of the Coral Reef Sponge Theonella swinhoei (Porifera, Demospongiae).* Invertebr Biol, 1999. **118**(3): p. 213-220.

93. Chan, G. W., *et al.*, *New leukotrine B4 receptor antagonist: leucettamine A and related imidazole alkaloids from the marine sponge Leucetta microraphis.* J Nat Prod, 1993. **56**: p. 116-121.

94. Sandler, J. S., Colin, P. L., Hooper, J. N. A and Faulkner, D. J., *Cytotoxic b-carbolines and cyclic peroxides from the Palauan sponge Plakortis nigra.* J Nat Prod, 2002. **65**: p. 1258-1261.

95. Williams, D. E., Lassota, P and Andersen, R. J., *Motuporamines A-C, Cytotoxic Alkaloids Isolated from the Marine Sponge Xestospongia exigua (Kirkpatrick).* J Org Chem, 1998. **63**: p. 4838-4841.

96. Müller, W. E. G., *Sponges (Porifera)*. 2003: Springer.

97. Müller, W. E. G., Schröder, H. C., Wiens, M., Perovic´-Ottstadt, S., Batel, R and Müller, I. M., *Traditional and Modern Biomedical Prospecting: Part II—the BenefitsApproaches for a Sustainable Exploitation of Biodiversity (Secondary Metabolites and Biomaterials from sponges).* Evid Based Complement Alternat Med, 2004. **1**(2): p. 133-144.

98. Faulkner, D. J., He, H. Y., Unson, M. D., Bewley, C. A and Garson, M. J., *New Metabolites from Marine Sponges - Are Symbionts Important.* Gazz Chim Ital, 1993. **123**(6): p. 301-307.

99. Pabel, C. T., *et al.*, *Antimicrobial Activities and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Bacillus Isolates from the Marine Sponge Aplysina aerophoba.* Mar Biotechnol, 2003. **5**(5): p. 424-434.

100. Oclarit, J. M., *et al.*, *Anti-Bacillus Substance in the Marine Sponge, Hyatella Species, Produced by an Associated Vibrio Species Bacterium.* Microbios, 1994. **78**(314): p. 7-16.

101. Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N and Matsunaga, S., *Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei.* Proc Natl Acad Sci U S A, 2004. **101**: p. 16222-16227.

102. Kocienski, P., Narquizian, R., Raubo, P., Smith, C., Farrugia, L. J., Muir, K and Boyle, F. T., *Synthetic studies on the pederin family of antitumour agents. Syntheses of mycalamide B, theopederin D and pederin.* J Chem Soc Perkin Trans 1, 2000(15): p. 2357-2384.

103. Brega, A., Falaschi, A., Carli, L and Pavan, M., *Studies on the mechanism of action of pederine.* J Cell Biol, 1968. **36**: p. 485-496.

104. Perry, N. B., Blunt, J. W., Munro, M. H. G and Pannell, L. K., *Mycalamide A, an antiviral compound from a New Zealand sponge of the genus Mycale.* J Am Chem Soc, 1988. **110**(14): p. 4850-4851.

105. Sakemi, S., Ichiba, T., Kohmoto, S., Saucy, G and Higa, T., *Isolation and structure elucidation of onnamide A, a new bioactive metabolite of a marine sponge, Theonella sp.* J Am Chem Soc, 1988. **110**(14): p. 4851-4853.

106. Cichewicz, R.H., Valeriote, F. A and Crews, P., *Psymberin, a potent sponge-derived cytotoxin from Psammocinia distantly related to the pederin family.* Org Lett, 2004. **6**(12): p. 1951-1954.

107. Abell, A. D., Blunt, J. W., Foulds, G. J and Munro, M. H. G., *Chemistry of the mycalamides: Antiviral and antitumour compounds from a New Zealand marine sponge .6. The synthesis and testing of analogues of the C(7)-C(10) fragment.* J Chem Soc Perkin Trans 1, 1997. **11**: p. 1647-1654.

108. Abbasipour, H and Taghavi, A. *Study of Rove beetles, Paederus spp., the agents of skin dermatitis in north of Iran*. in *The Fifth International Conference on Urban Pests*. 2005. Malaysia.

