

UNIVERSITA' DEGLI STUDI DI NAPOLI
"FEDERICO II"



DOTTORATO DI RICERCA IN "SCIENZA DEL FARMACO"
XXV CICLO 2010-2013

*Exploiting the potential of marine natural products:
structure elucidation and metagenomic approaches
to biotechnological production*

Dott. Gerardo Della Sala

Tutor
Prof.ssa V. Costantino

Coordinatore
Prof.ssa M. V. D'Auria

CONTENTS

ABSTRACT	I
ABSTRACT	VIII
INTRODUCTION	XVII
PART I	
CHAPTER 1	1
PORIFERA	1
<i>References</i>	6
CHAPTER 2	8
SPONGE-MICROBE ASSOCIATIONS	8
2.1 Diversity of uncultured bacteria associated with the marine sponge <i>P.simplex</i>	10
2.2 Physiology of Sponge-Associated Microorganisms	15
2.3 Cellular localization of metabolites from <i>P.simplex</i>	17
<i>References</i>	20
CHAPTER 3	23
PLAKORTIN	23
3.1 Polyketide synthases	28
3.2 Proposed biogenetic pathway of plakortin	34
<i>References</i>	38
CHAPTER 4	40
SWF, A NEW GROUP OF MONO-MODULAR TYPE-I PKS FROM SPONGE SYMBIONTS	40
4.1 Isolation of the clusters	41
4.2 <i>In silico</i> analysis of cluster sequences	45
4.2.1 Organization and genomic contexts of the <i>swf</i> gene clusters	45
4.2.2 Analysis of <i>swf</i> operons	47
4.3 Widespread diffusion of <i>swf</i> cluster	52

4.4 Functional study of the Swf enzymes	54
4.4.1 <i>In vivo</i> phosphopantetheinylation of SwfA _{ACP}	54
4.4.2 Heterologous expression and LC-MS analysis	56
4.4.3 Heterologous expression and fatty acid analysis	57
4.4.4 Heterologous expression of SwfA and SwfB	58
<i>References</i>	63
CHAPTER 5	67
DIVERSITY OF POLYKETIDE SYNTHASES GENES FROM <i>P. SIMPLEX</i>	67
5.1 454 pyrosequencing	67
5.2 Phylogenetic analysis of KS amplicons	69
5.3 Phylogenetic analysis of AT amplicons	78
<i>References</i>	81
CHAPTER 6	85
MATERIALS AND METHODS	85
6.1 Materials	85
6.1.1 Sponge collection	85
6.1.2 Chemicals	85
6.1.3 Chemical solutions	87
6.1.4 Enzymes	88
6.1.5 Vectors and Bacteria	88
6.1.5.1 Vectors	88
6.1.5.2 Bacteria	88
6.1.6 Antibiotics	89
6.1.7 Bacterial cultivation media	89
6.1.8 Others	89
6.1.9 Equipment	89
6.2 Methods	91
6.2.1 Isolation of sponge metagenomic DNA	91
6.2.2 DNA amplification by PCR	91
6.2.2.1 Standard PCR	91

6.2.2.2 Touchdown PCR	92
6.2.2.3 Gradient PCR	93
6.2.2.4 Colony PCR	93
6.2.2.5 Amplification of 16S gene products from <i>P. simplex</i>	93
6.2.2.6 PCR screening of metagenomic DNA for type I PKSs	94
6.2.3 Analysis of nucleic acids by gel electrophoresis	95
6.2.4 Recovery of DNA fragments from agarose gel	95
6.2.5 TA cloning	97
6.2.6 Phosphatase treatment of the linear vector	98
6.2.7 Preparation of electrocompetent <i>E. coli</i> XL1-blue cells	99
6.2.8 Transformation by electroporation	100
6.2.9 Blue white screening	100
6.2.10 Isolation of plasmid DNA	102
6.2.11 Construction of a fosmid library from the metagenome of <i>P. simplex</i>	103
6.2.12 The semi-liquid library method	109
6.2.13 Creation of fosmid library superpools	110
6.2.14 Screening and isolation of positive clones from a fosmid library	110
6.2.15 Induction of high copy number plasmid DNA	113
6.2.16 Overexpression of the SwfA _{ACP} protein	113
6.2.16.1 Cloning of the swfA _{ACP} domain	114
6.2.16.2 Ligation of the amplified ACP fragment with the vector pHIS8	116
6.2.16.3 Transformation of the recombinant DNA into XL1-blue	116
6.2.16.4 Transformation of the recombinant DNA into BL21 (DE3) RIPL	117
6.2.16.5 Protein expression and purification	117
6.2.16.6 Analysis of the expressed proteins by SDS PAGE	118
6.2.16.7 FTMS analysis of overexpressed proteins	121
6.2.17 Heterologous expression	121
6.2.17.1 Digestion of the expression vector pHIS8-Svp	122
6.2.17.2 Amplification of pPS2D9 fragments	122
6.2.17.3 Flanking of the digested pHIS8-Svp	123
6.2.17.4 Linearization of the flanked vector pHIS8-Svp	123

6.2.17.5 Recombination of the fosmid pPS2D9	124
6.2.17.6 Verification of the recombinant pGS38.....	127
6.2.17.7 Heterologous expression trials in <i>E. coli</i> BAP1.....	129
6.2.17.8 LC-HR-ESI-MS analyses	130
6.2.17.9 Lipid analysis	131
6.2.17.10 GC-MS analyses.....	131
6.2.18 Protein expression of SwfA and SwfB	132
6.2.18.1 Cloning of <i>swfA</i> and <i>swfB</i> genes	132
6.2.18.2 Verification of the recombinant pGS40.....	133
6.2.18.3 Expression of SwfA and SwfB proteins	134
6.2.19 Fosmid sequencing.....	136
6.2.20 Plasmid sequencing.....	137
6.2.21 Bioinformatics.....	137
6.2.22 454 sequencing of KS and AT amplicons from <i>P. simplex</i>	137
6.2.23 Accession codes	138

PART II

CHAPTER 7..... 142

ISOLATION PROCEDURES AND STRUCTURAL DETERMINATION METHODS..... 142

7.1 Isolation procedures

7.2 Structural elucidation

7.2.1 Mass spectrometry

7.2.2 Nuclear magnetic resonance.....

7.2.3 Circular dichroism.....

CHAPTER 8..... 152

CHALINULASTEROL, A CHLORINATED STEROID DISULFATE FROM *C. MOLITBA*..... 152

8.1 Sterols

8.2 Structure elucidation of chalinulasterol

8.3 Evaluation of chalinulasterol as PXR receptor modulator

8.4 Experimental Section

8.4.1 General experimental procedures

8.4.2 Collection, extraction, and isolation	164
8.4.3 Chalinulasterol	165
8.4.4 Cell culture	165
8.4.5 Transactivation experiments	165
8.4.6 Real-time PCR	166
8.4.7 Statistical analysis	166
8.4.8 Mass spectra	167
8.4.9 NMR data	168
<i>References</i>	172
CHAPTER 9	175
ANALYSIS OF THE SPONGE <i>PLAKORTIS</i> CF. <i>LITA</i>	175
9.1 Bacteriohopanoids	176
9.1.1 Biosynthesis of hopanoids	176
9.1.2 Function of hopanoids	177
9.1.3 Biohopanoids and Geohopanoids	178
9.2 Isolation and Structural determination of plakohopanoid	179
9.3 Hopanoids from <i>Plakortis</i> specimens	186
9.4 Experimental Section	187
9.4.1 General methods	187
9.4.2 Animal material	188
9.4.3 Extraction and isolation	188
9.4.4 Plakohopanoid peracetate	189
9.4.5 HPLC-ESIMS analysis	189
9.4.6 Degradation analysis of plakohopanoid peracetate	190
9.4.7 Methyl (22R)-33,34,35-trinorbacteriohopan-32-oate	191
9.5 Mass spectra	192
9.6 NMR data	193
<i>References</i>	200

LIST OF FIGURES

CHAPTER 1

Fig. 1.1 Sponge body structure.....	2
Fig. 1.2 Sponge classification.....	4
Fig. 1.3 The Sponges studied in this PhD Thesis	5

CHAPTER 2

Fig. 2.1 Neighbor joining tree of symbionts associated with <i>P. simplex</i>	15
Fig. 2.2 Transmission electron micrograph of the <i>P. simplex</i> mesohyl	17
Fig. 2.3 Cellular localization of secondary metabolites from <i>P. simplex</i>	18

CHAPTER 3

Fig 3.1 Pathway proposed for activation of artemisinin by ferrous ion	24
Fig. 3.2 Pathway proposed for activation of artemisinin by ferrous heme.....	25
Fig. 3.3 Plakortin.....	26
Fig. 3.4 a) Dihydroplakortin.....	26
Fig. 3.4 b) Plakortide E	26
Fig. 3.5 Deoxyeritronolide B synthase, a <i>cis</i> -AT type I modular PKS	29
Fig. 3.6 Doxorubicin type II PKS.....	30
Fig. 3.7 Chalcone synthetic pathway	30
Fig. 3.8 Macrolactin synthase, a <i>trans</i> -AT PKS	33
Fig. 3.9 Biogenetic dissection of plakortin	34
Fig. 3.10 Proposed biogenetic pathway of plakortin	35

CHAPTER 4

Fig. 4.1 Genomic contexts of the <i>swf</i> gene clusters	43
Fig. 4.2 True to scale sketches of the domain organization of the <i>swf</i> gene clusters	48
Fig. 4.3 UPGMA tree of KS-AT domains from different types of PKS and FAS	48
Fig. 4.4 Neighbor joining tree of full length KS domains	50
Fig. 4.5 Neighbor joining tree of radical SAM enzymes	52
Fig. 4.6 PCR detection of <i>swf</i> genes in “HMA” sponges.....	53
Fig. 4.7 ESI MS spectra of <i>apo</i> -SwfA _{ACP} and <i>holo</i> -SwfA _{ACP}	55
Fig. S1 Alignment of KS domains from type I PKSs and FASs.....	60
Fig. S2 Alignment of AT domains from type I PKSs and FASs	62

CHAPTER 5

Fig. 5.1 Neighbor joining tree of full length KS domains	73
Fig. 5.2 Neighbor joining tree of full length KS domains of known function	74
Fig. 5.3 Neighbor joining tree of AT amplicons from <i>P. simplex</i> metagenome	80

CHAPTER 6

Fig. 6.1 Metagenomic DNA library construction.....	104
Fig. 6.2 Fosmid library from the metagenomic DNA of <i>P. simplex</i>	109
Fig. 6.3 Flowchart for isolating positive clones from a complex fosmid library	110
Fig. 6.4 Semi-liquid cultures.....	112
Fig. 6.5 PCR screening of semi-liquid pools.....	112
Fig. 6.6 Diagram of protein expression.....	114
Fig. 6.7 Vector map of pHIS8-Svp	115
Fig. 6.8 Mini SDS-PAGE of apo-SwfA _{ACP} and holo-SwfA _{ACP}	120
Fig. 6.9 Homologous recombination of the <i>swf</i> gene cluster.....	122
Fig. 6.10 Agarose gel electrophoresis of checking pGS38.....	129
Fig. 6.11 PCR verification of the recombinant plasmid pGS40	133
Fig. 6.12 Mini SDS-PAGE of SwfA.....	135

CHAPTER 7

Fig. 7.1 Secondary metabolite isolation scheme	144
--	-----

CHAPTER 8

Fig. 8.1 Structures of chalinulasterol (1) and of solomonsterol A (2) and B (3)	153
Fig. 8.2 MEV and MEP pathways for IPP biosynthesis.....	154
Fig. 8.3 <i>De novo</i> biosynthesis of sterols.....	155
Fig. 8.4 Halistanol sulfate	156
Fig. 8.5 Weinbersterol disulfate A and B.....	157
Fig. 8.6 Contignasterol.....	158
Fig. 8.7 a) Theonellasterol B.....	158
Fig. 8.7 b) Conicasterol B	158
Fig. 8.8 Chalinulasterol. Positive-ion ESI MS/MS spectrum	159
Fig. 8.9 ¹ H-NMR spectrum of chalinulasterol (CD ₃ OD, 700 MHz).....	160
Fig. 8.10 a) Selected COSY and HMBC correlations of compound 1	160

Fig. 8.10 b) Selected ROESY correlations detected for 1.....	160
Fig. 8.11 Luciferase reporter assay.....	163
Fig. 8.12 Negative-ion ESI mass spectrum of chalinulasterol	167
Fig. 8.13 Positive-ion ESI mass spectrum of chalinulasterol.....	167
Fig. 8.14 COSY spectrum of chalinulasterol (CD ₃ OD, 700 MHz).....	169
Fig. 8.15 HSQC spectrum of chalinulasterol – high field region.....	169
Fig. 8.16 HSQC spectrum of chalinulasterol – low field region.....	170
Fig. 8.17 HMBC spectrum of chalinulasterol.....	170
Fig. 8.18 ROESY spectrum of chalinulasterol	171

CHAPTER 9

Fig. 9.1 Chemical structures of some metabolites from <i>P. simplex</i>	175
Fig. 9.2 Biosynthesis of the hopane skeleton by SHC (squalene-hopene cyclase)	177
Fig. 9.3 Proposed biogenetic pathway to bacteriohopanoids	177
Fig. 9.4 1a) Bacteriohopanetetrol	180
Fig. 9.4 1b) Bacteriohopanetetrol tetraacetate	180
Fig. 9.4 1c) 12-methylbacteriohopanetetrol.....	180
Fig. 9.4 2) 32,35-anhydro-bacteriohopanetetrol	180
Fig. 9.4 3a) Plakohopanoid	180
Fig. 9.4 3b) Plakohopanoid peracetate	180
Fig. 9.4 3c) Plakohopanoid peracetylated with (CD ₃ CO) ₂ O.....	180
Fig. 9.4 4a) Discoside.....	180
Fig. 9.4 4b) Discoside peracetate	180
Fig. 9.5 HPLC-ESIMS of the crude glycolipid fraction of <i>P. cf. lita</i>	181
Fig. 9.6 ¹ H-NMR spectrum of plakohopanoid peracetate (CDCl ₃).....	182
Fig. 9.7 Partial structure of 3b determined by long range proton carbon couplings	182
Fig. 9.8 Degradation procedure of plakohopanoid peracetate	186
Fig. 9.9 HR-ESI-MS spectrum of plakohopanoid peracetate	192
Fig. 9.10 EI-MS spectrum of hopanoic acid methyl ester	192
Fig. 9.11 ¹³ C-NMR spectrum of plakohopanoid peracetate (CDCl ₃).....	194
Fig. 9.12 COSY spectrum of plakohopanoid peracetate (CDCl ₃).....	194
Fig. 9.13 HSQC spectrum of plakohopanoid peracetate (CDCl ₃)-high field region	195

Fig. 9.14 HSQC spectrum of plakohopanoid peracetate (CDCl ₃)-low field region	195
Fig. 9.15 HMBC spectrum of plakohopanoid peracetate (CDCl ₃)	196
Fig. 9.16 HMBC spectrum of 3b (CDCl ₃) - methyl proton correlations	196
Fig. 9.17 ¹ H-NMR spectrum of plakohopanoid peracetate (C ₆ D ₆)	197
Fig. 9.18 COSY spectrum of plakohopanoid peracetate (C ₆ D ₆)	197
Fig. 9.19 HSQC spectrum of plakohopanoid peracetate (C ₆ D ₆)-high field region	198
Fig. 9.20 HSQC spectrum of plakohopanoid peracetate (C ₆ D ₆)-low field region	198
Fig. 9.21 HMBC spectrum of plakohopanoid peracetate (C ₆ D ₆)	199

LIST OF TABLES

Table 3.1 <i>In vitro</i> antiplasmodial activity against D10 and W2	28
Table 4.1 Putative genes identified on the genomic fragment pPSA11D7	46
Table 4.2 Putative genes identified on the genomic fragment pPS11G3	47
Table 6.1 Chemicals	85
Table 6.2 Chemical solution	87
Table 6.3 Enzymes	88
Table 6.4 Vectors	88
Table 6.5 Bacteria	88
Table 6.6 Antibiotics	89
Table 6.7 Equipment	89
Table 6.8 Composition of the stacking and resolving gels for SwfA _{ACP} SDS-PAGE ...	120
Table 6.9 Composition of the stacking and resolving gels for SwfA SDS-PAGE	136
Table 8.1 ¹ H (700 MHz) and ¹³ C NMR data of chalinulasterol in CD ₃ OD	168
Table 9.1 NMR spectroscopic data for plakohopanoid peracetate	193

ABSTRACT

Sponges represent the most prolific producers of novel marine bioactive secondary metabolites. In the last years, several drugs derived from marine natural products have appeared in the market, and others are in clinical trials. The aim of my research project was to exploit the unusual and often surprising chemistry of marine sponges, in the frame of the more general purpose of discovering and developing new drugs from natural products.

The research work presented in this PhD Thesis was directed to two different aspects of the study of marine secondary metabolites. On one hand, in parallel with the advent of environmental genomics from a drug discovery perspective, the largest part of my research activity focused on the metagenomic analysis of the Caribbean sponge *P. simplex*, and was aimed at the identification of new genes coding for polyketide synthases (PKSs), the giant enzyme complexes that produce polyketides, a large class of secondary metabolites that include many antibiotic and antitumor compounds.

On the other hand, the remaining part of the research described in this PhD Thesis was more related to the “core activity” of natural product chemistry, and directed to the isolation and structure elucidation of new bioactive compounds from different specimens of sponges living in tropical oceans, wonderful sources of unusual molecular architectures to be used as leads and scaffolds for the elaboration of new drugs.

Metagenomic investigations on *Plakortis simplex* (Demospongiae, Homosclerophorida, Plakinidae) was started because the sponge is known

for the production of large amounts of polyketide peroxides, of which plakortin is the most abundant. Plakortin is of special interest due to its anti-malarial activity, which is retained also against chloroquine-resistant strains of *Plasmodium falciparum*. Therefore, a study of the biosynthesis of plakortin was undertaken, with the final aim of its biotechnological production. Most non-aromatic polyketides are synthesized by type I polyketide synthases (type I PKSs), produced in a number of cases by bacterial symbionts. The bacterial origin of plakortin is therefore a reasonable hypothesis, and indeed cell fractionation of *P. simplex* has shown that plakortin is mainly present in the bacterial cells. Since cultivation of true sponge symbionts failed in most cases, the search for the plakortin genes had to rely on cultivation-independent techniques, such as the study of the sponge metagenome (collective genome of the sponge and its symbionts). While the putative genes implied in plakortin biosynthesis could not be identified, an unexpected result from the metagenomic library screening was the discovery of Swf, a new group of mono-modular type I PKS/FAS (“hybrid polyketide synthase/fatty acid synthase”), which appears to be specifically associated to sponge symbionts.

The putative *swf* operon consists of *swfA* (FAS/PKS I), *swfB* (R and ST domains), and *swfC* (radical SAM). SwfA contains a single PKS module, which builds the backbone of the acyl chain by recruiting iteratively malonyl units according to the substrate determining motif of its AT domain. The domain organization of SwfA is KS-AT-DH-ER-KR-ACP and from this architecture a saturated fatty acyl chain is expected, although a (poly)unsaturated and/or (poly)hydroxylated acyl chain cannot be excluded,

because in iterative PKSs the reduction domains can be optionally used during each of the elongation steps. SwfB [composed of R (thioester reductase) and ST (sulfotransferase) domains] and SwfC (a radical SAM), are expected to modify the acyl chain produced by SwfA in unknown ways.

As the R and ST domains are contiguous in SwfB, the expected product of elaboration of an acyl chain by SwfB would possibly be an alkyl sulphate or an alkylaminosulphonate: while the R domain can reductively release the assembled chain as a primary alcohol or amine, the ST domain can transfer the sulfate group to the hydroxyl or amino group. SwfC represents a radical SAM enzyme which can catalyze methylation of the substrate through a radical mechanism.

Two different examples of the *swf* cluster were found in the metagenome of *P. simplex*, PS11G3 and PSA11D7 (PSA11D7 lacks the *swfC* gene). In addition, PCR amplification of metagenomic DNA from three different and taxonomically distant “high microbial abundance” sponges, *Aplysina fulva*, *Smenospongia aurea* and *Pseudoceratina crassa*, with primers designed for *swf*, produced amplicons which showed high sequence similarities to the AT domain of *swfA*. Therefore, the *swf* cluster is widespread in marine sponges and presumably associated to ubiquitous sponge symbionts. It represents the second group of mono-modular PKS, after the *supA* family, to be ubiquitously present in marine sponges.

Preliminary studies of heterologous expression of *swf* genes were undertaken with the final aim of characterizing the unknown metabolite produced by the cluster. Activation of the ACP domain of the SwfA protein to its holo-form by co-expression with the phosphopantetheinyl transferase

Svp was the first functional proof of *swf* type genes in marine sponges. Furthermore, applying homologous recombination for expression vector engineering, *swfA* was clearly expressed at the protein level in *E.coli* BL21-CodonPlus[®](DE3)-RIPL cells by coexpression with the chaperone plasmid pTf16, which encodes for the molecular chaperone Trigger factor aiding the protein folding process.

After cloning the whole *swf* operon into the expression vector pHIS8-Svp by homologous recombination, the new recombinant construct was used for heterologous expression trials of the whole cluster in *E. coli* BL21-(DE3) BAP1. Methanol extracts of transformants and their culture broths were analysed by LC-HR-ESI-MS, but no compounds which were present in all the transformants and absent in all the negative controls could be detected. In addition, fatty acid composition of transformants and their culture broths was characterized by saponification of the lipid extract and derivatization to fatty acid methyl esters (FAMES) followed by GC/MS analysis. Even in this case, no new metabolite was detected, suggesting that the *swf* pathway is not functional in this expression system. As a consequence, the biosynthetic function of the *swf* cluster remains unknown at present.

In parallel, metagenomic investigations conducted using high-throughput sequencing based on massively parallel 454 pyrosequencing led to a comprehensive overview of the polyketide metabolism of *P.simplex* and its symbionts, shedding light on the existence of novel polyketide synthase pathways potentially involved in bioactive compound biosynthesis.

454 pyrosequencing was performed on complex and heterogeneous PCR products amplified from the metagenomic DNA of *P.simplex* with

degenerate probes targeting ketosynthase and acyltransferase domains of type I PKSs. Next generation sequencing of AT amplicon mixture generated 8995 reads; applying this modern approach, no PKS/FAS other than known SupA and SwfA could be found. Almost 51% of the total reads belonged to the Swf enzymes, while only the 4% was represented by AT belonging to SupA enzymes (the remaining reads appear not to be related to AT domains).

On the other hand, 454 pyrosequencing of KS amplicon PCR mixture generated 19333 reads. Besides the expected huge presence of KS forming parts of SupA enzymes (~ 80% of the total reads), BLASTx analyses led to the detection of 8 new KS fragments, not reported in genbank database.

All the eight putative KS fragments (which based on phylogenetic analysis appeared to be part of one hybrid NRPS-*cis*-AT PKS and seven *cis*-AT PKSs) are significantly different (E values $\geq 10^{-6}$) to each other, and BLASTx analysis as well as the rebuilt phylogenetic taxonomy revealed that they are only distantly related to PKSs of characterized function. In addition, phylogenetic analyses suggest that these KS fragments are mainly related to PKSs from *Cyanobacteria*, *Actinomycetes* and *Myxobacteria*, commonly known as precious sources of bioactive polyketides. These fragments may represent important starting points for further research towards the isolation of new PKS genes.

The second line of research of my project was directed to the isolation and structure elucidation of new secondary metabolites from two tropical sponges, *Chalinula molitba* and *Plakortis* cf. *lita*. Stereo structure determination of the new compounds was determined by a combination of

mass spectrometry, mono- and bidimensional NMR experiments and micro-scale chemical degradation.

The analysis of the organic extract of the Caribbean sponge *C. molitba* led to the identification of chalinulasterol, a new C-24 chlorinated sterol disulfate. On the basis of the structural similarity with the PXR agonist solomonsterol A, the possible role of chalinulasterol as modulator of the pregnane-X receptor activity was evaluated by carrying out a transactivation assay (luciferase assay) on HepG2 cells, a human hepatocarcinoma cell line. Despite the structural similarity, chalinulasterol failed in transactivating PXR. The possibility that chalinulasterol could act as potential PXR antagonist was investigated, thus, also in this case, failed to reverse the induction of luciferase caused by rifaximin. Although negative, these results have an important implication in terms of structure-activity relationship, because they suggest that the sulfate group (absent in chalinulasterol) present in the side chain at position C-24 of solomonsterol A is essential in the ligand-receptor binding and receptor transactivation, confirming a proposed binding model where a clear interaction of the 24-sulfate with the positively charged Lys210 is hypothesized.

Chemical analysis of the Indonesian sponge *P. lita* revealed, as first remarkable result, a secondary metabolite profile (glycolipids, hopanoids, polyketides) that, in spite of the geographical distance, was very similar to that of the Caribbean sponges of the genus *Plakortis*. Taking into account that secondary metabolites are often of bacterial origin, this indicates that bacterial communities associated to many species of *Plakortis* sponges are highly specific and consistently conserved in specimens collected in

different times and geographical areas, suggesting vertical transmission within their hosts. In addition, among the many known compounds, plakohopanoid, a novel type of hopanoid, was isolated. Plakohopanoid is composed of a C₃₂ hopanoid acyl ester-linked to a mannosyl-*myo*-inositol unit. It is interesting to note that C₃₂ hopanoic derivatives are commonly considered as geohopanoids, i.e. diagenetic products formed through abiotic degradation of the hopanoids biosynthesized by bacteria (biohopanoids). As a consequence, the presence of plakohopanoid in a marine living organism is worthy of note, because it shows that there is a biosynthetic pathway to C₃₂ hopanoic acids, which therefore should not be classified anymore as sure geohopanoids.

ABSTRACT

Nel mondo marino, i poriferi rappresentano una delle più prolifiche risorse di metaboliti secondari bioattivi. Negli ultimi anni sono comparsi sul mercato diversi farmaci derivanti da molecole di origine marina, mentre altri sono attualmente in fase di sperimentazione clinica e pre-clinica. Esplorare la sorprendente e inusitata chimica delle spugne marine è stato il filo conduttore del mio progetto di ricerca, in perfetta sintonia con i propositi che animano la scoperta e lo sviluppo di nuovi farmaci dai prodotti naturali.

Il lavoro presentato in questa Tesi di dottorato ha seguito due percorsi differenti della ricerca nel campo della chimica delle sostanze naturali marine. Da un lato, in seguito all'avvento della genomica ambientale finalizzata alla "drug discovery", gran parte della mia attività di ricerca si è focalizzata sull'analisi metagenomica della spugna caraibica *Plakortis simplex*, nell'intento di identificare nuovi geni di polichetide sintasi, complessi multienzimatici responsabili della biosintesi di polichetidi, una classe di metaboliti secondari che spesso ha rivelato interessanti proprietà farmacologiche (antibiotici, antitumorali).

Dall'altro lato, la seconda linea di ricerca presentata in questo elaborato di Tesi, ispirandosi alla tradizionale attività di ricerca della chimica delle sostanze naturali, si è incentrata sull'isolamento e la determinazione stereostutturale di nuovi composti bioattivi da diverse spugne marine provenienti dagli oceani tropicali, risorse di architetture molecolari davvero insolite da impiegare come "lead compounds" nella progettazione di nuovi farmaci.

Le indagini sul metagenoma (DNA della spugna e dei microorganismi simbiotici) di *P.simplex* sono state intraprese in quanto la spugna è nota per

la produzione di grandi quantità di polichetidi a nucleo endoperossidico, di cui la plakortina è il più abbondante (costituisce all'incirca il 20% dell'estratto lipofilico della spugna). La plakortina è una molecola di particolare interesse per la sua attività antimalarica, che si estende finanche sui ceppi di *Plasmodium falciparum* cloroquina-resistenti. Di conseguenza, è stato avviato uno studio sulla biosintesi di tale polichetide, con il fine ultimo di realizzarne la produzione biotecnologica. La plakortina, come è stato verificato per diversi altri polichetidi (onnamide, psimberina, swinholide A) è prodotta dai batteri che vivono in simbiosi con la spugna, secondo quanto testimoniato da studi sulla localizzazione cellulare dei metaboliti secondari isolati da *P. simplex*. Sulla base dell'architettura molecolare e della natura non aromatica del composto, la plakortina è presumibilmente biosintetizzata da una polichetide sintasi (PKS) di tipo I. Dal momento che la coltivazione dei microrganismi simbiotici delle spugne marine prevede la realizzazione di condizioni difficili da riprodurre in laboratorio, la ricerca del cluster genico della plakortina è stata affidata all'impiego di tecniche indipendenti dalla coltivazione cellulare: lo studio del metagenoma estratto dalla spugna.

Lo screening della libreria metagenomica di *P. simplex* non ha condotto fino ad ora all'isolamento del cluster genico responsabile della biosintesi della plakortina. Tuttavia, un risultato del tutto inaspettato è stato conseguito dall'analisi metagenomica, la scoperta di un nuovo gruppo di polichetide sintasi monomodulari, Swf ("sponge widespread fatty acid synthases"), che sembra sia strettamente associato ai batteri simbiotici delle spugne marine. Lo screening mediante PCR, impiegando primers specifici per il dominio

aciltransferasi di *swf*, ne ha, infatti, rivelato la presenza nel metagenoma di altre tre spugne ad elevato contenuto di batteri simbiotici (“high microbial abundance” sponges), *Smenospongia aurea*, *Aplysina fulva* e *Pseudoceratina crassa*. Swf rappresenta, dopo Sup, il secondo gruppo di PKS monomodulari ubiquitario nel metagenoma delle spugne marine.

L’architettura dell’operone *swf*, ricostruita mediante analisi bioinformatica, consta essenzialmente di tre geni: *swfA*, *swfB* e *swfC*.

La proteina SwfA rappresenta il modulo di una polichetide sintasi di tipo I iterativa, che attraverso il reclutamento sequenziale e reiterato di unità di malonil-CoA costruisce il backbone della catena acilica. Sulla base della contemporanea presenza dei domini DH, ER e KR (DH = deidratasi, ER = enoil-reduttasi, KR = cheto-reduttasi), si prevede la formazione di una catena acilica completamente satura, anche se l’ipotesi di una catena poli-insatura e/o poli-ossidrilata non può essere esclusa, dato che in alcune PKS iterative i domini preposti alla riduzione possono essere impiegati in maniera facoltativa in ogni step di allungamento della catena.

La proteina SwfB rappresenta un’ipotetica proteina di fusione di un dominio N-terminale tioestere-reduttasi (R) e un dominio C-terminale sulfotransferasi (ST). La combinazione dei domini R e ST nella proteina SwfB suggerisce che il prodotto finale di elaborazione della catena acilica assemblata da SwfA potrebbe essere rappresentato da un alchilsolfato oppure da un alchilamminosolfonato: se il dominio R catalizza, infatti, il rilascio riduttivo della catena acilica sotto forma di ammina o alcol primario, il dominio ST può trasferire il gruppo solfato sulla funzione amminica o ossidrilica generatasi. Il gene *swfC* codifica per una proteina

“radical SAM”, che, attraverso un meccanismo radicalico, è in grado di metilare il metabolita finale del pathway *swf*.

Studi preliminari di espressione eterologa dei geni *swf* sono stati intrapresi nell'intento di caratterizzare il cluster da un punto di vista funzionale e definire il metabolita finale del nuovo pathway. L'attivazione del dominio ACP (“acyl carrier protein”) di SwfA nella sua oloforma attraverso la coespressione della fosfopanteteneiltransferasi Svp costituisce la prima prova della funzionalità dell'operone *swf* nelle spugne marine. Inoltre, mediante l'applicazione della ricombinazione omologa per la costruzione di un adeguato vettore d'espressione, il gene *swfA* è stato espresso a livello proteico in linee cellulari di *E.coli* BL21-codonplus (DE3)- RIPL attraverso la coespressione del plasmide pTf16, che codifica per lo chaperone molecolare Trigger factor, coadiuvante del processo di folding delle proteine.

Al fine di eseguire tentativi di espressione eterologa dell'intero cluster, dopo aver realizzato il clonaggio del completo operone *swf* nel vettore di espressione pHIS8-Svp mediante ricombinazione omologa, il nuovo costrutto ricombinante è stato trasferito attraverso elettroporazione in linee cellulari di *E.coli* BL21-(DE3)-BAP1. L'estratto metanolico delle colture dei trasformanti e del brodo di coltura è stato analizzato mediante LC-HR-ESI-MS senza rivelare la presenza di alcun metabolita che fosse assente nei relativi controlli negativi. Inoltre, il contenuto di acidi grassi delle colture dei cloni ricombinanti è stato caratterizzato attraverso saponificazione dell'estratto lipidico e successiva derivatizzazione a esteri metilici. Anche in questo caso, l'analisi degli esteri metilici degli acidi grassi effettuata alla

GC-MS ha rivelato un profilo essenzialmente analogo ai relativi controlli negativi, suggerendo che il pathway *swf* è inattivo nel sistema eterologo prescelto. Di conseguenza, la funzione biosintetica del gene *swf* resta sconosciuta per il momento.

Parallelamente, indagini sul metagenoma di *P.simplex* condotte mediante l'impiego di tecniche di sequenziamento di ultima generazione, il "454 pyrosequencing", hanno consentito di realizzare un'analisi esauriente del metabolismo polichetidico della spugna, evidenziando la presenza di nuovi sistemi di polichetide sintasi, potenzialmente coinvolti nella biogenesi di molecole bioattive. Il "454 pyrosequencing" è stato eseguito su complessi ed eterogenei prodotti di PCR, amplificati direttamente dal metagenoma di *P. simplex*, mediante l'impiego di primers degeneri disegnati sulle sequenze amminoacidiche conservate dei domini aciltransferasi (AT) e chetosintasi (KS) delle PKS di tipo I.

L'applicazione di questo metodo innovativo ha generato una libreria di ampliconi AT di 8995 sequenze; l'analisi bioinformatica di queste ultime ha evidenziato esclusivamente la presenza di polichetide sintasi note delle famiglie SupA (4%) e SwfA (51%), ubiquitarie nel metagenoma dei poriferi (il restante 45% degli ampliconi è costituito da sequenze non correlate ai domini aciltransferasi).

Il pirosequenziamento della libreria di ampliconi KS ha rivelato dati più interessanti, generando ben 19333 sequenze. Accanto all'attesa presenza cospicua di frammenti di domini KS di SupA (circa l'80%), l'analisi bioinformatica ha condotto all'identificazione di otto nuovi frammenti di KS, non presenti nelle banche dati geniche. In particolare, sono state

individuate una sequenza che è parte di una PKS/NRPS (sistemi multienzimatici ibridi che si caratterizzano per l'alternanza di moduli di PKS e moduli di peptide sintetasi non ribosomiali) e 7 sequenze parziali di domini chetosintasi di *cis*-AT PKS (PKS che includono il dominio acil-transferasi all'interno del cluster genico). Le nuove sequenze individuate, implicate nella biosintesi di metaboliti secondari di *P.simplex*, sono sostanzialmente diverse tra loro (E value $\geq 10^{-6}$) e l'analisi eseguita mediante BLASTx come la tassonomia filogenetica ricostruita hanno indicato che sono soltanto lontanamente correlate alle PKS caratterizzate in letteratura. Inoltre, l'analisi filogenetica suggerisce che queste nuove sequenze parziali ricadono nel clade di PKS isolate da *Cyanobatteri*, *Actinomiceti* e *Mixobatteri*, comunemente riconosciuti come preziose risorse di polichetidi bioattivi. Questi otto frammenti di domini chetosintasi rappresentano il punto di partenza per l'isolamento di nuovi geni di PKS dal metagenoma di *P.simplex*.

La seconda linea di ricerca del mio progetto si è focalizzata sull'isolamento e la determinazione strutturale di nuovi metaboliti secondari dalle spugne *Chalinula molitba* e *Plakortis cf. lita*. La ricostruzione della struttura dei nuovi composti isolati è stata eseguita attraverso una combinazione di tecniche di spettroscopia NMR mono- e bidimensionale, spettrometria di massa ESI ad alta risoluzione, GC-MS e degradazione chimica in piccola scala.

L'analisi dell'estratto organico della spugna caraibica *C. molitba* ha condotto all'identificazione del chalinulasterolo, uno sterolo disolfato clorurato in catena laterale sul C-24. Sulla base dell'omologia strutturale

con il solomonsterolo A, agonista del recettore PXR, è stato valutato il possibile ruolo del chalinulasterolo come modulatore allosterico del recettore del pregnano X, attraverso un saggio di transattivazione sulle cellule HepG2, una linea cellulare di epatocarcinoma umano. Nonostante l'evidente similarità strutturale, il chalinulasterolo non ha indotto la transattivazione del recettore nucleare. Di conseguenza, è stata verificata la possibilità che il chalinulasterolo potesse agire da antagonista del recettore PXR; tuttavia, anche in questo caso, non ha prodotto alcuna interferenza nell'induzione dell'espressione del gene della luciferasi causata dalla rifaximina. Nonostante l'esito negativo, i risultati conseguiti hanno una rilevante implicazione nella definizione della relazione struttura-attività, suggerendo che il gruppo solfato sul C-24 della catena laterale del solomonsterolo A (assente nel chalinulasterolo) riveste un ruolo cruciale nell'interazione farmaco-recettore, in pieno accordo con il modello di binding proposto che ipotizza un'interazione tra il solfato in C-24 e la Lys210 del sito recettoriale.

L'analisi dell'estratto organico della spugna indonesiana *P. lita* ha evidenziato, come primo risultato degno di nota, un profilo metabolico (in particolare glicolipidi, opanoidi e polichetidi) analogo a quello delle specie caraibiche del genere *Plakortis*, nonostante la differente provenienza geografica. Partendo dal presupposto che spesso i metaboliti secondari sono prodotti dai microorganismi simbiotici dei poriferi, l'omologia della composizione dell'estratto organico rappresenta una prova evidente dell'esistenza di ceppi batterici associati in maniera specifica a spugne del genere *Plakortis*, che vengono conservati in maniera rilevante in esemplari

raccolti in tempi e luoghi differenti, suggerendone la trasmissione verticale nel corso dell'evoluzione della specie ospite. Inoltre, accanto agli altri composti noti isolati, un nuovo metabolita di tipo opanoideo è stato identificato nell'estratto lipofilo della spugna, il plakopanoide, costituito dall'acido opanoico C₃₂ esterificato con un residuo di mannosil-myo-inositolo. I derivati dell'acido opanoico C₃₂ sono comunemente considerati geo-opanoidi, prodotti diagenetici, che si formano dalla degradazione abiotica dei bio-opanoidi presenti nei batteri. Di conseguenza, la scoperta del plakopanoide in un organismo vivente marino costituisce un evento di particolare rilievo, indicando l'esistenza di un pathway biosintetico dell'acido opanoico C₃₂, che, quindi, non può essere sempre classificato come un geo-opanoide, sfuggendo talvolta agli schemi tradizionali.

INTRODUCTION

The chemistry of marine natural products has experienced enormous developments in the last five decades, leading to the discovery of a myriad of compounds with unusual and complex chemical architectures having no equivalent in the terrestrial environment. On land, communication between insects relies mainly on pheromones, which are generally structurally simple and easy to synthesize, because they need to be volatile. In the marine environment, communication between living organisms is strikingly related to the hydrophobicity, the natural consequence being more complex structures and larger molecular weights for the compounds implied in the regulation of the relationships between organisms. In addition, a larger number of organisms and less genetic homogeneity between separate populations of the same species exist in the marine compared to the terrestrial environment. This results in a higher structural diversity at the molecular level.

The huge biodiversity occurring in an aqueous environment finds its clearest evidence in the coexistence of a number of species which are able to interact with each other. Chemistry plays a crucial role in this interaction: all the marine living organisms can synthesize by themselves or receive from their microbial symbionts secondary metabolites, representing the main characters regulating communications between species or within a taxon. Secondary metabolites are the final result of the perpetual competitions which all forms of life are subjected to. Besides the adequate water solubility and the marine world biodiversity, the evolutionary pressure led

the organisms inhabiting seas and oceans to produce an enormous range of secondary metabolites as means to explicate different biological activities. Among the other functions, most of them being still unknown, these compounds are repellents for predators by their toxicity, attractive molecules to make reproduction more probable, or can even play a role in the complex food chain of the marine environment, determining what is edible and what is not.

Taking together all these factors, it is not surprising that marine organisms are a wonderful source of bioactive natural products.

Sponges represent the most prolific marine producers of novel compounds, with more than 200 new metabolites reported each year. In addition, more sponge-derived metabolites are in clinical and pre-clinical trials than metabolites from any other marine phylum. The frequent detection of structurally related compounds in taxonomically unrelated sponges has led to the hypothesis that these molecules are biosynthesized by specific sponge-microbe consortia. Considering that chemical synthesis of sponge derived chemicals is often difficult and expensive because of their structural complexity and the many chiral centers, the idea that at least some of the secondary metabolites from sponge are of bacterial origin is encouraging. Indeed, this would potentially make possible to obtain a sustainable and unlimited supply of compounds for clinical trials and large-scale drug production, for instance via cultivation of the relevant bacteria.

Sponge derived natural products belong to many different classes, including polyketides, terpenoids, peptides and alkaloids, and show a wide range of pharmaceutical activities (e.g. anticancer, antibacterial, antifungal,

antiviral, anti-inflammatory, and antifouling). Despite the large number of novel sponge-derived bioactive compounds, only two nucleoside analogs (Ara-A and Ara-C) have been approved as antiviral and anticancer drugs, respectively, and are in the market. These drugs are not directly isolated from sponges but are synthetic derivatives based on compounds from the Caribbean sponge *Cryptotethia crypta*. However, several promising sponge-derived compounds are in clinical and preclinical trials as antitumor drugs, as for example E7389, a synthetic analog of halichondrin B from the Japanese sponge *Halichondria okadai*.

The main obstacle to pharmaceutical applications of marine natural products is the so-called supply problem. Biologically active natural products are often produced in small amounts and/or by rare species, whose natural populations cannot sustain the extensive collections that would be required by clinical trials. Therefore, alternative means for producing large amounts of these compounds are required.

The most obvious way to guarantee a sufficient supply is chemical synthesis, but a cost-effective, industrial-scale synthesis of a complex natural compound is usually impossible. Alternative approaches have therefore been proposed, such as the cultivation of the organism of interest under controlled conditions (aquaculture) simulating the natural environment, and the laboratory production of metabolites in bioreactors from cell cultures. However, the scale-up of the operation to the levels necessary for commercial production is not easy to realize, usually because of the small amounts in the sponge tissues of the compound of interest, which make the aquaculture option economically unfeasible.

An alternative method for accessing the hidden chemistry of marine sponges is represented by the cultivation of metabolite-producing microorganisms. The potential advantages from the cultivation approach are substantial: if the metabolite producers can be isolated on artificial media and grown to significant cell numbers (while continuing to produce the bioactive metabolite), then this obviates the need for massive collection of natural sponge individuals, with its environmentally and economically negative implications. However, only a small fraction of the sponge-associated microorganisms can be cultivated using the current techniques.

Recently, the advent of metagenomics provided an interesting and culture-independent approach to investigate the biosynthetic potential of marine sponges. The analysis of genome fragments (“genomic libraries”) from complex sponge-microbe consortia can lead to the isolation of the biosynthetic gene clusters of bioactive metabolites, paving the way for large-scale, sustainable production in heterologous hosts (e. g. *E. coli*).

The aim of my research project was to exploit the unusual and often surprising chemistry of marine sponges, in the frame of the more general purpose of discovering and developing new drugs from natural products.