109. Lopanik, N., Lindquist, N and Targett, N., *Potent cytotoxins produced by a microbial symbiont protect host larvae from predation.* Oecologia, 2004. **139**(1): p. 131-139.

110. Kellner, R. L. L., *What is the basis of pederin polymorphism in Paederus riparius rove beetles? The endosymbiotic hypothesis.* Entomol Exp Appl, 1999. **93**(1): p. 41-49.

111. Piel, J., Höfer, I and Hui, D., *Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of Paederus fuscipes beetles.* J Bacteriol, 2004. **186**: p. 1280-1286.

112. Kellner, R. L. L., *Molecular identification of an endosymbiotic bacterium associated with pederin biosynthesis in Paederus sabaeus (Coleoptera: Staphylinidae).* Insect Biochem Mol Biol, 2002. **32**(4): p. 389-395.

113. Bull, A. T., *Microbial Diversity and Bioprospecting*. 2004: The American Society for Microbiology Press.

114. *Discalamide compounds and their use as anti-proliferative agents.* US Patent 6476065, 2002.

115. Richter, A., Kocienski, P., Raubo, P and Davies, D. E., *The in vitro biological activities of synthetic 18-O-methyl mycalamide B, 10-epi-18-O-methyl mycalamide B and pederin.* Anti-Cancer Drug Des, 1997. **12**(3): p. 217-227.

116. Lee, K. H., Nishimura, S., Matsunaga, S., Fusetani, N., Horinouchi, S and Yoshida, M., *Inhibition of protein synthesis and activation of stress-activated protein kinases by*

onnamide A and theopederin B, antitumor marine natural products. Cancer Sci, 2005. **96**(6): p. 357-364.

117. Hart, J. B., Lill, R. E., Hickford, S. J. H., Blunt, J. W and Munro, M. H. G., *The halichondrins: Chemistry, biology, supply and delivery*. Drugs from the sea. 2000. 134-153.

118. Kaeberlein, T., Lewis, K and Epstein, S. S., *Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment.* Science, 2002. **296**: p. 1127-1129.

119. Hugenholtz, P and Pace, N. R., *Identifying microbial diversity in natural environment: a molecular phylogenetic approach.* Trends Biotechnol, 1996. **14**: p. 190-197.

120. Nobutada, K., *Metagenomics: Access to Unculturable Microbes in the Environment.* Microbes Environ, 2006. **21**(4): p. 210-215.

121. Rappé, M. S and Giovannoni, S. J., *The uncultured microbial majority.* Annu Rev Microbiol, 2003. **57**: p. 369-394.

122. Lorenz, P., Liebeton, K., Niehaus, F and Eck, J., *Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space.* Curr Opin Biotechnol, 2002. **13**(6): p. 572-577.

123. Gillespie, D. E., *et al., Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA.* Appl Environ Microbiol, 2002. **68**(9): p. 4301-4306.

124. Courtois, S., *et al., Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products.* Appl Environ Microbiol, 2003. **69**: p. 49-55.

125. Pace, N. R., Stahl, D. A., Lane, D. J and Olsen, G. J., *The analysis of natural microbial populations by ribosomal RNA sequences.* Adv Microb Ecol, 1986. **9**: p. 1-55

126. Handelsman, J., *Metagenomics: Application of Genomics to Uncultured Microorganisms.* Microbiol. Mol. Biol. Rev., 2004. **68**(4): p. 669-685.

127. Brady, S. F., Chao, C. J., Handelsman, J and Clardy, J., *Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA.* Org Lett, 2001. **3**(13): p. 1981-1984.

128. Hildebrand, M., Waggoner, L. E., Lim, G. E., Sharp, K. H., Ridley, C. P and Haygood, M. G., *Approaches to identify, clone, and express symbiont bioactive metabolite genes.* Nat Prod Rep, 2004. **21**(1): p. 122-142.

129. Handelsman, J., Liles, M., Mann, D., Riesenfeld, C and Goodman, R. M., *Cloning the metagenome: Culture-independent access to the diversity and functions of the uncultivated microbial world*, in *Functional Microbial Genomics*. 2002. p. 241-255.