The present PhD thesis describes the research work performed at the Department of Pharmacy, University of Naples “Federico II”, in collaboration with the NeaNAT group. NeaNAT is a multidisciplinary research group founded by Professor Ernesto Fattorusso. The group has been active for over 40 years in the isolation and identification of natural products of marine origin expanding this “core activity” to the study of the biological and pharmacological properties of the isolated compounds as well

as to the understanding of the biosynthesis of secondary metabolites at a genetic level. The largest part of my research activity focused on the metagenomic analysis of the Caribbean sponge *P. simplex*, and was aimed at the identification of new genes coding for polyketide synthases (PKSs), the multi-domain enzymes or enzyme complexes that produce polyketides, a large class of secondary metabolites that include many antibiotic and antitumor compounds (e. g. the antibiotic erythromycin). Marine sponges of the genus *Plakortis* (Demospongiae, Homosclerophorida, Plakinidae) are known for the production of large amounts of polyketide peroxides, of which plakortin is the most abundant. Plakortin is of special interest due to its anti-malarial activity, which is retained also against chloroquine-resistant strains of *Plasmodium falciparum*. Therefore, my research group has undertaken a study of the biosynthesis of plakortin, with the final aim being its biotechnological production. While the putative genes implied in plakortin biosynthesis could not be identified, an unexpected result from the metagenomic library screening was the discovery of *Swf*, a new group of mono-modular type I PKS/FAS (“hybrid of polyketide synthase and fatty acid synthase”), which appears to be widespread in marine sponge metagenomes and specifically associated to sponge symbionts. Preliminary trials of heterologous expression of *swf* genes were performed, outlining interesting perspectives for the biotechnological production of sponge symbiont metabolites.

In parallel, metagenomic investigations conducted using high-throughput sequencing based on massively parallel 454 pyrosequencing led to a comprehensive overview of the polyketide metabolism of *P. simplex* and its

symbionts, shedding light on the existence of novel polyketide synthase pathways potentially involved in bioactive compound biosynthesis.

A second line of research was directed to the “core activity” of natural product chemistry, the isolation and structure elucidation of new bioactive compounds from different specimens of sponges living in tropical oceans, wonderful sources of unusual molecular architectures to be used as leads and scaffolds for the elaboration of new drugs. In this respect, the analysis of the organic extracts of the sponges *Chalinula molitba* and *Plakortis cf. lita* was undertaken, allowing the isolation of new interesting secondary metabolites. Chalinulasterol, a new chlorinated sterol disulfate, was isolated from the Caribbean sponge *Chalinula molitba* and its possible role as modulator of the pregnane-X-receptor was investigated, giving useful information about the structural requirements for the interaction with this nuclear receptor. On the other hand, an unusual C₃₂ biohopanoid, plakohopanoid, was isolated from the Indonesian sponge *P. lita*. The presence of plakohopanoid in a marine living organism shows that there is a biosynthetic pathway to C₃₂ hopanoic acids, which should not be considered anymore as sure geohopanoids.

Therefore, the results related to a three-year research work have been organized in two main sections:

- **Metagenomic analysis of polyketide metabolism of the Caribbean sponge *Plakortis simplex* and biotechnological prospects.**
- **Isolation and structure elucidation of new secondary metabolites from the marine sponges *Chalinula molitba* and *Plakortis cf. lita*.**

PART I

*Metagenomic analysis of the Caribbean sponge
Plakortis simplex
and biotechnological prospects*

Chapter 1

Porifera

Marine sponges are among the most representative phyla of the benthic community, not only in terms of biomass but also for their biological role in benthic and pelagic ecosystems. Sponges (phylum Porifera)^[1] are the simplest and most ancient phylum among multicellular animals known as metazoans. They are sessile, filter-feeding organisms which obtain their nourishment (the tiny and floating organic particles known as plankton) and oxygen from the surrounding water that they filter through their body. Sponges are ubiquitous animals and inhabit a wide variety of marine and freshwater environments because of their ability to adapt in tropical, polar and temperate areas. There are from 5,000 to 10,000 known species of sponges, most of them living in salt water; only about 150 species live in fresh water.

The phylum Porifera is a paraphyletic grouping consisting of three main classes: the Hexactinellida (glass sponges), Calcarea (calcareous sponges), and Demospongiae (demosponges), with the last sublineage including the majority of the known sponge species. Sponge morphology is strikingly related to many biological functions of these organisms. The histological architecture is made up of several cell layers. The outer layer is formed by pinacocytes, epithelial plate-like cells which digest food particles too large to enter the pores (ostia) located on sponge surface. The body plan consists also of specialized tube-like cells called porocytes which control ostia,

channels leading to the interior of the sponge. The internal side of the sponge consists of a series of chambers constituted by choanocytes, cylindrical flagellated cells. In these chambers, collectively called choanoderm, choanocytes pump in water through the ostia by the flagellar beat, and filter out food particles from the water (such as bacteria and microalgae) which are then digested in the mesohyl. The mesohyl is the gelatinous matrix within the sponge mainly made of collagen, where phagocytosis of food particles by the archaeocytes takes place. Archaeocytes are basal cells able to differentiate into any other sponge cell line and are located into the mesohyl together with several symbiont microorganisms. Once filtered in the choanocyte chambers, water flows out of the sponge through large openings called oscula, carrying out also waste products. It has been estimated that up to 24,000 liters of water can be pumped through a 1-kg sponge in a single day.

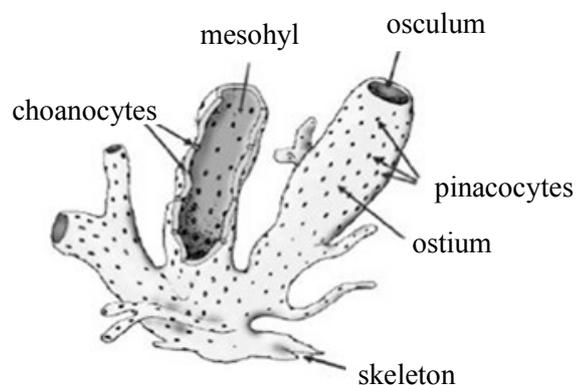


Figure 1.1 Sponge body structure

The basic body plan of sponges is reinforced by the skeleton, composed of collagen fibers and spicules. Calcareous sponges produce spicules made of silica calcium carbonate, while the larger class (90%) of Demospongiae produce a special form of collagen called spongin together with silica spicules; glass sponges, common in polar water and in the depths of temperate and tropical seas contain syncytia in their structure which enable them to extract food from these resource-poor waters with the minimum of effort.

The simplest body structure in sponges is a tube or vase shape known as “asconoid”^[2], but this severely limits the size of the animal. The body structure is indeed characterized by a stalk-like spongocoel surrounded by a single layer of choanocytes. The limited choanocyte layer extension affects negatively the pumping activity which is essential to supply food and oxygen to sponge tissues, and influences the growth of the organism. Asconoid sponges seldom exceed 1 mm in diameter.

Some sponges overcome this limitation by adopting the “syconoid”^[2] structure, in which the body wall is pleated. The inner pockets of the pleats are lined with choanocytes, which connect to the outer pockets of the pleats by ostia. This increase in the number of choanocytes and hence in pumping capacity enables syconoid sponges to grow up to a few centimeters in diameter.

Leuconoid^[2] sponges contain a network of chambers lined with choanocytes and connected to each other and to the water intakes and outlets by tubes. Thanks to this network, the pumping capacity is so amplified that these

organisms can grow to over 1 meter in diameter and in any direction acquiring the most different shapes.

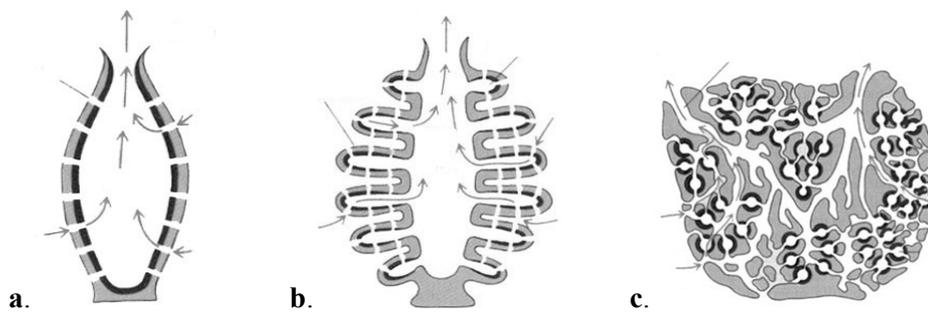


Figure 1.2 Sponge classification: a. asconoid; b. siconoids; c. leuconoids

Most sponges are hermaphrodites (each adult can act as either the female or the male in reproduction).^[2,3] Fertilization is internal in some species; other individuals release sperm (produced by choanocytes) randomly which floats to another sponge with the water current. If a sperm is caught by another choanocyte of a sponge of the same species, fertilization of an egg (eggs are formed by transformation of archeocytes or choanocytes) by the travelling sperm takes place inside the sponge.

The resulting tiny larva is released and is free-swimming; it uses tiny cilia (hairs) to propel itself through the water. The larva eventually settles on the sea floor, becomes sessile and grows into an adult.

Sponges also reproduce asexually through three distinct ways: by fragmentation, by budding, and by producing gemmules.^[2,3]

Fragments of their body are broken off by water currents and carried to another location, and, by using the mobility of their pinacocytes and

choanocytes as well as the remodelling of the mesohyl, they can re-attach themselves to an appropriate surface where the sponge will grow into a clone of the parent sponge (its DNA is identical to parent's DNA).

Very few individuals can reproduce by budding: buds are formed outside the sponge and are similar to “miniature” sponges which, once they are detached from the “mother” sponge can generate a new organism.

Some sponges are able to produce gemmules inside their body. Gemmules are mainly made up of archeocytes, contained in wrapped shells of spongin, often reinforced by spicules. Once germination takes place in a suitable environment, gemmules release archeocytes through an opening located on the surface, called micropilum, in order to generate a new individual.



Figure 1.3 The sponges studied in this PhD Thesis: *Plakortis simplex*, *Chalinula molitba* and *Plakortis lita*.

References

1. Michael W., Taylor R. R. (2007). Sponge-Associated Microorganisms: Evolution, Ecology, and Biotechnological Potential. *Microbiology and Molecular Biology Reviews* , 295-347.
2. Ruppert E. E. (2004). *Invertebrate Zoology*. 76–97.
3. Smith D. G. (2001). Pennak's Freshwater Invertebrates of the United States: Porifera to Crustacea. 47-50.

Chapter 2

Sponge-microbe associations

Different types of Demospongiae, such as the Caribbean sponge *Plakortis simplex*, are associated with endosymbiotic micro-organisms and these symbionts contribute to 38-57% of the total sponge biomass.^[1] Bacteria and fungi are mostly present in the mesohyl, which contains heterotrophic (eubacteria, archaea) and some autotrophic bacteria. As sponge-microbe interactions are widespread and some bacteria seem to be specific and permanently associated with the sponges, the existence of sponge-bacteria symbiosis is well established. The sponge-symbiont relationship can be categorized into obligatory mutualism (i.e. the symbiont play an essential role in the metabolism of their host), facultative mutualism (they have a beneficial effect on their host, but the host will survive without the symbiont) or commensalism (they are present without obvious beneficial effects to their host).

According to the literature, possible aims of sponge-microbe symbiosis are:

- providing nutrition through intracellular digestion or translocation of metabolites;^[2]
- access of new products via nitrogen fixation;^[3]
- stabilization of the sponge skeleton;^[4]
- prevention against predation or fouling through secondary metabolite production.^[5]

The microbiota from a variety of marine sponges have been studied using cultured independent methods. Hentschel and colleagues^[1] performed phylogenetic analyses with sponge-derived 16S rRNA sequences, most of them being from *Aplysina aerophoba*, *Rhopaloeides odorabile*, and *Theonella swinhoei*. Even if these three sponges were phylogenetically distantly related and were collected from different geographical areas (Mediterranean Sea, the Great Barrier Reef, and Micronesia/Japan/Red Sea, respectively), they showed to share largely their microbial community.^[6] Comparative studies of Antarctic sponges have shown that sponge symbionts are species specific, with sponges from the same species being inhabited by similar microbiota that are distinct from those of other species of sponges from the same location.^[7] In addition, both cultivation-based and molecular methods have provided evidence for distinct microbial communities between sponges and the surrounding seawater. Subsequent studies led to similar results, giving further weight to these notions. The existence of widespread microbial associations within marine sponges, some of which being also species specific, has become something of a paradigm. In total, 17 bacterial phyla and both major archaeal lineages, *Crenarchaeota* and *Euryarchaeota*, have been reported in the census of sponge-associated microorganisms: *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Nitrospira*, *Planctomycetes*, “*Poribacteria*,” *Proteobacteria* (*Alpha-*, *Beta-*, *Delta-*, and *Gammaproteobacteria*), *Spirochaetes*, *Verrucomicrobia*, *Chlorobi*, *Lentisphaerae* and the candidate phylum *TM6*.^[6] Besides the high diversity of the microbial communities associated within *Porifera*,

biological surveys revealed the simultaneous presence of the three domains of life, *Bacteria*, *Archae* and *Eukarya* inhabiting marine sponges.

2.1 Diversity of uncultured bacteria associated with the marine sponge *P. simplex*

The microbiota associated with the marine sponge *Plakortis simplex* have been studied using a culture independent approach in order to realize a general survey about the identity of the sponge symbionts, representing a precious source as producers of lots of bioactive metabolites. 16S rRNA amplicon library was prepared from metagenomic DNA extracted from sponge tissue. 16S rRNA genes were amplified by PCR using 16S rRNA primers specific for bacteria, yielding a band of the expected size of ~ 1450 bp. PCR products were subcloned via T/A cloning into the vector pBluescriptII SK(+) and representative plasmids were sent for single read sequencing. 42 unique partial 16S rRNA gene sequences were obtained. The rRNA sequences were analysed using the RDP (RIBOSOMAL DATABASE PROJECT) SEQ MATCH^[8] tool as well as a BLASTn search, and were shown to belong for the most part to seven phyla of bacteria, which comprised *Proteobacteria* (15 strains), *Acidobacteria* (5 strains), *Chloroflexi* (12 strains), *Actinobacteria* (2 strains), *Nitrospira* (3 strains), *Firmicutes* (1 strain), *Gemmatimonadetes* (1 strain); the remaining three strains (PSB2, PSB25 and PSB33) could not be classified because they share at most only 79% sequence identity with bacteria of the phylum *Chloroflexi* as shown by a BLASTn search. RDP analyses of partial 16S

rRNA fragments from PSB2, PSB25 and PSB33 confirmed the previous matches supported by low scores (S_{ab} scores were, respectively, 0.434, 0.391 and 0.410). None of the sequences corresponded exactly to any known bacterial species, and overall the sequence analysis clearly indicates the high diversity of bacterial phylotypes of the microbes associated with *P. simplex*, that fall in the 17 sponge specific symbiont phyla detected by Taylor *et al.*. A neighbor joining tree of the 16S rRNA fragments (figure 2.1) was built applying maximum composite likelihood as statistical method, and it displayed the same seven phyla detected through the RDP SEQ MATCH tool. The same clade formation and branching patterns were observed in minimum evolution tree using either p-distance or maximum composite likelihood statistics. Bacterial topology were basically supported by high bootstrap values in all phylogenetic analyses.

Proteobacteria. The group of *Proteobacteria* (figure 2.1) in the phylogenetic tree displays three main clades corresponding to α , γ and δ -*Proteobacteria*. 7 strains (PSB6, PSB7, PSB12, PSB23, PSB29, PSB36 and PSB37) cluster with phototroph γ -*Proteobacteria* and are purple sulfur bacteria according to the rebuilt phylogenetic taxonomy. PSB7, PSB12 and PSB29 are closely related with haloalkaliphilic bacteria such as *Alkalispirillum mobile*, *Thioalkalivibrio thiocyanodenitrificans* and *Thioalkalivibrio denitrificans*, members of the family *Ectothiorhodospiraceae* while clones PSB6 and PSB37 find the autotrophic bacteria *Halochromatium glycolicum* and *Halochromatium roseum* (*Chromatiaceae*) as their closest homologs among the BLASTn hits. PSB1 and PSB9 are included in the subgroup of α -*Proteobacteria* and are members

of the phototroph heterotrophic family *Rhodospirillaceae* as underscored by the detailed hierarchy traced by RDP and the high bootstrap value (99%) in the phylogenetic tree, where they are closely related with an uncultured *Rhodospirillaceae bacterium* (figure 2.1). Four strains (PSB5, PSB28, PSB16 and PSB35) are affiliated within the δ -subdivision of *Proteobacteria*; particularly, PSB16 and PSB35 belong to the anaerobic order of *Desulfovibrionales* as shown in the taxonomy attributed by RDP.

Chloroflexi. The second predominant bacterial group found in the *P. simplex* 16S rRNA amplicon library was the phylum *Chloroflexi*, subdivision of the green non-sulfur bacteria. Clones PSB19, PSB20 and PSB21 cluster together with *Caldilinea aerophila* (bootstrap value 99%). On the other hand, the other nine 16S rRNA partial sequences (PSB8, PSB10, PSB14, PSB17, PSB18, PSB24, PSB31, PSB40, PSB41) present in this clade are only distantly related to the bacterium *Dehalogenimonas lykanthroporepellens* as it can be deduced from the phylogenetic tree.

Acidobacteria. The distinct clade of *Acidobacteria*, supported by a bootstrap value of 98%, included five clones (PSB13, PSB27, PSB30, PSB34 and PSB38). Clones PSB27 and PSB30 are members of the subdivision Gp6 within the phylum *Acidobacteria* while PSB38 is affiliated with the subdivision Gp9 as suggested by high scored RDP matches.

Gemmatimonadetes*, *Firmicutes* and *Nitrospira. Among the microbial community associated with *Plakortis simplex*, one strain of the phylum *Gemmatimonadetes* was detected; its representative clone, PSB3, has *Gemmatimonas aurantiaca strain T27* as its closest known relative (99%

bootstrap value ; max identity 86% in the BLASTn search). This strain, recently discovered (2003) is a gram-negative rod-shaped aerobe that appears to replicate by budding.^[9] The *Firmicutes* clone PSB39 was found to have the highest sequence identity with Gram-positive lactic acid bacteria from the genus *Granulicatella*. The phylum *Nitrospira* in the phylogenetic tree includes one clone sequence (PSB32) strikingly related with *Nitrospira moscoviensis strain NSP M-1* and the relationship is clearly supported by a bootstrap value of 100% as well as by the BLASTn scores (89% max identity). PSB11 and PSB15 appear to be affiliated with the phylum *Nitrospira* according to the reported taxonomy, even if BLASTn searches and RDP seq match tool showed conflicting results.

Actinobacteria. PSB22 and PSB42 are representative clones of the suborder *Acidimicrobineae* within *Actinobacteria* according to RDP reconstructed hierarchy and share high sequence identity (89% and 91% respectively) with *Iamia majanohamensis strain NBRC 102561* designed as the first hit in the NCBI 16S ribosomal RNA sequences (Bacteria and Archaea) database.^[10]

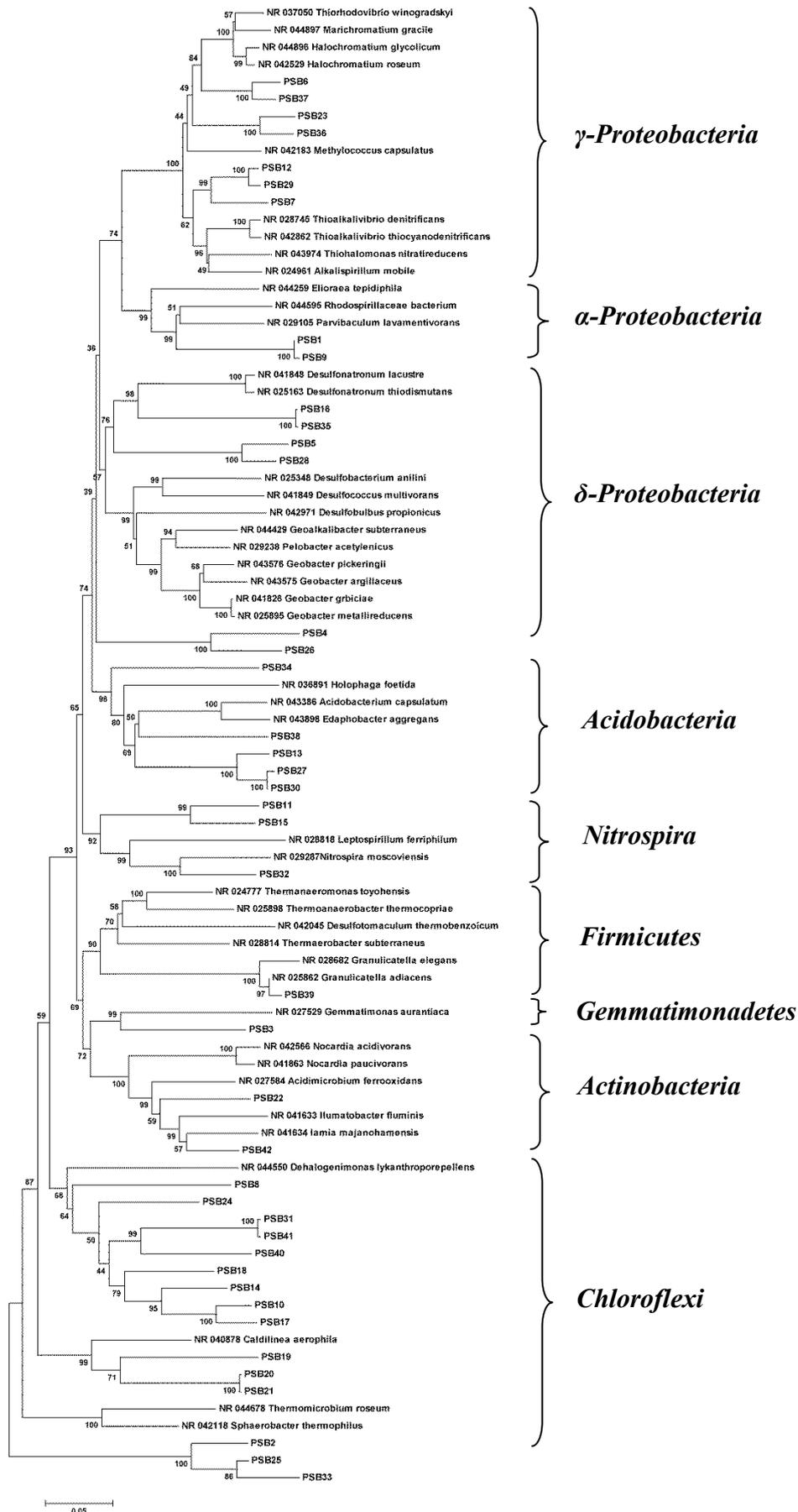


Figure 2.1 Neighbor joining tree displays phylogenetic taxonomy of symbionts associated with *P. simplex*.

2.2 Physiology of Sponge-Associated Microorganisms

16S rRNA amplicon library screening showed a great deal of diversity in the bacteria associated with *P. simplex*, which was found to be the host of bacterial groups previously identified as symbionts of other marine sponges. Most amplicons (67%) matched mainly with partial 16S rRNA fragments amplified from metagenomes of the marine sponges *Xestospongia muta* (Florida, USA), *Xestospongia testudinaria* (Manado, Indonesia) and *Geodia barretti* (Norway). It is also interesting to note that 19% of the 16S rRNA fragments from *P. simplex* metagenome are closely related to uncultured bacteria inhabiting the Caribbean coral *Montastraea faveolata* as clearly shown by the high scores (>0.855) in the RDP search. Taken together, these data confirm that even unrelated sponges with nonoverlapping geographic ranges can share a common core of bacterial associates.

The lack of knowledge about the physiological features of sponge symbionts is due to the unavailability of their pure cultures. Then, the detection and analysis of 16S rRNA genes supported by the few existing-pure culture studies represent one of the easiest ways to infer metabolic properties of sponge-associated bacteria. Generally, among other functions, microbes are capable of photosynthesis, methane oxidation, nitrification, nitrogen fixation, sulphate reduction, and dehalogenation. α -*Proteobacteria* affiliated with the family *Rhodospirillaceae* and γ -*Proteobacteria* from the family *Chromatiaceae* occurred among the microbial community associated with *P. simplex*. Members of *Rhodospirillaceae* and *Chromatiaceae* were already isolated from sponges *Ircinia sp.* and *Euspongia officinalis* in the 1970s and play a key role in the sulphur cycle as sulphur oxidizing bacteria,

capable of oxidize reduced sulphur compounds such as H₂S. Furthermore, besides the presence of sulphur oxidizing bacteria, sulphate reducing microbes from the phylum *δ-Proteobacterium* were observed within *P. simplex*, and are apparently members of the order *Desulfovibrionales*. *Desulfovibrio*-related organisms were already found in enrichment cultures grown on *Aplysina aerophoba*-derived brominated phenolic compounds. The coexistence of sulphate reducing bacteria and sulphur oxidizing bacteria suggested that a possible endosymbiotic sulphur cycle can take place in *P. simplex*, as it has been demonstrated for an oligochaete and hypothesized for other sponges. Strains strikingly related to the genus *Thioalkalivibrio* (*γ-Proteobacteria*) were found to inhabit *P. simplex*. Members of this genus are typical chemolithotrophic and sulphur oxidizing bacteria, and as most purple phototrophic bacteria of the family *Ectothiorhodospiraceae* they are capable of nitrogen fixation. Amplification of *nifH* genes encoding for a subunit of the dinitrogen-reducing enzyme nitrogenase was reported for some *Ectothiorhodospiraceae*, including genera *Thioalkalivibrio* and *Alkalispirillum*, but also for *α*- and *γ*-*proteobacteria* inhabiting several Caribbean sponges.^[6, 11] 16S rRNA phylogenetic analysis revealed the presence of bacteria from the phylum *Nitrospira*, known as nitrite oxidizing microorganisms. They represent essential characters in the nitrogen cycle of sponges for the biological conversion of the toxic waste metabolic product ammonia into nitrate, which can be recycled by sponges. The widespread presence of *Nitrospira* in sponges may indicate low nitrite availability in these hosts, as members of the *Nitrospira* typically favor low-nitrite habitats.^[6]

2.3 Cellular localization of metabolites from *P. simplex*

The bacterial content in *Plakortis simplex* is particularly high, around 90 bacteria per sponge cell, which are permanently associated with the host. The isolation of bacteriohopanoids^[12] that are typical bacterial metabolites, from the sponge led to the hypothesis that not only these compounds, but also the other characteristic metabolites isolated from *P. simplex*, such as the antimalarial polyketide plakortin^[13] and several unusual glycolipids,^[14, 15] may be produced by its symbionts.

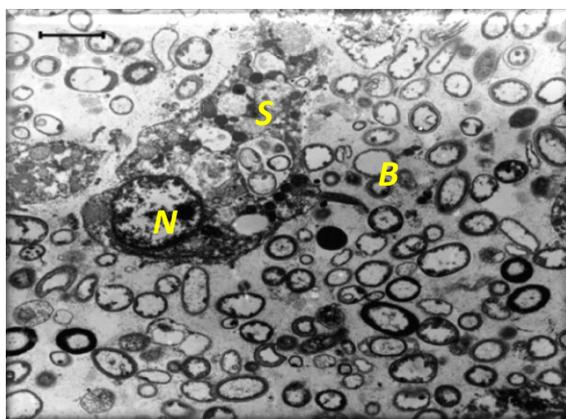


Figure 2.2 Transmission electron micrograph (TEM) of the *P. simplex* mesohyl. S, sponge cell; N, sponge cell nucleus; B, bacteria. Scale bar 2 μm .

This hypothesis was first tested by localizing specific metabolites in cells using physical separation of sponge cells, bacteria and extracellular matrix by differential centrifugation.^[16] The obtained fractions were separately analyzed for the typical *Plakortis simplex* metabolites by NMR and mass spectrometry. The results suggested that bacterial symbionts are the producers of most of the secondary metabolites isolated from *Plakortis*.

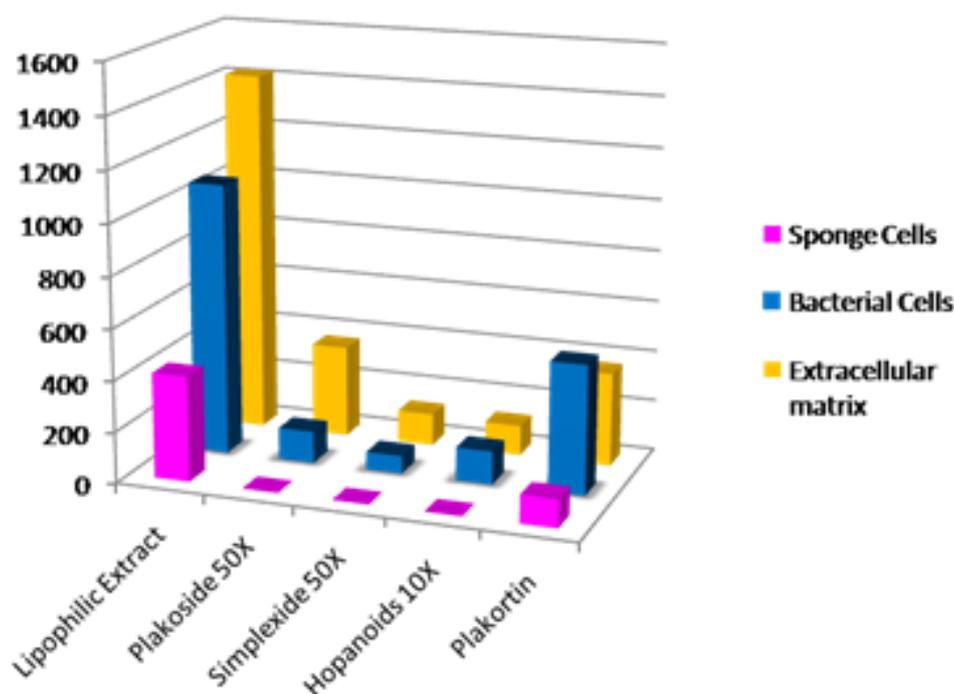


Figure 2.3 Cellular localization of secondary metabolites from *P. simplex*.

In fact, hopanoids and glycolipids were found only in bacterial cells, while the polyketide plakortin and its derivatives were present in both bacteria and sponge cells. However, the amounts were considerably different, as the bacteria fraction contained five times as much plakortin as the sponge cell fraction. All the metabolites were also present in the extracellular matrix, either because the metabolites are excreted, or because part of the cells producing them are lysed during the cell separation process.

The bacterial origin of these metabolites may be the key to overcome the problem of the limited availability of sponge material, which rises every time the commercial production of a marine compound of pharmaceutical interest from a sponge is hypothesized. If the metabolite is produced by a bacterium, then large-scale laboratory culture is possible, and the need to harvest sponges from their natural environment is eliminated.

In the attempt to identify the microorganisms producing the metabolites found in *P. simplex*, over 150 strains were isolated from the pool of its symbionts; 30 representative species were cultivated and subjected to chemical analysis in search of glycolipids and plakortin. Unfortunately, we could not find the metabolites in any strains. These results were not surprising, because it has been estimated that less than 1% of marine microbes are cultivable in the laboratory today, and this is even more true for symbiotic bacteria. In addition, it is not uncommon that biosynthetic gene clusters remain silent unless some particular culture conditions are met.

Therefore, a culture-independent approach was explored, the search for genes responsible for the production of the antimalarial polyketide plakortin. The final aim of this research is to transfer the plakortin genes into a bacterial host, which would become capable of biosynthesizing plakortin, and would allow its production on large scale through a fermentative process.

References

1. Hentschel U., Hopke J., Horn M., Friedrich A. B., Wagner M., Moore B.S. (2002). Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* 68: 4431–4440.
2. Wilkinson C. R., Garrone R. (1980). Nutrition in marine sponges. Involvement of symbiotic bacteria in the uptake of dissolved carbon. In: Smith D, et al. (eds) *Nutrition in lower Metazoa*. Pergamon, Oxford, pp 157-161.
3. Wilkinson C. R., fay P. (1979). Nitrogen fixation in coral reef sponges with symbiotic bacteria. *Nature*. 279:527-529.
4. Wilkinson C. R., Nowak M., Austin B., Cilwell R. R., (1981). Specificity of bacterial symbionts in Mediterranean and Great Barrier Reef Sponges. *Microbiol. Ecol.* 7:13-21.
5. Unson M. D., Holland N. D., Faulkner D. J. (1994). A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar. Biol.* 119:1-11.
6. Michael W., Taylor R. R. (2007). Sponge-Associated Microorganisms: Evolution, Ecology, and Biotechnological Potential. *Microbiology and Molecular Biology Reviews* , 295-347.
7. Webster N. S., Negri, A. P., Munro M. M. H. G., Battershill C. N. (2004). Diverse microbial communities inhabit Antarctic sponges. *Environ Microbiol* 6: 288–300.
8. Cole J. R., Wang Q., Cardenas E., Fish J., Chai B., Farris R. J., Kulam-Syed-Mohideen A. S., McGarrell D. M., Marsh T., Garrity G. M., Tiedje J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37 (Database issue): D141-145.
9. Zhang H., Sekiguchi Y., Hanada S., Hugenholtz P., Kim H., Kamagata Y., Nakamura K. (2003). "Gemmatimonas aurantiaca gen. nov., sp. nov., a gram-negative, aerobic, polyphosphate-accumulating micro-

- organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov.". *Int J Syst Evol Microbiol* 53 (Pt 4): 1155–63.
10. Zhang Z., Schwartz S., Wagner L., and Miller W. (2000). "A greedy algorithm for aligning DNA sequences". *J Comput Biol* 2000; 7(1-2):203-14.
 11. Tourova T. P., Spiridonova E. M., Berg I. A., Slobodova N. V., Boulygina E. S. and Sorokin D. Y. (2007). Phylogeny and evolution of the family *Ectothiorhodospiraceae* based on comparison of 16S rRNA, *cbbL* and *nifH* gene sequences. *IJSEM*. 2387-2398.
 12. Costantino V., Fattorusso E., Imperatore C., Mangoni A. (2001). A biosynthetically significant new bacteriohopanoid present in large amounts in the Caribbean sponge *Plakortis simplex*. *Tetrahedron*. 57:4045–4048.
 13. Higgs M. D., Faulkner D. J. (1978). Plakortin, an antibiotic from *Plakortis halichondroides*. *J Org Chem*. 43:3454–3457.
 14. Costantino V., Fattorusso E., Mangoni A., Di Rosa M., Ianaro A. (1999). Simplexides, novel immunosuppressive glycolipids from the Caribbean sponge *Plakortis simplex*. *Bioorg Med Chem Lett*. 9:271–276.
 15. Costantino V., Fattorusso E., Mangoni A., Di Rosa M., Ianaro A. (1997). A unique prenylated glycosphingolipids with immunosuppressive activity from the marine sponge *Plakortis simplex*. *J Am Chem Soc*. 119:12465–12470.
 16. Laroche M., Imperatore C., Grozdanov L., Costantino V., Mangoni A., Fattorusso E. (2007). "Cellular Localization of Secondary Metabolites Isolated from the Caribbean Sponge *Plakortis simplex*". *Mar Biol*. 151:1365–1373.

Chapter 3

Plakortin

Malaria is one of the major causes of mortality in the tropical regions. Based on documented cases, the WHO estimates that there were 219 million cases of malaria in 2010 resulting in 660,000 deaths, the majority of whom are young children in Sub-Saharan Africa. This is equivalent to roughly 2000 deaths every day. Pregnant women are also especially vulnerable, and in Sub-Saharan Africa maternal malaria is associated with up to 200,000 estimated infant deaths yearly. Although some are under development, no vaccine is currently available for malaria; preventive drugs must be taken continuously to reduce the risk of infection. These prophylactic drug treatments are often too expensive for most people living in endemic areas. Chloroquine is very cheap and, until recently, was so effective, that it has been administered as the antimalarial drug of choice for many years in most parts of the world. However, resistance of *Plasmodium falciparum* to chloroquine has spread quickly from Asia to Africa, making the drug ineffective against the most dangerous *Plasmodium* strain in many affected regions of the world. In those areas where chloroquine is still effective it remains the first choice. Unfortunately, chloroquine-resistance is also associated with reduced sensitivity to other drugs, such as quinine and amodiaquine.

The more recently discovered artemisinin, a sesquiterpene endoperoxide from *Artemisia annua*, is active against *Plasmodium falciparum* and chloroquine resistant strains. Artemisinin and its semisynthetic liposoluble

(artemether and artether) and water soluble (artesunate) derivatives are characterized by a 1,2,4-trioxane (6-membered ring containing 3 oxygen atoms) that is the essential pharmacophoric group for their activity. Even though its mechanism of action has not been exactly determined yet, there is experimental evidence that the artemisinin endoperoxide function is crucial for the pharmacological activity. Artemisinin and its derivatives produce carbon-centred free radicals in the presence of catalytic quantities of free Fe^{2+} and/or by reacting with the heme iron atom (figure 3.1 – 3.2). The reaction yields an oxygen radical, and after a rearrangement a carbon radical. This alkylating species could be responsible for the antimalarial activity by forming covalent adducts with parasite proteins. Particularly, recent studies showed that the inhibition of the SERCA orthologue (PfATP6) of *Plasmodium falciparum* represents, among several and sometimes conflicting hypotheses, one of the main mechanisms of action for the parasiticidal activity of artemisinin.^[1,2]

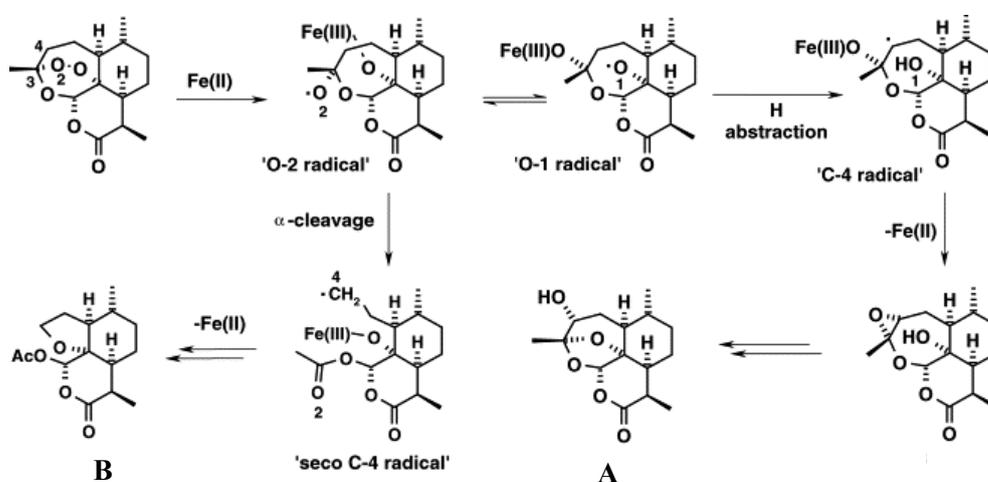


Figure 3.1 Pathways proposed for activation of artemisinin by ferrous iron leading to the seco C-4 radical and C-4 radical, and the end-products A and B. Neither A or B possesses antimalarial activity, and thus antimalarial activity is claimed to reside in the radicals by their reacting with “sensitive biomolecules” in the parasite.

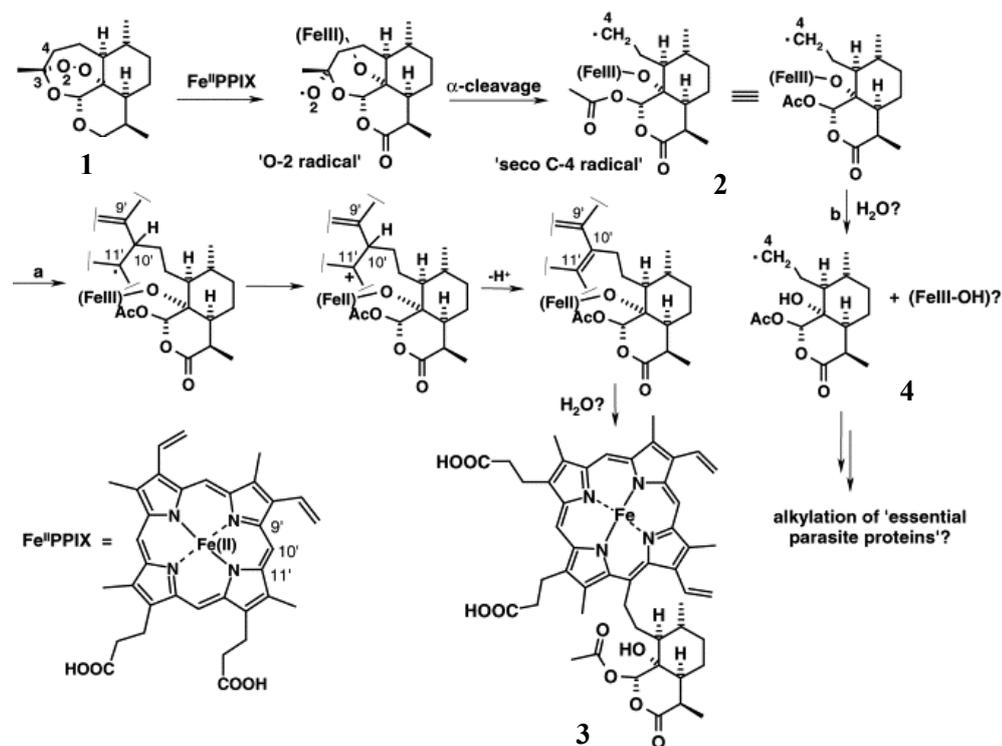


Figure 3.2 Proposed activation pathway for artemisinin 1 by ferrous heme (Fe(II) protoporphyrin IX) leading to the seco C-4 radical 2 and alkylated adduct 3 (pathway a), or by presumed loss of the seco C-4 radical 4 from the heme (pathway b) to alkylate “essential parasite protein” ((FeII)- and (FeIII)-: ferrous and ferric heme residue).

The problem with these compounds is their neurotoxic activity which is related to the endoperoxide ring; for this reason new antimalarial compounds are required that are safer and can be synthesized and preserved more easily. In this respect, several secondary metabolites from sponges show an endoperoxide ring similar to that of artemisinin. Therefore, these compounds are being tested in order to evaluate their possible antimalarial activity. The first active compound, isolated in 1978 by Faulkner *et al.* from the lipophilic extract of the sponge *Plakortis halicondroides*, is plakortin, a 6-membered cyclic peroxide.

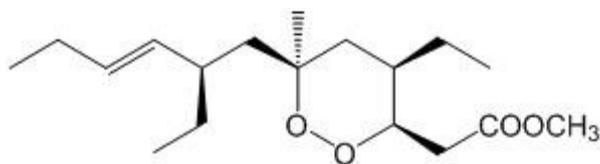


Figure 3.3 Plakortin

Most polyketides that are appearing in the literature have some common key features: they all contain a 1,2-dioxane ring with an acetate residue at position 3 and alkyl chains at positions 4, 6, and 6. Related molecules with a 5-membered peroxide ring (1,2-dioxolane), such as Plakortide E, were also discovered. The analysis of the lipophilic extracts of the Caribbean sponge *Plakortis simplex* showed that plakortin is the major constituent of this fraction (20%), together with 9,10-dihydroplakortin and the 5-membered Plakortide E.

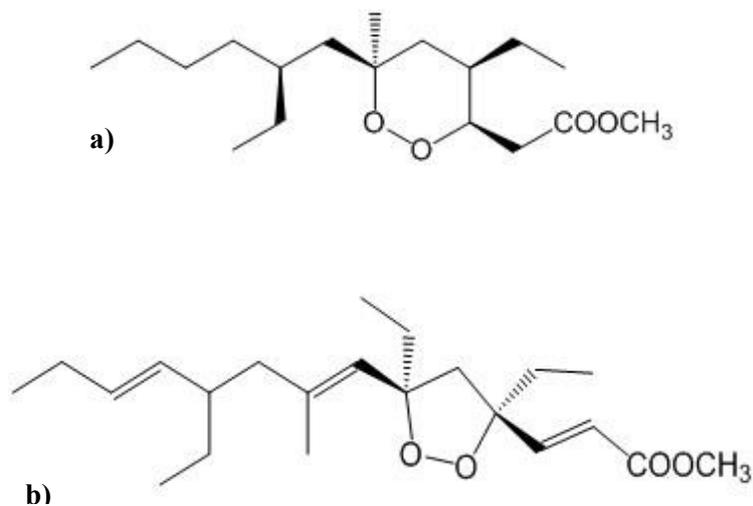


Figure 3.4 Dihydroplakortin (a) and Plakortide E (b)

These three metabolites were tested for their antimalarial activity against D10 chloroquine-sensitive and W2 chloroquine-resistant strains of *Plasmodium falciparum* (table 1). Parasite growth was determined spectrophotometrically (OD650) by measuring the activity of the parasite lactate dehydrogenase (pLDH), according to the method of Makler *et al.* (1993), in control and treated cultures.^[3] Both plakortin and dihydroplakortin have a good antimalarial activity, while plakortide E is inactive, proving that a six-membered peroxide ring is essential for activity. Plakortin is very active on the chloroquine-resistant W2 strain because endoperoxide compounds are active against *Plasmodium falciparum* with a different mechanism compared to chloroquine. The biological target of plakortin is similar to that of artemisinin, but the former shows a lower activity, probably related to its 1,2-dioxane ring instead of the 1,2,4-trioxane of artemisinin. In fact, it seems that the oxygen not involved in the peroxide moiety can make the radical production easier, increasing the antimalarial activity.^[4] In spite of its lower activity, its simple structure and insignificant cytotoxicity make plakortin a very interesting compound.

Compound	D10(CQ-S) IC ₅₀ μM	W2(CQ-R) IC ₅₀ μM
Plakortin	0.80	0.38
Dihydroplakortin	0.86	0.44
Plakortide E	inactive	inactive
Artemisinin	0.05	0.01

Table 3.1 *In vitro* antiplasmodial activity against D10 and W2.

3.1 Polyketide synthases

The studies about cellular localization showed plakortin to be present mostly in the bacterial cells. Because any attempt to isolate the plakortin producer bacterial strain was unsuccessful, a culture independent approach appeared to be the best alternative, and screening of the metagenomic (sponge + symbiont) DNA from *P. Simplex* was conducted with the aim to find the gene responsible for the biosynthesis of plakortin.

On the basis of its chemical structure, plakortin resembles the architecture of a polyketide, and polyketides are biosynthesized by polyketide synthases (PKS), enzyme complexes which are widespread in bacteria and have recently raised great interest due to their involvement in the biosynthesis of many bioactive secondary metabolites.

PKSs are large (100-10000 KDa) multienzymatic systems responsible for the production of extremely complex natural products from simple building blocks with 2, 3 or 4 carbon atoms such as acetyl-CoA, propionyl-CoA, butyryl-CoA and their activated derivatives malonyl-, methylmalonyl- and ethylmalonyl-CoA. They show strong genetic, structural, and mechanistic similarity with fatty acid synthases.

The key step in the biosynthesis of polyketides is chain elongation, obtained through Claisen condensation and decarboxylation. Unlike fatty acid biosynthesis, in which every chain elongation step is followed by a fixed sequence of ketoreduction, dehydration, and enoylreduction, the reaction intermediate obtained during the polyketide biosynthesis are subjected to none, some or all these modifications, yielding very complex products. An added degree of complexity comes from the possible use of

different starter and elongation groups as well as from the formation of new stereogenic centres.

In the microbial kingdom, three types of PKSs were identified that differ in the architecture:

- Type I PKSs (figure 3.5): large multifunctional proteins, which can be either modular (erythromycin, rapamycin and rifamycin) or iterative (lovastatin). Iterative PKSs are involved in the biosynthesis of polyketides such as 6-methyl salicylic acid and aflatoxin in fungi; these multidomain proteins include all the active sites for the biosynthesis of the polyketide and they reuse domains in a cyclic fashion like the fatty acid synthases.

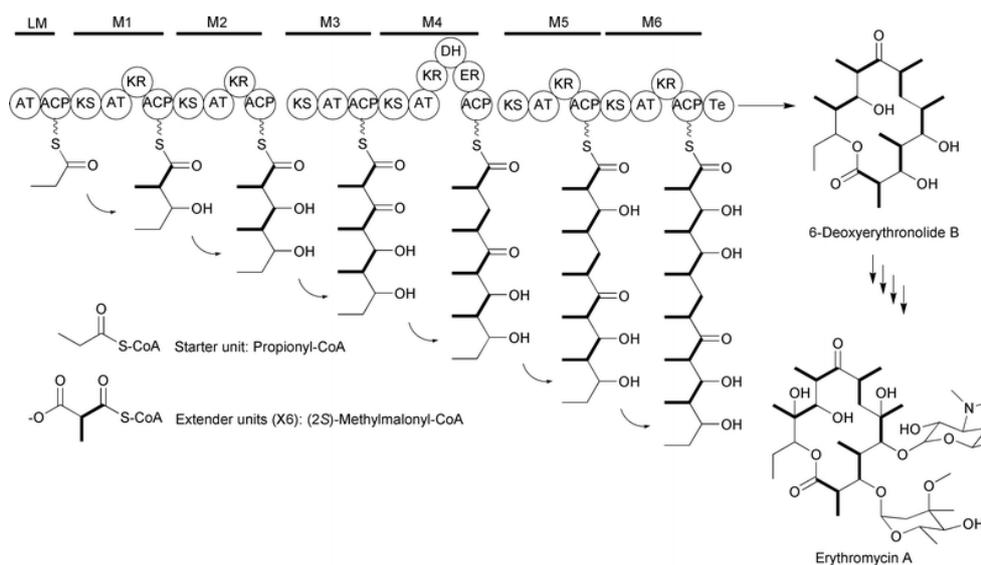


Figure 3.5 Deoxyerythronolide B synthase, a *cis*-AT type I modular PKS.

- Type II PKSs: they show active sites similar to those of type I PKSs, but they are distributed on different small monofunctional peptides.

They are involved in the biosynthesis of bacterial aromatic compounds such as doxorubicin (figure 3.6).

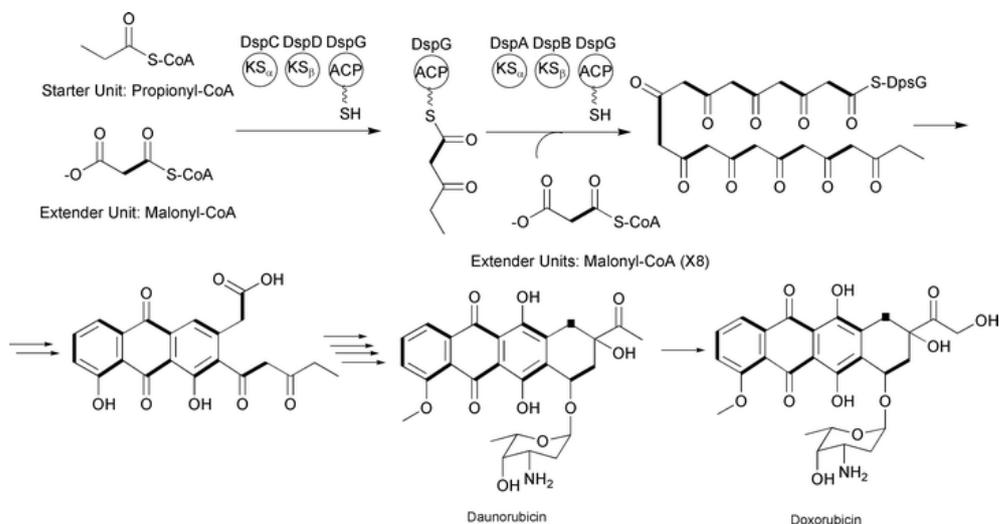


Figure 3.6 Doxorubicin type II PKS.

- Type III PKSs: iterative PKSs responsible for the production of chalcones and stilbenes in plants and polyhydroxy phenols in bacteria. Chalcone synthases consisting of a single polypeptide chain do not show ACP domain, and their substrates are directly CoA derivatives (figure 3.7).

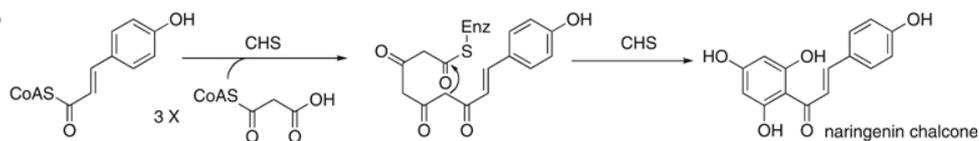


Figure 3.7 Chalcone synthetic pathway. CHS, chalcone synthase.

Modular PKSs are very interesting enzymes; they contain a sequence of separate modules; a module consists of non-repeated catalytic domains with defined function separated by short spacer regions. Each module is

responsible of one step of polyketide elongation and of modification of functional groups. The most common catalytic domains found in PKSs are:

- Acyltransferase (AT): this enzyme transfers the elongation group (malonyl or a malonyl derivative) from CoA to the ACP module. This module is responsible for the selection of the proper elongation group.
- Acyl Carrier Protein (ACP): this domain binds the elongation group and, after the Claisen condensation, the growing chain through the distal –SH group of the 4'-phosphopantetheine arm. This protein is biosynthesized as an inactive form and after translation binds to 4'-phosphopantetheine through a serine residue.
- Ketosynthase (KS): catalyzes the displacement of the growing polyketide chain from the ACP domain of the previous module to an –SH in the KS domain itself, and then the decarboxylative Claisen condensation with the elongation group linked to the ACP domain of the current module, leading to a β -ketoacyl.
- Ketoreductase (KR): this NADPH-dependent domain catalyzes the reduction reaction of the β -ketoacyl to a β -hydroxy acyl.
- Dehydratase (DH): this enzymatic subunit catalyzes the removal of a water molecule from the β -hydroxy acyl with the formation of a double bond.
- Enoylreductase (ER): catalyzes the reduction of the double bond produced by the DH domain.
- Thioesterase (TE): this domain is present only in the last module, and catalyzes the release of the polyketide from the last ACP domain.

All the modules are at least composed of a KS, an AT, and an ACP domain.

Moreover, specific combinations of KR, DH, ER and also epimerase (E)

and methyltransferase (MT) domains can be present in each module. This type of PKS, where each forming module contains the AT domain, is called *cis*-AT PKS. Recently, an architecturally unusual type of “AT-less” PKS has been discovered during studies about the secondary metabolism of bacteria and has been called *trans*-AT PKS.^[5,6] While revealing a multimodular organisation, this distinct type of PKS is composed by modules lacking the AT and receives building acyl blocks by a free-standing AT. There are only one to three AT enzymatic functions encoded by each gene cluster, either encoded on individual genes, fused as tandem ATs or fused with an oxidoreductase domain that serves as a *trans*-acting ER. The same ATs are therefore used iteratively to acylate each module, and due to this mechanism each PKS module is usually loaded with the same type of acyl unit: malonyl-CoA. In spite of the superficial similarities to *cis*-AT systems, *trans*-AT PKSs exhibit an impressive variety of module types and combinations. *Cis*-AT PKSs display only eight different domain sets (KS AT ACP, KS AT KR ACP, KS AT DH KR ACP, KS AT DH ER KR ACP and their corresponding combinations with MT domains), and due to the non-iterative nature of these enzymes, it is possible to predict the structure of the polyketide produced by a PKS on the basis of the number of modules of the PKS and the domains present in each module (“colinearity rule”). In contrast, more than 50 module variants have been reported for *trans*-AT PKSs. Modules with unusual domain orders, novel types of domains or repeated domain sets, modules split in various ways between two proteins as well as redundant and apparently useless modules occur in *trans*-AT PKSs,

breaking the colinearity rule so that it is more difficult to predict the polyketide structure from the domain organisation.

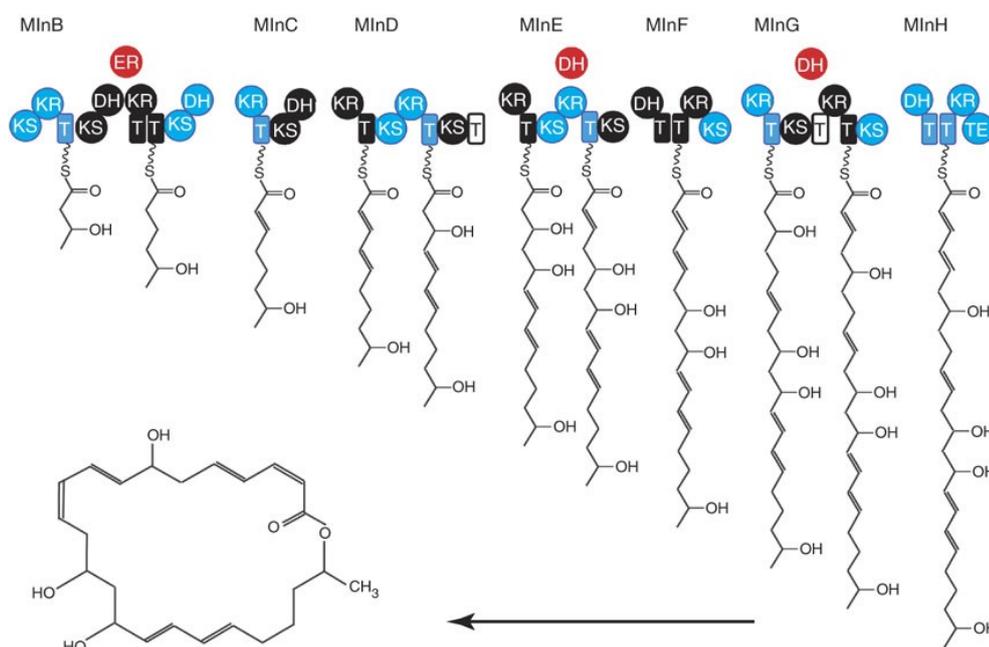


Figure 3.8 Macrolactin synthase, a *trans*-AT PKS.

Type I PKS modules can also work together with non-ribosomal peptide synthetases (NRPS) to yield complex biosynthetic pathways capable of producing a range of peptide metabolites with exotic chemistries. These mixed type I PKS-NRPS systems often occur in sponge metagenomes, and 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.

3.2 Proposed biogenetic pathway of plakortin

If, on the basis of the rule of colinearity, polyketide core structures can be deduced from gene sequences with high confidence, conversely it is also possible to predict the modular architecture of PKSs breaking up the chemical structures of a polyketide into its building acyl blocks. As discussed in previous section, plakortin is presumably biosynthesized by a *cis*-AT type I PKS.

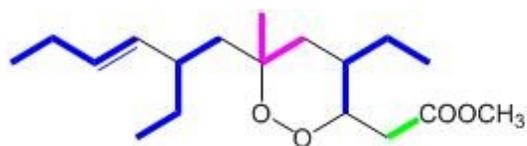


Figure 3.9 Biogenetic dissection of plakortin.

Looking at the structure of plakortin, it is quite apparent that carbon skeleton should be produced by condensation of three butyrate, one propionate, and one acetate units. Alternatively, the C₄ unit at the end of the molecule could arise from two acetate units instead of one butyrate. The reactions that lead to the formation of plakortin can be grouped into four cycles (corresponding to four modules of a PKS), each composed of several steps catalyzed by different catalytic domains. As in all modular and non-iterative PKS, the synthesis proceeds through a cascade mechanism and the reaction product of each reaction is the substrate for the next reaction.

Based on the above discussion, the reaction sequence for the biosynthesis of plakortin can be hypothesized as follow:

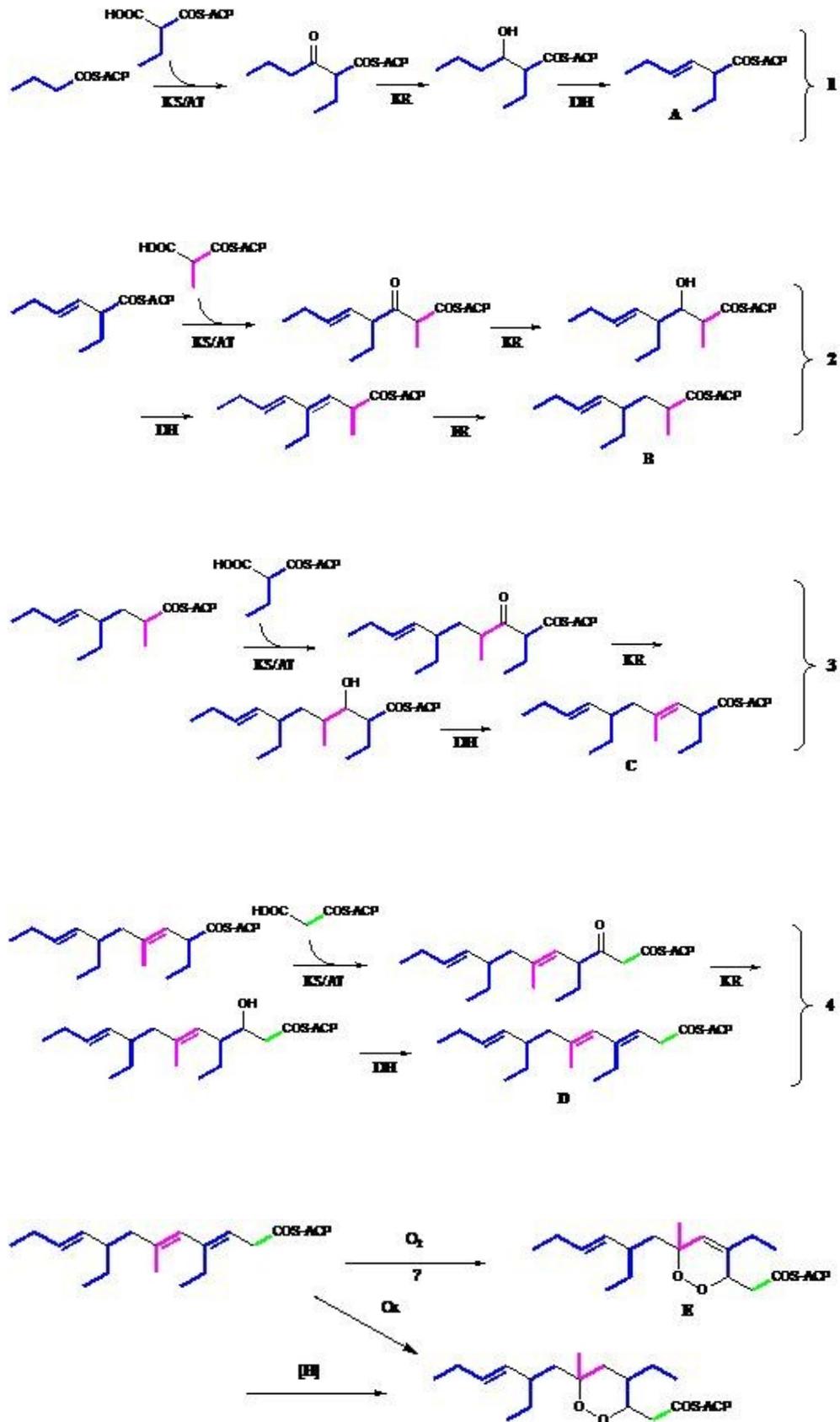


Figure 3.10 Proposed biogenetic pathway of plakortin.

Loading

The loading stage consists of three steps. First, the loading AT domain transfers the starter group from ethylmalonyl-CoA to the –SH group of the ACP domain, followed by decarboxylation to give butyryl-ACP.

Cycle 1

The elongation group is loaded by the current AT domain onto the ACP domain, yielding ethylmalonyl-ACP. The ACP-bound elongation group reacts with the butyryl residue bound the ACP of the starting module in a Claisen condensation and decarboxylation, catalyzed by the KS domain. The obtained β -ketoacyl group is reduced to β -hydroxy by the KR domain. Then, the DH domain splits off a water molecule, resulting in the α,β -unsaturated thioester (intermediate A) which is the substrate for the subsequent enzymatic reactions.

Cycle 2

The second cycle is similar to the first one, but in this case the AT domain loads a methylmalonyl-CoA molecule instead of ethylmalonyl-CoA onto the ACP module. The decarboxylative Claisen condensation (KS) is followed by reduction (KR), dehydration (DH) reactions and by the α,β double bond reduction, catalyzed by ER domain (intermediate B).

Cycle 3

The KS domain catalyzes the condensation between compound 2 and ethylmalonyl residue on the ACP domain. The β -ketoacyl group of the growing chain is reduced and the subsequent β -hydroxyacyl dehydration yields the intermediate C.

Cycle 4

The AT domain loads a malonyl-CoA onto the ACP domain. The Claisen condensation and decarboxylation (KS) between malonyl-ACP and the intermediate C followed by reduction (KR) and dehydration (DH) produces the intermediate D.

Oxidative Cyclisation

As in literature there are not comparable examples, it is difficult to hypothesize an accurate mechanism of oxidative cyclisation. The reaction could take place in the presence of singlet oxygen through a Diels-Alder type mechanism. This reaction yields the intermediate E and, after the reduction of the double-bond and hydrolysis of the thioester, plakortin.

References

1. Krishnaa S., Uhlemanna A., Haynes R. K. (2004). Artemisinin: mechanisms of action and potential for resistance. *Drug Resistance Updates* 7. 233–244.
2. Eckstein-Ludwig U., Webb1 R. J., van Goethem I. D. A., East J. M., Lee A.G., Kimura M., O’Neill P.M., Bray P.G., Ward S.A., & Krishna S. (2003). Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature*. Volume 424.957-961.
3. Makler M. T., Hinrichs D. J. (1993). Measurement of the lactate dehydrogenase activity of *Plasmodium falciparum* as an assessment of parasitemia. *Am J Trop Med Hyg*. 205-210.
4. Fattorusso E., Parapini S., Campagnuolo C., Basilico N., Taglialatela-Scafati O. and Taramelli D. (2002). Activity against *Plasmodium falciparum* of cycloperoxide compounds obtained from the sponge *Plakortis simplex*. *Journal of Antimicrobial Chemotherapy*. 883-888.
5. Nguyen T., Ishida K., Jenke-Kodama H., Dittmann E., Gurgui C., Hochmuth T., Taudien S., Platzer M., Hertweck C., Piel J. (2008). Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nature Biotechnology*. 225-233.
6. Piel J. (2010). Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Natural product reports*. 996-1047.

Chapter 4

Swf, a New Group of Mono-Modular Type-I Polyketide Synthases from Sponge Symbionts

Since cultivation of true sponge symbionts has failed in most cases, the search for the genes for plakortin had to rely on cultivation-independent techniques, such as the study of the collective genome of the sponge and its symbionts (e.g. the metagenome).

While the putative gene for plakortin biosynthesis could not be identified, an unexpected result from this study was the discovery of Swf, a new group of mono-modular type I PKS/FAS, which appears to be specifically associated to sponge symbionts.

Only one example of the cluster is present in the literature, but has been considered to be orthologous to the *wcb/rkp* cluster, which is not the case. The putative *swf* operon consists of *swfA* (FAS/PKS I), *swfB* (R and ST domains), and *swfC* (radical SAM). Activation of the ACP domain of the SwfA protein to its holo-form by co-expression with Svp is a first functional proof of *swf* type genes in marine sponges. However, the biosynthetic function of the *swf* cluster remains unknown.

Two different examples of the *swf* cluster were found in the metagenome of *P. simplex*, both very similar to a PKS previously identified in the genome from a single cell of a *Poribacterium* from the sponge *Aplysina aerophoba*. The *swf* cluster was then shown to be also present in three different, taxonomically distant sponge species. It represents the second group of

mono-modular PKS, after the *supA* family, to be ubiquitously present in marine sponges.

4.1 Isolation of the clusters

For the screening of a metagenome, degenerate PCR primers targeting conserved motifs of the ketosynthase (KS) domain of modular PKS (so-called type I PKS) have been successfully used in the past.^[1] However, in sponge metagenomes the search for PKS biosynthetic genes for secondary metabolites is severely hampered by a group of type I PKS that is present in sponge metagenomes at very high quantities and diversity. These dominant genes — named *sup* for sponge symbiont ubiquitous *pks* — usually compose ~85% or more of the total PCR amplicons obtained with the degenerate KS primers.^[2,3]

Therefore, a different approach was adopted, and degenerate primers AT1F and AT3R2 were designed, targeting the conserved regions FPGQGsQW and QGEIAAA, respectively, of the acyltransferase (AT) module of a type I PKS. Using these AT primers, we amplified and analyzed twelve AT domain DNA sequences (ca. 290 bp) from the metagenome of *P. simplex*. Even with the use of the AT primers, eight of these sequences showed high similarity with *sup* genes. However, the remaining four sequences (PSAT_PCR01, PSAT_PCR14, PSAT_PCR20, PSAT_PCR28) were not part of a *sup* cluster, and appeared closely related to each other. In a BLASTp search, the proteins deduced from these four sequences were all highly identical (77% to 84% identical) to the POR_0547 open reading frame (orf) from the genome of a sponge symbiont

(see below).^[4] All of the other BLASTp hits showed a remarkably lower similarity to any proteins in the databases. A large-insert 8,000-clone fosmid library was then constructed from the metagenome of *P. simplex* and screened by PCR using the AT primers. Two positive clones were detected, of which one (pPS23H10) was readily shown to be of the *sup* type, and not further studied, while the other (pPS11G3) was different and therefore completely shotgun sequenced. The PKS gene cluster was located on 2 contigs interrupted by a small gap, which was closed by primer walking. Because a portion of an outer ORF on the pPS11G3 insert, encoding a putative radical S-adenosyl methionine (radical SAM) enzyme, was missing, a larger metagenomic fosmid library of ~245,000 clones was constructed and screened for a fosmid containing the whole cluster using the specific primers 11G3_SAM. Two clones containing the PS11G3 cluster were found, and the sequence of the cluster was completed by primer walking.

Screening of this large library also led to the isolation of another fosmid, pPSA11D7, containing a second, different cluster (PSA11D7). This was also shotgun sequenced, and contained a PKS gene cluster truncated at the end by the vector sequence. The sequence of PSA11D7 was completed by primer walking after isolating further clones containing the cluster.

Conserved domain analysis of the PS11G3 and PSA11D7 inserts clearly showed that they encoded a type I PKS system. Recently, Hentschel and coworkers published a very similar PKS operon (ORFs POR_0547–POR_0550) isolated from a single bacterial cell of a ubiquitous sponge symbiont of the candidate phylum "Poribacteria". The sequence similarities

between homologous parts of the three PKS clusters are moderate to high (58-73% identity over ~1000 aa, Tables 2 and 3), and also the PKS domains are the same, although the arrangement of the ORFs is slightly different (Figure 4.1).

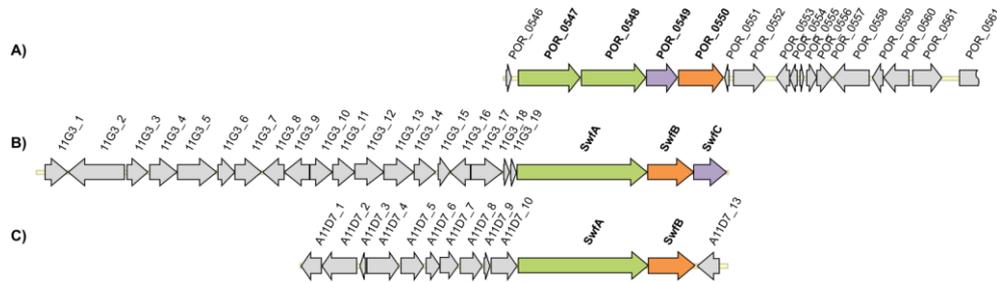


Figure 4.1 Genomic contexts of the *swf* gene clusters: A) POR_0547–POR_0550, B) PS11G3, and C) PSA11D7

In the absence of any close homologue, POR_0547 and POR_0548 had been designated in the original study as belonging to a WcbR-type PKS.

In spite of this classification, POR_0547 and POR_0548 share only low similarity with WcbR-type PKS and, in addition, the two ORFs reveal comparable low identity values with several characterized PKSs present in the genbank database.

The *wcb* clusters (and their orthologous *rkp* clusters)^[5,6,7] are involved in lipopolysaccharide biosynthesis, and are found in α -Proteobacteria which can induce root nodule formation in plants (for instance bacteria of the taxon Rhizobiales). Additionally, other bacteria living in symbiosis with plants contain these genes, especially β -Proteobacteria: the clusters in *Burkholderia* are designated as *wcb*, while those in *Nitrosomonas* (and

Rhizobium) are called *rkp*, most of the genes of *wcb* and *rkp* clusters being homologs.

Beside the type I PKS gene (*wcbR/rkpA*), the cluster comprises several genes for the biosynthesis and the export of capsular polysaccharides (e. g., *WcbA/rkpZ1/rkpZ2*, *WcbC/rkpU*, *WcbD/rkpR/kpsE*, *WcbO/rkpJ*, *wzm2/rkpT1*, *wzt/rkpS*), one acyl-CoA- or aminotransferase (*wcbT/rkpG*), one sulfatase (*WcbQ*, *rkpI*), one oxidoreductase/dehydrogenase (*WcbP/rkpH*), and in some cases a deacetylase (*WcbS*), glycosyltransferases (*wcbBEH*), and GDP-D-mannose dehydratase (*wcbK*).

However, the discovery of PS11G3 and PSA11D7 shows that the assignment of POR_0547 and POR_0548 to a *WcbR*-type PKS needs to be revised. The similarity to *WcbR* is only moderate (POR_0547: 96% coverage, 33% identity with NP_841435; POR_0548: 65% coverage, 35% identity with EGD01311). Furthermore, comparable similarities are observed to very different type I PKS enzymes, such as those involved in the secondary metabolism (e.g. 32% identity of POR_0548 with *AnaF* from the anatoxin-a synthetase gene cluster).

These data, together with the close homology between the three gene clusters, the analogy of their domains, and the absence of the other genes normally present in the *wcb* clusters, suggest that the clusters belong to a separate group of PKSs, with no particular relationship to the *wcb* cluster. The name *swf* (“sponge (symbiont) widespread fatty acid synthases”) has been proposed, in contrast to the *sup* genes, the other group of type I PKS/FAS that are widespread and abundant in sponge metagenomes.

4.2 *In silico* analysis of cluster sequences

4.2.1 Organization and genomic contexts of the swf gene clusters

Cluster PS11G3. The overall GC content of the contig containing the *swf* cluster is 67%. The variation of the GC content is from 55 to 75% suggesting multiple origins of the genes, although no genes for transposases/integrases were found. Only sequences upstream of PKS operon are known, which hampers a direct comparison with the POR contig where (with the exception of POR_0546) only ORFs downstream the PKS operon are known. BLASTp searches (Table 4.2) revealed no functional correlation of those ORFs with genes of the original *wcb/rkp* cluster. Solely, ORF15 and ORF18 (both anti anti sigma factors) were found to be frequently involved in capsular exopolysaccharide biosynthesis, for instance in the *Bacteroidetes/Chlorobi* group. The PKS gene (SwfA) has a very high homology to a fusion of POR_0547 and POR_0548. It is followed by SwfB, which encodes for thioester-reductase (R) and sulfotransferase (ST) domains (high homology to POR_0550), and by SwfC which encodes for a radical SAM (high homology to POR_0549). Interestingly, the two latter ORFs have changed their relative positions compared to the POR operon

Cluster PSA11D7. On the library fosmid pPSA11D7, the *swf* cluster was located on a contig containing 12 ORFs (Figure 4.1), with the *swf* cluster at the end. Like for PS11G3, no further ORFs orthologous to *wcb/rkp* were present (Table 4.1). Unlike the two other contigs, three integrases (ORF5-7) were present upstream the cluster, which is consistent with the variance from 55% to more than 80% in the GC content throughout

the contig. This *swf* operon contains only two ORFs (SwfA and SwfB), but a gene for a radical SAM, that would be homolog to SwfC (PS11G3) and POR_0549, is not present in PSA11D7. More sequence information about downstream ORFs is not available (Figure 4.1).

POR cluster. The cluster is contained in the contig c00157_POR, from the shotgun sequencing of the whole genome of a single Poribacterial cell by the group of Ute Hentschel, and comprises the ORFs POR_0546 to POR_0562 (Figure 4.1). The GC content of c00157 is on average 59.4% (with a variation from ~40-60% in a 200 bp window size). Only the sequences downstream the putative WcbR-like operon (POR_0546 to 0550) are known. All those ORFs have no correlation either with genes from the *wcb/rkp* clusters or with those from the contigs on PS11G3 and PSA11D7.

Table 4.1 Putative genes identified on the genomic fragment pPSA11D7 (*P. simplex*). Genes encoding enzymes in the “sponge (symbiont) widespread fatty acid synthases” (Swf) cluster are bolded.

ORF	Position [nt]	No. of aa	Putative function	Closest homolog (accession#) organism	expect value	identity/positives [% aa]
ORF1	1564-488	358	glycosyl transferase	ZOD2009_15156 (ZP_08045396), <i>Haladaptatuspaucihalophilus</i> DX253	5e-63	36/53
ORF2	3534-1570	654	hypothetical protein	Y11_36911 (CBY28839), <i>Yersinia enterocolitica</i> subsp. <i>palaearctica</i> Y11	3e-78	31/49
ORF3	3985-3695	96	hypothetical protein	sce5011 (YP_001615654), <i>Sorangium cellulosum</i> 'So ce 56'	2e-17	50/69
ORF4	4123-5823	566	HNH endonuclease	Anae109_1700 (YP_001378888), <i>Anaeromyxobacter</i> sp. Fw109-5	1e-14	54/71
ORF5	5926-7167	413	integrase/recombinase	ROS217_01170 (ZP_01038402.1), <i>Roseovarius</i> sp. 217	2e-114	49/66
ORF6	7161-8072	303	integrase/recombinase	XerC (P_002540039), <i>Agrobacterium vitis</i> S4	3e-103	55/70
ORF7	8069-9061	330	integrase/recombinase	KKY_3614 (YP_004901348), <i>Pelagibacterium halotolerans</i> B2	6e-171	72/86
ORF8	10431-9385	348	NAD dependent epimerase/dehydratase	Hoch_1681(YP_003266124), <i>Haliangium ochraceum</i> DSM	8e-36	35/48
ORF9	10469-10784	104	hypothetical protein	SeloA3_010100011807(ZP_09956060) <i>Sphingomonas elodea</i> ATCC 31461	7e-11	36/60
ORF10	10837-12291	484	sugar transporter	ED21_29266(ZP_0186498), <i>Erythrobacter</i> sp. SD-21	3e-79	37/53
SwfA	12302-19387	2361	type I PKS (KS AT DH ER KR ACP)	POR_0547 (ZP_06385951), Candidatus Poribacteria sp. WGA-A3 (aa 3-1111) POR_0548 (ZP_06385952), Candidatus Poribacteria sp. WGA-A3 (aa 1118-2345)	0.0	73/84 65/77
SwfB	19396-21894	832	R + ST domains	POR_0550 (ZP_06385954), Candidatus Poribacteria sp. WGA-A3	0.0	58/73
ORF13	22917-21928	329	Oxidoreductase	GobsU_010100005149 (ZP_02731160), <i>Gemmata obscuriglobus</i> UQM 2246	1e-49	35/52

Table 4.2 Putative genes identified on the genomic fragment pPS11G3 (*P. simplex*). Genes encoding enzymes in the “sponge (symbiont) widespread fatty acid synthases” (Swf) cluster are bolded.

ORF	Position [nt]	No. of aa	Putative function	Closest homolog (accession#) organism	expect value	identity/positives [% aa]
ORF1	515-1783	423	aminotransferase	HNE_2588 (YP_761278), <i>Hyphomonas neptunium</i> ATCC 15444	4e-44	31/51
ORF2	4873-1796	1025	TonB-dependent receptor	Acid_6465(YP_827674), <i>Candidatus Solibacter usitatus</i> Ellin6076	0.0	43/60
ORF3	5016-6143	375	hypothetical protein	SRM_00098 (YP_003569971) , <i>Salinibacter ruber</i> M8	4e-75	41/53
ORF4	6241-7743	500	D-glutamate deacylase	Acid345_3040 (YP_592115), <i>Candidatus Koribacter versatilis</i> Ellin345	4e-140	48/63
ORF5	7792-9927	711	Peptidase S9	A3SI_03433 (YP_002762158), <i>Nitritalea halalkaliphila</i> LW7	0.0	52/71
ORF6	10017-10904	295	β -lactamase	Rhom172_1530(YP_004825287), <i>Rhodothermus marinus</i> SG0.5JP17-172	4e-81	48/62
ORF7	10922-12394	490	β -lactamase	SinacDRAFT_5672 (ZP_09567976), <i>Singulisphaera acidiphila</i> DSM 18658	4e-79	43/56
ORF8	13621-12410	403	β -lactamase	SinacDRAFT_1225 (EHO65342), <i>Singulisphaera acidiphila</i> DSM 18658	4e-44	31/49
ORF9	14889-13618	423	hypothetical protein	HMPREF0765_1491 (ZP_03967296), <i>Sphingobacterium spiritivorum</i> ATCC 33300	5e-71	35/55
ORF10	15102-16220	372	FAD dependent oxidoreductase	Acid_5446 (YP_826678.1), <i>Candidatus Solibacter usitatus</i> Ellin6076	7e-99	51/62
ORF11	16245-17453	402	aminomethyltransferase	Acid345_1269 (YP_590345), <i>Candidatus Koribacter versatilis</i> Ellin345	2e-122	49/65
ORF12	17453-19066	537	FAD dependent oxidoreductase	Acid345_1270 (YP_590346), <i>Candidatus Koribacter versatilis</i> Ellin345	0.0	66/82
ORF13	19063-20667	534	FAD dependent oxidoreductase	Acid345_1270 (YP_590346), <i>Candidatus Koribacter versatilis</i> Ellin345	2e-130	41/60
ORF14	20708-21874	388	aminomethyltransferase	GB2207_03824 (ZP_01224676), <i>gamma proteobacterium</i> HTCC2207	2e-146	55/73
ORF15	21969-22625	218	anti-anti-sigma regulatory factor	(ACY25442), uncultured microorganism from <i>Aplysina aerophoba</i> (aa 13-111)	1e-15	40/61
			hypothetical protein	POR_0546 (ZP_06385950) <i>Candidatus Poribacteria</i> sp. WGA-A3 (aa 136-195)	1e-15	44/72
ORF16	23738-22662	358	alcohol dehydrogenase	HMPREF0017_00813 (ZP_06068892), <i>Acinetobacter lwoffii</i> SH145	2e-131	51/71
ORF17	23842-25515	557	permease	SupE (ABE03914), <i>Aplysina aerophoba</i> bacterial symbiont clone pAPKS18	2e-136	44/64
ORF18	25597-25944	115	anti-anti-sigma regulatory factor	(ACY25442), uncultured microorganism from <i>Aplysina aerophoba</i>	3e-13	35/58
ORF19	25987-26265	92	hypothetical protein	POR_0546 (ZP_06385950), <i>Candidatus Poribacteria</i> sp. WGA-A3	8e-21	48/72
SwfA	26422-33471	2349	type I PKS (KS AT DH ER KR ACP)	POR_0547 (ZP_06385951), <i>Candidatus Poribacteria</i> sp. WGA-A3 (aa 1-1108)	0.0	70/83
				POR_0548 (ZP_06385952), <i>Candidatus Poribacteria</i> sp. WGA-A3 (aa 1105-2326)	0.0	63/76
SwfB	33468-35987	839	R + ST domains	POR_0550 (ZP_06385954), <i>Candidatus Poribacteria</i> sp. WGA-A3	0.0	59/74
SwfC	36003-37760	585	radical SAM	POR_0549 (ZP_06385953), <i>Candidatus Poribacteria</i> sp. WGA-A3	0.0	65/79

4.2.2 Analysis of swf operons

The putative *swf* FAS/PKS operon consists of *swfA* (FAS/PKS I), *swfB* (R and ST domains), and *swfC* (radical SAM, not present in PSA11D7).

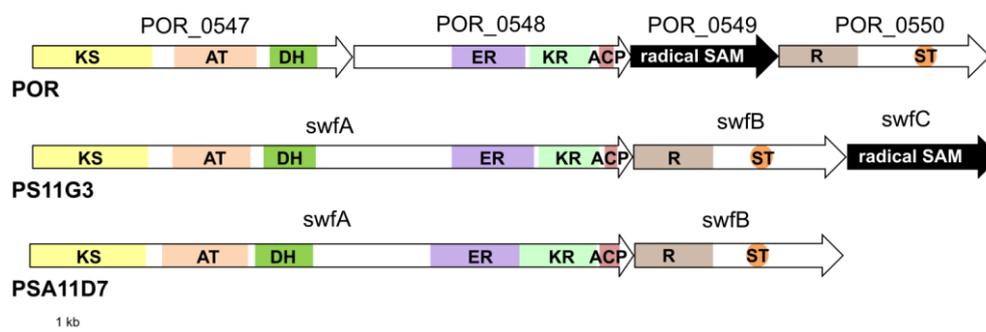


Figure 4.2 True to scale sketches of the domain organization of the *swf* gene clusters: A) POR_0547–POR_0550, B) PS11G3, and C) PSA11D7.

SwfA. The SwfA proteins are each composed of only one complete PKS module, indicative of either an iterative mode of action or a single elongation. The domain organization of SwfA is KS-AT-DH-ER-KR-ACP. From the domain organization of SwfA, a saturated acyl chain product is expected, although a (poly)unsaturated and/or (poly)hydroxylated acyl chain cannot be excluded, because in iterative PKSs the reduction domains can be optionally used during each of the elongation steps.^[1]

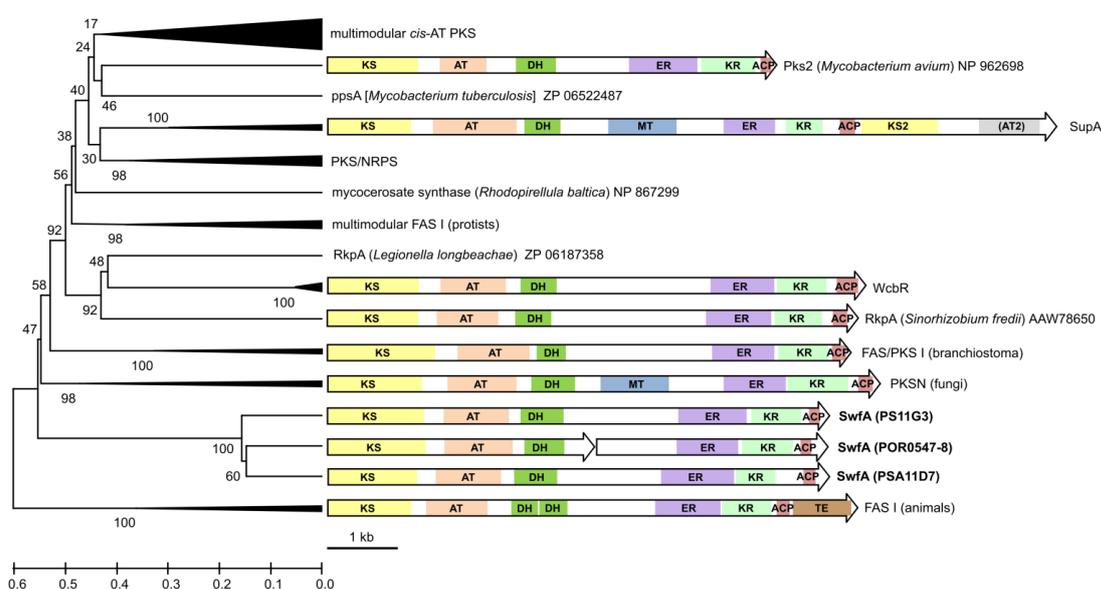


Figure 4.3 UPGMA tree showing phylogenetic relationships of KS-AT domains from different types of PKS and FAS enzymes, and in-scale sketches of domain architectures. Branches comprising multiple sequences have been collapsed and are illustrated as triangles; bootstrap values are shown at the nodes.