130. Schloss, P. D and Handelsman, J., *Biotechnological prospects from metagenomics.* Curr Opin Biotechnol, 2003. **14**: p. 303-310.

131. Wenzel, S. C and Müller, R., *Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways.* Curr Opin Biotechnol, 2005. **16**: p. 594-606.

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

133. Corton, J. C., Ward, J. E. Jr and Lutkenhaus, J., *Analysis of cell division gene ftsZ (sulB) from Gram-negative and Gram-positive bacteria.* J Bacteriol, 1987. **169**: p. 1-7.

134. Kelly, T. M., Padmalayam, I and Baumstark, B. R., *Use of the Cell Division Protein FtsZ as a Means of Differentiating among Bartonella Species.* Clin Diagn Lab Immunol, 1998. **5**(6): p. 766-772.

135. *Bacteriologists: Webster's Quotations, Facts and Phrases,* ed. I.G. International. 2008: ICON Group International, Inc.

136. Jain, R., Rivera, M. C and Lake, J. A., *Horizontal gene transfer among genomes: The complexity hypothesis.* Proc Natl Acad Sci U S A, 1999. **96**: p. 3801-3806.

137. Yunt, Z., *et al.*, *Cleavage of four carbon-carbon bonds during biosynthesis of the griseorhodin a spiroketal pharmacophore.* J Am Chem Soc, 2009. **131**(6): p. 2297-2305.

138. Kowtoniuk, W. E., Calderone, C. T., Kelleher, N. L., Walsh, C. T and Dorrestein, P. C., *Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8977-8982.

139. Schroeder, F. C., Butcher, R. A., Fischbach, M. A., Straight, P. D., Kolter, R., Walsh, C. T and Clardy, J., *The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1506-1509.

140. Matsunaga, S., Fusetani, N and Nakao, Y., *8 new cytotoxic metabolites closely related to onnamide A from 2 marine sponges of the genus Theonella.* Tetrahedron, 1992. **48**(39): p. 8369-8376.

141. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A and Gerwick, W. H., *Structure and biosynthesis of the jamaicamides, new mixed*

polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. Chem Biol, 2004. **11**(6): p. 817-833.

142. Qiagen, *The QIAexpressionist™ A handbook for high-level expression and purification of 6xHis-tagged proteins.* 5th ed. 2003.
143. Sambrook, J and Russell, D., *Molecular Cloning: A Laboratory Manual*. 3rd ed. 2001: CSHL Press.
144. *Chaperone Competent Cell BL21 Series* http://www.takara-bio.com.
145. Moldenhauer, J., Chen, X. H., Borriss, R and Piel, J., *Biosynthesis of the Antibiotic Bacillaene, the Product of a Giant Polyketide Synthase Complex of the trans-AT Family.* Angew Chem Int Ed Engl, 2007. **46**(43): p. 8195-8197.
146. Magarvey, N. A., Ehling-Schulz, M and Walsh, C. T., *Characterization of the Cereulide NRPS α-Hydroxy Acid Specifying Modules: Activation of α-Keto Acids and Chiral Reduction on the Assembly Line.* J Am Chem Soc, 2006. **128**(33): p. 10698-10699.
147. Lutz, S and Bornscheuer, U. T., *Protein Engineering Handbook*. 2009: Wiley-VCH.
148. Tillett, D., Dittmann, E., Erhard, M., Döhren, H., Börner, T and Neilan, B. A., *Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system.* Chem Biol, 2000. **7**(10): p. 753-764.
149. Michael, B. G., *Nucleic acids in chemistry and biology*. 3rd ed. 2006: Royal Society of Chemistry.
150. Turner, P. C., McLennan, A., Bates, A and White, M., *Molecular biology*. 3rd ed. 2005: Taylor & Francis.
151. Yarnell, A., *Enzymes responsible for installing halogen atoms in natural products with exquisite selectivity reveal their catalytic logic.* Chem Eng News 2006. **84**(21): p. 12-18
152. Vaillancourt, F.H., Yeh, E., Vosburg, D. A., Tsodikova, S. G and Walsh, C. T., *Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate.* Chem. Rev., 2006. **106**(8): p. 3364-3378.
153. Dong, C., Kotzsch, A., Dorward, M., van Pée, K. H and Naismith, J. H., *Crystallization and X-ray diffraction of a halogenating enzyme, tryptophan 7-halogenase, from Pseudomonas fluorescens.* Acta Crystallogr D Biol Crystallogr, 2004. **D60**: p. 1438-1440.
154. Murphy, C. D., *A Review - New frontiers in biological halogenation.* J Appl Microbiol, 2003. **94**: p. 539-548.
155. Gribble, G. W., *Naturally occurring organohalogen compounds.* Acc Chem Res, 1998. **31**: p. 141-152.