A phylogenetic comparison between SwfA and some related FAS/type I PKS proteins is shown in Figure 4.3. SwfA starts with a KS domain (~420 aa) with moderate homology to type I PKSs from various organisms. Phylogenetic analysis (KS tree rooted with the type II KS FabB from *E. coli*) shows a clear separation of the new sponge-derived SwfA enzymes from the WcbR/RkpA sequences (Figure 4.4). The KS domain of SwfA shows features that are not shared by any other KS domains, including WcbR/RkpA (Figure S1). In addition, the second “Q” in the DPQQR motif is replaced by either “I” or “V”, which is similar to animal FAS I and not to PKSs (all *cis*-AT and WcbR/RkpA sequences possess an intact DPQQR motif), and the motif HGTGT of *cis*-AT and WcbR/RkpA sequences is changed to HATGT. This explains why *swf* sequences may be underrepresented or absent in previous PCR screenings for PKSs in sponge metagenomes^[8, 9] which targeted the standard DPQQR and HGTGT motifs.

For the AT domain, the sequence of aa #722-725 of all SwfA homologs illustrate the motif IAFH, which suggests malonyl-CoA as the substrate^[10] (Figure S2), but the other substrate determining motif GHSSGE, present at aa 610-615, is unusual, because, according to Smith and Tsai^[11] an amino acid larger than Ser is normally present in malonyl-CoA specific AT domains. The QCALVEL motif (aa #591-597) is unique to SwfA_{AT}, and is also present (with some variation in the first and last aa) in the four *swf* AT fragments amplified from the metagenome of *P. simplex* with the AT1F/AT3R2 primers. It appears to be a well-conserved signature motif for *swfA* genes and was used to design primers to screen for the presence of the *swf* cluster in the metagenomes of other species of sponges. No useful

homology information could be obtained for the DH, ER, KR, and ACP domains. The BLASTp hits were heterogeneous and displayed only moderate to low homology.

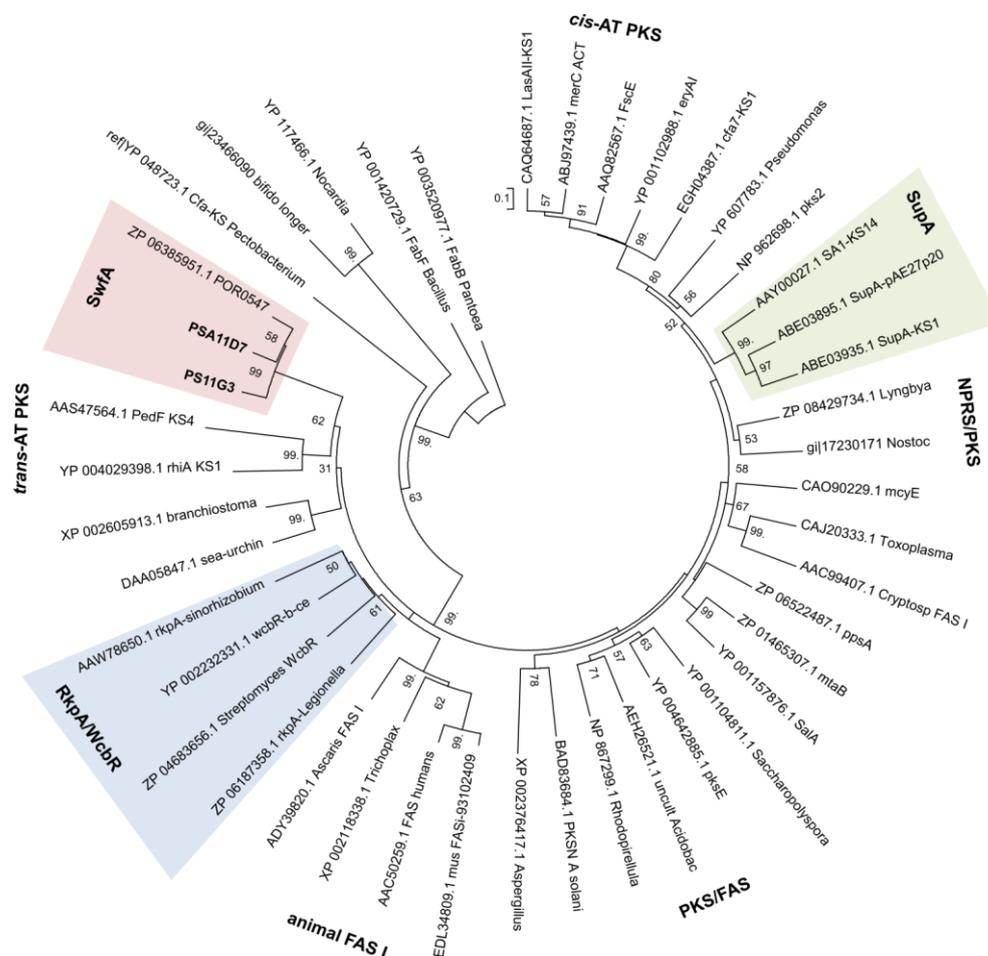


Figure 4.4 Neighbor-joining tree of full length KS domains from type I FAS, *cis*-AT PKS, *trans*-AT PKS, Sup (sponge symbiont ubiquitous pks), PKS/FAS, PKS/NRPS, RkpA/WcbR. The KS tree is rooted with the type II KS FabB from *E. coli*. Bootstrap values are given at the nodes.

SwfB. The ORF encodes a predicted fusion of an N-terminal thioester reductase (R) domain and a C-terminal sulfotransferase (ST) domain (figure 4.2). SwfB shows high similarity to the C-terminal two domains of two multimodular type I FAS of protists of the class *Coccidia* (Table 4.1 and 4.2), the only known homologues in which the R and ST domains are

contiguous as in SwfB. Because no functional studies have been conducted with this gene, the function of SwfB cannot be inferred.

R domains are frequently present in modular enzymes of fungi, myxobacteria, and cyanobacteria.^[12] They can reductively release the assembled chain as an aldehyde (often further elaborated to primary alcohol or amine),^[13] or they can be redox-inactive catalysts of carbon chemistry, for instance to form heterocycles.^[14] ST domains catalyze transfer of a sulfonate group to a hydroxy or amino group.^[15] The resulting sulfate group is usually found in the final metabolite, but sulfonation may also be a means to generate a good leaving group as reported for curacin A biosynthesis,^[16] in which sulfonation of a β -hydroxy-acyl-ACP is followed by thioesterase (TE) mediated hydrolysis, decarboxylation, and sulfate elimination to give a terminal alkene. For SwfB, the combination of R and ST modules could suggest conversion of the terminal carboxylate to a primary sulfate group.

SwfC. SwfC shows high similarity to a number of radical SAM enzymes. It has a predicted N-terminal (aa #12-129) vitamin B12 binding domain and a central (aa #170-346) radical SAM superfamily domain, including the characteristic CX₃CX₂C motif. Radical SAM enzymes can catalyze an array of different reactions.^[17, 18] The first event is always the extraction of a hydrogen radical from an unreactive C—H bond, but the final outcome of the reaction can vary greatly, including isomerizations, complex rearrangements, oxidations, and methylations. In a phylogenetic tree constructed using radical SAM proteins which have been functionally characterized, SwfC clusters with proteins that act as methyltransferases (Figure 4.5). Therefore, SwfC might act as a methyltransferase as well. The

closest homologue of SwfC in the tree is an enzyme (HpnP) responsible for 2-methylation of hopanoids.^[19] While a similar enzymatic activity must be present in some of the bacterial symbionts of *P. simplex*, because 12-methylhopanoids have been found in large amounts in the extract of this sponge,^[20] it appears unlikely that SwfC is involved in this specific activity.

Summarizing, SwfA is expected to assemble an acyl chain which is modified by SwfB and SwfC in unknown ways. While the expected product of elaboration of an acyl chain by SwfB would possibly be an alkyl sulphate that might be methylated by SwfC, no such metabolites are known to be present in *P. simplex*.

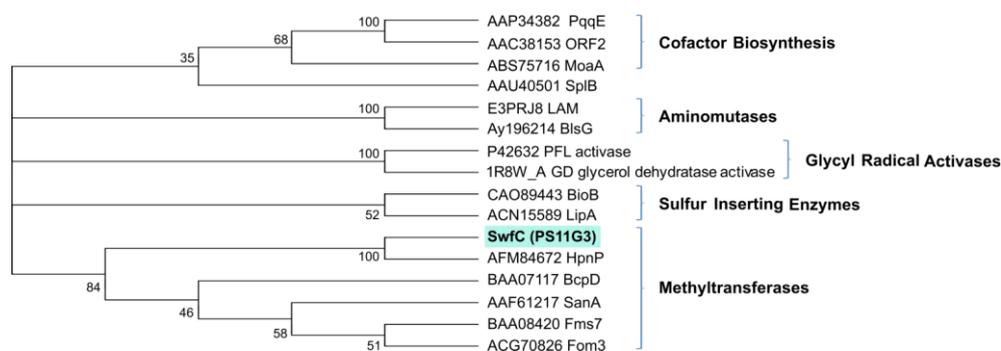


Figure 4.5 Neighbor-joining tree (condensed tree, cutoff value is 35%) displaying topology of radical SAM enzymes with known biological function. SwfC (PS11G3) clusters with radical SAM methyltransferases. Bootstrap values are given at the nodes.

4.3 Widespread diffusion of the *swf* cluster

For a preliminary analysis of the diffusion and diversity of the *swf* cluster, we designed the primers SWF_ATF and SWF_ATR from the conserved regions FSGQGTQW and QCALVEL, respectively, of the AT

domain of SwfA proteins. Using these primers, we amplified by PCR this fragment from the metagenome of *P. simplex* as well as from three "high microbial abundance sponges",^[21] namely *Aplysina fulva*, *Pseudoceratina crassa*, and *Smenospongia aurea*. In all cases, the PCR product showed the expected length (about 220 bp) for a *swfA* fragment (Figure 4.6). The PCR products were subcloned, and six clones from each species were cloned and sequenced. All the deduced protein fragments showed a very high similarity to SwfA; each amplicon showed 68% to 88% identity to each of the three SwfA homologues described above. In addition, a BLASTp search found POR_0547 as the first hit for all the sequences (E value < 10^{-26}), while all the other hits were heterogeneous and showed a much lower similarity (at most 51% identity, E value > 10^{-13}).

These data clearly show that all the amplicons were from *swfA* genes, and therefore, that the *swf* cluster was present in all the sponges studied. Therefore, the *swf* cluster appears to be widespread in marine sponges. Furthermore, in *Aplysina aerophoba*, *swf* has been shown to be hosted by *Poribacteria*, and the same may hold true for *P. simplex* (which is known to contain *Poribacteria*) and for the other three species studied, although no positive evidence about this exists at present.

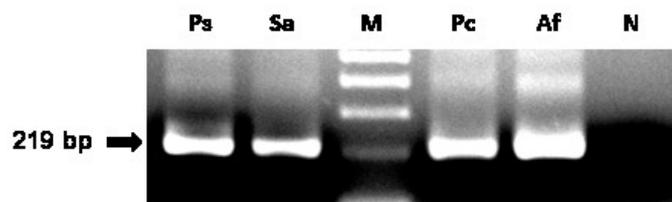


Figure 4.6 PCR detection of *swf* genes in four different "high microbial abundance" marine sponges. Primers SWF_ATF and SWF_ATR designed on SWF_{AT} were used. Abbreviations: Ps, *Plakortis simplex*; Sa, *Smenospongia aurea*; M, DNA size marker; Pc, *Pseudoceratina crassa*; Af, *Aplysina fulva*; N, negative control.

4.4 Functional study of the Swf enzymes

4.4.1 *In vivo* phosphopantetheinylation of SwfA_{ACP}

The ACP domain of SwfA (SwfA_{ACP}) was expressed with and without co-expression of a gene for the 4'-phosphopantetheinyl transferase (PPTase) Svp from the bleomycin-producing *Streptomyces verticillus* ATCC15003. The high-resolution ESI mass spectra of the two purified enzyme preparations were compared, to test whether the *apo*-SwfA_{ACP} could be activated to its *holo*-form *in vivo* (Figure 4.7). An array of peaks with different charges for each protein were observed; the $[M+6H]^{6+}$ peaks were the most intense, and were used for the subsequent measurements. HRMS allowed the accurate measurement of the monoisotopic m/z values for the two expressed proteins, each showing a cluster of isotopic peaks in which the $M+6$ ion was the most abundant. The monoisotopic m/z values measured for the $M+6$ isotopic peaks carrying a +6 charge were 1720.1949 and 1776.7079, respectively, for the *apo*- and *holo*-form, corresponding to masses of 10321.1694 amu for the *apo*-form and 10660.2474 amu for the *holo*-form (Figure 4.7). The measured mass of the *apo*-form is in excellent agreement with the calculated mass ($C_{443}^{13}C_6H_{710}N_{134}O_{140}S_3$ gives 10321.1888 amu, error 1.8 ppm). The mass of the *holo*-form is higher by 339.0780, which is exactly the expected change for a phosphopantetheine adduct. In this respect it should be noted that, while the experimentally well-established mass change for phosphopantetheinylation is 339^[22,23], the expected mass change should be 340, if one considers the molecular mass of phosphopantetheine (358 in the neutral dihydrogenophosphate form) and the

loss of a water molecule. A mass change by 339 corresponds to an additional $C_{11}H_{20}N_2O_6PS$ neutral fragment, which for the nitrogen rule must be an odd-electron fragment, and therefore implies a radical.

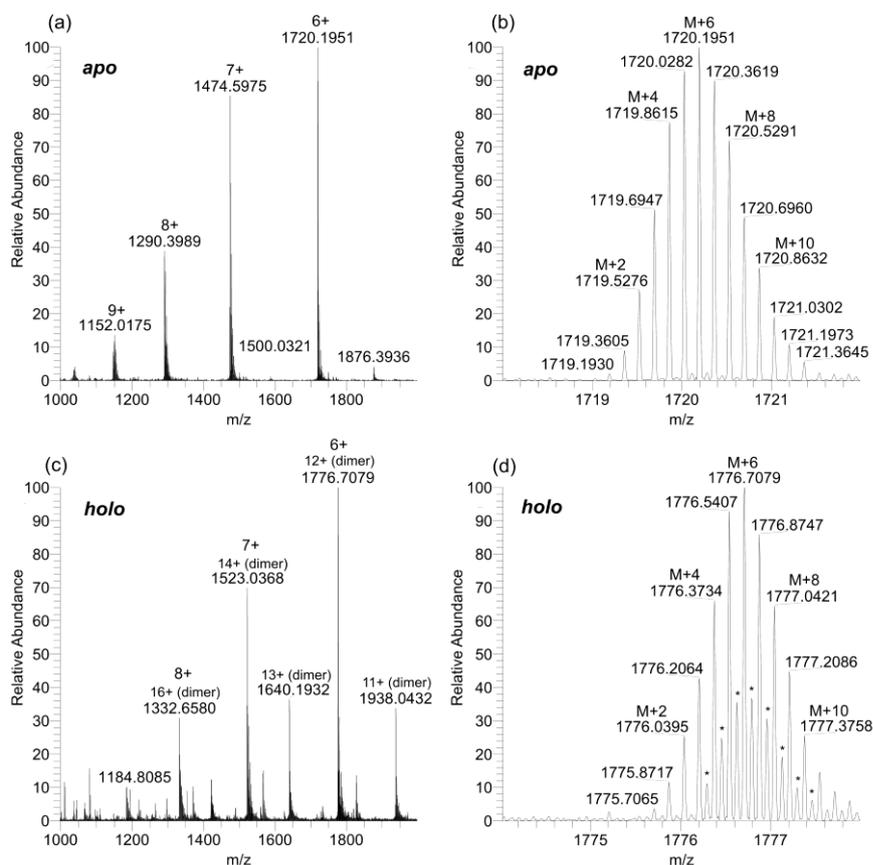


Figure 4.7 (a) High-resolution ESI MS spectrum of *apo*-SwfA_{ACP}. (b) Expansion of the 6-fold charged ion peak of *apo*-SwfA_{ACP} showing individual monoisotopic peaks. (c) High-resolution ESI MS spectrum of *holo*-SwfA_{ACP}. The peaks of the *holo*-protein are overlapped with those of its dimer (originating from a disulfide bond between phosphopantetheine S atoms). The dimer shows exactly the same *m/z* as the monomer when it has twice as many charges. Odd-charged peaks of the dimer are also present. (d) Expansion of the 6-fold charged ion peak of *holo*-SwfA_{ACP} showing individual monoisotopic peaks; the weaker peaks, denoted with asterisks, are the monoisotopic peaks of the 12-fold charged dimer. The difference between corresponding peaks of *apo*- and *holo*-SwfA_{ACP} accounts for a mass change of 339.0780, indicating a phosphopantetheine adduct.

Until now, in the literature, this discrepancy has never been previously noticed, nor explained. A possible explanation is that the terminal SH of the phosphopantetheine group may be oxidized yielding a disulfide dimer (this dimer was indeed observed in the ESI mass spectrum of *holo*-SwfA_{ACP},

(Figures 4.7). In the ESI source, the usual (poly)protonation reactions may be accompanied by homolytic cleavage of the disulfide bond, yielding two sulfur radicals.

The comparison of the mass spectra of SwfA_{ACP} with and without Svp co-expression showed that in the absence of Svp only the *apo*-form of the ACP was present. In contrast, the co-expression with Svp resulted in a complete conversion of the *apo*- into the *holo*-form. This is the first functional information about the *swf* cluster, demonstrating that the predicted ACP domain folds correctly, is recognized by PPTase, and is therefore presumably functional. In addition, these results demonstrate that ACPS (the PPTase of the *E. coli* expression host) is unable to catalyze phosphopantetheinylation, and therefore, that in functional studies of *swf* in *E. coli*, co-expression of the Svp PPTase is essential.

4.4.2 Heterologous expression and LC-MS analysis

For heterologous expression of the *swf* cluster, the whole cluster (PS11G3 type) was cloned in the expression vector pHIS8-Svp,^[24,25] yielding the recombinant plasmid pGS38. Homologous recombination was performed using *E. coli* BW25113 with the red helper plasmid pKD46, coding for the phage λ Red recombinase.^[26] For this purpose, the host strain was transformed with the donor fosmid pPS2D9 containing the whole PS11G3 sequence (see Section 6.2.17 for details) and the linear vector pHIS8-Svp modified at both ends by adding extensions homologous to the initial and the final regions of the *swf* cluster. Recombination in *E. coli* BW25113 was promoted by inducing expression of the λ Red recombinase

under the control of araBAD promoter, and allowing transfer of the *swf* cluster from the donor fosmid into pHIS8-Svp, resulting in the plasmid pGS38. After cloning, the new recombinant plasmid was transferred into *E. coli* BAP1^[27] by electroporation in order to perform heterologous expression under control of the T7 promoter.

Transformed cultures were grown in the presence of the inducer IPTG, harvested, and subsequently subjected to MeOH extraction. Additionally, the culture broths were collected, freeze dried, and extracted with MeOH. The MeOH extracts of the transformants containing the *swf* cluster and of their culture broths were analyzed by LC-HR-ESI-MS, and compared to the methanol extracts from negative controls (i.e. clones grown in presence of the inducer, but containing only the expression vector pHIS8-Svp without the *swf* insert). All the experiments were performed in triplicate, and both positive- and negative-ion mass spectra were recorded. Although even remarkable differences were observed in the relative amounts of some metabolites, the LC-MS data revealed no compounds which were present in all the transformants and absent in all the negative controls.

4.4.3 Heterologous expression and fatty acid analysis

E. coli BAP1 cultures transformed with the expression plasmid pGS38 were grown as described in the previous paragraph 4.4.2, harvested, and then subjected to lipid extraction. Fatty acids were characterized by saponification of the lipid extract and derivatization to fatty acid methyl esters (FAMES) followed by GC/MS analysis. Additionally, the culture broths were collected, freeze dried, and dissolved in nanopure water to be

subjected to fatty acid analysis following the protocol observed for bacterial pellets. The fatty acid composition of transformants containing the *swf* cluster was compared both to the lipid extract of clones grown in absence of the inducer IPTG and to that of clones including only the vector pHIS8-Svp (without the *swf* insert) and grown in the presence of IPTG. GC/MS measurements revealed no compounds which were present in all the transformants and absent in all the negative controls.

4.4.4 Heterologous expression of SwfA and SwfB

With the aim to perform heterologous expression of SwfA and SwfB, the relevant genes were cloned in the expression vector pHIS8-Svp by λ RED-mediated homologous recombination, yielding the recombinant plasmid pGS40. Homologous recombination was successfully realized in *E.coli* BW25113- pKD46 transformed with the donor fosmid pPS3I10 (containing *swfA* and *swfB* genes of the cluster PSA11D7) and the linearized vector pHIS8-Svp flanked at the ends by two fragments homologous respectively to the initial part of *swfA* and the final part of *swfB*. After cloning, the new recombinant construct pGS40 was transferred by electroporation into *E.coli* BL21-CodonPlus[®] (DE3)-RIPL cells transformed with the chaperon plasmid pTf16 in order to perform protein expression under control of the T7 promoter. Coexpression of pGS40 with the arabinose-inducible pTf16 was carried out as previous trials failed because of protein inclusion bodies formation. The chaperone plasmid encodes indeed for the Trigger factor molecular chaperone aiding the protein folding and consequently increasing protein solubility.

After purification of the protein extract from transformants applying the Ni-NTA technology, only weak expression of SwfA was observed through SDS-PAGE analysis. While being expressed as N-terminal His8-tag fusion protein (the vector pHIS8-Svp carries an octahistidyl tag) and showing increased solubility, SwfA was not retained by the Ni-NTA matrix and was discarded in the flow-through and washing elution fractions during chromatographic separation. In some cases, the His tag is partially or completely hidden by the tertiary structure of the native protein preventing its binding to the resin.

Thus, an incontrovertible result was that *swfA* gene was transcribed and translated in the heterologous host, outlining new perspectives for the heterologous expression of genes from sponge symbionts.

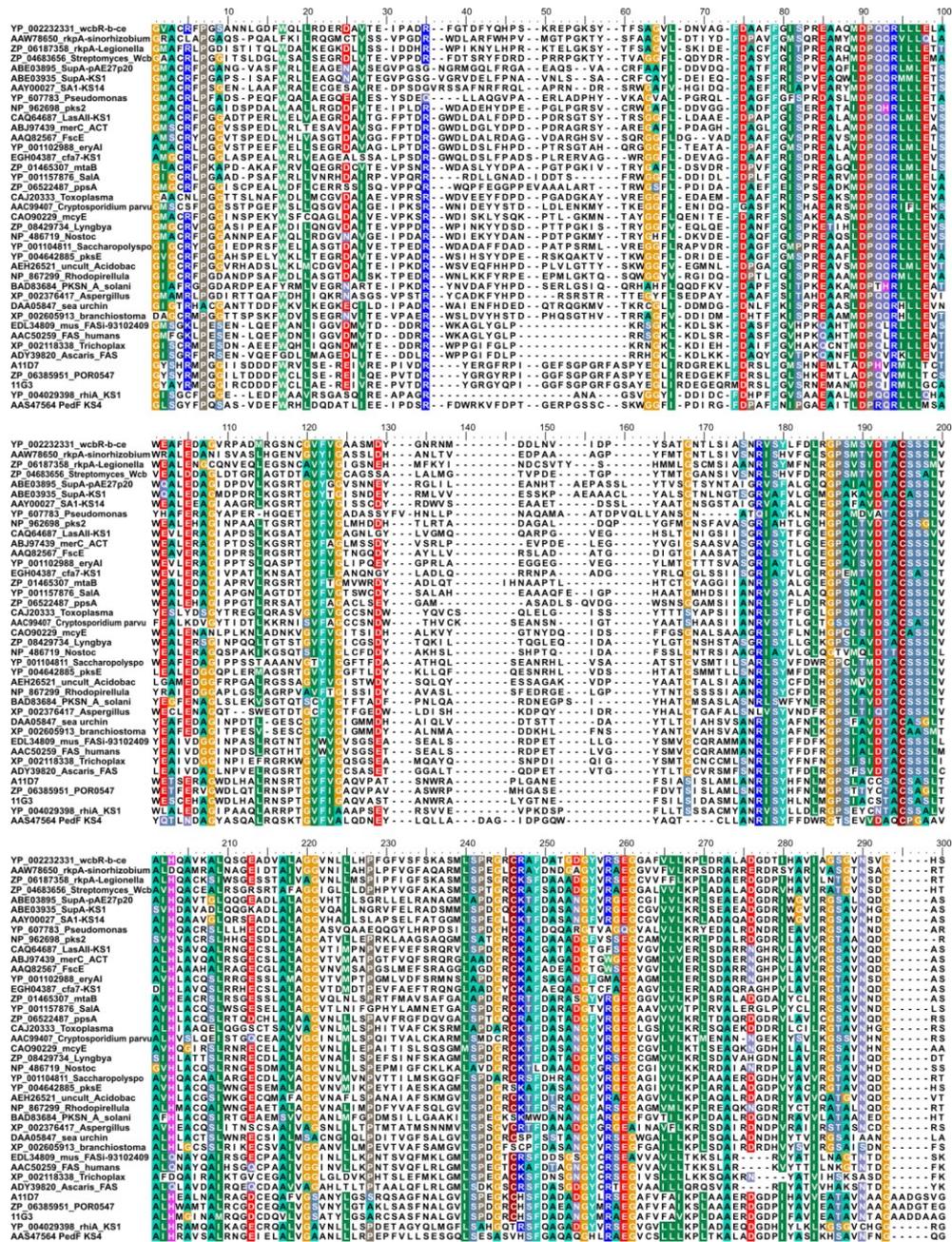


Figure S1. Alignment of KS domains from type I PKs and FASs. The KS domain of SwfA shows features that are not shared by any other KS domains, including WcbR/RkpA. These are, for instance, the triple “D” at position 8-10 aa, the “EATAVN” motif (aa 250-255) and 8 additional amino acids after aa 275, and the NGHCVVVR motif at the end (aa 413-419). In addition, the second “Q” in the DPQQR motif is either “I” or “V”, which is similar to animal FAS I and not to PKs (all *cis*-AT and WcbR/RkpA sequences possess an intact DPQQR motif), and the motif HGTGT of *cis*-AT and WcbR/RkpA sequences is changed to HATGT, with the A being unique among all KS sequences.

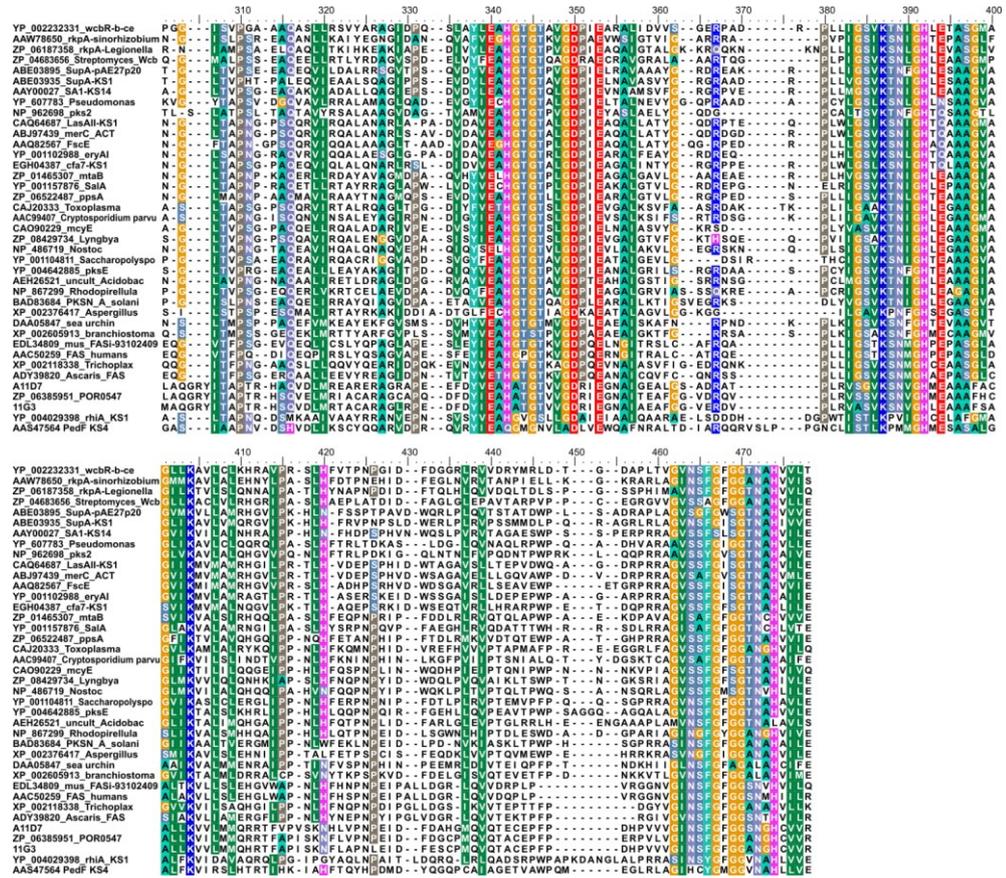


Figure S1 (continued)

References

1. Hertweck C. (2009). The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed.* 48: 4688-4716.
2. Piel J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc Nat Acad Sci U S A.* 99: 14002-14007.
3. Piel J., Hui D., Fusetani N., Matsunaga S. (2004). Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ Microbiol.* 6: 921-927.
4. Siegl A., Kamke J., Hochmuth T., Piel J., Richter M., Liang C. *et al.* (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J.* 5: 61-70.
5. Kiss E., Reuhs B. L., Kim J. S., Kereszt A., Petrovics G., Putnoky P. *et al.* (1997). The rkpGHI and -J genes are involved in capsular polysaccharide production by *Rhizobium meliloti*. *J Bacteriol.* 179: 2132-2140.
6. Parada M., Vinardell J. M., Ollero F. J., Hidalgo A., Gutierrez R., Buendia-Claveria A. M. *et al.* (2006). *Sinorhizobium fredii* HH103 mutants affected in capsular polysaccharide (KPS) are impaired for nodulation with soybean and *Cajanus cajan*. *Mol. Plant Microbe Interact.* 19: 43-52.
7. Donadio S., Monciardini P., Sosio M. (2007). Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep.* 24: 1073-1109.
8. Fieseler L., Hentschel U., Grozdanov L., Schirmer A., Wen G., Platzer M. *et al.* (2007). Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl. Environ. Microbiol.* 73: 2144-2155.

9. Schirmer A., Gadkari R., Reeves C. D., Ibrahim F., DeLong E. F., Hutchinson C. R. (2005). Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl. Environ. Microbiol.* 71: 4840-4849.
10. Reeves C. D., Murli S., Ashley G.W., Piagentini M., Hutchinson C. R., McDaniel R. (2001). Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry.* 40: 15464-15470.
11. Smith S., Tsai S. C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* 24: 1041-1072.
12. Zhu G., Shi X., Cai X. (2010). The reductase domain in a Type I fatty acid synthase from the apicomplexan *Cryptosporidium parvum*: restricted substrate preference towards very long chain fatty acyl thioesters. *BMC Biochem.* 11: 46.
13. Du L., Lou L. (2010). PKS and NRPS release mechanisms. *Nat. Prod. Rep.* 27: 255-278.
14. Liu X., Walsh C. T. (2009). Cyclopiazonic Acid Biosynthesis in *Aspergillus* sp.: Characterization of a reductase-like R* domain in cyclopiazonate synthetase that forms and releases cyclo-acetoacetyl-L-tryptophan. *Biochemistry.* 48: 8746–8757.
15. Chapman E., Best M. D., Hanson S. R., Wong C. H. (2004). Sulfotransferases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed.* 43: 3526-3548.
16. Gu L., Wang B., Kulkarni A., Gehret J. J., Lloyd K. R., Gerwick L., Gerwick W. H. *et al.* (2009). Polyketide decarboxylative chain termination preceded by O-sulfonation in curacin A biosynthesis. *J. Am. Chem. Soc.* 131: 16033-16035.
17. Wang S. C., Frey P. A. (2007). S-Adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends Biochem Sci.* 32: 101-110.

18. Roach P. L. (2011). Radicals from S-adenosylmethionine and their application to biosynthesis. *Curr. Opin. Chem. Biol.* 15: 267-275.
19. Welander P. V., Coleman M. L., Sessions A. L., Summons R. E., Newman D. K. (2010). Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proc. Natl. Acad. Sci. U S A.* 107: 8537-8542.
20. Costantino V., Fattorusso E., Imperatore C., Mangoni A. (2010). The first 12-methylhopanoid: 12-methylbacteriohopanetetrol from the marine sponge *Plakortis simplex*. *Tetrahedron.* 56: 3781-3784.
21. Hentschel U., Usher K. M., Taylor M.W. (2006) Marine sponges as microbial fermenters. *FEMS Microb. Ecol.* 55: 167-177.
22. Byers H.L., Ward M. A. (2003). Mass spectrometry. A powerful analytical tool. In *Proteomic and Genomic Analysis of Cardiovascular Disease*. Van Eyk J.E. and Dunn M.J. (eds.). Weinheim: Wiley-VCH Verlag. p. 206.
23. Sigma-Aldrich (2012). Mass changes resulting from typical post-translational modifications of proteins and peptides. [WWW document].
24. Jez J. M., Ferrer J. L., Bowman M. E., Dixon R. A., Noel J. P. (2000). Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry.* 39: 890-902.
25. Izumikawa M., Cheng Q., Moore B. S. (2006). Priming Type II Polyketide Synthases via a Type II Nonribosomal Peptide Synthetase Mechanism. *J. Am. Chem. Soc.* 128: 1428-1429.
26. Datsenko K. A., Wanner B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U S A.* 97: 6640-6645.
27. Pfeifer B. A., Admiraal S. J., Gramajo H., Cane D. E., Khosla C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science.* 291: 1790-1792.

Chapter 5

Diversity of polyketide synthase genes from *P.simplex*

Since metagenomic library screenings have proven to be ineffective in finding plakortin genes and/or other bioactive polyketide pathways, a new strategy was adopted to access the PKS metabolism of the Caribbean sponge *P. simplex*. Metagenomic analyses were undertaken to determine the polyketide synthase (PKS) diversity and explore the biosynthetic potential of the sponge associated microorganisms by next generation sequencing (454 pyrosequencing) of complex and heterogeneous PCR products amplified from the metagenomic DNA with degenerate probes targeting ketosynthase and acyltransferase domains.

5.1 454 pyrosequencing

The 454 pyrosequencing^[1] technology was developed by 454 Life Sciences as a new, highly parallel two step DNA sequencing system with significantly greater throughput than the Sanger sequencing systems. It is based on the “sequencing by synthesis” principle which involves utilizing single strand DNA, to be sequenced, and sequencing its complementary strand with enzymatic action. It is a system integrated with PCR amplification of numerous DNA fragments linked to high throughput parallel pyrosequencing in picolitre-sized wells. The “sequencing by synthesis” principle relies on the detection of pyrophosphate (PPi) released on nucleotide incorporation, generating a light signal, rather than chain

termination with dideoxynucleotides. 454 pyrosequencing is a highly parallel two step approach. The DNA is firstly prepared by cutting the DNA into blunt ends and attaching oligonucleotides adaptors to both ends of the cut DNA molecules. These adaptors provide priming sequences for both amplification and sequencing of the sample-library fragments. One adaptor (Adaptor B) contains a 5'-biotin tag for immobilization of the DNA library onto streptavidin-coated beads. The fragments are then individually attached to a bead, which is then amplified via PCR in droplets of an oil-water emulsion, generating multiple copies of the same DNA sequence on each bead. The beads are then captured in picotitre wells on a fabricated substrate and then pyrosequenced.

The sequencing occurs in several steps, using the Genome Sequencer FLX system. Firstly, the template, a single stranded PCR amplicon, is hybridized to a sequencing primer and then incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase as well as the substrates adenosine 5' phosphosulfate (APS) and luciferin. The first dNTP is then added and if complementary, it is bound to the DNA strand by DNA polymerase and this is accompanied by the release of a PPi. The PPi is converted to ATP by ATP sulfurylase in the presence of APS and then, a luciferase catalysed reaction generates visible light proportional to the amount of the generated ATP. Apyrase continuously degrades unincorporated nucleotides and ATP. After the degradation process, another nucleotide is added. As the process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the pyrogram trace. The strengths of pyrosequencing technology

are represented by the high sensitivity such as by the amount of sequence information obtainable. The system sequences 400-600 million bp in 10 hours and gives short-length reads of 300-500 nucleotides.

454 pyrosequencing technology enables a variety of applications including sequencing of whole genomes, target DNA regions, amplicons, metagenomes and transcriptomes.

5.2 Phylogenetic analysis of KS amplicons

The metagenomic DNA of *P. simplex* was used as template for PCR screening with the degenerate primers KSDPQQF/KSHGTGTR targeting high conserved motifs in KS domains of polyketide synthases, with the aim to study in depth the polyketide metabolism of the marine sponge.

The research of polyketide genes from sponge metagenome applying the KSDPQQF/KSHGTGTR degenerate primers, as widely documented in the literature, is critically hampered by a huge amount of widespread, highly sponge-specific group of polyketide synthases called SupA (“sponge (symbiont) ubiquitous polyketide synthases”).^[2] Then, *supA* KS fragments are definitely predominant among the PCR amplicons, with the consequent risk to overlook underrepresented KS amplicons if traditional techniques for the analysis of PCR products are applied. Therefore, the 454-pyrosequencing-method was adopted in order to gain a comprehensive view of the complex and heterogeneous PCR amplicon mixture.

Next generation sequencing of the PCR products generated 19333 reads; sequences shorter than 200 nucleotides were excluded from the BLASTx

analysis because not long enough to allow a correct identification. As expected, BLASTx analysis of KS amplicons revealed as first remarkable result the high abundance of *sup* genes in *P. simplex* metagenome: almost 68% of the analyzed sequences were clearly KS fragments belonging to this sponge-specific group of PKSs. The data led to an estimation of 246 different *sup* KS fragments, with duplicates and close orthologues eliminated at a 5% similarity cut off. In addition, a high number (~12.4%) of sequences containing one or more frameshifts could be included into the *sup* PKS group. It is not known so far whether these mutations actually occur in the metagenome of *P. simplex*, or they are simple artifacts.

Five amplicons were obviously orthologous to *wcb/rkp*^[3,4,5] cluster as deduced from the top-matched sequences in a BLASTx search, represented by KS belonging to putative *wcb*-PKS from different sponges such as *A. aerophoba* (DQ996390_uncultured bacterium clone 2m14 ketosynthase gene) and *D. dissoluta* (AY897162_spT-KS63g). Two representative fragments of this group, PS_W3FZ and PS_W8E9, were included in the phylogenetic analysis (figure 5.1) confirming BLASTx matches and showing close relationship with *wcbR*-KS from bacteria of the order Rhizobiales (*Sinorhizobium meliloti* and *Burkholderia pseudomallei*), as supported by high bootstrap value (99%).

Further examples of type-I-PKS-like FAS were amplified from *P. simplex* applying the degenerate primers KSDPQQF/KSHGTGTR. Two amplicons, PS_X0DW (present in duplicate) and PS_O68N shared high similarity with FAS respectively from *Saccoglossus kowalevskii* (Hemichordata) and *Branchiostoma floridae* (Cephalochordata), as

underscored by E values (respectively, $1e-30$ and $9e-39$) of the BLASTx search. Besides, one amplicon, PS_THOB, present in triplicate in the PCR amplicon mixture, found its closest homologue (E value $4e-64$; max identity 75%) in the partial ketosynthase domain from an uncultured bacterium of the marine sponge *D. dissoluta* (AY897165_spT-KS37Ai), which clusters with trans-AT PKSs according to the phylogenetic analysis reported in the relevant paper.^[6] This data is in contrast with the phylogenetic tree (figure 5.1) reported below where the sequence appears to be related to type-I-PKS-like FAS and is closely related to FAS from protists as underscored by a bootstrap value of 92%. Two other KS fragments, PS_T094 and PS_O23H, shared high homology with KS belonging to 3-oxoacyl carrier protein synthases of the type-II fatty acid synthase system. Particularly, BLASTx search showed that PS_T094 found its closest relatives in *fabH* genes from α - and γ -*Proteobacteria*; on the other hand, PS_O23H matched among the BLAST hits mainly with *fabF* genes from marine γ -*Proteobacteria*.

Besides the presence of *sup*, *fas* and *wcb/rkp* genes, involved in the primary metabolism of sponge symbionts, the 454-pyrosequencing-method allowed the detection of eight KS fragments possibly responsible for the biosynthesis of secondary metabolites in *P. simplex*. Among these eight amplicons, one is a partial KS domain of an hybrid NRPS-PKS, as deduced from the BLASTx analysis where the first hit was a beta-ketoacyl synthase from a bacterium of the order Rhizobiales, *Methylosinus trichosporium* OB3b (NZADVE01000087; E value $1e-68$, identity 87%), which appeared located in the genome between two regions coding for aminoacid

adenylation domain proteins. Running sequence analysis in the NAPDOS^[7] database (where only PKSs of known function are included), the previous result was confirmed: the query sequence matched the JamM (AAS98784) hybrid KS domain (E value $1e-35$, identity 58%) of the jamaicamide biosynthetic pathway.^[8] The remaining seven partial type I PKS fragments resembled basically modular *cis*-AT PKSs, as shown by the phylogenetic trees (figure 5.1 and 5.2) and by a BLASTx search. The sequence PS_WUGN shared high identity with putative partial beta-ketoacyl synthases from bacteria of the genera *Streptomyces* (best E value $5e-53$) and *Sorangium* (best E value $3e-52$). Interesting similarities were also found with different actinobacterial and myxobacterial biosynthetic gene clusters, including the neuroprotectant meridamycin^[9] from *Streptomyces* (E value $1e-47$; max identity 63%) and the antimycobacterial macrolide thuggacin A from *Sorangium cellulosum*^[10] (E value $6e-48$; max identity 62%). Five further partial KS sequences (PS_SA3U, PS_WH8C, PS_V7ZG, PS_W828 and PS_UZ2Z), among the other matches, provided similarities to various PKS genes from the phylum *Cyanobacteria*. Significant hits were also found for all the five fragments, with many different cyanobacterial and myxobacterial biosynthetic gene clusters such as the antitubulin metabolite curacin A from the cyanobacterium *Lyngbya majuscula*^[11], the electron transport chain inhibitors myxalamid^[12] and stigmatellin from the myxobacterium *Stigmatella aurantiaca*^[13], the hepatotoxin nodularin from the marine cyanobacterium *Nodularia spumigena*^[14] and the cryptophycins which are potent anticancer agents from *Nostoc cyanobionts*.^[15] The identity with the matched *cis*-AT KS domains of these PKS genes was between 37%

and 71%. Among these five KS fragments, PS_WH8C and PS_W828 appeared to be related also to putative KS forming parts of PKS genes from Eukaryota, with the top matched sequences in the BLASTx results being respectively the MlnB PKS from the protist *Salpingoeca sp. ATCC 50818* (E value= $1e-28$, max identity= 46%; accession no EGD79036) and an hypothetical protein from the red algae *Aureococcus anophagefferens* (E value = $1e-36$, max identity = 56%; accession no EGB07663).

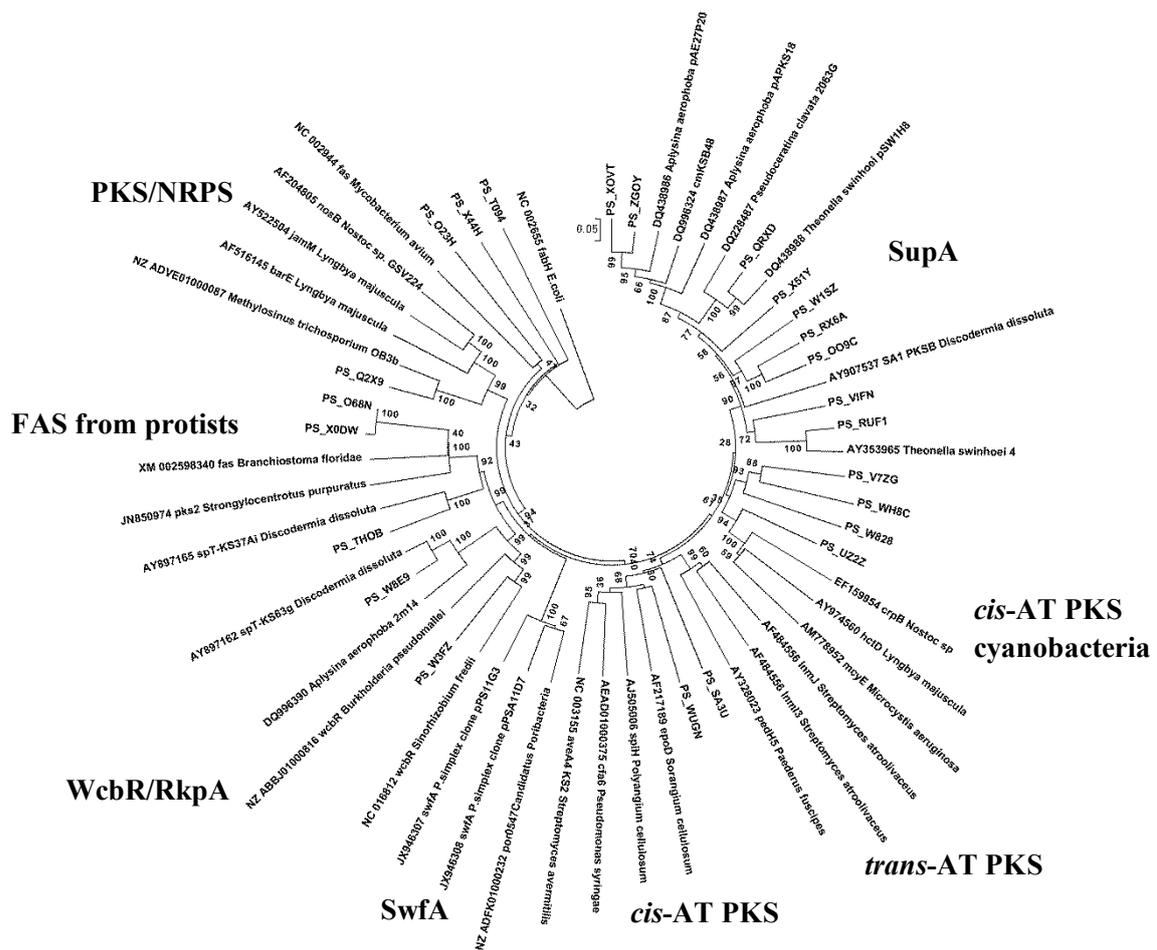


Figure 5.1 Neighbor-joining tree of full length KS domains from type I FAS, *cis*-AT PKS, *trans*-AT PKS, Sup, PKS/FAS, PKS/NRPS, RkpA/WcbR, Swf. KS fragments from *P. simplex* are included in the tree. The KS tree is rooted with the type II KS FabH from *E. coli*. Bootstrap values are given at the nodes.

The last fragment, PS_X44H, must belong to an unusual KS, because BLASTx analysis showed only low-score matches (best E value $2e-19$) with

putative beta-ketoacyl synthases from various strains of the genus *Streptomyces*. Moreover, BLAST hits conflicted with the phylogenetic tree where the PS_X44H is included in the clade of cyanobacterial *cis*-AT modular KS domains with low bootstrap values (figure 5.2).

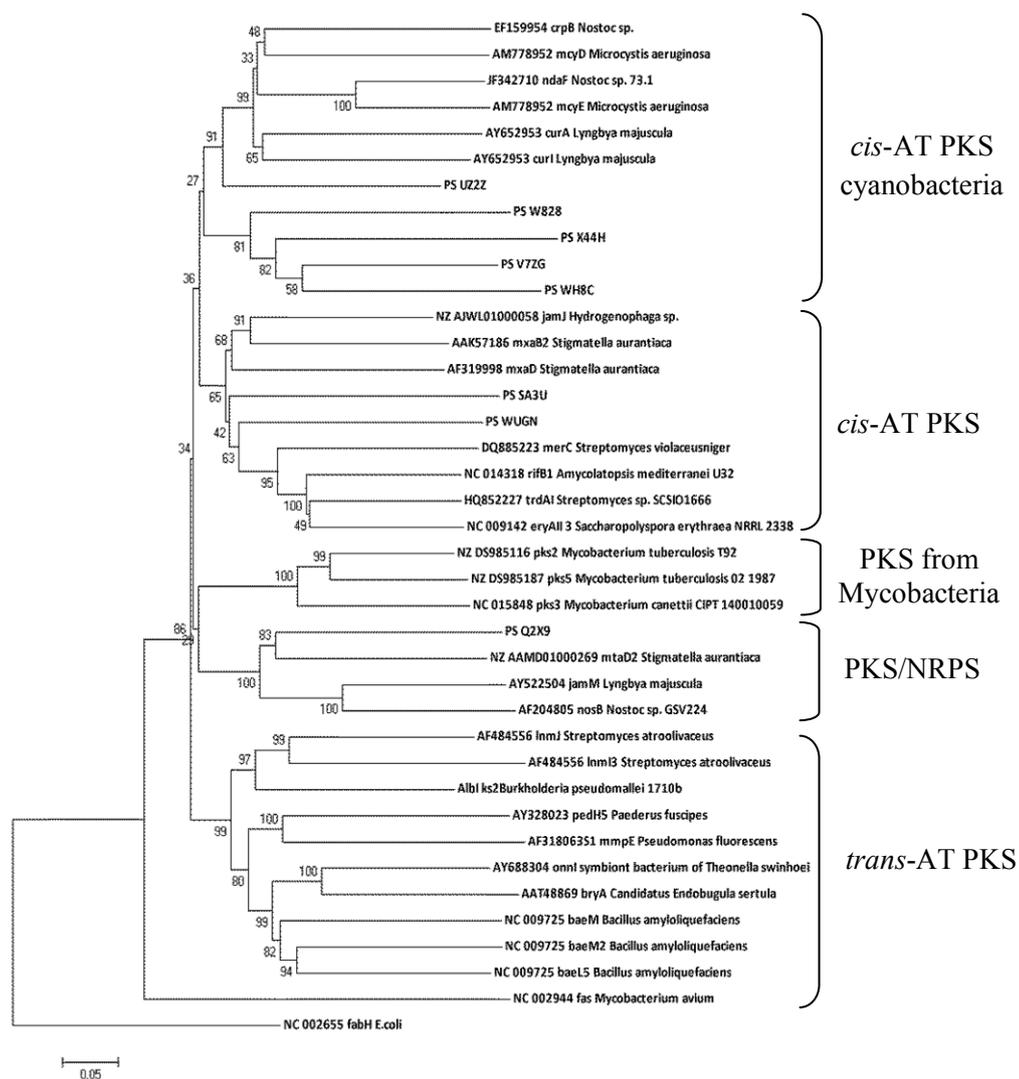


Figure 5.2 Neighbor-joining tree of full length KS domains from PKS of known function. KS fragments from *P.simplex* are included in the tree. The KS tree is rooted with the type II KS FabH from *E. coli*. Bootstrap values are given at the nodes.

The vast majority of PKS genes detected through PCR screenings of many sponge metagenomes is represented by Sup enzymes, unusual small monomodular polyketide synthases, widely and exclusively distributed in

“high microbial abundance” (HMA) sponges where up to half biomass can consist of microbial symbionts. *P. simplex* is a HMA sponge, and indeed its metagenome was characterized by the widespread occurrence of *sup* genes, as observed also for the marine sponges *Aplysina aerophoba*, *Theonella swinhoei*, *Agelas dilatata*, *Ircinia felix*, *Discodermia dissoluta* and *Xestospongia muta*.^[16]

Although the general architecture of Sup enzymes traces typical features of mammalian fatty acid synthases, they are phylogenetically related to PKSs. The domain organization of *sup* PKS is KS-AT-DH-MT(methyltransferase)-ER-KR-ACP-KS-AT. From the domain organization of SupA, a saturated acyl chain product is expected which is methylated by the methyltransferase, and so *sup* genes are presumably involved in the biosynthesis of methyl branched fatty acids (MBFAs). Sponges are indeed one of the richest known sources for mid-chain-branched FAs (MBFAs), and Sup enzymes have been proposed as candidates for the biosynthesis of these lipids, commonly recognized as bacterial metabolites. Furthermore, recent studies revealed on one hand the presence of *sup* genes in the genome from a single cell of a *Poribacterium* symbiontically associated with the sponge *Aplysina aerophoba*,^[17] and on the other hand a positive correlation between the presence of *Poribacteria* and *sup* genes in HMA sponges rich in MBFAs.^[16] These data gave further support to the hypothesis of poribacterial Sup enzymes being involved in biosynthetic pathway of methyl branched fatty acids.

The biological functions of methyl-branched fatty acids in sponge symbionts are unknown. Interestingly, *Mycobacterium tuberculosis* mutants

lacking mycocerosate (2,4,6,8-tetramethyl C32 fatty acid) and other cell wall lipids exhibited an attenuated growth in various animal hosts, indicating that these compounds are crucial for the infection process and adaptation to the host. It is an interesting question whether the *sup* genes may play a similar role, in that they might be essential for the establishment and maintenance of symbiosis in sponges.^[18]

Besides the widespread diffusion of *sup* genes, a second group of PKS genes was detected in the metagenome of *P. simplex*, generally classified as lipopolysaccharide type-I-PKS-like FASs, occurring in various marine sponges, and orthologous to *wcbR/rkpA* genes. Prototypical Wcb/Rkp clusters are indeed known from those α -*Proteobacteria* that are capable to induce root nodules in plants (for instance bacteria of the taxon *Rhizobiales*) and also from other bacteria (especially β -*Proteobacteria*) living in symbiosis with plants. The produced capsular polysaccharides can be regarded as virulence factors not only for the infection of plant roots, but also of animal and human cells. An example for this is the disease melioidosis caused by *Burkholderia pseudomallei*.^[19]

All the eight putative KS fragments (one hybrid NRPS-*cis*-AT PKS and seven *cis*-AT PKSs) involved in the biosynthesis of secondary metabolites from *Plakortis simplex* are significantly different to each other (E values $\geq 10^{-6}$) and BLASTx analysis as well as the rebuilt phylogenetic taxonomy revealed that they are only distantly related to PKSs of characterized function. In addition, phylogenetic analyses suggest that these KS fragments are mainly related to PKSs from *Cyanobacteria*, *Actinomycetes* and *Myxobacteria*, commonly known as precious sources of bioactive

polyketides. In spite of the huge number of KS amplicons obtained from the metagenome of *P. simplex*, apparently no *trans*-AT KS domain could be identified.

Marine sponges of the genus *Plakortis* are known for the production of large amounts of polyketide peroxides, of which the anti-malarial plakortin is the most abundant. Although plakortin is an abundant metabolite in the Caribbean sponge *Plakortis simplex*, a high represented KS amplicon could not be detected within the heterogeneous PCR mixture deep sequenced by 454 next generation sequencing. However, these eight sequences represent KS belonging to new PKSs, and therefore potential candidates for the biosynthesis of novel pharmaceutically bioactive compounds. Therefore, they surely deserve further studies aimed at the identification of the whole clusters which they are part of, and at the investigation of their biosynthetic role. The diversity of PKS genes from the metagenome of *P. simplex* represents one of the most important aspects of the sponge PKS metabolism, and even if these PKS genes may not produce bioactive drugs, the discovery of new PKS modules is useful for providing modules for combinatorial polyketide synthesis via gene recombination techniques.

Symbiont bacteria are commonly recognized as the main producers of a wide array of polyketides. However, considering that several KS fragments amplified from *P. simplex* as well as from other marine sponges show, among the other BLAST hits, matches with PKS genes from protists and algae, it can be argued that bacteria are not always the unique and ultimate source of sponge polyketides.

5.3 Phylogenetic analysis of AT amplicons

With the aim to find novel PKS genes, hopefully including the plakortin gene cluster, as well as to detect acyltransferases of type-I PKS specifically recruiting unusual substrates, the metagenomic DNA from *P. simplex* was used as template for PCR-based screening with degenerate primers AT1F/AT3R2 targeting conserved motifs of acyltransferase domains of type-I PKS enzymes. According to the hypothesis about the biosynthesis of plakortin (see paragraph 3.2), this polyketide indeed could be synthesized by a PKS system including AT domains selective for ethylmalonyl-CoA. Therefore, the 454-pyrosequencing-method was adopted also in this case to gain a detailed survey of the partial AT fragments amplified from *P. simplex* metagenome.

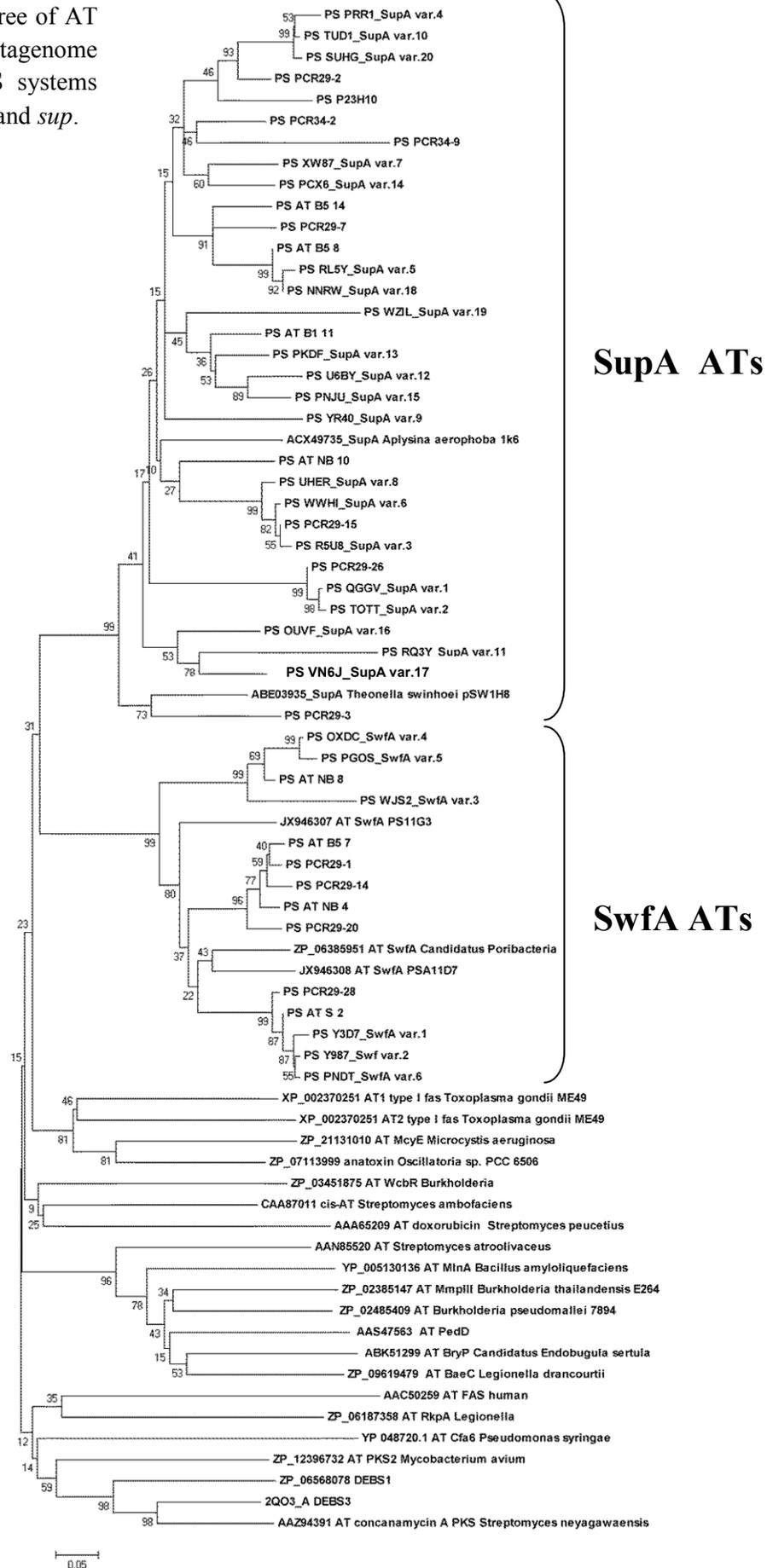
Next generation sequencing of the PCR mixture generated 8995 reads; however, also with this modern approach no other PKS/FAS than the known SupA and SwfA could be found. Almost 51% of the analyzed sequences belonged to the Swf enzymes; six different sequences related to this monomodular PKS were found, with duplicates and close orthologues eliminated at a 5% similarity cut off. Only the 4% of the total reads was represented by AT forming parts of SupA enzymes; particularly, 20 different sequences orthologous to the *supA* AT group were detected. The remaining reads appeared not to be related to AT domains (~ 45% of the amplicons). Obviously, the target amino acid motifs FPGQGsQW and QGEiAAA, recognized by the primers AT1F/AT3R2, are not specific only to PKS/FAS genes, and sequences that shared similar motifs or highly

abundant sequences with less similar motifs could also be amplified by PCR.

In the phylogenetic tree where acyltransferases from PKSs of known function are included together with representatives of the AT amplicons from *P. simplex*, the formation of two distinct clades referred to the two major ubiquitous sponge PKS systems was clearly observed (figure 5.3). In the phylogenetic tree, two AT forming parts of *supA* genes from other HMA sponges (*Theonella swinhoei* and *Aplysina aerophoba*) were included, and they cluster together with *supA* AT fragments from *P. simplex* in spite of the taxonomic divergence of the three species. Furthermore, the shallow branching topology of the SupA subclade was worthy of note, suggesting a common organism source of these genes. So far, this feature has been noticed only for the SupA KS-amplicons.^[18] So, these findings confirm what had been deduced from phylogenetic analyses of *supA* KS domains: a group of ubiquitous sponge symbionts common to all HMA sponges are the host of these PKSs and, as previously discussed, *Poribacteria* are the main candidates for this role.

In addition, it is worth noting that the Swf subclade shows the same features as the SupA subclade, displaying not deep branching in the phylogenetic tree. Therefore, if we consider also the widespread occurrence in HMA sponges of *swf* genes (see section 4.3) as well as the detection of the *swf* genes in *Poribacteria* from *Aplysina aerophoba*, the analogy between the *swf* and *sup* clusters appears to be more than evident.

Figure 5.3 Neighbor joining tree of AT amplicons from *P. simplex* metagenome displays the two major PKS systems within sponge symbionts, *swf* and *sup*.



References

1. Next Generation Sequencing Technologies: 454 Pyrosequencing. <http://www.biotecharticles.com>.
2. Fieseler L., Hentschel U., Grozdanov L., Schirmer A., Wen G., Platzer M. *et al.* (2007). Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl. Environ. Microbiol.* 73: 2144-2155.
3. Kiss E., Reuhs B. L., Kim J. S., Kereszt A., Petrovics G., Putnoky P. *et al.* (1997). The rkpGHI and -J genes are involved in capsular polysaccharide production by *Rhizobium meliloti*. *J. Bacteriol.* 179: 2132-2140.
4. Parada M., Vinardell J. M., Ollero F. J., Hidalgo A., Gutierrez R., Buendia-Claveria A. M. *et al.* (2006). *Sinorhizobium fredii* HH103 mutants affected in capsular polysaccharide (KPS) are impaired for nodulation with soybean and *Cajanus cajan*. *Mol. Plant Microbe Interact.* 19: 43-52.
5. Donadio S., Monciardini P., Sosio M. (2007). Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep.* 24: 1073-1109.
6. Schirmer A., Gadkari R., Reeves C. D., Ibrahim F., DeLong E. F., Hutchinson C. R. (2005). Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl. Environ. Microbiol.* 71: 4840-4849.
7. Ziemert N., Podell S., Penn K., Badger J. H., Allen E., Jensen P. R. (2012). The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One*.
8. Edwards D. J., Marquez B. L., Nogle L. M., McPhail K., Goeger D. E., Roberts M. A., Gerwick W. H. (2004). Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* 6: 817-833.

9. He M., Haltli B., Summers M., Feng X., Hucul J. (2006). Isolation and characterization of meridamycin biosynthetic gene cluster from *Streptomyces* sp. NRRL 30748. *Gene*. 377: 109-18.
10. Buntin K., Irschik H., Weissman K. J., Luxenburger E., Blöcker H., Müller R. (2010). Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. *Chem. Biol.* 4: 342-56.
11. Chang Z., Sitachitta N., Rossi J. V., Roberts M. A., Flatt P. M., Jia J., Sherman D. H., Gerwick W. H. (2004). Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* 8: 1356-67.
12. Silakowski B., Schairer H. U., Ehret H., Kunze B., Nordsiek G., *et al.* (1999). New lessons for combinatorial biosynthesis from *Myxobacteria*. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* 274: 37391–37399.
13. Gaitatzis N., Silakowski B., Kunze B., Nordsiek G., Blocker H., Hofle G., and Muller R. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* 277: 13082–13090.
14. Moffitt M. C., Neilan B. A. (2004). Characterization of the Nodularin Synthetase Gene Cluster and Proposed Theory of the Evolution of Cyanobacterial Hepatotoxins. *Appl. Environ. Microbiol.* 70: 6353-6362.
15. Magarvey N. A., Beck Z. Q., Golakoti T., Ding Y., Huber U., Hemscheidt T. K., Abelson D., Moore R. E., Sherman D. H. (2006). Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts. *ACS Chem. Biol.* 1 (12): 766-79.
16. Hochmuth T., Niederkruger H., Gernert C., Siegl A., Taudien S., Platzer M., Crews P., Hentschel U., and Piel J. (2010). Linking Chemical and Microbial Diversity in Marine Sponges: Possible Role for Poribacteria as Producers of Methyl-Branched Fatty Acids. *ChemBioChem.* 11:2572-2578.

17. Siegl A., Kamke J., Hochmuth T., Piel J., Richter M., Liang C. *et al.* (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J.* 5: 61-70.
18. Fieseler L., Hentschel U., Grozdanov L., Schirmer A., Wen G., Platzer M. *et al.* (2007). Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl. Environ. Microbiol.* 73: 2144-2155.
19. Warawa J. M., Long D., Rosenke R., Gardner D., Gherardini F. C. (2009). Role for the *Burkholderia pseudomallei* capsular polysaccharide encoded by the web operon in acute disseminated melioidosis. *Infect. Immun.* 77(12): 5252-61.

Chapter 6

Materials and Methods

6.1 Materials

6.1.1 Sponge collection

The sponges *P. simplex*, *A. fulva*, *P. crassa*, and *S. aurea* were collected by scuba diving at depths of 5–15 m offshore Little San Salvador Island and/or Grand Bahamas Island, Caribbean Sea, Bahamas. Individuals were cut into pieces and immediately stored in five volumes of RNA later (Life Technologies) stabilization solution. The samples were kept at –20 °C until shipped to the laboratory, then the stabilization solution was removed and the samples were kept frozen at –80 °C until used.

6.1.2 Chemicals

Table 6.1 Chemicals

Chemical	Source
1 kb DNA extension ladder	Invitrogen
1 kb DNA ladder	Roth
5-Bromo-4-chloro-3-indoxyl- β -D-galactopyranoside (X-Gal)	Roth
Acetonitrile	Sigma
Acrylamide-bisacrylamide 37, 5:1(Rotiphoresegel 30)	Roth
Adenosine triphosphate	5 PRIME
Agar	Roth
Ammonium acetate	Sigma
Ammonium chloride	Sigma
Boric Acid	Sigma
Brilliant-Blue R250	Roth
Bromophenol blue sodium salt	Roth
Calcium chloride	Sigma
Chloroform	Sigma
Copper (II) sulfate	Roth
CTAB	Fluka
D(+)-Glucose anhydrous	Fluka

Chemical	Source
Desoxynucleotide (dNTPs)	5 PRIME
Disodium hydrogenphosphate dehydrate	Sigma
Ethanol	Sigma
Ethylenediaminetetraacetic acid (EDTA)	Sigma
Formic acid	Sigma
Glacial acetic acid	Sigma
Glutammic acid	Roth
Glycerin	Sigma
Glycine	Roth
Hexane	Sigma
Imidazole	Sigma
Iron (III) chloride	Roth
Isopropanol	Sigma
Isopropylthiogalactoside (IPTG)	Sigma
L-Arabinose	Sigma
Magnesium chloride hexahydrate	Roth
Magnesium sulfate heptahydrate	Roth
Mercaptoethanol	Sigma
Methanol	Sigma
Midori Green	Genetics
MTBE (methyl <i>tert</i> -butyl ether)	Roth
N, N-Dimethylformamide (DMF)	Sigma
Ni-NTA Agarose	5 PRIME
Peptone	Roth
Potassium acetate	J.T. Baker
Potassium chloride	Roth
Potassium hydroxide	Sigma
Protein marker RotiMark Standard	Roth
PVP (Polyvinylpyrrolidone)	Sigma
Seaprep agarose	Lonza
Silica gel	Roth
Sodium acetate	Roth
Sodium chloride	Roth
Sodium dihydrogenphosphate dihydrate	Roth
Sodium dodecylsulfate (SDS)	Sigma
Sodium hydroxide	Roth
Sodium molybdate dehydrate	Fluka
Sodium sulphate anhydrous	Roth
Strontium chloride hexahydrate	Sigma
TBE buffer	Bio-Rad
Tetramethylethylenediamine (TEMED)	Roth
Tryptone	AppliChem
Xylencyanol	AppliChem
Yeast Extract	Fluka

6.1.3 Chemical solutions

Table 6.2 Chemical solutions

Solution and buffer	Composition	Store
Destaining protein solution	30% methanol 10% acetic acid	room temperature
Gel loading dye	0.05% bromophenol blue 0.05% xylene cyanol	4°C
Protein gel electrophoresis buffer 10X	250 mM Tris-HCl 2.5 M glycine 1% SDS	room temperature
Protein gel loading dye 3X	2.4 mL Tris-HCl 1 M 3 mL SDS 20% 3 mL glycerol 1.6 mL mercaptoethanol 0.006g bromophenol blue	4°C
Staining protein dye	30% methanol 10% acetic acid 1.25 g Brilliant-Blue R250	room temperature
TAE 50X	242 g Tris-base 57.1 mL Glacial acetic acid 100 ml EDTA (0.5 M) Ad 1 L H ₂ O	room temperature
TE	10 mM Tris-HCl (pH 7.5) 1 mM EDTA (pH 8.0)	room temperature
IPTG	20 mg/mL	-20°C
X-gal	20 mg/mL	-20°C
SB 20X	50 g boric acid add sodium hydroxide until pH 8.3 Ad 1 L H ₂ O	room temperature
M9 salts	64 g Na ₂ HPO ₄ ·7 H ₂ O 15 g KH ₂ PO ₄ 2.5 g NaCl 5.0 g NH ₄ Cl Ad 1 L H ₂ O	room temperature
Trace elements 2000X	1.2 g FeCl ₃ · 6 H ₂ O 1.4 g MnSO ₄ 1.6 g CuSO ₄ Ad 500 mL H ₂ O	room temperature
Bacteria cell lysis buffer pH8	50 mM Tris-HCl 500 mM NaCl 10 mM MgCl ₂	room temperature

6.1.4 Enzymes

Table 6.3 Enzymes

Enzyme	Source
Kapa T4 DNA ligase	Kapa Biosystems
Lysozyme	Sigma
Proteinase K	Sigma
Restriction enzymes	Takara, Promega
RNase A	ApplyChem
Shrimp alkaline phosphatase (SAP)	Promega
<i>Taq</i> DNA polymerase	RBC Bioscience

6.1.5 Vectors and bacteria

6.1.5.1 Vectors

Table 6.4 Vectors

Vector	Source
pBluescript SK II (+)	Stratagene
pCC1FOS	Epicentre
pHIS8	Derivative of pET28a, Novagen
pKD46	Nature Technology Corporation

6.1.5.2 Bacteria

Table 6.5 Bacteria

Bacteria	Genotype
<i>E. coli</i> BL21 (DE3) BAP1	<i>[F' ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) λ(DE3) [lacI lacUV5-T7 gene1 ind1sam7nin5] ΔprpRBCD::T7prom-sfp, T7prom-prpE</i>
<i>E. coli</i> BL21 CodonPlus (DE3) RIPL	<i>E. coli B F⁻ ompT hsdS(rB⁻ mB⁻) dcm⁺ Tetr gal λ(DE3) endA Hie [argU proL Cam^r] [argU ileY leuW Strep/Sp^e]</i>
<i>E. coli</i> BW25113	<i>lacIqrrnBT14ΔlacZWJ16hsdR514 ΔaraBADAH33ΔrhaBADLD78</i>
<i>E. coli</i> EPI300™-T1 ^R	<i>F-mcrAA(mrr-hsdRMS-mcrBC) Φ80dlacZAM15 ΔlacX74 recA1</i>
<i>E. coli</i> XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZAM15 Tn10(Tetr)]</i>

6.1.6 Antibiotics

Table 6.6 Antibiotics

Antibiotic	Concentration	Source
Ampicillin	100 µg/mL	Sigma
Carbenicillin	100 µg/mL	Roth
Chloramphenicol	12.5 µg/mL	Fluka
Kanamycin	50 µg/mL	Roth
Tetracycline	12.5 µg/mL	Roth

6.1.7 Bacterial cultivation media

LB: 10 g tryptone, 10 g NaCl, 5 g yeast extract. Added 1 L H₂O, pH 7.

LB agar: 10 g tryptone, 10 g NaCl, 5 g yeast extract, 15 g agar. Added 1 L H₂O, pH 7.

Semi-liquid SeaPrep medium: 10 g tryptone, 10 g NaCl, 5 g yeast extract, 800 mL H₂O, 5 g SeaPrep agarose. The mixture was put in microwave for 5 min. Added up to 1 L H₂O, pH 7.

All the previous media were autoclaved at 121°C for 20 min.

MMGAGtr: 200 mL M9 salts and 500 µL 2000 X trace element solution were added to 800 mL H₂O. The solution was autoclaved at 121°C for 20 min. Then, 2 mL MgSO₄ (1 M, sterile), 100 µL di CaCl₂ (1 M, sterile), 10 g filter-sterilized glucose in 50 mL H₂O, 5 g filter-sterilized glutamic acid in 150 mL H₂O were added to the autoclaved solution. Sterile water was added to get 1L final volume.

6.1.8 Others

- CopyControl- Fosmid Library Production Kit (Epicentre)
- QIAquick[®] Gel Extraction Kit (Qiagen)
- QIAprep[®] Spin miniprep kit (Qiagen)

6.1.9 Equipment

Table 6.7 Equipment

Equipment	Manufactory
Agarose/acrilamide gel electrophoresis	Bio-Rad
Autoclave	Smeg
Balance	Gibertini

Equipment	Manufactory
Barnstead Easy Pure II	Thermo Scientific
Blue Light Transilluminator	Bio View
C-1000 Thermal Cycler	Bio-Rad
Cellophan papers	Roth
Centrifuge, cool, type Z 36 HK	Hermle
CHEF-DR®III Pulsed Field Electrophoresis Systems	Bio-Rad
Clean bench	Thermo Scientific
Electroporation cuvettes 0.2 cm	Bio-Rad
Electroporator Micropulser	Bio-Rad
Gel Doc™ XR	Bio-Rad
Glass wool	Roth
Incubator	Fratelli Galli, Carlo Erba
Incubator Shakers	innova® 42, Thermo Scientific
Microcentrifuge Z 233 M2	Hermle
Microcentrifuge, cool, type Z 216 MK	Hermle
Microwave	DeLonghi
Nanodrop 2000 C	Thermo Scientific
Pipettes	Socorex, Gilson
Poly-prep columns	Bio-Rad
SpeedDry Rotational vacuum concentrator	Christ
Thermomixer Comfort	Eppendorf
Thermostatic bath, PID system	Instruments s.r.l.
Vivaspin500, 5000 MWCO	Sartorius
Vortex mixer MS525-20	Heidolph (REAX)

6.2 Methods

6.2.1 Isolation of sponge metagenomic DNA

To ~ 40 mg of frozen sponge (in RNA later) 700 μ L of lysis buffer I (200 mM Tris-Cl, 50 mM EDTA, 1.4 M NaCl, 2% CTAB, 0.5% PVP, all in milliQ[®]-H₂O) were added and incubated at 37 °C for 1 h in a thermomixer (1,400 rpm). After addition of 2.8 μ L β -mercaptoethanol, 70 μ L 10% SDS, 2 μ L RNase A (100 mg/mL), and 40 μ L proteinase K (10 mg/mL) the tube was incubated at 55 °C for a further hour in a thermomixer (1,400 rpm). At this time, the microcentrifuge tube was spun 4 min at 5,000 rpm. The clear middle phase was transferred to a new microcentrifuge tube containing 750 μ L CHCl₃ and centrifuged 10 min at 15,000 rpm. After repetition of the CHCl₃ wash, the supernatant was transferred to a new microcentrifuge tube containing 750 μ L of 70% aqueous isopropyl alcohol containing 10% (v/v) 3M NaOAc (pH 5.5) at room temperature. The precipitated DNA was spun down at top speed for 20 min, washed with ice-cold ethanol, dried and dissolved in ~60 μ L elution buffer (10 mM Tris-Cl, pH 8.5). The amount of one tube was enough for PCR screening, for library construction the protocol was upscaled to ~500 mg of frozen sponge.

6.2.2 DNA amplification by PCR

6.2.2.1 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°C-95°C. Next is the annealing step, when primers anneal to the single-stranded DNA at lower temperature, 50°C-65°C. 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°C. In addition, the final extension at 72°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. The reaction mixture for standard PCR was prepared as follows: 14.65µL H₂O, 1.5µL DMSO, 0.5µL dNTP, 2.5µL forward primer (10µM), 2.5µL reverse primer (10µM), 2.5µL 10X Taq Buffer advanced with 15 mM Magnesium (5 PRIME), 0.35µL RBC Taq DNA polymerase (5u/µL). Standard PCR was performed according to this program:

➤ 95°C for 30 sec, 54°C for 30 sec, 72°C for 45 sec, 72°C for 5min.

The annealing temperature was normally set to 2 degrees below the lowest melting temperature calculated for the primer pair to be used; elongation time was adjusted to the PCR product size, assuming 1 min for 1000 bp.

6.2.2.2 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°C higher than the optimal primer melting temperature (T_m) and this temperature is decreased by 0.5-2°C every cycle

until the T_m 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.

6.2.2.3 Gradient PCR

During the gradient PCR, an annealing temperature gradient, which is programmed between 1°C and 10°C, is built up across the thermoblock so that each row has a different annealing temperature. This allows the most stringent parameters for every primer set to be calculated with the aid of only one single PCR reaction.

6.2.2.4 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 200 µL LB medium with the proper antibiotic. After growing at 37°C for 1 hour, 1 µL aliquot was subjected to PCR and the remaining colony suspension can be stored at 4°C for some days for cultivation. In a colony PCR, the initial denaturation step is prolonged for 10 min instead of 45 sec as in a normal DNA PCR, to promote cell wall and membrane disruption and DNA release.

6.2.2.5 Amplification of 16S gene products from *P.simplex*

Metagenomic DNA was used for PCR amplification of 16S rRNA genes with the prokaryote specific primers F27/R1492. Polymerase chain reaction was carried out under the following conditions: initial denaturation at 94°C

for 1 min, followed by 30 cycles of 95°C for 30 sec, 48°C for 30 sec and 72°C for 3 min, with a final extension 72°C for 5 min. The reaction mixture (25µL) contained: 14.65 µL H₂O, 0.5 µL DMSO, 1.5 µL dNTP (10mM), 2.5 µL Taq buffer advanced (Eppendorf), 2.5 µL primer F27 (10µM), 2,5 µL primer R1492 (10µM), 0.35 µL RBC Taq DNA polymerase (5 U/µL, RBC Bioscience), 0.5 µL DNA (primers: F27 5'-AGAGTTTGATCMTGGCTCAG – 3'; R1492 5'-TAC GGY TAC CTT GTT ACG ACTT).

6.2.2.6 PCR screening of metagenomic DNA for type I PKSs

In general, three primer pairs for type I PKSs were used: KSDPQQF (5'-MGN GAR GCN NWN SMN ATG GAY CCN CAR CAN MG-3') and KSHGTGR (5'-GGR TCN CCN ARN SWN GTN CCN GTN CCR TG-3') for the KS domains, AT1F (5'-TTY CCN GGN CAR GGN NSS CAG TGG-3', binding to the motif FPGQGsQW) and AT3R2 (5'-GC IGC IGC NAT CTC NCC C-3', binding to the motif QGEIAAA) for the AT domains, and SWF_ATF (5'-TTC TCC GGG CAG GGC ACG CAG TG-3', binding to the motif FSGQGTQW) and SWF_ATR (5'-CAG TTC CAC CAG CGC GCA CTG-3', binding to the motif QCALVEL) designed to be specific for the AT domain of *swfA*. To obtain PCR products, 0.5 µL of sheared (by pipetting 100 times up and down) metagenomic DNA was used in a 50 µL reaction [27 µL H₂O, 2 µL MgCl₂ (25 mM), 3 µL DMSO, 1.5 µL dNTP (10 mM), 5 µL primers (10 µM), 5 µL Taq buffer advanced (Eppendorf), 1 µL RBC Taq DNA polymerase (5 U/µL, RBC Bioscience)]. The cycler program was 1) 94 °C for 45 s, 2) 94 °C for 1 min, 3) 54 °C/ 56 °C/ 58 °C

for 1 min in case of KS primers, and 58 °C/ 60 °C/ 62 °C for 1 min in case of AT primers, respectively, 4) 72 °C for 45 s, repetition of 2), 3), and 4) 30 times, 5) 72 °C for 7 min, 6) 4 °C forever.

6.2.3 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 molecular DNA and RNA 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 or SB and dissolving by boiling. The concentration of agarose gel depended on the size of analyzed DNA, normally ranged from 0.8 - 2%. Since DNA is colourless, it can only be visible under UV or blue light by adding gel green into agarose gel to a final concentration of 1X. 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 density of the sample.

6.2.4 Recovery of DNA fragments from agarose gel

The extraction and purification of DNA fragments (PCR amplicons and digestion fragments) from agarose gel were carried out following the

QIAquick Gel Extraction Kit protocol. QIAquick Kit contains a silica membrane assembly for binding of DNA in high-salt buffer and elution with low-salt buffer or water. The purification procedure removes primers, nucleotides, enzymes, mineral oil, salts, agarose, gel green, and other impurities from DNA samples.

The DNA fragment was excised from the agarose gel with a clean and sharp scalpel, minimizing the size of the gel slice by removing extra agarose. After weighing the gel slice in a tube, 3 volumes of Buffer QG were added to 1 volume of gel (100 mg ~ 100 μ L). Generally, 600 μ L of Buffer QG were used for 200 mg of gel. The mixture is incubated at 50°C for 10 min (or until the gel slice had completely dissolved) in a thermomixer (1400 rpm). After the gel slice had dissolved completely, 1 gel volume of isopropanol was added to the sample which is then mixed by vortexing. This step increased the yield of DNA fragments <500 bp and >4 kb. For DNA fragments between 500 bp and 4 kb, addition of isopropanol had no effect on yield. Then, a QIAquick spin column was placed in a provided 1.5 mL collection tube. To bind DNA, the sample was applied to the QIAquick column, and centrifuged for 1 min at 13000 rpm. The flow-through was discarded and the QIAquick column was placed back in the same collection tube. To wash, 0.75 mL of Buffer PE were added to QIAquick column and centrifuged for 1 min. After discarding the flow-through, the QIAquick column was centrifuged for an additional 1 min (13,000 rpm). After the washing step, QIAquick column was placed into a clean 1.5 mL microcentrifuge tube. To elute DNA, 30 μ L of Buffer EB (10 mM Tris·Cl,

pH 8.5) or H₂O were added to the center of the QIAquick membrane. After 1 min, the column was centrifuged for 1 min at maximum speed.

6.2.5 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 *EcoRV* 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.

1. *Digestion of pBluescript SK II (+) with EcoRV.* 10 µL of pBluescript SK II (+) (concentration between 150 ng/µL and 300 ng/µL), 2.5 µL of *EcoRV* (15 u/µL), 5 µL buffer H 10X (Takara- 500mM Tris-HCl, pH 7.5; 100mM MgCl₂; 10 mM DTT; 1000mM NaCl) and 32.5 µL H₂O were incubated at 37°C for 2 hours. The linearized vector was gel purified and extracted with the QIAquick Gel Extraction Kit, eluting in 39.5 µL of H₂O.

2. *Making T-vector.* 0.5 µL Taq Polymerase, 5 µL Taq buffer 10X and 5 µL dTTP 10 mM were added to the linearized vector and the mixture was incubated at 72°C for 2 hours. The T-vector is purified by chloroform extraction (100 µL) centrifuging at 13000 rpm for 5 min. The upper phase was transferred to a new tube containing 70 µL of isopropanol to

allow DNA precipitation. After centrifuging at maximum speed at 4°C for 20 min, the DNA pellet was washed twice with ice-cold ethanol 70%. Then, the T-vector was dried at 50°C (10 min) and resuspended in 10 µL of H₂O.

3. *TA ligation*. The 3' overhang PCR product could be introduced into the T-vector by a sticky-end ligation. 7.5 µL of PCR product and 1 µL of the T-vector were pre-mixed at 65°C for 5 min. Then, after cooling on ice, 1 µL of ligation buffer and 0.5 µL of KAPA T4 ligase were added to the mixture. The reaction was incubated at 16°C overnight. On the next day, the ligation was heat inactivated at 65°C for 15 min.

Finally, the modified genetic construct was transferred into competent *E. coli* XL1 blue cells by electroporation and selected by blue/white screening and antibiotic selection.

6.2.6 Phosphatase treatment of the linear vector

In order to remove 5' phosphate group and prevent vector self-ligation, the linear vectors were generally treated with Shrimp alkaline phosphatase (SAP).

➤ Phosphatase treatment formula of the linearized vector

40 µL linear vector (in water)
5 µL SAP 10X reaction buffer (50 mM Tris-HCl, 10mM MgCl₂)
5 µL SAP (1u/µL - Promega)

50 µL total volume

The reaction was incubated at 37°C for 30 min. After that, further 5 µL were added to the mixture, which was again incubated for 30 min. SAP was heat inactivated at 65°C for 15 min.

6.2.7 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 °C. Usually, cells were harvested in the ‘log-phase growth’ when the culture reached an OD₆₀₀ of 0.4-0.8. This was done because an excessive number of living cells in an electrical field could be the cause of apoptosis and necrosis phenomenon in which the cell morphology was changed.

5 mL LB medium with *E. coli* XL1 blue was inoculated via toothpick. The culture was incubated overnight on a shaker (250 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 (250 rpm) at 37 °C till OD₆₀₀ was around 0.4-0.8. The culture was placed immediately on ice and the cells were 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.

6.2.8 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. 1.5 μL plasmid DNA or 4 μL of inactivated ligation mixture were mixed with 50-70 μL of electrocompetent cells on ice. The mixture was transferred to an ice-cold 0.2 cm electroporation cuvette and an electroporation is performed using a Bio-Rad electroporator set at 2.5 kV. The expected time constant should have been close to 5 milliseconds. Immediately after electroporation, 900 μL 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 (250 rpm) at 37 °C for 45 min in order to ensure the expression of the antibiotic resistance gene. Afterwards, 200 μL (the whole mixture in case of TA ligation product) 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.

6.2.9 Blue white screening

Blue white colony screening is a strategy to quickly and easily distinguish between recombinant and non-recombinant colonies. It requires

a special vector, such as pBluescript SK II (+), and a special strain of *E. coli*, such as XL1 blue.