156. Wynands, I and van Pée, K. H., *A novel halogenase gene from the pentachloropseudilin producer Actinoplanes sp. ATCC 33002 and detection of in vitro halogenase activity.* FEMS Microbiol Lett, 2004. **237**: p. 363-367.

157. Van Pée, K. H., *Microbial biosynthesis of halometabolites.* Arch Microbiol, 2001. **175**: p. 250-258.

158. Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Karunakaran, P and Valla, S., *Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria.* Plasmid, 1997. **38**(1): p. 35-51.

159. Timmis, K. N., *Pseudomonas putida: a cosmopolitan opportunist par excellence.* Environ Microbiol, 2002. **4**(12): p. 779-781.

160. Ballerstedt, H., *et al., Genomotyping of Pseudomonas putida strains using P. putida KT2440-based high-density DNA microarrays: implications for transcriptomics studies.* Appl Microbiol Biotechnol, 2007. **75**(5): p. 1133-1142.

161. Nelson, K. E., *et al., Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440.* Environ Microbiol, 2002. **4**: p. 799-808.

162. Martinez, A., Kolvek, S. J., Yip, C. L., Hopke, J., Brown, K. A., MacNeil, I. A and Osburne, M. S., *Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts.* Appl Environ Microbiol, 2004. **70**(4): p. 2452-2463.

163. Wenzel, S. C., Gross, F., Zhang, Y., Fu, J., Stewart, A. F and Müller, R., *Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering.* Chem Biol, 2005. **12**: p. 349-356.

164. Rocky, S. T., *Recombinant gene expression protocols.* 2nd ed. 1997: Humana Press.

165. Gust, B., Challis, G. L., Fowler, K., Kieser, T and Chater, K. F., *PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1541-1546.

166. Tropp, B. E and Freifelder, D., *Molecular Biology: Genes to Proteins*. 3rd ed. 2007: Jones & Bartlett Publishers.

167. Hansen, M. C., Nielsen, A. K., Molin, S., Hammer, K and Kilstrup, M., *Changes in rRNA Levels during Stress Invalidates Results from mRNA Blotting: Fluorescence In Situ rRNA*

Hybridization Permits Renormalization for Estimation of Cellular mRNA levels. J Bacteriol, 2001. **183**(16): p. 4747-4751.