The first gene in the *E. coli* lac operon is lacZ, which encodes β -galactosidase (β -gal). The active form of β -gal is a tetramer and hydrolyses lactose into glucose and galactose. Deleting amino acids 11-41 of β -gal (called the lacZ Δ M15 mutation) means the enzyme is unable to form a tetramer and is non-functional. Supplying amino acids 1-59 (the α -peptide) of β -gal in trans (separately) allows the truncated β -gal to form tetramers and function again. Rescuing β -gal by supplying the α -peptide in this way was termed α -complementation.

The α -complementation can be used to screen *E. coli* colonies for the presence of inserts by cloning the α -peptide coding region into a pBluescript (or similar) plasmid and introducing a multiple cloning site (MCS) into the middle of that region. When a piece of DNA is ligated into the MCS, it disrupts the α -peptide, rendering the β -gal non-functional. 5-bromo-4-chloro-indolyl- β -D-galactopyranoside (X-gal) is a colourless analogue of lactose. When β -galactosidase hydrolyses X-gal, it creates a blue product (5,5'-dibromo-4,4'-dichloro-indigo). In blue white screening, an *E. coli* strain is transformed with a ligation reaction and spread onto agar plates containing x-gal. A blue coloured colony indicates that the α -peptide in the plasmid is intact (no insert) whereas a white colony indicates that the α -peptide is disrupted (insert present).

Induction of β -galactosidase expression in the host strain is obtained by adding IPTG onto agar plates.

The transformed colonies were selected by blue/white screening on an LB agar plate with ampicillin (100 µg/mL) and also containing 100 µL IPTG (10mM) and 100 µL of X-gal (2% p/v) which were plated freshly.

6.2.10 Isolation of plasmid DNA

Plasmid DNA isolation from recombinant clones is performed using the QIAprep[®] Spin miniprep kit. The QIAprep miniprep procedure is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high salt. 1.5 mL of overnight bacteria culture was centrifuged at 10000 rpm for 30 sec. After discarding the supernatant, the pelleted bacterial cells were resuspended in 250 µL Buffer P1 (containing RNase A). Then, 250 µL Buffer P2 were added for alkaline lysis and the tube was gently inverted 4–6 times to mix, not allowing the reaction to proceed more than 5 minutes. 350 µL Buffer N3 were added to the mixture and the tube was again gently mixed by inverting 4-6 times. Then, the mixture was centrifuged at 13000 rpm for 10 min to precipitate a white compact pellet. The supernatant was applied on a QIA prep spin column by decanting and pipetting, and centrifuged at 13000 rpm for 1 min. The flow-through was discarded, and 750 µL Buffer PE containing ethanol were added to wash the column by centrifugation for 1 min. The flow-through was discarded again and the QIA prep spin column was centrifuged for 1 min to remove residual washing buffer. Finally, the QIA prep spin column was placed into a clean tube and the plasmid DNA which bound the silica membrane was eluted by adding 50 µL of water or buffer EB (10 mM Tris-

Cl, pH 8.5). After 1 min, the QIA prep spin column was centrifuged for 1 min at 13000 rpm so that the plasmid DNA flew through.

6.2.11 Construction of a fosmid library from the metagenomic DNA of the sponge *P. simplex*

After extraction, metagenomic DNA was used to construct a fosmid library using the CopyControl™ Fosmid Library Production Kit (Epicentre). A gene library is a population of bacterial colonies, each carrying a different DNA fragment that is inserted in a cloning vector. There are many types of cloning vectors which can be used to build genomic library. Plasmids are circular and double-stranded extra-chromosomal DNA molecules that are capable of replicating independently of the chromosomal DNA. Cosmids are hybrid plasmids. They contain the *cos* sequences from the Lambda phage, which allow the recognition by the phage capsid proteins. In this way cosmids, packaged into capsids, can be easily transferred into bacterial cells where they can replicate as plasmids, so that many copies of cosmids can be present in each bacterial cell. Cosmids are able to contain 37 to 52 kbp of DNA while normal plasmids carry only 1-20 kbp. Fosmids are more sophisticated than cosmids; they are derived from the bacterial plasmid F, and contain a partitioning and replication system which allows the bacterial host (*E. coli*) to contain a single copy of the fosmid, offering a higher stability to the library. The CopyControl pCC1FOS vector, which was used to build metagenomic DNA fosmid library from *Plakortis simplex*, has distinctive features, such as chloramphenicol-resistance for selection of the transformed bacterial cells, and the inducible high-copy *oriV* origin of

replication in addition to the single-copy origin. Initiation of replication from *oriV* requires an inducer, and can be activated during the fosmid isolation to increase its amounts.

The following steps were involved in *P. simplex* DNA library construction:

- DNA extraction and purification (see section 6.2.1)
- Random shearing of the DNA to approximately 40 Kb fragments
- End-repair reaction of the sheared DNA to blunt, 5'-phosphorylated ends
- Size selection of end-repaired DNA by Pulsed Field Gel Electrophoresis (PFGE)
- Ligation of blunt end DNA to the CopyControl pCC1FOS vector
- Packaging of ligated DNA and transformation of EPI 300-T1^R competent cells
- Semi-liquid library preparation

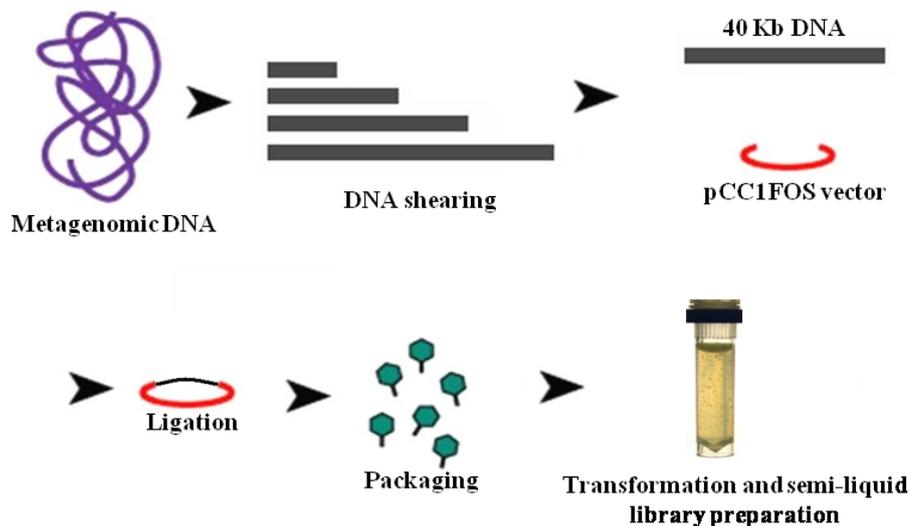


Figure 6.1 Metagenomic DNA library construction.

End-Repair. During this step DNA fragments ends, which could have been damaged during DNA extraction and shearing, were repaired and 5'-phosphorylated. *End-repair Enzyme Mix* contains T4 polynucleotide kinase

which catalyses γ -phosphate transfer from ATP to free –OH groups at DNA 5'-position and T4 polymerase with both 5'→3' polymerase and 3'→5' exonuclease activities. 100 times sheared DNA was subjected to *End-Repair* reaction according to the *CopyControl Fosmid Library Production Kit* (Epicentre) procedure. The following reagents were thawed on ice and mixed:

- 8 μ L 10X *End-Repair* Buffer
- 8 μ L 2.5 mM dNTP Mix
- 8 μ L 10 mM ATP
- 52 μ L sheared DNA (~0.5 mg/mL)
- 4 μ L *End-Repair* Enzyme Mix

Mixture was incubated at room temperature for 45 min and then at 70 °C for 10 min to inactivate the *End-Repair Enzyme Mix*. Reaction was incubated at room temperature and then at 70° C for 10 minutes to inactivate *End-repair Enzyme mix*.

Size Selection of the End-Repaired DNA. Size selection of DNA fragments is required before ligation. The optimal fragment size is around 40 Kb according to the capacity of fosmid cloning vector. It is important that the DNA recovered is ≥ 25 Kb in order to avoid unwanted chimeric clones.

The end-repaired metagenomic DNA was separated by Pulsed Field Gel Electrophoresis (PFGE). PFGE resolves DNA by alternating the electrical field between spatially distinct pairs of electrodes, causing DNA molecules as large as several megabases to reorient and move at different speeds through the pores in an agarose gel. PFGE systems create homogeneous electrical fields, employing the technology of the contour-clamped homogeneous electric field (CHEF) based on the application of controlled electric fields that change direction at a predetermined angle to samples of DNA that have been embedded in an agarose gel matrix .

The CHEF-DR[®] III Pulsed Field Electrophoresis System was set as follows:

(Chiller) Temperature: 14 °C
(Electrical Field) Switch Time: 1-6 seconds
Run Time: 11 hours
Angle: 120°
Voltage Gradient: 6 V/cm

The end-repaired DNA (80 µL) was prestained with 10X Gel Loading Buffer (Invitrogen) and loaded into a pulsed field certified agarose gel (Bio-Rad). This gel was prepared by melting 1g of agarose in 100 mL 0.5X TBE buffer (Bio-Rad). Next day, once run was completed, the gel was stained with a Gel Green DNA stain solution (25 µL of Gel Green in 250 mL 0.5X TBE buffer) and visualized on a blue light transilluminator to excise a gel slice containing DNA of the desired size (35-45 kb).

Recovery of the Size-Fractionated DNA from Agarose Gel. The gel slice, containing the DNA of correct size, was transferred in a 1.5 mL tube and incubated at 70 °C for 15 min to melt the agarose and then at 45 °C. The appropriate volume of pre-warmed (at 45°C) GELase 50X Buffer was added to 1X final concentration. 1 U (1 µL) of GELase Enzyme Preparation was added to the tube for each 100 µL of melted agarose. The melted agarose solution was kept at 45°C and gently mixed. The solution was incubated at 45°C for 1 hour. After incubation, the solution was transferred to 70 °C for 10 min to inactivate the GELase enzyme. The solution was chilled in an ice bath for 5 min and then centrifuged at 13000 rpm for 20 min to pellet any insoluble oligosaccharides. A gelatinous pellet formed. 90-95% of the supernatant, containing DNA, was removed into a clean microcentrifuge tube. 1/10 volume of 3 M sodium acetate (pH 7.0) and 2.5 volumes of ethanol were added to the supernatant. The mixture was inverted several times and DNA was allowed to precipitate for 30 min at -20 °C. Then it was

centrifuged for 20 min at 13000 rpm at 4°C. DNA pellet was washed twice with 70% EtOH, then dried under laminar flow for 10 min and redissolved in 30 µL of H₂O.

Ligation. The following reagents were thawed on ice, combined and mixed:

- 0 µL ultrapure sterile water
- 1 µL 10X Fast-Link Ligation Buffer
- 1 µL 10 mM ATP
- 1 µL CopyControl pCC1FOS Vector (0.5 µg/µL)
- 6 µL insert DNA (0.25 µg of ~40 Kb)
- 1 µL Fast-Link DNA Ligase

(10 µL total reaction volume). The mixture was incubated for 2 h at room temperature and then transfer to 70 °C for 10 min to inactivate Fast-Link DNA Ligase.

Packaging. After ligation the new fosmids were mixed with the capsid proteins (*MaxPlax Lambda Packaging Extract*) of bacteriophages, giving phage particles capable to infect bacterial cells. 50 mL of LB + 10 mM MgSO₄ were inoculated with a single EPI 300-T1^R colony and kept under shaking (250 rpm) overnight at 37 °C. The next day 5 mL of EPI 300-T1^R overnight culture were used to inoculate 45 mL of LB + 10 mM MgSO₄. Bacterial growth at 37 °C was monitored by UV spectroscopy to an OD₆₀₀ = 0.8. One tube of *MaxPlax Lambda Packaging Extract* was thawed on ice; 25 µL were transferred in a clean eppendorf tube and all the 10 µL of the ligation reaction were added. The mixture was mixed by pipetting several times and incubated at 30 °C for 90 min. After the 90 min packaging reaction was completed. The remaining 25 µL of *MaxPlax Lambda Packaging Extract*, previously thawed on ice, were added and the reaction was incubated for additional 90 min at 30 °C. At the end of the second incubation, Phage Dilution Buffer (PDB) was added to 1 mL final volume

and after mixing 25 μL of chloroform was added too. A viscous precipitate formed, which didn't interfere with library production.

Transformation and semi-liquid library preparation. Transformation is the process by which competent bacterial cells take up foreign DNA molecules and allow its replication and expression. When DNA is transferred to one bacterium by a virus, transformation is called transduction. During transduction bacteriophages attach to a specific receptor on the bacterial cell surface through tail fibers. Lysozyme dissolves a hole in the cell wall, the phage contracts and the nucleic acid is injected into the cell.

1 mL of phage particles (packaged fosmid clones) into *Phage Dilution Buffer* (PDB) were mixed with 10 mL EPI 300-T1^R host cells and incubated at 37° C for 20min. After transduction, 100 μL of 1:10² dilution of infected EPI300-T1^R in LB medium were spread on LB plates + 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol for fosmid library titering. After overnight incubation at 37°C, the colonies on 1:10² dilution plate were 23, therefore the titer of the entire fosmid library was approximately 250,000 cfu. After transfection, the resulting 11 mL containing the entire fosmid library were mixed with 250 mL liquid LB medium containing a supplement of 5 g/L SeaPrep Agarose (Lonza) and chloramphenicol (12.5 $\mu\text{g}/\text{mL}$) for selection. SeaPrep Agarose allows the culture medium to gelify at 4°C. 250 aliquots of 1 mL (referred as pools) were transferred to 2 mL screw cap vials, stored on wet ice for 1 h to allow the gelling process of the medium, and incubated at 37°C overnight. After overnight growth, if the predicted titer of the entire fosmid library was 250,000 cfu, correspondingly, every pool possessed about 1,000 clones. These pools were labeled with three numbers: number of the box, of

the row and of the column. For example, the pool 2D9 means the pool from box 2, row D and column 9. After overnight growth, suspended colonies (semi-liquid or 3D cultures) had formed and each vial was vortexed shortly to homogenize the mixture; 25 μ L were removed for PCR screening, and 1 mL glycerol 30% was added to the remainder for storage at -80 °C.



Figure 6.2 Fosmid library from the metagenomic DNA of *P. simplex*.

6.2.12 The semi-liquid library method

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 (*Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools*, Piel et al., 2007) 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 underrepresented clones in the doubling period. 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.

6.2.13 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. Therefore, 29 row superpools and 29 column superpools were constructed from the 261 pools of the fosmid library of *P.simplex*. 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 25 μ L 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. For example, the row superpool 2D comprised nine pools in the row D, D2-D10 of the box 2; the column superpool 2/9 included nine pools in the column 9, B9-J9 of the box 2.

6.2.14 Screening and isolation of positive clones containing target genes from a fosmid library

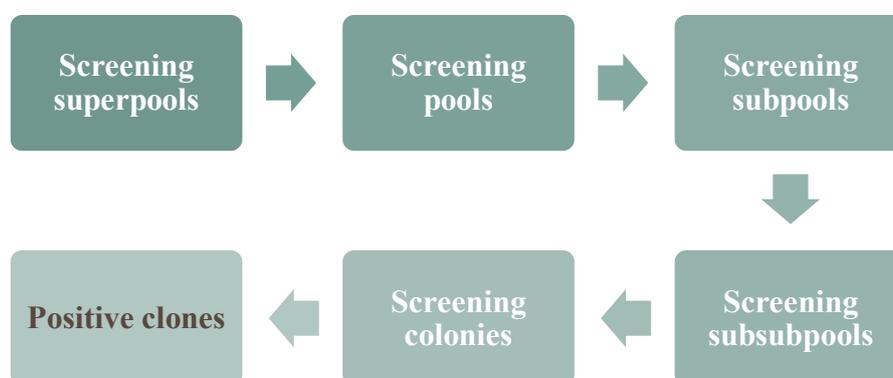
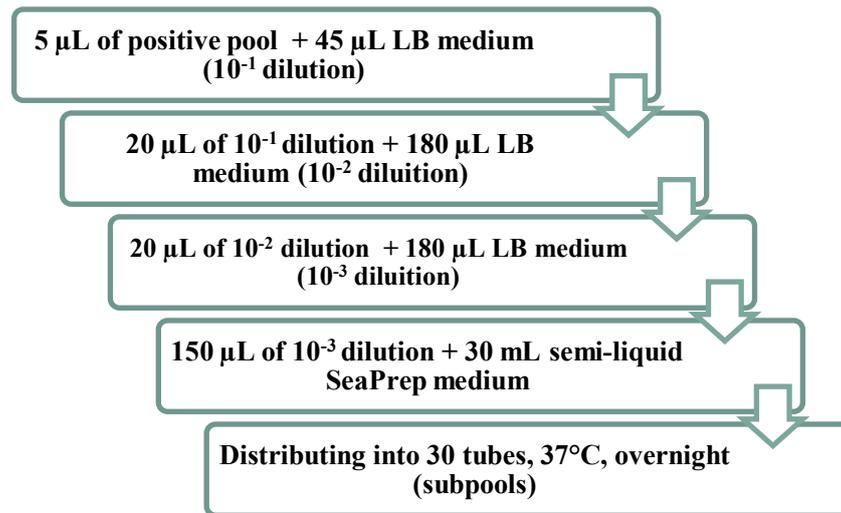
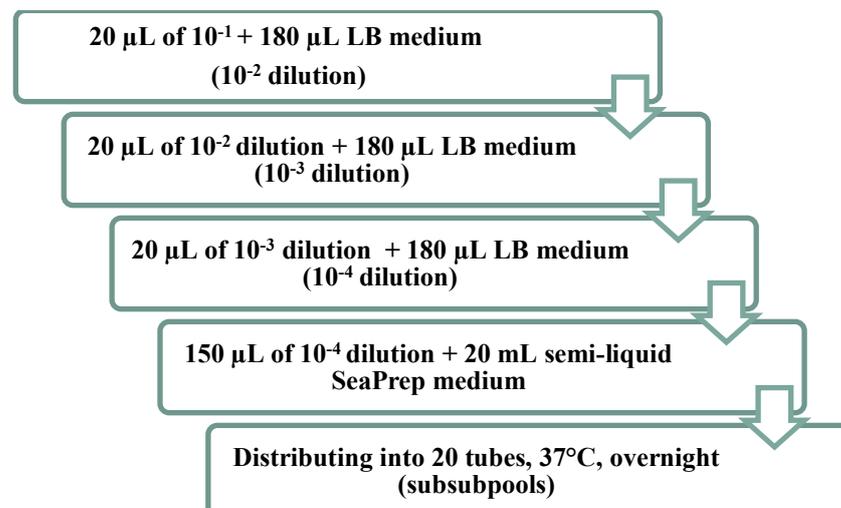


Figure 6.3 Flowchart for isolating positive clones from a complex fosmid library.

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 as follows:



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. First, 5 μL of the positive subpool were diluted with 45 μL of LB medium (1:10 dilution), then dilution series were prepared as follows:



Once the positive subsubpool was recognized, the final dilution was prepared to generate 20-30 separate colonies on LB agar plate. After overnight incubation, individual colonies were picked and resuspended with 200 μ L LB medium and vortexed. At the end, the positive clone was identified by colony PCR (Figure 6.4 – 6.5).



Figure 6.4 Semiliquid cultures. Bacterial pools containing 1000 clones per each.

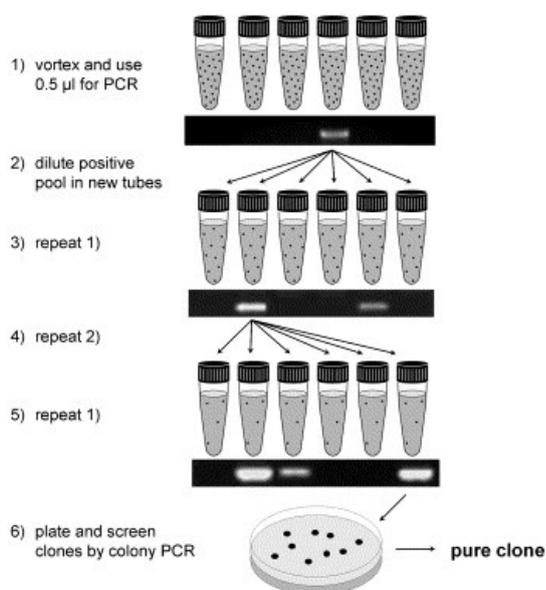


Figure 6.5 Isolation of positive clones from a fosmid library by PCR screening of semi-liquid pools.

6.2.15 Induction of high copy number plasmid DNA

The fosmid library of the metagenomic DNA of *P.simplex* 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°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 shaken vigorously at 37°C for 5 hours. Afterwards, fosmid DNA was isolated following the protocol described in Section 6.2.10 (“Isolation of plasmid DNA”).

6.2.16 Overexpression of the SwfA_{ACP} protein

The ACP domain of the SwfA protein was expressed both in its *apo*-form and its *holo*-form by co-expression with Svp as described in the following diagram.

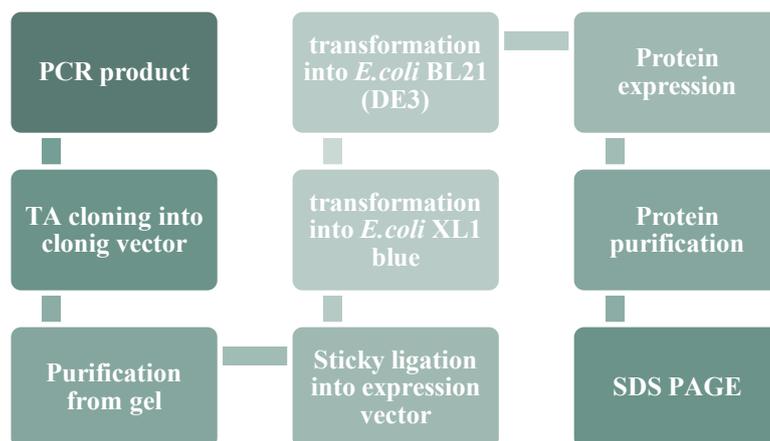


Figure 6.6 Diagram of protein expression.

6.2.16.1 Cloning of the *swfA*_{ACP} domain

The *swfA*_{ACP} domain was expressed using the expression vector pHis8 (Figure 6.7). Therefore, in preparation for sticky-end ligation, the sequences of restriction enzymes belonging to the MCS of the vector pHis8 had to be introduced into the sequences of the ACP domain. Obviously, the nucleotide sequence of the chosen restriction enzymes should not be present in the ACP domain to avoid cleaving them during the genetic engineering process. As a further requirement, for preserving ORFs of the target protein, the nucleotide sequences of these restriction enzymes were obligated to be present in the same triplet codon frame with the start codon of pHis8.

In addition, to dominate the target protein synthesis, the presence of the start and stop codon in the expressed ACP domain had to be controlled. The start codon ATG was already present in the expression vector pHis8, so only the stop codon TGA was required to be added on the reverse primer. The sequence of the ACP domain of *swfA* was amplified by PCR using primers 11G3_ACP_up (5'-AAA GGA TCC ctg acg ctc gaa ggt gtg gt-3') and

11G3_ACP_low (5'-AAA AAG CTT TCA gtc ggc ccc gtg cac gat tgc c-3') (introduced restriction sites are underlined and stop codon is in bold). The 234 bp long PCR product was subcloned via T/A-cloning in pBluescript II SK(+) (Stratagene) to yield pGS21, and cut out from the vector with *Bam*HI and *Hind*III. The plasmid pGS21 was digested with BamHI (Promega) according to the following formula:

- 20 μ L PGS21 (150 ng/ μ L)
- 5 μ L 10X buffer E (Promega – 60mM Tris-HCl pH 7.5, 1M NaCl, 60mM MgCl₂ and 10 mM DTT)
- 0.5 μ L BSA (10 μ g/ μ L)
- 2.5 μ L *Bam*HI (10u/ μ L)
- 22 μ L H₂O

(50 μ L total volume). The reaction was incubated for 2 hours. After gel electrophoresis, DNA was gel extracted and purified, eluting in 42.5 μ L of H₂O. Then, the linearized pGS21 vector was cut with *Hind*III (Takara) to obtain the ACP coding fragment, according to the following protocol:

- 42,5 μ L of linearized pGS21
- 5 μ L of 10X buffer M (Takara - 60mM Tris-HCl pH 7.5, 100mM MgCl₂, 500mM NaCl and 10mM DTT)
- 2,5 μ L *Hind*III (15u/ μ L)

(50 μ L total volume). The double digested ACP coding fragment was gel extracted and purified, eluting in 30 μ L.

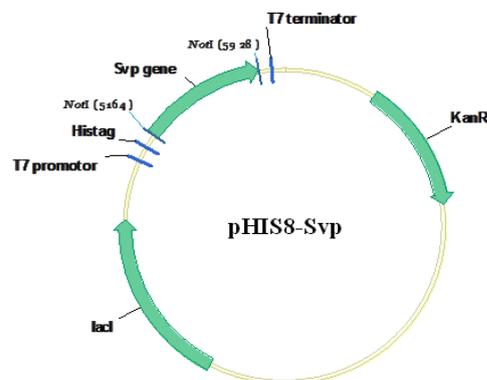


Figure 6.7 Vector map of pHIS8-Svp

6.2.16.2 Ligation of the amplified ACP fragment with the expression vector pHis8

The swfA_{ACP} fragment was then ligated into two different vectors by a sticky-end ligation: (i) a pHis8 vector with an additional gene for the PPTase Svp inserted into its NotI site (i.e., pHis8-Svp, to yield pGS30), and (ii) a pHis8 vector without this gene (i.e., pHis8, to yield pGS34).

In preparation for sticky-end ligation, the vectors pHis8 and pHis8-Svp were also digested with the corresponding enzymes used for pGS21 cutting, *Bam*HI and *Hind*III. After purification from agarose gel using the Purification Gel Extraction kit of Qiagen, the ligation between the digested vectors pHis8, pHis8-Svp and the ACP fragment were carried out to form pGS34 and pGS30, respectively.

➤ Digestive formula of vector pHis8 and pHis8-Svp: *see scheme for double digestion of pGS21* (Paragraph 6.2.16.1)

➤ Ligation formula:

7.5 µL ACP fragment
1 µL vector (pHis8 or pHis8-Svp)
1 µL 10X Kapa T4 ligase buffer
0.5 µL Kapa T4 ligase
10 µL total volume
Incubation at 16°C overnight. Heat inactivation at 65°C for 15 min.

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

The two new expression plasmids pGS34 and pGS30 were transformed into competent *E. coli* XL1 blue cells by electroporation (Section 6.2.7). The transformed cells were selected on LB agar plates containing kanamycin (resistance gene on pHis8).

6.2.16.4 Transformation of the recombinant DNA into *E.coli* BL21 CodonPlus[®] (DE3)-RIPL cells

The two plasmids pGS34 and pGS30 from the transformed *E. coli* XL1 blue were introduced into electrocompetent BL21-CodonPlus[®](DE3)-RIPL (Stratagene) cells by electroporation and selected on LB agar plates containing 50 µg/mL kanamycin (Kan₅₀).

6.2.16.5 Protein expression and purification

Proteins from plasmids pGS30 and pGS34 were expressed as N-terminal His8-tag fusion proteins and purified by immobilization on Ni-NTA matrices based on the tight bond between imidazole ring of histidine residue and nickel ion.

To do so, *E. coli* BL21-CodonPlus[®] (DE3)-RIPL electrocompetent cells (Stratagene) were transformed with the expression plasmid and spread out on LB agar + Kan₅₀ plates and grown o/n at 37 °C. From this plate, 4 clones were picked to prepare o/n cultures in LB + Kan₅₀. 5 mL of this o/n cultures were used to inoculate 200 mL LB medium without antibiotics. The cultures were incubated at 37 °C for 2-3 h shaking with 250 rpm (OD₆₀₀ ~ 0.6-0.8). After cooling to 16 °C, the cultures were induced with 1 mM IPTG and incubated for further 24 h at 16 °C. After harvesting by centrifugation (5,000 rpm, 5 min), the cell pellets were dissolved in 3 mL lysis buffer (50 mM Tris-Cl, 500 mM NaCl, 10 mM MgCl₂, pH 8.0) and frozen o/n.

After thawing on ice and ultrasonication, the lysates were obtained by centrifugation for 20 min at 15,000 rpm in a table top centrifuge at 4 °C.

The supernatants were transferred to 15 mL falcon tubes and after addition of 300-600 μ L PerfectPro[®] Ni-NTA agarose (5PRIME) incubated on ice for 1 h under horizontal shaking; the pellets were (partly) resuspended in 1 mL of lysis buffer by vortexing.

To obtain purified soluble proteins, the lysate/Ni-NTA agarose mixtures were transferred to Poly-Prep[®] chromatography columns (Bio-Rad) that fit to a SPE chamber (Supelco). Elution fractions (500 μ L each) were obtained by stepwise increasing the imidazole concentration in the lysis buffer from 0 (= wash) to 300 mM.

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.

6.2.16.6 Analysis of the expressed 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. 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 5-18.5% depending on protein sizes. 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.

In the performed experiments (figure 6.8), aliquots of 50 μ L of the eluted fractions were mixed with 25 μ L of 3 \times SDS loading buffer [2.4 mL Tris-Cl (1 M, pH 6.8), 3 mL SDS (20%), 3 mL glycerol, 1.6 mL β -mercaptoethanol, 6 mg bromophenol blue, and 10 mL H₂O], incubated for 5 min at 99 $^{\circ}$ C, spun down shortly and \sim 20 μ L of elution fractions were loaded on mini SDS-PAGE (18.5% resolving gel; 6% stacking gel). 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 200 mV for about 90

min, the proteins were separated in the resolving gel. Since almost proteins are colourless, their bands on acrylamide gel are detected by staining with coomassie, the most popular protein gel stain. This dye (1.25 g) was dissolved in a mixture of methanol (150 mL), acetic acid (50 mL) and dH₂O (300 mL) which fixed proteins on the gel. The gel was stained in about 250 mL dye solution for 20 min and shaken at 45 rpm at room temperature. The dye was then removed by destaining with a solution comprising methanol (150 mL), acetic acid (50 mL) and dH₂O (300 mL) till the blue bands of proteins appeared on a clear background.

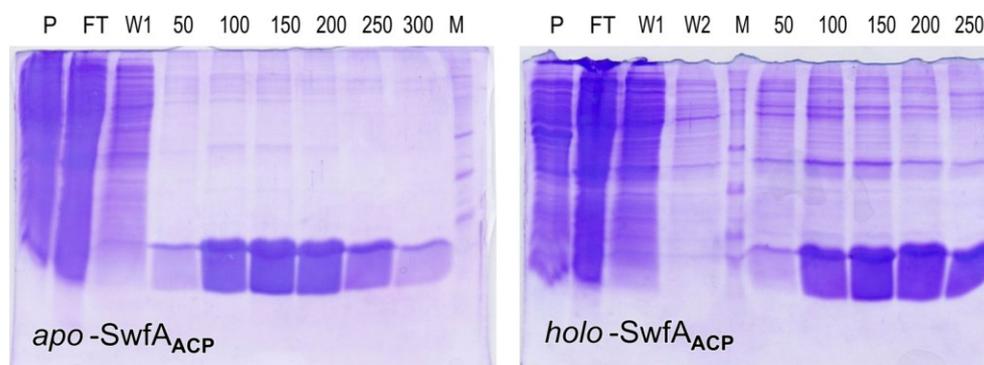


Figure 6.8 The proteins *apo*-SwfA_{ACP} and *holo*-SwfA_{ACP} were expressed as N-terminal His8-tag fusion proteins and purified over Ni-NTA columns. Elution fractions (500 μ L each) were obtained by increasing the imidazole concentration in the lysis buffer from 0 to 300 mM stepwise. Mini SDS-PAGE was loaded with resuspended cell pellet (5 μ L, lane P), column flow through, containing unbound proteins (12 μ L, lane FT), washing fractions eluted with lysis buffer (20 μ L each, lanes W1 and W2), and fractions eluted with increasing imidazole concentrations (20 μ L each, lanes 50,100,150, 200, 250, and 300). All the fractions were previously mixed with 3X SDS loading buffer. Lane M is a pre-stained protein marker (Roti[®]-Mark STANDARD).

Table 6.8 Composition of the stacking and resolving gels for SwfA_{ACP} SDS-PAGE.

Composition	6% stacking gel	18.5% resolving gel
30:0.8% w/v acrylamide:bisacrylamide	1mL	5mL
1M Tris-Cl pH8.8	-	3mL
1M Tris-Cl pH6.8	630 μ L	-
20% SDS	25 μ L	38 μ L
dH ₂ O	3.6 mL	-
10% APS	25 μ L	36 μ L
TEMED	5 μ L	6 μ L

6.2.16.7 FTMS analysis of overexpressed proteins

The purified proteins *holo*-ACP and *apo*-ACP were desalted 3 times with milli-Q[®]-H₂O in VivaSpin500 centrifugation units (Sartorius, 5000 MWCO), diluted 1:2 with MeOH, containing 0.2% formic acid, and subjected to FTMS analysis. The protein mass spectra were obtained with LTQ Orbitrap XL (Thermo scientific) using the preset MS method "ubiquitin" in the FTMS + p ESI full ms mode.

6.2.17 Heterologous expression

The library fosmid pPS2D9 containing the whole *swf* cluster (PS11G3 type) was used for heterologous expression trials in an expression system consisting of the vector pHIS8-Svp and the bacterial host *E. coli* BAP1, which is a derivative of *E. coli* BL21-(DE3) including a gene for the 4'-phosphopanthetheinyltransferase Sfp.

To introduce the gene cluster into the vector, the cloning vector pCC1FOS was replaced by the expression vector pHIS8-Svp by λ RED-mediated homologous recombination (figure 6.9). 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. 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. In preparation for homologous recombination, pHIS8-Svp had to be modified with two PCR fragments that

were homologous to the two end regions of the *swf* cluster carried by pPS2D9.

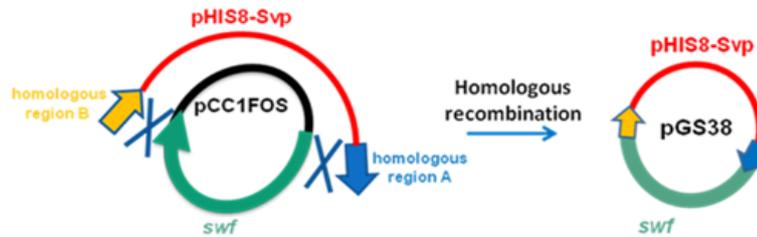


Figure 6.9 Insertion of the *swf* gene cluster into the vector pHIS8-Svp via homologous recombination

6.2.17.1 Digestion of the expression vector pHIS8-Svp

In order to prepare the linear vector pHIS8-Svp modified at both ends by adding with extensions homologous to the initial and the final regions of the *swf* cluster, the vector pHIS8-Svp was subjected to a double restriction enzyme digestion with *Bam*HI and *Hind*III, following the formula reported in the paragraph 6.2.16.2 for pGS21 cutting.

The vector was then loaded on an agarose gel and gel extracted and purified, eluting in 30 μ L of H₂O.

6.2.17.2 Amplification of pPS2D9 fragments

For homologous recombination, the expression vector pHIS8-Svp had to be flanked by two short PCR fragments that were homologous, respectively, to the initial and final regions of the *swf* gene cluster (figure 6.9).

To do so, a region upstream the *supE* permease (22091-22720, ~1000 bp) was amplified by standard PCR from the fosmid pPS2D9 using primers 5'-AAA GGA TCC tga cca cgc cct cgt gct ca-3' and 5'-AAA GAA TTC tgg acc act tcc ggc aac tac-3' (introduced restriction sites for *Bam*HI and *Eco*RI are underlined) and ligated into pBluescript II SK (+) via T/A cloning to

yield pGS15. The insert of pGS15 was cut out from the vector with *Bam*HI and *Eco*RI (see procedure described in Paragraph 6.2.16.2).

Another region at the end of the *swf* cluster (i.e. *swfC* 36006-37813) was amplified from pPS2D9 with the primers 5'- AAA GAA TTC tcg aac gcc etc etc atc tac c-3' (*Eco*RI restriction site is underlined) and 5'- AAA AAG CTT gcc cac tac gtg ctg cat cgg-3' (*Hind*III restriction site is underlined), ligated into pBluescript II SK(+) via T/A cloning to yield pGS29, and cut out from the vector with *Eco*RI and *Hind*III (see procedure described in Paragraph 6.2.16.2).

6.2.17.3 Flanking of the digested pHIS8-Svp

Inserts cut out from pGS21 and pGS29 were ligated in a three-point ligation into the *Bam*HI/*Hind*III sites of pHIS8-Svp. The resulting plasmid was named pGS27.

➤ Three point ligation formula:

4 μL pGS21 insert B/E
4 μL pGS29 insert E/H
1.5 μL pHIS8-Svp B/H
1 μL 10X KAPA T4 ligase buffer
0.5 μL KAPA T4 ligase

10 μL total volume. 16 °C overnight, then heat inactivation at 65°C for 15 min.

6.2.17.4 Linearization of the flanked vector pHIS8-Svp

The flanked vector pHIS8-Svp was introduced into electrocompetent *E. coli* XL1 blue cells, after which isolated DNA was linearized with the restriction enzyme *Eco*RI. The linear construct was purified from agarose gel using the QIAquick Gel Extraction Kit and dephosphorylated.

➤ Digestion formula

42.5 μ L pGS27

5 μ L 10X buffer H (Takara- 500mM Tris-HCl pH7.5, 100mM MgCl₂, 10mM DTT, 1M NaCl)

2.5 μ L *Eco*RI (Takara)

50 μ L total volume. 37°C for 2 hours.

6.2.17.5 Recombination of the fosmid pPS2D9

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). Homologous recombination was performed in the host strain *E. coli* BW25113 with the *red* helper plasmid pKD46 (Amp^R).

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 pKD46 was used, which involves the bacteriophage λ recombination system in order to inhibit any exonuclease activity, so that linear DNA molecules are preserved. In addition, this plasmid also has the promoter *araBAD* that can be regulated by arabinose and *repA101ts*, a temperature - sensitive origin of replication (30°C is the optimal temperature for growing cells transformed with pKD46; higher temperature causes loss of the plasmid).

In preparation for homologous recombination, the fosmid pPS2D9 (containing the whole *swf* cluster) was transformed into electrocompetent *E. coli* BW 25113 cells containing the plasmid pKD46. The transformed cells were selected on LB agar containing ampicillin (resistance gene on pKD46)

and chloramphenicol (resistance gene on the pPS2D9 fosmid). Subsequently, the linearized pGS27 (Section 6.2.17.4) was transformed into the electrocompetent *E. coli* BW25113/pKD46 cells containing pPS2D9. In this step, a homologous recombination took place.

The cloning vector pCC1FOS in pPS2D9 was replaced by the expression vector pHIS8-Svp.

Upon this recombination, a modified genetic construct was obtained, namely pGS38 (figure 6.9).

Recombinant cells were selected on LB plate with kanamycin due to the presence of a *kanR* gene inside the expression vector pHIS8-Svp.

1. Preparation of electrocompetent E. coli BW25113/pKD46 cells.

Electrocompetent *E. coli* BW25113/pKD46 cells were always prepared freshly before use. 50 μ L of *E. coli* BW25113/ pKD46 were used to inoculate with 6 mL LB containing ampicillin (100 μ g/mL – Amp₁₀₀) overnight at 30°C. In the next day, 5 mL of *E. coli* BW25113/ pKD46 from the overnight culture were inoculated in 200 mL LB medium + Amp₁₀₀. After growing for 3-4 hours at 30 °C with shaking at 250 rpm to an OD₆₀₀ of ~ 0.6, the cells were recovered by centrifugation at 5,000 rpm for 5 min at 4 °C. The medium was discarded and the pellet was resuspended by gentle mixing in 100 mL of cold 10% glycerol.

Then it was centrifuged and the pellet was resuspended in 50 mL of cold 10% glycerol, centrifuged and decanted. 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 μ L of cells were pipetted to pre-chilled 1.5 mL eppendorf tubes. The aliquots were stored at -80 °C.

2. Transformation of pPS2D9. Electrocompetent *E. coli* BW 25113/pKD46 cells were mixed with 1.5 μ L DNA of the donor fosmid pPS2D9 (containing the whole *swf* cluster). Immediately, electroporation was carried out in an ice-cold electroporation cuvette using a Bio-Rad electroporator set at 2.5 kV. 900 μ L of cold LB was added to the shocked cells and incubated for 1 hour at 30 °C shaking at 250 rpm. The cells were spread onto LB agar plate containing chloramphenicol (12.5 μ g/mL) and ampicillin (100 μ g/mL). The plate was incubated overnight at 30 °C to maintain plasmid pKD46 inside the strain.

3. Preparation of electrocompetent *E. coli* BW25113/pKD46 cells containing the donor fosmid pPS2D9. Electrocompetent *E. coli* BW25113/pKD46 cells containing the donor fosmid pPS2D9 were prepared freshly before use. One isolated colony of *E. coli* BW25113/pKD46 including pPS2D9 was picked, cultured overnight at 30 °C in 5 ml LB with ampicillin (100 μ g/mL) and chloramphenicol (12.5 μ g/mL). 80 mL of LB medium containing the two above antibiotics and 5% of the overnight culture of *E. coli* BW25113/pKD46 including pPS2D9 were used for inoculation. 800 μ L of 1 M L-arabinose stock solution was added for induction of the λ *red* genes on pKD46. It was grown for 3-4 hours at 30 °C with shaking at 250 rpm to an OD600 of ~ 0.6 .

The cells were recovered by centrifugation at 5,000 rpm for 5 min at 4 °C. The medium was discarded and the pellet was resuspended by gentle mixing in 40 mL of cold 10% glycerol. Then it was centrifuged and the pellet was resuspended in 20 mL of cold 10% glycerol, centrifuged and decanted. After the last wash in 10 mL of 10% cold glycerol, the cell pellet was resuspended in ~ 70 µL of cold 10% glycerol.

4. Transformation of the linear flanked *pHIS8-Svp* (*pGS27*). *E. coli* BW25113/pKD46 cells containing the donor fosmid pPS2D9 were mixed with 3 µL DNA of the linear flanked vector *pHIS8-Svp* (*pGS27*). Electroporation was carried out in ice-cold electroporation cuvette using a Bio-Rad electroporator set at 2.5 kV. Immediately, 900 µL of cold LB was added to the shocked cells and incubated shaking for 1 hour at 37 °C. In this step, a homologous recombination took place: the cloning vector pCC1FOS in pPS2D9 was replaced by the expression vector *pHIS8-Svp*, yielding the new recombinant expression plasmid pGS38 which contains the target gene cluster *swf*. The transformed cells were spread onto LB agar plate containing kanamycin (50 µg/mL). Because no further gene replacement continued to be carried out, the cells were incubated overnight at 37 °C to promote the loss of pKD46.

6.2.17.6 Verification of the recombinant pGS38

Homologous recombination normally does not occur in all copies of the fosmid in one cell. Thus, there were transformed cells containing both the

recombinant pGS38 and pPS2D9. 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 pGS38 (pHIS8-Svp + *swf* genes) were resistant to one antibiotic, kanamycin (resistance gene on the expression vector pHIS8-Svp). In contrast, the cells containing both pGS38 and the donor fosmid pPS2D9 were resistant to two antibiotics *i.e.*, kanamycin (resistance gene on the expression vector pHIS8-Svp inside pGS38) and chloramphenicol (resistance gene on the vector pCC1FOS inside pPS2D9). So, forty single transformed *E. coli* BW25113 colonies were picked to examine antibiotic resistances and were spread on both LB plate containing kanamycin and LB plate containing kanamycin and chloramphenicol. These two plates were incubated at 37°C overnight and cell growth was observed. One clone grew on LB plate + kan₅₀ but did not grow on LB plate + kan₅₀ + caf_{12.5} indicating that it was the strongest candidate possessing only the recombinant plasmid pGS38. O/n culture of this individual was prepared and subjected to DNA plasmid isolation. 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 the colony of *E. coli* BW25113 was introduced into electrocompetent *E. coli* XL1 blue cells and then spread on LB + kan₅₀ plates. 4 colonies were picked from the plate to prepare four different o/n cultures. Plasmids from the 4 colonies were isolated and characterized by restriction enzyme digestion with *Kpn*I (figure 6.10). All the 4 plasmids displayed the expected restriction pattern for the recombinant

construct pGS38. This result confirmed that the pGS38 construct was successfully obtained. Therefore, it was chosen for heterologous expression trials of *swf* genes in *E. coli* BAP1.

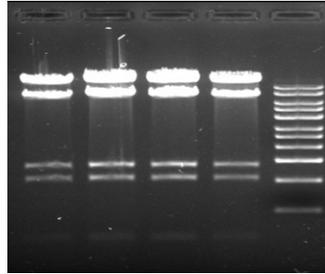


Figure 6.10 Agarose gel electrophoresis of checking pGS38 in four *E. coli* XL1 blue colonies. Restriction analysis was performed with *KpnI*.

➤ Digestion formula

6 μ L plasmid
 2.5 μ L dH₂O
 1 μ L 10X buffer L (Takara- 100 mM Tris-HCl pH7.5, 100mM MgCl₂, 10mM DTT)
 0,5 μ L *KpnI* (Takara – 10u/ μ L)

10 μ L total volume. The reaction was incubated for 2 hours at 37°C.

6.2.17.7 Heterologous expression trials in *E. coli* BAP1

The plasmid was introduced by electroporation into *E. coli* BAP1, which is a derivative of *E. coli* BL21-(DE3) including a gene for the 4'-phosphopanthetheinyltransferase Sfp.

The strains were grown in baffled 500 mL Erlenmeyer flasks in 100 mL MMGAGTr medium supplemented with 50 μ g mL⁻¹ kanamycin. MMGAGTr was prepared as follows: 200 mL M9 salts (64 g Na₂HPO₄ x 7 H₂O, 15 g KH₂PO₄, 2.5 g NaCl, 5.0 g NH₄Cl, *ad* 1000 mL milliQ H₂O), 500 μ L of 2000X trace element solution (1.2 g FeCl₃ x 6 H₂O, 1.4 g MnSO₄, 1.6 g CuSO₄, and 500 mL milliQ H₂O), 2 mL 1 M MgSO₄, 100 μ L 1 M CaCl₂, 10 g glucose, 5 g L-glutamic acid, *ad* 1000 mL milliQ H₂O.

Four clones of transformants were picked from the LB plates to inoculate 2.5 mL MMGAGTr medium in one test tube for overnight cultures (37 °C, 250 rpm) in order to inoculate the 100 mL medium in EM-flasks next day. After 2-3 h of shaking (250 rpm) at 37 °C the *T7* promoter was induced with 0.1 mM IPTG, and 5-10 nM cobalamin (vitamin B12) was added as a co-factor of SwfC.

The cells and the culture broths were harvested separately after 3-4 h of further growth at 30 °C, 250 rpm. The cell pellet was dissolved in 1 mL of nanopure water, frozen o/n, and after thawing on ice and ultrasonication, the lysates were subjected to extraction with 4 mL of MeOH. Culture broths were freeze dried o/n and extracted with 5 mL of MeOH/H₂O (8:2).

All the experiments were performed in triplicates.

6.2.17.8 LC-HR-ESI-MS analyses

The experiments were performed using a Thermo LTQ Orbitrap XL high-resolution ESI mass spectrometer coupled to an Agilent model 1100 LC, which included a solvent reservoir, in-line degasser, binary pump, and refrigerated autosampler.

A 2.6 μm Kinetex C18 column (50 × 2.1 mm), maintained at room temperature, was used. It was eluted at 150 μL min⁻¹ with H₂O and CH₃CN, using as a gradient elution 70–95% CH₃CN over 23 min and hold 13 min. Crude extracts from transformants and their culture broths were filtered and injected (5 μL) without any further workup.

Both positive-ion and negative-ion mass spectra were recorded in separate HPLC runs.

6.2.17.9 Lipid analysis

After growth the cultures were centrifuged in falcon tubes (5,000 rpm, 5 min) and the pellets were dissolved in 2 mL nanopure water and transferred to 15 mL glass vials with Teflon lined caps (Supelco). Lipid extraction was performed using a method adapted from the MIDI protocol. Nanopure water was prepared by adding 45 mL CHCl_3 to 900 mL autoclaved milliQ H_2O to remove lipid contaminations from the water. To the culture pellets dissolved in 2 mL nanopure water, 6 mL of a modified reagent MIDI 1 (32.3 g NaOH dissolved in 288.5 mL MeOH + 141.5 mL nanopure water) were added; the mixture was vortexed, and incubated in the 15 mL vials with tightly closed caps o/n at 50 °C. After cooling, the samples were divided on two tubes and 8 mL of reagent MIDI 2 (325 mL 6 N HCl + 275 mL MeOH) were added to each tube and vortexed. After incubation at 80 °C for exactly 10 min and cooling down to room temperature, the FAMES were extracted with 3 mL of reagent MIDI 3 (200 mL hexane + 200 mL MTBE). The upper layer with the FAMES was purified by passing a self made silica gel column [0.5 g silica gel 60 (Roth) transferred to a standard Pasteur pipette with 5 mL MeOH + 0.02 M ammonium acetate]. The FAMES were dried and subsequently dissolved in hexane (150 μL) for GC/MS measurements.

6.2.17.10 GC-MS analyses

Fatty acid methyl esters (FAMES) were analyzed by GC/MS (Agilent, 6850 series II/ 5973 Network MSD) on a HP-5MScapillary column (Agilent, 5% Phenyl Methyl Siloxane) (30 m, 0.25 mm \varnothing , 0.25 μm). Helium was used as a carrier gas, injection was in split mode, the program

(MOLBIO55) was as follows: hold 150°C for 15 min, heat to 300°C with 5°C/min, hold 300°C for 10 min.

6.2.18 Protein expression trials of SwfA and SwfB

The library fosmid pPS3I10 containing the *swfA* and *swfB* genes (PSA11D7 type) was used for heterologous expression trials in an expression system consisting of the vector pHIS8-Svp and the bacterial host *E. coli* BL21-CodonPlus[®](DE3)-RIPL.

To introduce the target genes into the vector, the cloning vector pCC1FOS was replaced by the expression vector pHIS8-Svp by λ RED-mediated homologous recombination, following the procedure described in details for the construction of the plasmid pGS38 (Paragraphs 6.2.17.1 – 5).

6.2.18.1 Cloning of the *swfA* and *swfB* genes

The upstream region of *swfA* starting directly after the start codon was amplified from pPS3I10 by standard PCR using primers 5'-AAA GGA TCC cgc gga gac gcg gtc gcg atc gtc ggc tac-3' (forward) and 5'-AAA GAA TTC cga cgg ccc cat cag gtt gaa gtg gta gga ga-3' (reverse), ligated into pBluescript II SK(+) via T/A cloning to yield pTH151, and cut out from pTH151 with *Bam*HI and *Eco*RI (introduced restriction sites are underlined). Another region downstream the *swfB* ORF was amplified from pPS3I10 with the primers 5'- AAA GGA TTC gct gca ggg gcg ctg gta cga-3' (forward) and 5'- C CC AAG CTT gcg caa tgg gcg tga ttc c-3' (reverse), ligated into pBluescript II SK (+) via T/A cloning to yield pGS13, and cut out from the vector with *Eco*RI and *Hind*III (introduced restriction sites are

underlined). Both inserts were ligated in a three-point ligation into the *Bam*HI/*Hind*III sites of pHIS8-Svp. The resulting plasmid (pGS7) was linearized with *Eco*RI and dephosphorylated with SAP (Promega). Homologous recombination was performed with this linear flanked vector (Kan^R) and the donor fosmid pPS3I10 (Cam^R) in the host strain *E. coli* BW25113 with the *red* helper plasmid pKD46 (Amp^R). Transformants containing the *swfA* and *swfB* genes cloned in the vector pHIS8-Svp, i. e. pGS40, were selected on LB + Kan₅₀ agar plate.

6.2.18.2 Verification of the recombinant pGS40

With the aim to verify the outcome of homologous recombination in the *E. coli* BW25113 – pKD46 cells, 20 clones were picked from the LB + Kan₅₀ agar plate and subjected to colony PCR screening using the primers T7 (forward) 5'- TAATACGACTCACTATAGGG -3', complementary to a region of the T7 promoter on the vector pHIS8-Svp, and KSPER (reverse) 5'-CGGTCGCCCACGACGGTGCCGGTCGCATG- 3', complementary to a region of the KS domain of *swfA* (nu 13249 – 13278). After screening, only one clone gave the band of the expected size (1063 bp) for the recombinant construct pGS40 (Figure 6.11).

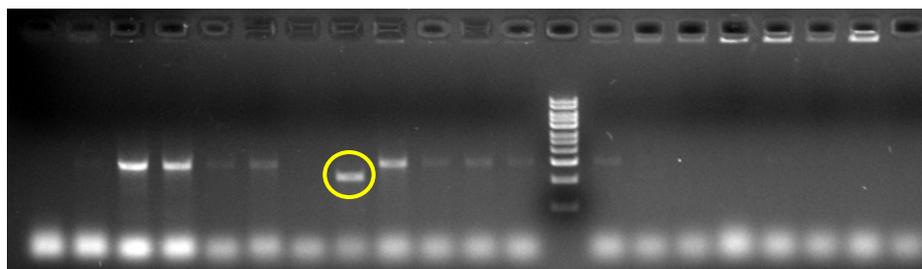


Figure 6.11 PCR verification of the recombinant plasmid pGS40 in *E. coli* BW25113.

As homologous recombination normally does not occur in all copies of the fosmid in one cell and there is the chance that transformed cells contain both the recombinant pGS40 and pPS3I10 (unwanted clones), plasmid DNA was isolated from the o/n culture of the single positive clone to transform by electroporation XL1-Blue cells, subsequently selected on LB + Kan₅₀ agar plate. Electroporation usually leads to stable integration of a single copy of the exogenous DNA, and therefore it was very likely that transformants growing on the plate contained only the recombinant plasmid pGS40.

3 colonies were picked from the plate to prepare three different o/n cultures. Plasmids from the 3 cultures were isolated and characterized by double restriction enzyme digestion with *Xho*I and *Xba*I.

➤ Digestion formula

7.9 μ L plasmid
1 μ L 10X buffer D (Promega - 60 mM Tris-HCl pH7.9, 60 mM MgCl₂, 10 mM DTT, 1.5 M NaCl)
0.1 μ L BSA (Promega – bovine serum albumin)
0.5 μ L *Xho*I (Promega - 10 u/ μ L)
0.5 μ L *Xba*I (Promega - 10 u/ μ L)

10 μ L total volume. The reaction was incubated for 2 hours at 37°C.

All the 3 plasmids displayed the expected restriction pattern for the recombinant construct pGS40. This result confirmed that the pGS40 construct was successfully obtained. Therefore, it was chosen for protein expression trials of *swfA* and *swfB* genes in *E. Coli* BL21-CodonPlus[®] (DE3)-RIPL.

6.2.18.3 Expression of SwfA and SwfB proteins

Proteins from plasmid pGS40 were expressed by IPTG induction of the T7 promoter in *E. Coli* BL21-CodonPlus[®] (DE3)-RIPL, previously transformed

with the L-arabinose inducible chaperone plasmid pTf16 (Cam^R). Culture conditions for *swfA* and *swfB* expression were the same as described for SwfA_{ACP} overexpression (see Paragraph 6.2.16.5), except for the additions to the culture broth of chloramphenicol 20 µg/mL (resistance gene on pTf16) and of L-arabinose 2 mg/mL in order to induce transcription of the *tig* gene carried by pTf16 encoding for the molecular chaperone Trigger factor. After cell lysis of transformants, the lysate was purified by immobilization on Ni-NTA matrix eluting with increasing concentration of imidazole in the lysis buffer, as for ACP protein purification (see Paragraph 6.2.16.5).

Eluted fractions were analyzed by SDS PAGE and only weak expression of SwfA was detected in the column flow-through and washing elution fractions, showing that SwfA was not retained by the agarose matrix although being expressed as His-tag fusion protein (Figure 6.12, Table 6.9).

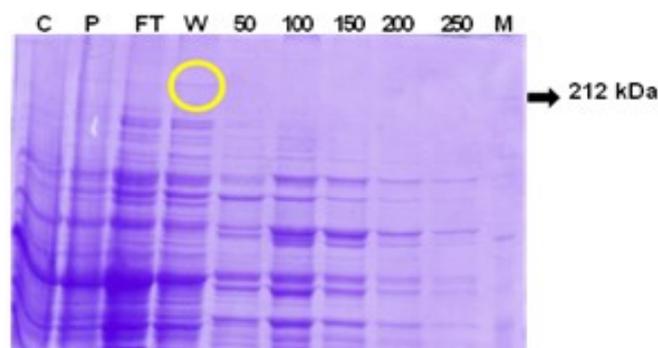


Figure 6.12 The protein SwfA was expressed as N-terminal His8-tag fusion protein and purified over Ni-NTA columns. Elution fractions (500 µL each) were obtained by increasing the imidazole concentration in the lysis buffer from 0 to 250 mM stepwise. Mini SDS-PAGE was loaded with an unpurified sample of cells and lysate (5 µL, lane C), resuspended cell pellet (5 µL, lane P), column flow through, containing unbound proteins (12 µL, lane FT), washing fractions eluted with lysis buffer (20 µL, lanes W), and fractions eluted with increasing imidazole concentrations (20 µL each, lanes 50, 100, 150, 200, and 250). All the fractions were previously mixed with 3X SDS loading buffer. Lane M is a pre-stained protein marker (Roti[®]-Mark STANDARD).

A band corresponding to a ~250 KDa protein in agreement with the expected size of SwfA (*apo*-SwfA 255,495 KDa, *holo*-SwfA 255,835 KDa) and absent in the negative control, was indeed observed in the SDS PAGE gel from protein extract of pGS40 transformants.

On the other side, no evidence of SwfB expression could be detected.

Table 6.9 Composition of the stacking and resolving gels for SwfA SDS-PAGE.

Composition	4% stacking gel	8% resolving gel
30:0.8% w/v acrylamide:bisacrylamide	660 μ L	2mL
1M Tris-Cl pH8.8	-	3mL
1M Tris-Cl pH6.8	630 μ L	-
20% SDS	25 μ L	38 μ L
dH ₂ O	3.6 mL	2.43 mL
10% APS	25 μ L	36 μ L
TEMED	5 μ L	5 μ L

6.2.19 Fosmid sequencing

The fosmid pPS11G3 was found with AT1F/AT3R2 primers and shotgun sequenced. A new fosmid pPS2D9, was isolated from the library with 11G3_SAM_up (5'-ACC CGA AGC AGC CTC CCA CCT ACT-3') and 11G3_SAM_low (5'- CCC CGC GAG AAC TGC AGA CAC ATC-3') primers and was shown to contain the complete cluster by end sequencing of the insert. To close the gap of PS11G3 cluster, 11G3_close_gap_up (5'-CGA AGA CCG CTC CTT CCT C-3') binding to EDRSFL and 11G3_close_gap_low (5'-GCC GCT GAC CGG CAC TCT-3') binding to PLTGT were used on the template of pPS2D9. For primer walking to obtain the end of 11G3 the primer 2D9_restSAMup (5'-AGC AGA TGG TGG CCG AGT TCG ACT-3') was used. The second completely sequenced fosmid (pPSA11D7) was isolated from the library using ATPEF (5'-ATG

GTG TTT TCG GGG CAG GGC ACG CA-3'; MVFSGQGTQ) and ATPER (5'-GGC GGC GGC CAC CTC GCC CGA GCT GTG TCC-3'; GHSSGEVAAA). More representatives were isolated from the library using A11D7_54F (5'-GTT TCC TGG GAC ACC TTC AG-3') and A11D7_530R (5'-GGT GAG CTT TGC GTT GTT G-3').

To obtain the missing final part of the sequence PSA11D7, 3i10E_lastF (5'-CGA GCC TCG CGA GTT CAC C-3') was used.

6.2.20 Plasmid sequencing

Plasmids including pBluescript II SK(+) as vector were sent to GATC Biotech AG (Konstanz, Germany) for single read sequencing using the T7 primer and sequencing was performed on ABI 3730XL (Applied Biosystems).

6.2.21 Bioinformatics

Sequences were analyzed using BLASTp and BLASTx (Altschul *et al.*, 1997) and aligned with BioEdit (Hall, 1999). Phylogenetic analyses (Neighbor Joining, 111 replicates, 1000 bootstraps) were performed using ClustalX (Larkin *et al.*, 2007) or the MEGA 5.05 software package (Tamura *et al.*, 2011).

6.2.22 454 sequencing of KS and AT amplicons from *P. simplex*

KS and AT amplicon mixtures were obtained by PCR amplification of metagenomic DNA from *P. simplex* using respectively two primers pairs, designed on the *signature regions* of type I PKSs: KSDPQQF (5'-MGN

GAR GCN NWN SMN ATG GAY CCN CAR CAN MG-3') and KSHGTGR (5'-GGR TCN CCN ARN SWN GTN CCN GTN CCR TG-3') for the KS domains, AT1F (5'-TTY CCN GGN CAR GGN NSS CAG TGG-3', binding to the motif FPGQGsQW) and AT3R2 (5'-GC IGC IGC NAT CTC NCC C-3', binding to the motif QGEIAAA) for the AT domains. PCR conditions are described in details in the Paragraph 6.2.2.6, *PCR screening of metagenomic DNA for type I PKSs*. KS and AT PCR products were separated by gel electrophoresis, and gel extracted and purified using the QIAquick Gel Extraction Kit. The concentrations of KS and AT amplicon libraries were adjusted to 20 ng/ μ L and 50 μ L of both mixtures were sent to GATC Biotech AG European Genome and Diagnostics Centre (Konstanz, Germany) to be subjected to 454 pyrosequencing. Sequencing was done on a 70 x 75 FLX picotiter plate of a Roche GS FLX sequencer.

6.2.23 Accession codes

The pPS11G3 and pPSA11D7 complete nucleotide sequences were deposited into GenBank respectively under the accession numbers JX946307 and JX946308. The AT gene partial sequences (cited in Chapter 4) from PCR screening were deposited into GenBank under the accession numbers KC424641 through KC424644 (primers AT1F/AT3R2) and JX946309 through JX94633 (primers SWF_ATF/SWF_ATR).

PART II

*Isolation and structure elucidation of new
secondary metabolites from the marine sponges
Chalinula molitba and Plakortis cf. lita*

Chapter 7

Isolation procedures and structural determination methods

The core activity of natural product chemistry has traditionally been, and still is, the isolation of new compounds from natural sources. In the frame of the more general purpose of discovering and developing new drugs from natural products, the research work described in the following chapters is focused on the identification of new secondary metabolites from different specimens of sponges living in tropical oceans, wonderful sources of unusual molecular architectures.

The work leading to the isolation of new compounds consists mainly of two subsequent steps:

- purification and isolation of the target compound(s) from the animal tissues
- stereostructural determination of the new molecule(s).

The methods used for the isolation and structural determination are a matter that deserves preliminary discussion. In the next two sections, an overview will be given about the main techniques and procedures involved in the identification of novel marine secondary metabolites.

7.1 Isolation procedures

Through the experience gained along years of research job in the isolation of marine natural products, my research group has developed a

general efficient strategy for purification of crude extracts from marine organisms.

Basically, the homogenate of the tissue is extracted subsequently with methanol, mixtures of methanol and chloroform in different proportions, and finally with pure chloroform, yielding a complex extract containing virtually all the low molecular weight metabolites. The combined extracts (except for the chloroform extract) are then partitioned between water and butanol, most secondary metabolites being recovered in the butanol phase. The crude butanol extract is combined with the chloroform extract and then subjected to reversed-phase column chromatography as the first step of the purification, according to the following elution scheme based on increasing apolarity of the mobile phases:



The next steps of the separation procedure consists in purifying the eluted fractions A1, A2, A3, A4 and A5 through reversed phase HPLC. In contrast, fractions B and C from the reversed-phase column are usually subjected to a further column chromatography separation on SiO₂ with eluents of increasing polarity, usually *n*-hexane/ethyl acetate mixtures.

Finally, normal-phase and/or reversed-phase HPLC separations of these partially purified fractions are performed in order to isolate pure compounds, the particular procedure used depending on the particular metabolite to be isolated.

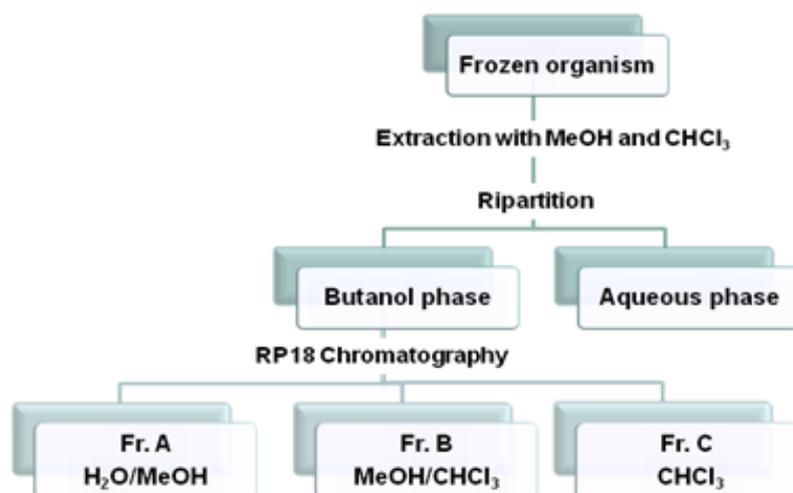


Figure 7.1 Secondary metabolite isolation scheme.

7.2 Structural elucidation

Until a few decades ago, structural determination of new organic compounds was only pursued through the use of chemical techniques (degradation and interconversion of functional groups). The development of spectroscopic techniques dramatically changed this approach. Today, it is possible to determine complex organic structures completely in a non-destructive way, with submilligram samples. Structural determination described in this thesis is largely based on spectroscopic techniques, mostly mass spectrometry (MS) and nuclear magnetic resonance (NMR), even if degradation methods were used in some cases.

7.2.1 Mass Spectrometry

The first step in the study of a new bioactive compound is the determination of its molecular formula using high resolution mass spectrometry.

Mass spectrometry can measure the molecular mass of a compound on the basis of the mass-to-charge ratio (m/z ratio) of ions produced from the molecules. A very accurate measurement of the molecular mass (high resolution mass spectrometry) can also provide the molecular formula of the molecule under study.

The *source* is the component of the mass spectrometer which produces ions from the molecule, while the *analyzer* measures the mass-to-charge ratio of the ions. There are many different types of sources, as well as of analyzers. During or after ionization the molecule may fragment, and the mass of the fragments provides information on the structure of the molecule under examination. If the ions do not fragment by themselves, they may be induced to fragment by letting them collide with gas molecules. In this case, a second analyzer is used to measure the mass of the fragments. This is known as tandem mass spectrometry or MS/MS.

Most of compounds described in the following sections were analyzed by ESI mass spectrometry. The ESI source is a widely used technique for polar and/or charged macromolecules. The sample is dissolved in a volatile solvent like H₂O, MeOH, and CH₃CN; volatile acids, bases or buffers are often added to the solution. This solution is pumped through a charged metal capillary, and it forms a spray while coming out of the capillary. Because of the electric potential of the capillary, each droplet of the spray carries an excess positive or negative charge, and this causes extensive protonation or deprotonation of the molecules of the sample, which become ions. An uncharged carrier gas such as nitrogen is used to help the liquid to nebulize and the neutral solvent in the droplets to evaporate.

As the solvent evaporates, the ionized analyte molecules become closer and closer, until they can escape from the droplet by electrostatic repulsion. For molecules with a high molecular weight, the ions may take more than one proton (up to some tens), and therefore may have multiple charge. Formation of multiply charged ions allows the analysis of high molecular weight molecules such as proteins, because it reduces the m/z ratio of the ions, which is therefore easier to measure.

7.2.2 Nuclear Magnetic Resonance

The most important spectroscopic technique used for structure elucidation of the isolated secondary metabolites was nuclear magnetic resonance (NMR). In addition to standard ^1H and ^{13}C NMR spectra, a large use of 2D NMR experiments was made. They are superior to their 1D NMR counterparts both for the shorter acquisition times, and for the easier assignment of nuclei resonating in crowded regions of the spectra (signal overlapping is much less likely in two dimensions than in one).

The COSY (Correlation SpectroscopY) experiment is one of the simplest and yet most useful 2D NMR experiment. It allows determination of the connectivity of a molecule by identifying which protons are scalarly coupled. In spite of the many modifications which have been proposed along the years, the very basic sequence composed of two $\pi/2$ pulses separated by the evolution period t_1 is still the best choice if one is simply dealing with the presence or the absence of a given coupling, but not with the value of the relevant coupling constant.

The TOCSY (Total Correlation SpectroscopY) experiment is a 2D NMR experiment very useful in the analysis of molecules composed of many separate spin systems, such as oligosaccharides or peptides. The TOCSY spectrum shows correlation peaks between nuclei that may be not directly coupled, but are still within the same spin system. The appearance of a TOCSY spectrum resembles in all aspects a COSY; the difference is that the cross peaks in a COSY result from coupled spins, whereas in the TOCSY spectra they arise from coherence transfer through a chain of spin-spin couplings, and therefore any pair of protons within a spin system may give rise to a peak. The range of the coherence transfer (i.e. through how many couplings the coherence may be transferred) increases with increasing mixing times (Δ), but a mixing time too long may reduce sensitivity.

The HSQC (Heteronuclear Single Quantum Correlation) and HMQC (Heteronuclear Multiple Quantum Correlation) experiments are 2D NMR heteronuclear correlation experiments, in which only one-bond proton-carbon couplings ($^1J_{\text{CH}}$) are observed. In principle, the HSQC experiment is superior to HMQC in terms of selectivity and additionally allows DEPT-style spectral editing. However, the sequence is longer and contains a larger number of π pulses, and is therefore much more sensitive to instrumental imperfections than HMQC.

The HMBC (Heteronuclear Multiple Bond Correlation) experiment is a heteronuclear two- and three-bond ^1H - ^{13}C correlation experiment; its sequence is less efficient than HSQC because the involved $^{2,3}J_{\text{CH}}$ couplings are smaller (3-10Hz). Moreover, while $^1J_{\text{CH}}$ are all quite close to each other, $^{2,3}J_{\text{CH}}$ can be very different, making impossible to optimize the experiment

for all couplings. As a result, in most HMBC spectra not all of the correlation peaks which could be expected from the structure of the molecule are present.

The ROESY (Rotating-frame Overhauser SpectroscopY) experiment is a chemical shift homonuclear correlation which can detect ROEs (Rotating-frame Overhauser Effect). ROE is similar to NOE, being related to dipolar coupling between nuclei, and depending on the geometric distance between the nuclei. While NOE is positive for small molecules and negative for macromolecules, ROE is always positive. Therefore, the ROESY experiment is particularly useful for mid-size molecules, which would show a NOE close to zero. The ROESY sequence is similar to the TOCSY sequence, and unwanted TOCSY correlation peaks may be present in the ROESY spectra. Fortunately, these unwanted peaks can be easily recognized, because their phase is opposite compared to ROESY correlation peaks. It is important to acquire ROESY spectrum in phase-sensitive mode for a correct interpretation of the spectrum.

7.2.3 Circular Dichroism

Circular dichroic (CD) spectroscopy of optically active compounds is a powerful method for studying the three-dimensional structure of organic molecules, and can provide information on absolute configurations, conformations, reaction mechanisms, etc. The CD spectroscopy takes advantage of the different absorption shown by chiral compounds of left and right circularly polarized UV/Vis light. In circularly polarized light, the electric field vector rotates about its propagation direction forming a helix in

the space while propagating. This helix can be left-handed or right-handed, hence the names left and right circularly polarized light.

At a given wavelength, circular dichroism of a substance is the difference between absorbance of left circularly polarized and right circularly polarized light:

$$\Delta A = A_L - A_R$$

Since circular dichroism uses asymmetric electromagnetic radiations, it can distinguish between enantiomers. Two enantiomers have the same CD spectra, but with reversed sign.

Of course, in order to show a *differential* absorbance, the molecules need to absorb the UV/Vis light, and therefore must possess at least one chromophore. If the molecule does not have a chromophore, this can be introduced using a derivatization reaction. This is why methyl glycosides are benzoylated to determine their absolute configuration.

One of the most important methods to establish the absolute configuration of a molecule is the exciton chirality method. This method is based on the interaction between two chromophores. When two or more strongly absorbing chromophores are located nearby in space and constitute a chiral system, their electric transition moments interact spatially (exciton coupling) and generate a circular dichroism. Because the theoretical basis of exciton coupling are well understood, it is possible to correlate the CD spectrum of an exciton-coupled chromophore system with the spatial orientation of the chromophores, which in turn can be related to the absolute configuration of the molecule. It is important to point out that, unlike for example optical rotation, the exciton chirality method does not require any

reference compound to provide the absolute configuration of the molecule under study (if its conformation is known).

Chapter 8

Chalinulasterol, a Chlorinated Steroid Disulfate from *Chalinula molitba*

Sulfated sterols are a well-known class of secondary metabolites, often found in sponges and echinoderms,^[1,2] that are emerging as a new potential class of lead compounds in the research for new drugs. A recent paper^[3] reported the isolation from the sponge *Theonella swinhoei* of solomonsterols A (**2**) and B (**3**), two tri-sulfated sterols based on the cholane skeleton. They differ from each other in the length of the side chain, and are among the few examples of truncated steroid derivatives isolated from marine sources. Both compounds have shown an important pharmacological activity, in that they are agonists of the PXR nuclear receptor.^[4] PXR receptor is a transcription factor which is able to bind to a wide spectrum of structurally distinct endobiotic substrates and xenobiotic compounds, and is involved in innate immunity, xenobiotic metabolism, and detoxification.^[5,6] PXR is proving to be an attractive target for small molecule drug discovery. In recent years, potential applications for exogenous PXR ligands have emerged in the treatment of important pathologies such as cancer^[7] and inflammatory diseases.^[8]

As part of a research program focused on the search for new anti-inflammatory and anti-cancer lead compounds from marine sponges, the chemical composition of the Caribbean sponge *Chalinula molitba* (de

Laubenfels, 1949), a light purple sponge that habits the mangroves of Little San Salvador (Bahamas Islands), was analyzed. This study led to isolation and structural identification of chalinulasterol (**1**) a new chlorinated sterol disulfate (Figure 8.1). Compound **1** has a close structural relationship with **2**, differing from the latter compound in having a chlorine atom instead of a sulfate function at position C-24 of the side chain. This relationship prompted us to investigate the possible role of **1** as modulator of the PXR receptors.

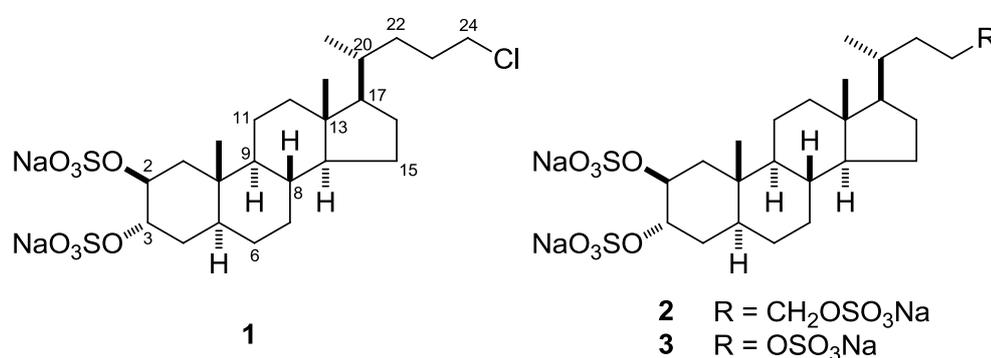


Figure 8.1 Structures of chalinulasterol (**1**) and of solomonsterol A (**2**) and B (**3**).

8.1 Sterols

Sterols are one of the possible end products of isoprenoid biosynthetic pathway, in which a huge variety of cyclic and acyclic compounds are built up from the universal C₅ building block isopentenyl diphosphate (IPP or IDP). IPP can be formed by 2 different pathways (figure 8.2):

- MVA pathway: IPP is formed from three molecules of acetyl-CoA via mevalonate (this pathway is found in non-photosynthetic eukaryotes)

- MEP pathway: IPP is formed from pyruvate and glyceraldehyde which form 1-deoxyxylulose 5-phosphate (DOXP) converted by intramolecular rearrangement and reduction into 2-C-methylerythritol 4-phosphate (MEP).

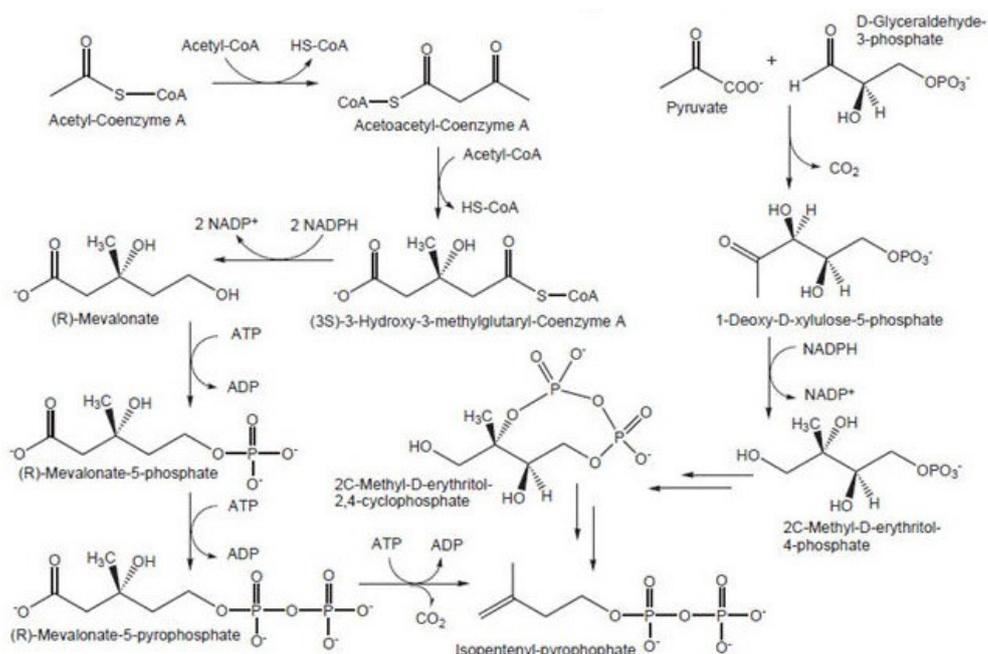


Figure 8.2 MEV and MEP pathways for IPP biosynthesis.

A key intermediate in the biosynthesis of sterols is the oxygenated C_{30} isoprenoid oxidosqualene (figure 8.3). Animals, fungi, and dinoflagellates cyclize oxidosqualene to lanosterol as the first cyclic intermediate in sterol biosynthesis. Higher plants, most microalgae and many Protozoa convert oxidosqualene to cycloartenol.

Porifera possess the most diverse and unique collection of sterols present in the entire metazoan kingdom. Sponge sterols originate by *de novo* synthesis, dietary intake, transformation of dietary sterols and interaction with associated microorganisms. Unique sterols emerge

primarily from modifications (oxygenation, alkylation, degradation) to the nucleus and/or the side chains, but also from fundamental structural transformations of the basic skeleton.

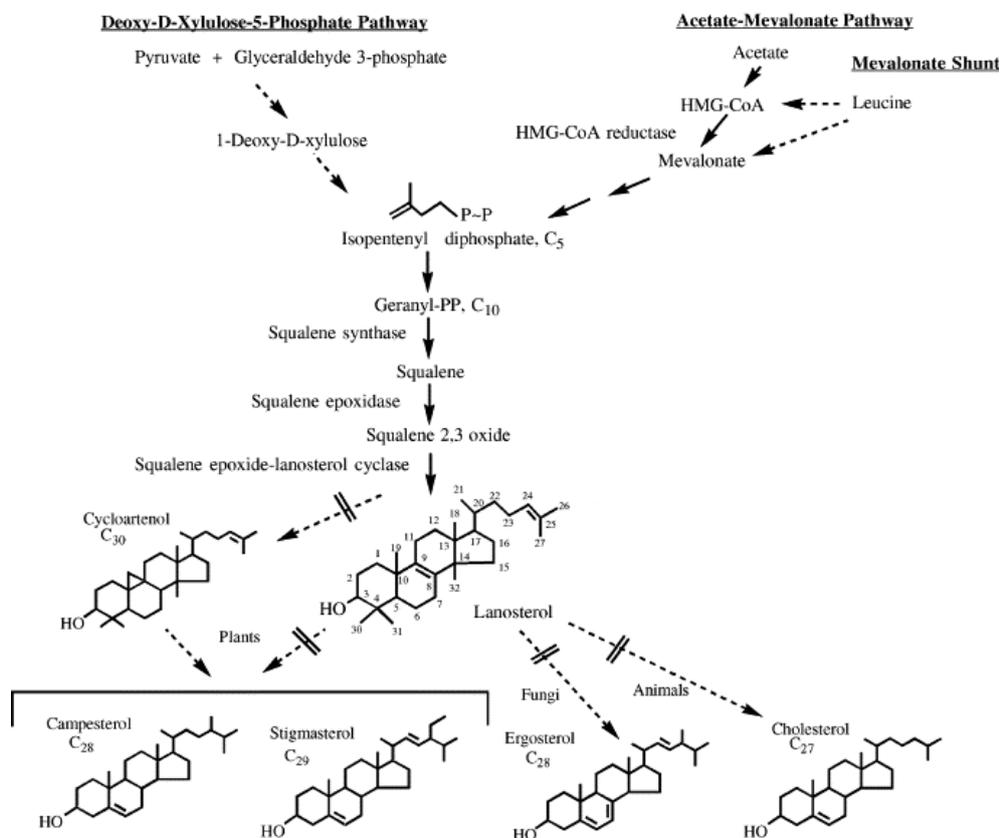


Figure 8.3 *De novo* biosynthesis of sterols.

For many years, the essential carbon backbone of sterols has been considered to range from C₂₇ to C₂₉, with variations occurring exclusively at C₂₄ in the side chain. Later, widespread diffusion of C₂₆ sterols in marine invertebrates and phytoplankton was detected (C₂₆ sterols were first isolated in 1970 from the mollusk *Placopecten Magellanicus*), and sterols with unprecedented structures were isolated from marine sources from then on. It's widely documented in the literature^[9] that in some sponge species unconventional steroids are predominant in the cell membranes compared to cholesterol or other

conventional 3- β -hydroxysterols. The peculiarity of sterols occurring in the cell membranes of sponges is probably related to the unique fatty acid composition of their phospholipids. It has been hypothesized that the structural modifications exhibited by the sponge sterols may be a sort of structural adjustments for a better fit with other membrane components which are so different from those of other higher animals.^[10] “Unconventional” sponge sterols include compounds having unusual nuclei, displaying a variety of oxygenated functionalities such as polyhydroxy, epoxide, epidioxy, and mono or polyenone systems. In addition, sponge sterols often include in their atypical structures side chains modified by the apparent loss of carbon atoms or by the addition of extra carbon atoms at biogenetically unprecedented positions. Quaternary alkyl groups, cyclopropane and cyclopropene rings, allenes and acetylenes are the commonest features occurring in the side chain of marine steroids.^[9]

Highly functionalized steroids have been the focus of marine sterol research because of their considerable biological and pharmaceutical activities.

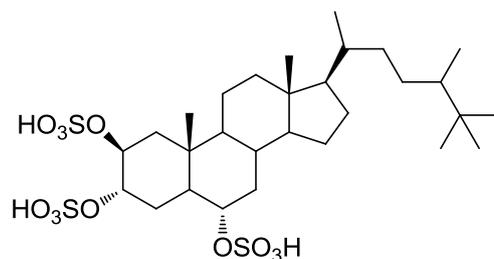


Figure 8.4 Halistanol sulfate.

Halistanol sulfate (figure 8.4) was the first example of polyoxygenated sterol, isolated from the sponge *Halichondria moorei*. Besides its biogenetically intriguing *tert*-butyl moiety in the side chain and the 2 β ,3 α ,6 α -trisulfoxy functions, this compound revealed interesting pharmaceutical properties against HIV virus.^[9]

Weinbersterol disulfate A and B (figure 8.5), with an unusual cyclopropane side chain, exhibited *in vitro* activity against the feline leukemia virus. In addition, the former compound showed pharmacological effects against HIV.

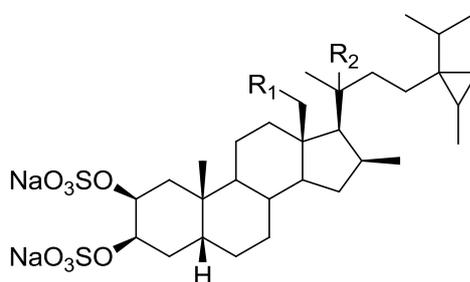


Figure 8.5 Weinbersterol disulfate A ($R_1=H$; $R_2=OH$) and B ($R_1=OH$; $R_2=H$).

Another remarkable example was the 9,11-secosterol herbasterol, a polyhydroxylated sterol responsible for the ichthyotoxic and antimicrobial effects of the methanol extract of the sponge *Dysidea herbacea*. Contignasterol (figure 8.6) represents the first marine steroid with a *cis* C/D ring junction, as well as a cyclic hemiacetal functionality at C-29 in the side-chain; it showed to be a potent inhibitor of histamine release from rat mast cells induced by anti-IgE.

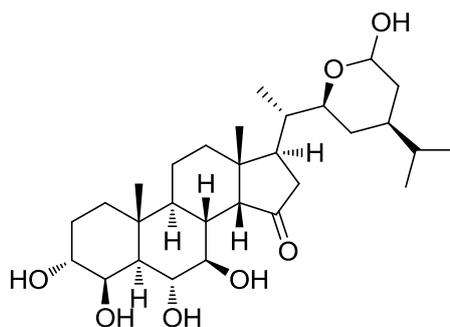


Figure 8.6 Contignasterol.

Other examples of sterols with unconventional nuclei are theonellasterols and conicasterols^[11,12], isolated from *Theonella swinhoei* (figure 8.7). These compounds are known to act as FXR (farnesoid-X-receptor) antagonists and PXR (pregnane-X-receptor) agonists, role of pharmacological and clinical relevance in cholestasis and in several human disorders, including cancers of the esophagus, stomach and pancreas.

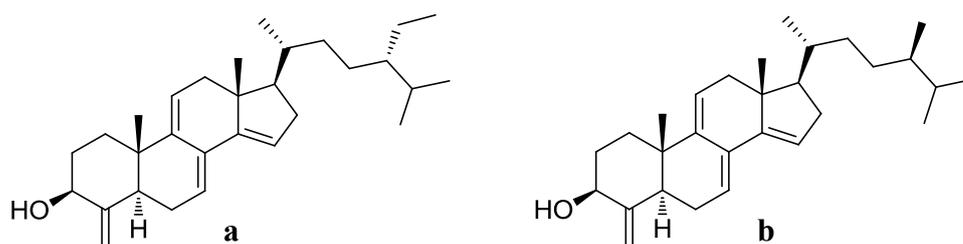


Figure 8.7 Theonellasterol B (a); Conicasterol B (b).

8.2 Structure elucidation of chalinulasterol

The structure of chalinulasterol was elucidated using mass spectrometry and NMR experiments.