168. Hillis, D. M and Dixon, M. T., *Ribosomal DNA: Molecular Evolution and Phylogenetic Inference.* Q Rev Biol, 1991. **66**(4): p. 411-453.
169. Rastogi, S. C., *Cell And Molecular Biology*. 2003: New Age International.
170. Wu, W., Welsh, M. J and Zhang, H. H., *Gene biotechnology*. 2nd ed. 2003: CRC Press.
171. Fuchs, J and Podda, M., *Encyclopedia of Medical Genomics and Proteomics*. 2004: CRC Press.
172. Stratagene, *Manual: XL1-Blue Electroporation-Competent Cells*. 2004.
173. Hanahan, D., *Studies on transformation of Escherichia coli with plasmids* J Mol Biol, 1983. **166**(4): p. 557-580
174. Chang, D. C., Chassy, B. M., Saunders, J. A and Sowers, A. E., *Guide to electroporation and electrofusion*. 1992: Academic Press.
175. Shmaefsky, B., *Biotechnology 101*. 2006: Greenwood Publishing Group.
176. Goeddel, D. V., *Gene Expression Technology*. Vol. 185. 1990: Academic Press.
177. Demain, A. L., Davies, J. E and Atlas, R. M., *Manual of Industrial Microbiology and Biotechnology*. 2nd ed. 1999: ASM Press.
178. Ratledge, C and Kristiansen, B., *Basic biotechnology*. 2nd ed. 2001: Cambridge University Press.
179. Rapley, R., *The Nucleic Acid Protocols Handbook*. 2000: Humana Press.
180. Trun, N. J and Trempy, J. E., *Fundamental bacterial genetics*. 2004: Wiley-Blackwell.
181. McNally, E. J and Hastedt, J. E., *Protein formulation and delivery*. 2nd ed. 2007: CRC Press.
182. Scheppler, J. A., Cassin, P. E and Gambier, R. M., *Biotechnology explorations: applying the fundamentals*. 2000: ASM Press.
183. Allison, J. H., Agrawalb, H. C and Moore, B. W., *Effect of N,N,N′,N′-tetramethylethylenediamine on the migration of proteins in SDS polyacrylamide gels.* Anal Biochem, 1974. **58**(2): p. 592-601.
184. Scopes, R. K., *Protein purification: principles and practice*. 3rd ed. 1994: Springer.
185. Farrell, R. E., *RNA methodologies: a laboratory guide for isolation and characterization*. 3rd ed. 2005: Academic Press.
186. Wall, P. E., *Thin-layer chromatography: a modern practical approach*. 2005: Royal Society of Chemistry.

187. Jork, H., *Thin-layer Chromatography: Reagents and Detection Methods*. 1990: VCH.

RESEARCH PROJECTS

1. Tu Anh, N., *et al*, Department of Microbiology (2005-2006) Set up the Biotechnology Center of School of Pharmacy. Funding from Ministry of Health, Vietnam
2. Tu Anh, N.**,** Dong, T. C., Lan, H. T. N., Tam, N. K. M., Mai, V. T., Linh, H. T. Y., Thanh, N. V. (2002-2006) Study on lactic bacteria in order to produce probiotic products which prevent antibiotic associated diarrhea. Funding from Ministry of Health, Vietnam
3. Tu Anh, N., Thuy, N. T. T., Thanh, N. V (2004-2005) Research on the combination of probiotic microorganisms. Funding from School of Pharmacy
4. Tu Anh, N.**,** Binh, N. T. M., Thanh, N. V (2003-2004) Optimization of the freeze-drying conditions of *Lactobacillus casei* and preservation of the freeze-dried bacteria. Funding from School of Pharmacy
5. Tu Anh, N., Hai, T. M. (2003-2004) Optimization of the freeze-drying conditions of *Saccharomyces boulardii* and test on its ability to prevent antibiotic associated diarrhea. Funding from Ho Chi Minh City University of Natural Sciences
6. Tu Anh, N.**,** Linh, L. T. M., Thanh, N. V. (2002-2003). Optimization of the cultivation conditions of *Lactobacillus casei* and test on its ability to prevent antibiotic associated diarrhea. Funding from School of Pharmacy

PUBLICATIONS

1. Tu Anh Nguyen, Keishi Ishida, Holger Jenke-Kodama, Elke Dittmann, Cristian Gurgui, Thomas Hochmuth, Stefan Taudien, Matthias Platzer, Christian Hertweck & Joern Piel, Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection**.** *Nature Biotechnology*, 2008. **26**(2): p. 225-233
2. Tu Anh Nguyen, Thuy Nguyen Thi Thu, Thanh Nguyen Van, Research on the combination of some probiotic microorganisms. *Proceeding of The Fourth Indochina Conference on Pharmaceutical Sciences* 2005
3. Tu Anh Nguyen, Linh Le Thi My, Thanh Nguyen Van, Optimization of the cultivation conditions of *Lactobacillus casei* and test on its ability to prevent antibiotics associated diarrhea. *Medicine of HCMC*, 2003. **7**(4): p. 195-197

AWARDS

2003	First Prize at National Conference of Science & Technology of Youth of Universities in Medicine & Pharmacy.
2003	Third Prize at National Conference of Science & Technology of Youth of Universities in Medicine & Pharmacy.