The positive-ion high-resolution ESI mass spectrum of chalinulasterol (**1**) displayed a $[M + Na]^+$ pseudomolecular ion peak at

m/z 623.1447, in accordance with the formula $C_{24}H_{39}ClS_2O_8Na_3^+$ for this ion (calcd. 623.1462) (figure 8.8).

The intensity of the (M + 2) isotopic peak in the MS spectrum (45%, calcd. 46.5%) and the peak at m/z 587.1675 ($[M - HCl + Na]^+$) in the HRMS/MS spectrum confirmed the presence of a chlorine atom in the molecule.

The MS/MS fragmentation pattern of **1** also revealed the presence of two sulfate groups from the peaks at m/z 503.1946 $[M - NaHSO_4 + Na]^+$, 262.8870 $[2NaHSO_4 + Na]^+$, and 244.8765 $[Na_2S_2O_7 + Na]^+$.

The molecular formula was confirmed by the pseudomolecular ion peak at m/z 577.1669 observed in the negative-ion ESI mass spectrum, accounting for $C_{24}H_{39}ClS_2O_8Na^-$ (calcd. 577.1678).

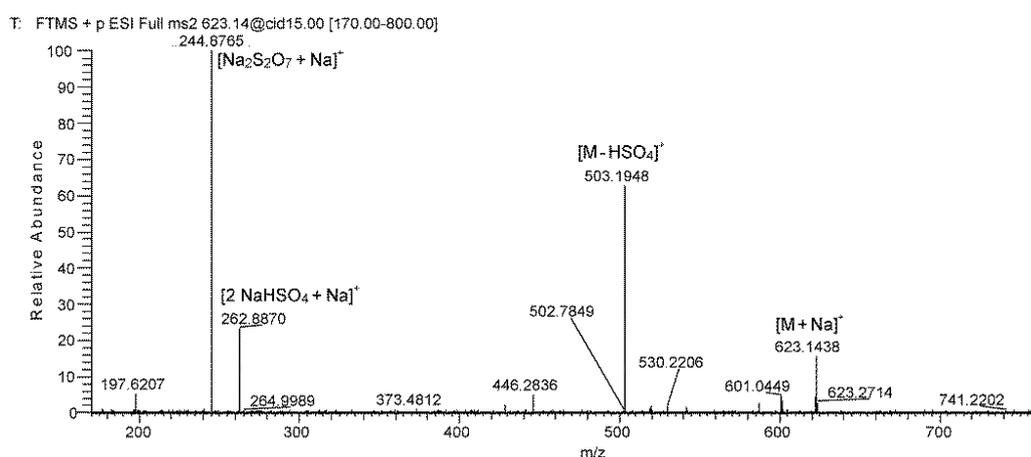


Figure 8.8 Chalinulasterol.
Positive-ion ESI MS/MS spectrum, parent ion at m/z 623.

Inspection of the 1H NMR spectrum of compound **1** showed two methyl singlets at δ 0.69 (H_3 -18) and δ 1.01 (H_3 -19) and one methyl doublet at δ 0.95 (H_3 -21) suggesting its steroidal nature (figure 8.9).

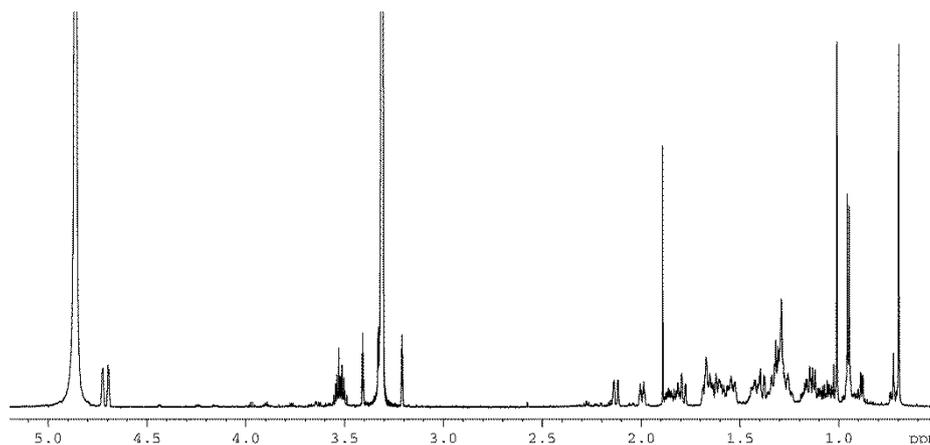


Figure 8.9 ^1H NMR spectrum of chalinulasterol (CD_3OD , 700 MHz).

The steroidal backbone could be assembled through the interpretation of COSY, TOCSY, HSQC and HMBC 2D NMR experiments. Analysis of the COSY and TOCSY spectra allowed the sequential assignment of all the protons of the tetracyclic system. The sulfate groups were located at position C-2 and C-3 because of the low-field resonances of H-2 and H-3 (δ 4.72 and 4.69) and of the respective carbon atoms C-2 and C-3 (δ 76.5 and 76.3). The HMBC correlation peaks of the methyl protons H₃-19 with C-1, C-5, C-9, and C-10 and of H₃-18 with C-12, C-13, C-14, and C-17 located the A/B and C/D ring junctions and completed the planar structure determination of the steroid ring system (figure 8.10).

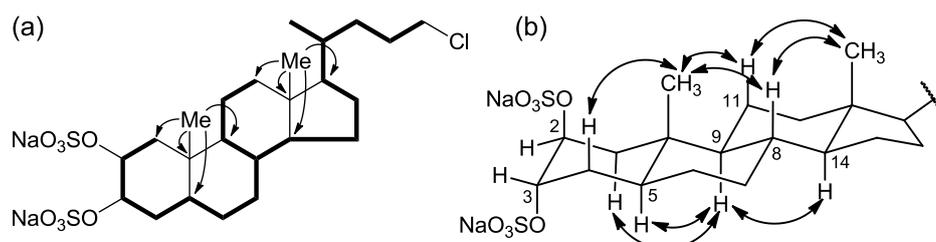


Figure 8.10 (a) Selected COSY and HMBC correlations of compound **1**, represented as bold bonds and arrows, respectively. (b) Selected ROESY correlations detected for **1**.

Information on the side chain was also provided by analysis of 2D NMR data. The COSY correlation between H-17 and the multiplet at δ 1.44 identified H-20; the latter was coupled with the methyl H₃-21 and the protons at δ 1.15 and 1.56 (H-22a and H-22b), in turn coupled with the methylene protons at δ 1.65 and 1.81 (H-23a and H-23b). The coupling of H-23a and H-23b and the two protons at δ 3.51 and 3.53 (H-24a, and H-24b) could be also evidenced from the COSY spectrum. The linkage to this latter methylene group of the chlorine atom required from the molecular formula was shown by its ¹H (δ 3.51 and 3.53) and ¹³C (δ 46.4) chemical shifts.

Analysis of the ROESY spectrum, supported by coupling constant analysis, defined the A/B, B/C and C/D *trans* ring junctions of a 5- α -cholane skeleton. The axial orientation of H-5, H-8, H-9, and H-14 was apparent from their respective coupling constants (Table 8.1), that of the angular methyl groups from the ROESY correlations of both H₃-18 and H₃-19 with H-8 and the axial H-11 β . On this skeleton, the small coupling constants showed by H-2 and H-3 illustrated their equatorial orientation, and therefore the diaxial (*i.e.*, 2 β ,3 α) orientation of the two sulfate groups. According to this information, the structure of chalinulasterol (**1**) was established as sodium 24-chloro-5 α -cholane-2 β ,3 α -diyl 2,3-disulfate.

8.3 Evaluation of chalinulasterol as PXR Receptor Modulator

Considering the structural similarity with the PXR agonist solomonsterol A (**2**), we have investigated a possible role of chalinulasterol (**1**) in regulating the pregnane-X receptor activity and carried out a transactivation assay on HepG2 cells, a human hepatocarcinoma cell line as described in the Experimental Section. As shown in Figure 8.11 A, despite the structural similarity with **2**, compound **1** failed in transactivating PXR. We have also investigated the possibility that **1** could act as potential PXR antagonist. As shown in Figure 8.11 B, **1** failed to reverse the induction of luciferase caused by rifaximin, indicating that it was not a PXR antagonist. Similar results have been obtained by analyzing the effect exerted by **1** in terms of regulation of PXR mediated induction of Cyp3A4 gene. Indeed, **1** also failed to induce Cyp3A4. Although negative, these results have an important implication in terms of structure-activity relationship, because they suggest that the sulfate group present at position C-24 of compound **2** is essential in the ligand-receptor binding. This can be rationalized by the binding model of **2** to the PXR receptor proposed in [3], in which a clear interaction of the 24-sulfate with the positively charged Lys210 is observed, and further supports this model.

Although halogen-containing secondary metabolites are well-known and abundant in nature, particularly in marine organisms, only a few natural chlorinated steroids have been reported so far,^[13,14] and there is only one example in the literature of a sulfated and chlorinated steroid.^[15]

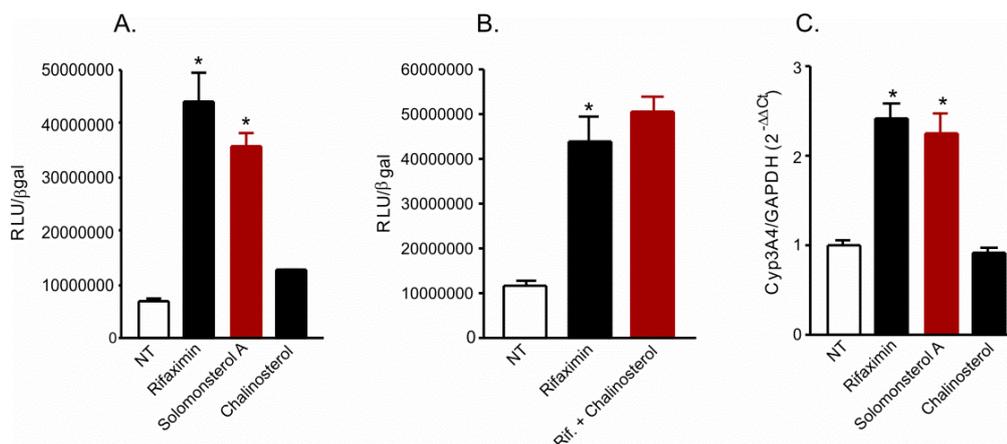


Figure 8.11 (A,B) Luciferase reporter assay. HepG2 cells, a hepatocarcinoma cell line, were transiently transfected with pSG5-PXR, pSG5-RXR, pCMV-βgalactosidase and p(CYP3A4)-TK-Luc vectors and then stimulated with (A) 10 μM rifaximin, solomonsterol A (2) or chalinulasterol (1) for 18 h, or (B) 10 μM rifaximin alone or in combination with 50 μM of compounds 1. Relative Luciferase Units were normalized with β-galactosidase Units (RLU/βgal). NT, not treated. * $P < 0.05$ versus NT cells. Data are mean ± SE of three experiments. (C) Real-Time PCR of Cyp3A4. HepG2 cells were stimulated with 10 μM rifaximin, 2 or 1 for 18 h. Total RNA was extracted and the relative mRNA levels of PXR target gene Cyp3A4, was measured. Values were normalized relative to GAPDH mRNA and are expressed relative to those of untreated mice, which are arbitrarily set to 1. Analysis was carried out in triplicate and the experiment was repeated twice. NT, not treated. * $P < 0.05$ versus NT cells.

8.4 Experimental Section

8.4.1 General experimental procedures

High-resolution ESI-MS spectra were performed on a Thermo LTQ Orbitrap XL mass spectrometer. The spectra were recorded by infusion into the ESI source using MeOH as the solvent. Optical rotations were measured at 589 nm on a Jasco P-2000 polarimeter using a 10-cm microcell. CD spectra were recorded on a Jasco J-710 spectrophotometer using a 1-cm cell. NMR spectra were determined on Varian Unity Inova spectrometers at 700 and 500 MHz; chemical shifts were referenced to the residual solvent signal (CD₃OD: δ_H 3.31, δ_C 49.00). For an accurate

measurement of the coupling constants, the one-dimensional ^1H NMR spectra were transformed at 64K points (digital resolution: 0.09 Hz). Through-space ^1H connectivities were evidenced using a ROESY experiment with a mixing time of 450 ms. The HSQC spectra were optimized for $^1J_{\text{CH}} = 142$ Hz, and the HMBC experiments for $^{2,3}J_{\text{CH}} = 8.3$ Hz. High performance liquid chromatography (HPLC) separations were achieved on a Varian Prostar 210 apparatus equipped with a Varian 350 refractive index detector.

8.4.2 Collection, Extraction, and Isolation

Specimens of *Chalinula molitba* were collected in the mangroves of Little San Salvador (Bahamas Islands) during the 2010 Pawlik expedition. The samples were frozen immediately after collection and stored at -20 °C until extraction. The sponge (424 g of dry weight after extraction) was homogenized and extracted with MeOH (5×4 L) and then with CHCl_3 (2×4 L). The MeOH extracts were partitioned between H_2O and *n*-BuOH; the BuOH layer was combined with the CHCl_3 extract and concentrated *in vacuo*.

The organic extract (10, 70 g) was chromatographed on a column packed with RP-18 silica gel. A fraction eluted with MeOH/ H_2O (8:2, 150 mg) was subjected to HPLC separation on a preparative RP-18 column [MeOH/ H_2O (6:4), Ascentis[®] C18—25 cm \times 10 mm, 5 μm -SUPELCO[®]], thus affording a fraction (1.4 mg) mainly composed of compound **1**. Final purification was achieved by HPLC on an analytical RP-18 column (Ascentis[®] C18—25 cm \times 4.6 mm, 5 μm -

SUPELCO[®]), using MeOH/H₂O (6:4) as eluent, which gave 1 mg of pure chalinulasterol.

8.4.3 Chalinulasterol (1)

Colorless amorphous solid, $[\alpha]_D^{25} +11.4$ (*c* 0.1, MeOH); HRESIMS (positive ion mode, MeOH) *m/z* 623.1447 ($[M + Na]^+$, calcd. for C₂₄H₃₉ClS₂O₈Na₃⁺ 623.1462); MS isotope pattern: M (100%), M + 1 (27.5%, calcd. 27.8%), M + 2 (45.5%, calcd. 46.4%), M + 3 (11.5%, calcd. 12.2%), M + 4 (4.9%, calcd. 5.4%); ¹H and ¹³C NMR: Table 8.1.

8.4.4 Cell Culture

HepG2 cells were maintained at 37 °C in E-MEM containing 10% fetal bovine serum (FBS), 1% L-glutamine and 1% penicillin/streptomycin. HepG2 cells were stimulated 18 h with 10 μM rifaximin, **1** and compound **2** and relative mRNA levels of CYP3A4 were analyzed by Real-Time PCR.

8.4.5 Transactivation Experiments

HepG2 cells were transfected using Fugene HD transfection reagent (Roche). The plasmids used for luciferase assay^[16] were pSG5-PXR, pSG5-RXR, pCMV-βgalactosidase and the reporter vector p(CYP3A4)-TK-Luc. 48 h post-transfection cells were stimulated 18 h with 10 μM rifaximin, solomonsterol A, compound **1** or with the combination of 10 μM rifaximin plus 50 μM compound **1**. Cells were lysed in 100 μL

diluted reporter lysis buffer (Promega). 20 μ L of cellular lysates were read using the Luciferase Substrate (Promega). Luminescence was measured using the Glomax 10/10 luminometer (Promega). Luciferase activities were normalized for transfection efficiencies by dividing the relative light units by β -galactosidase activity expressed from cotransfected pCMV- β gal.

8.4.6. Real-Time PCR

Total RNA was extracted using the TRIzol reagent (Invitrogen), purified of the genomic DNA by DNAase I treatment (Invitrogen) and random reverse-transcribed with Superscript II (Invitrogen). 50 ng template was amplified using the following reagents: 0.2 μ M of each primer and 12.5 μ L of 2 \times SYBR Green qPCR master mix (Invitrogen). All reactions were performed in triplicate and the thermal cycling conditions were: 2 min at 95 $^{\circ}$ C, followed by 40 cycles of 95 $^{\circ}$ C for 20 s, 55 $^{\circ}$ C for 20 s and 72 $^{\circ}$ C for 30 s in iCycler iQ instrument (Bio-Rad). Primers used for qRT-PCR are reported in the following scheme:

Template	Forward primer	Reverse primer
hGAPDH	5'-GAAGGTGAAGGTCGGAGT-3'	5'-CATGGGTGGAATCATATTGGAA-3'
hCYP3A4	5'-CAAGACCCCTTTGTGGAAAA-3'	5'-CGAGGCGACTTTCTTTCATC-3'

8.4.7. Statistical Analysis

All values are expressed as means \pm standard error (SE) of n observations/group. Comparisons of 2 groups were made with a one-

way ANOVA with post hoc Tukey's test. Differences were considered statistically significant at values of $P < 0.05$.

8.4.8 Mass spectra

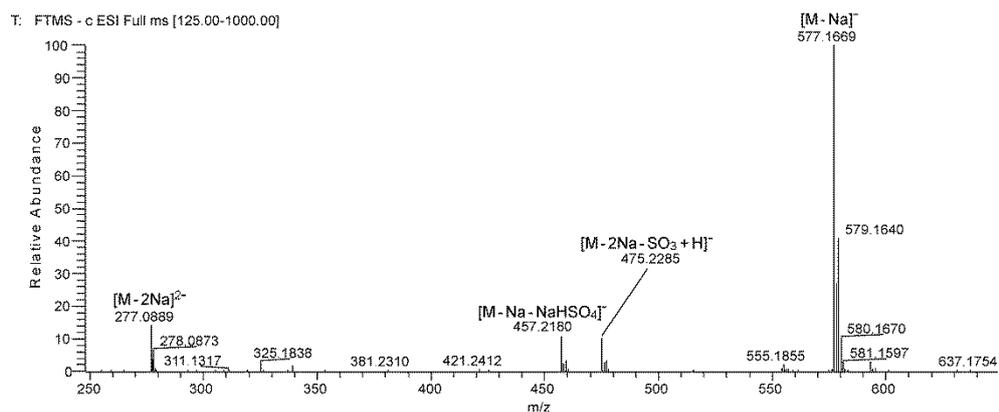


Figure 8.12 Negative-ion ESI mass spectrum of chalinulasterol (1).

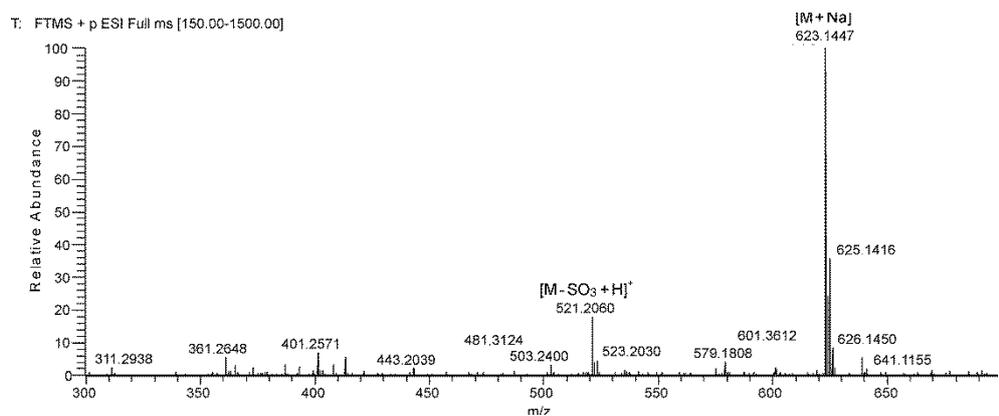


Figure 8.13 Positive-ion ESI mass spectrum of chalinulasterol (1).

8.4.9 NMR data

Table 8.1 ^1H (700 MHz) and ^{13}C (175 MHz) NMR data of chalinulasterol in CD_3OD .

Pos.		δ_{H} [mult. ^a , J (Hz)]	δ_{C} [mult.]	Pos.		δ_{H} [mult. ^a , J (Hz)]	δ_{C} [mult.]
1	α/ax	1.38(dd, 14.6, 3.5)	39.1 (CH_2)	13	-		43.8 (C)
	β/eq	2.12(dd, 14.6, 2.1)		14		1.05 (m)	57.8 (CH)
2		4.72(q, 2.7)	76.5 (CH)	15	α	1.59 (m)	25.2 (CH_2)
3		4.69(q, 2.7)	76.3 (CH)		β	1.08 (m)	
4	α/eq	1.66(dt, 14.6, 2.7)	30.4 (CH_2)	16	α	1.86(ddd, 14.6, 9.4, 3.9)	29.3 (CH_2)
	β/ax	1.79(ddd, 14.6, 12.6, 2.7)			β	1.28 ^a	
5		1.62(tt, 12.6, 2.7)	40.3 (CH)	17		1.12(q, 9.7)	57.5 (CH)
6	α/eq	1.25(br. d, 14.5)	29.2 (CH_2)	18		0.69 (s)	12.8 (CH_3)
	β/ax	1.29(qd, 12.7, 3.5)		19		1.01 (s)	14.2 (CH_3)
7	α/ax	0.95(qd, 12.6, 4.6)	33.3 (CH_2)	20		1.44 (m)	36.6 (CH)
	β/eq	1.68(dq, 13.0, 3.3)		21		0.95 (d, 6.5)	19.2 (CH_3)
8		1.41 (qd, 11.2, 3.5)	36.5 (CH)	22	a	1.56(ddd, 13.4, 10.6, 5.6, 2.9)	34.4 (CH_2)
9		0.72(ddd, 13.2, 10.5, 3.8)	56.7 (CH)		b	1.15(ddd, 13.4, 10.6, 8.8, 4.3)	
10		-	36.4 (C)	23	a	1.81 (ddtd, 13.9, 11.2, 7.1, 4.5)	30.6 (CH_2)
11	α/eq	1.53 (dq, 14.1, 3.5)	22.1 (CH_2)		b	1.65 (ddq, 13.9, 10.9, 5.7)	
	β/ax	1.33 (qd, 13.2, 3.7)		24	a	3.53 (dt, 10.7, 6.5)	46.4 (CH_2)
12	α/ax	1.14 (td, 12.6, 3.8)	41.4 (CH_2)		b	3.51 (ddd, 10.7, 7.1, 6.4)	
	β/eq	1.99 (dt, 12.2, 3.5)					

^a Multiplicity and coupling constants of overlapping signals were determined using sections of the zTOCSY spectrum; dt = doublet of triplets, td = triplet of doublets, dq = doublet of quartets, *etc.*

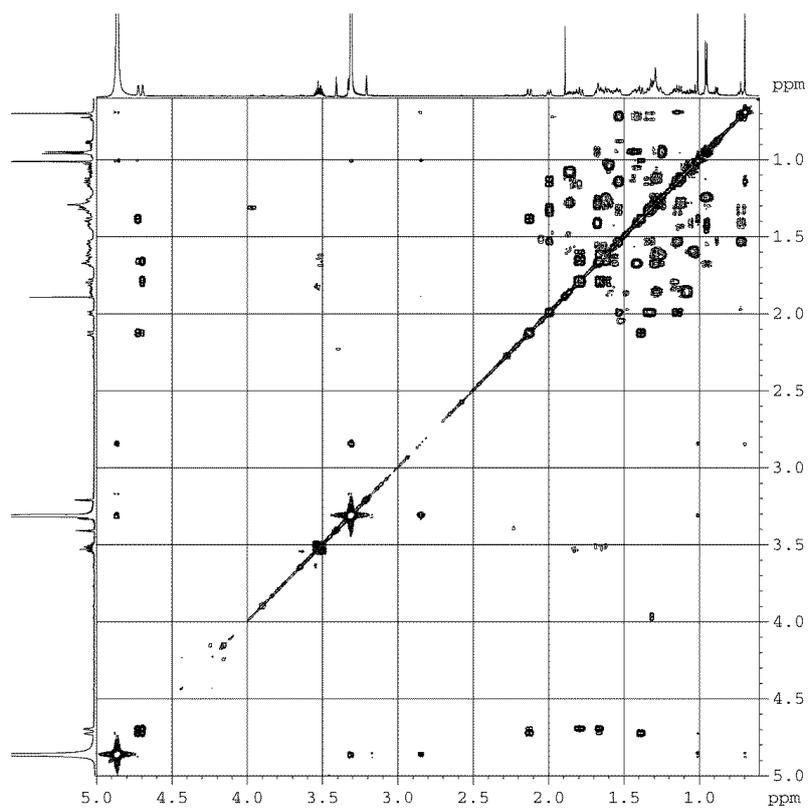


Figure 8.14 COSY spectrum of chalinulasterol (**1**) (CD_3OD , 700 MHz).

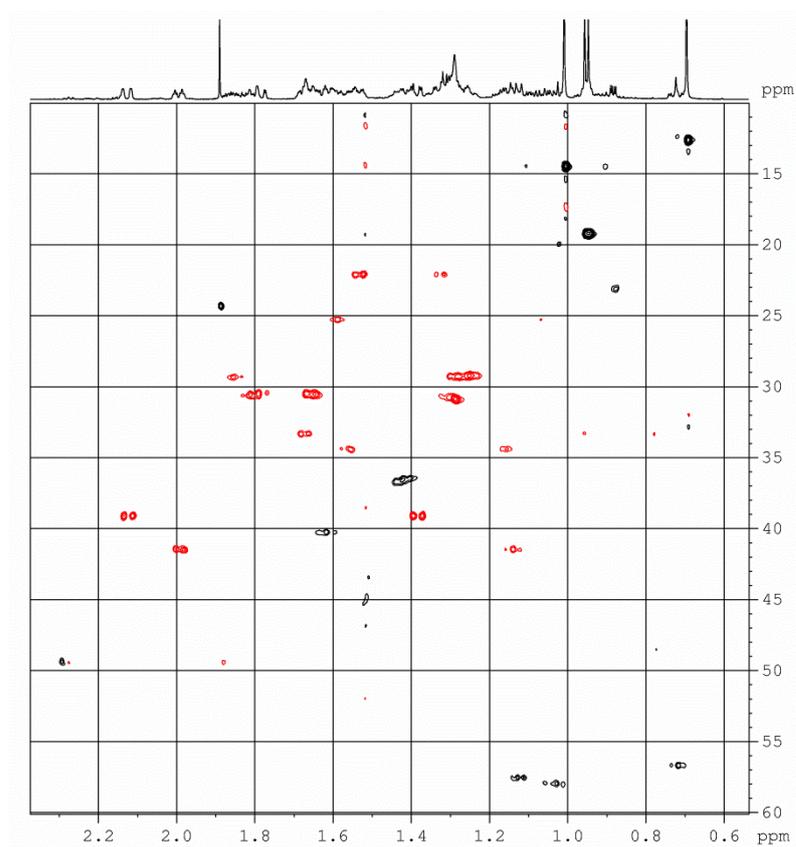


Figure 8.15 HSQC spectrum of chalinulasterol (**1**) – high-field region.

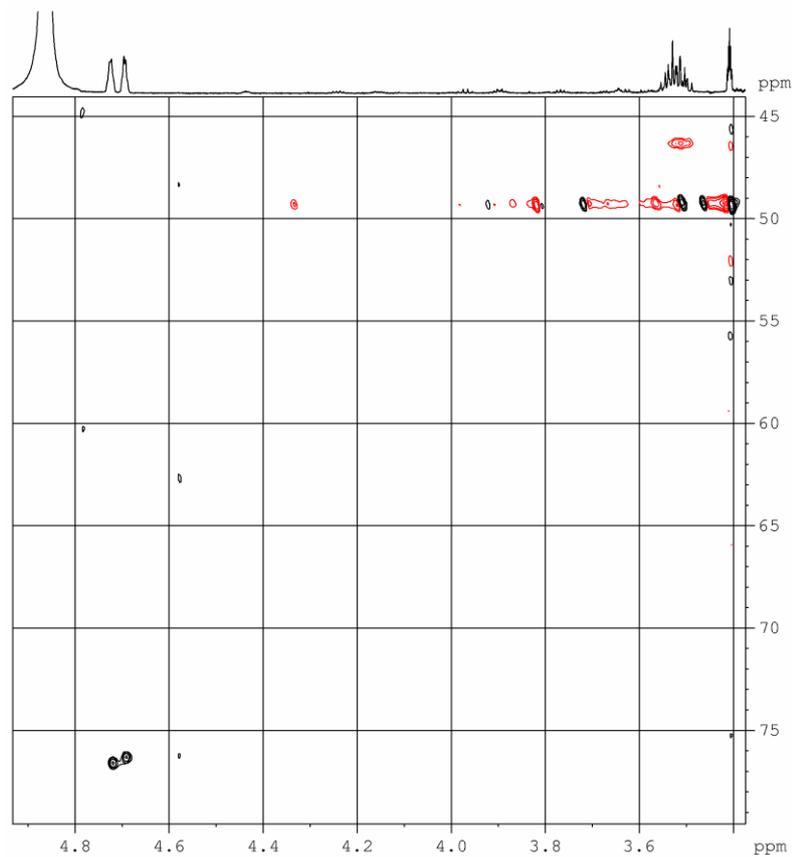


Figure 8.16 HSQC spectrum of chalinulasterol (1) – low-field region.

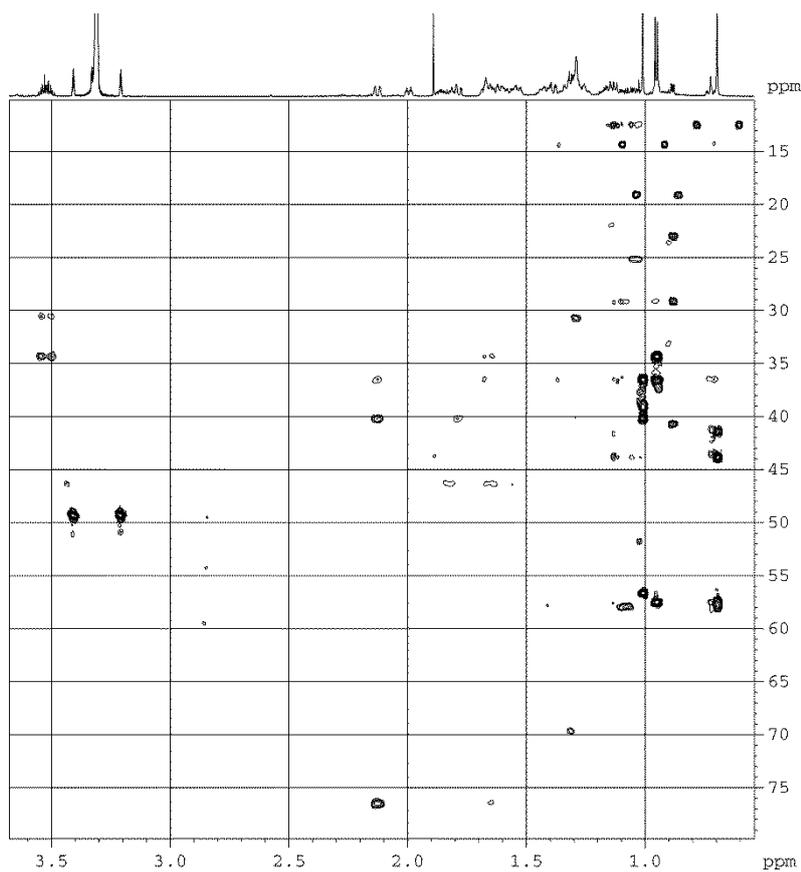


Figure 8.17 HMBC spectrum of chalinulasterol (1).

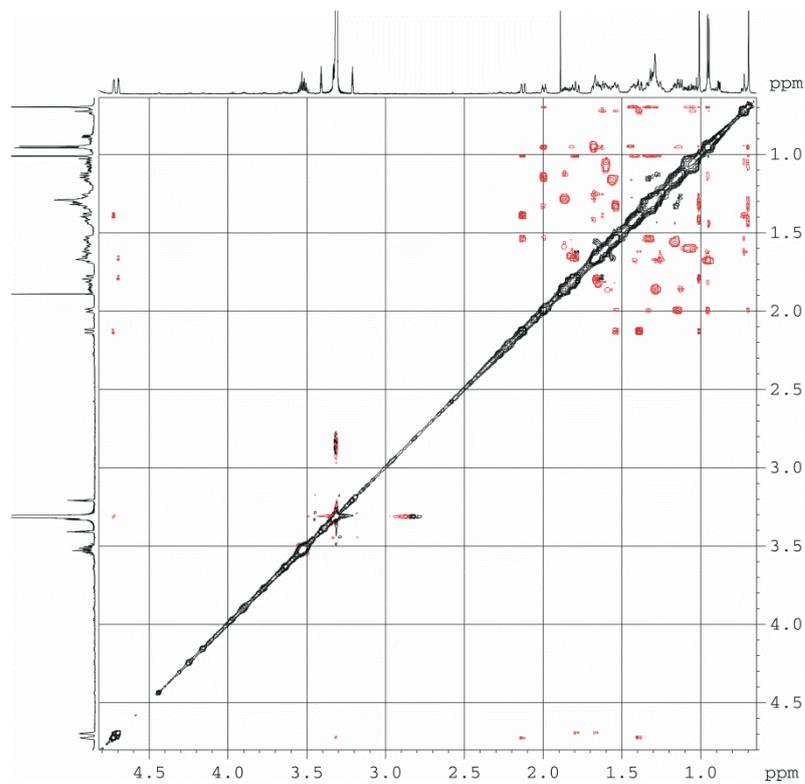


Figure 8.18 ROESY spectrum of chalinulasterol (**1**).

References

1. Sperry S., Crews P. (1997). Haliclostanoone sulfate and halistanol sulfate from an Indo-Pacific *Haliclona* sponge. *J. Nat. Prod.* 60, 29–32.
2. Aiello A., Menna M., Fattorusso E. (1999). Steroids from sponges: Recent reports. *Steroids.* 64, 687–714.
3. Festa C., De Marino S., D’Auria M. V., Bifulco G., Renga B., Fiorucci S., Petek S., Zampella A. (2011). Solomonsterols A and B from *Theonella swinhoei*. The first example of C-24 and C-23 sulfated sterols from a marine source endowed with a PXR agonistic activity. *J. Med. Chem.* 54, 401–405.
4. Sepe V., Ummarino R., D’Auria M. V., Mencarelli A., D’Amore C., Renga B., Zampella A., Fiorucci S. (2011). Total synthesis and pharmacological characterization of solomonsterol A, a potent marine pregnane-X-receptor agonist endowed with anti-inflammatory activity. *J. Med. Chem.* 54, 4590–4599.
5. Rock K. L., Latz E., Ontiveros F., Kono H. (2010). The sterile inflammatory response. *Annu. Rev. Immunol.* 28, 321–342.
6. Matic M., Mahnsa A., Tsoli M., Corradin A., Polly P., Robertson G. R. (2007). Pregnane X Receptor: Promiscuous regulator of detoxification pathways. *Int. J. Biochem. Cell Biol.* 39, 478–483.
7. Chen Y., Tang, Y., Guo, C., Wang, J., Boral, D., Nie, D. (2012). Nuclear receptors in the multidrug resistance through the regulation of drug-metabolizing enzymes and drug transporters. *Biochem. Pharmacol.* 83, 1112–1126.
8. Mencarelli A., Migliorati M., Barbanti M., Cipriani S., Palladino G., Distrutti E., Renga B., Fiorucci S. (2010). Pregnane-X-receptor mediates the anti-inflammatory activities of rifaximin on detoxification pathways in intestinal epithelial cells. *Biochem. Pharmacol.* 80, 1700–1707.
9. D’Auria M. V., Minale L., Riccio R. (1993). Polyoxygenated sterols of marine origin. *Chem. Rev.* 93, 1839–1895.

10. Carballera N., Thomeon J. E., Djerassi C. (1986) *J. Org. Chem.* 51, 2761 and referencee cited there.
11. De Marino S., Ummarino R., D'Auria M. V., Chini M. G. , Bifulco G., Renga B., D'Amore C., Fiorucci S., Debitus C., Zampella A. (2011). Theonellasterols and Conicasterols from *Theonella swinhoei*. Novel Marine Natural Ligands for Human Nuclear Receptors *Journal of Medicinal Chemistry*. 54 (8), 3065-3075.
12. Sepe *et al.*(2012) Preliminary Structure-Activity Relationship on Theonellasterol, a New Chemotype of FXR Antagonist, from the Marine Sponge *Theonella swinhoei*. *Mar. Drugs*. 10(11): 2448–2466.
13. Teruya T., Nakagawa S., Koyama T., Arimoto H., Kita M., Uemura D. (2004). Nakiterpiosin and nakiterpiosinone, novel cytotoxic C-nor-D-homosteroids from the Okinawan sponge *Terpios hoshinota*. *Tetrahedron*. 60, 6989–6993.
14. Fattorusso E. *et al.* (2004). Polychlorinated androstanes from the burrowing sponge *Cliona nigricans*. *Org. Lett.* 6, 1633–1635.
15. Guzii A. G., Makarieva T. N., Denisenko V. A., Dmitrenok P. S., Burtseva Y. V., Krasokhin V. B., Stonik V. A. (2008). Topsentiasterol sulfates with novel iodinated and chlorinated side chains from the marine sponge *Topsentia* sp. *Tetrahedron Lett.* 49, 7191–7193.
16. Smale St. (2010) Luciferase assay. *Cold Spring Harb Protoc.*

Chapter 9

Analysis of the sponge *Plakortis cf. lita*

Sponges of the genus *Plakortis* are known to be a rich source of secondary metabolites. Chemical analysis of the organic extract of *Plakortis simplex* led indeed to the isolation of novel glycolipids, such as plakosides A (**b**) and B (**c**),^[1] simplexides (**d**),^[2] crasseride^[3] (**e**) and discoside (**f**) (isolated for the first time from the Caribbean sponge *Discodermia dissoluta*, but also found in specimens of *Plakortis simplex*),^[4] alkaloids, and the polyketide plakortin discussed in Section 2.3.

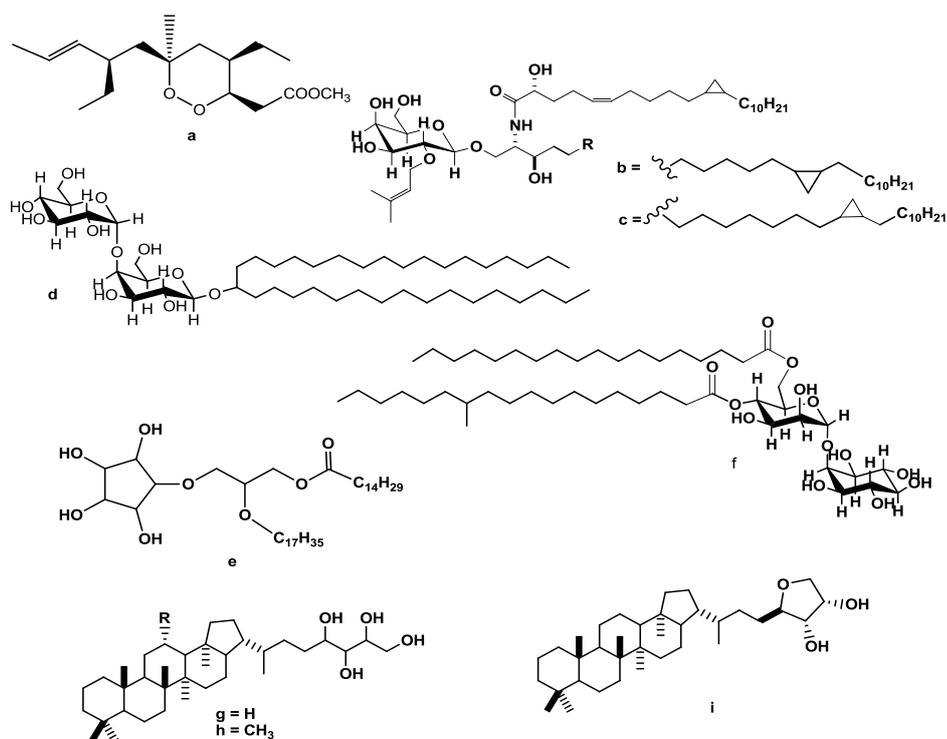


Figure 9.1 Chemical structures of some metabolites from *P. simplex*: plakortin (**a**), plakoside A (**b**) and plakoside B (**c**), simplexide (**d**), crasseride (**e**), discoside (**f**), bacteriohopanetetrol (**g**), 12-methyl-bacteriohopanetetrol (**h**) and 32,35-anhydro-bacteriohopanetetrol (**i**).

In addition, large amounts (about 50% of the total sterol fraction in weight) of three bacteriohopanoids, namely bacteriohopanetetrol (**g**), 12-methylbacteriohopanetetrol (**h**)^[5] and 32, 35-anhydro bacteriohopanetetrol (**i**)^[6] were detected.

This Chapter describes the chemical analysis of an Indonesian specimen of another genus of *Plakortis*, i.e., *Plakortis* cf. *lita*. Most of the peculiar metabolites previously reported from *P. simplex* were also present in *Plakortis* cf. *lita*.

Interestingly, this study led to the isolation of a further bacteriohopanoid, plakohopanoid, which represents a novel type of hopanoid.

9.1 Bacteriohopanoids

Bacteriohopanoids are unique mixed-biogenesis compounds, occurring in bacteria, in which a triterpenoid with the hopane skeleton is linked through the isopropyl group to a sugar-derived polyfunctionalized C₅ side chain.

9.1.1 Biosynthesis of hopanoids

Hopane is a pentacyclic triterpenoid which, like sterols, is synthesized from isopentenyl units. All hopanoids originate from isopentenyl units produced by the non-mevalonate pathway described above (Section 8.1 and Figure 8.2).

Six isoprene units are joined to form squalene, the latest common precursor in hopanoid and sterol synthesis. In a highly complex cyclization reaction that shares considerable similarities with that of oxidosqualene to sterols

(see Figure 8.3), the hopane skeleton is formed from squalene by the squalene-hopene cyclase (Figure 9.2).

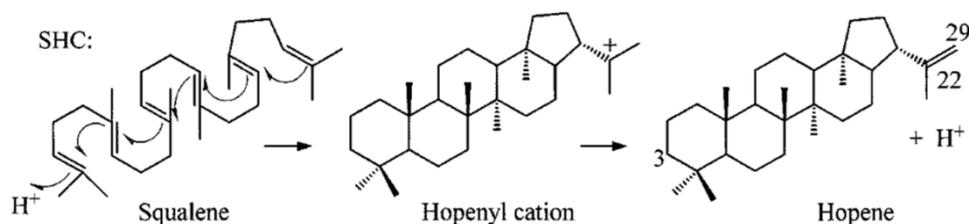


Figure 9.2 Biosynthesis of the hopane skeleton by SHC (squalene-hopene cyclase).

Although the mechanism by which hopanes are converted into bacteriohopanoids in bacteria is not known in detail, it has been demonstrated that the C₅ side chain of most bacteriohopanoids derives from ribose, whose stereochemistry is kept in bacteriohopanetetrol.^[7] A C-adenosyl derivative, which has been isolated from *Rhodopseudomonas acidophila*,^[8] and a C-ribosyl derivative, never isolated from a natural source, have been proposed to be biosynthetic intermediates (Figure 9.3).^[9]

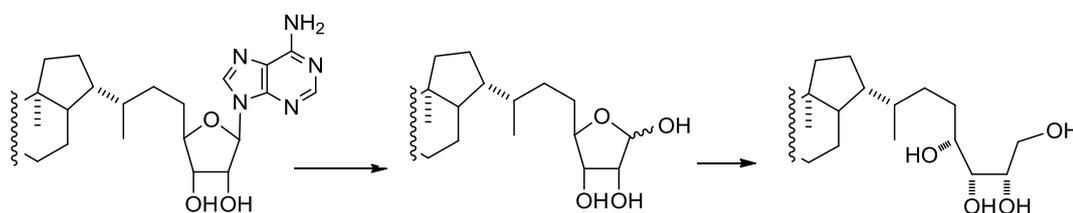


Figure 9.3 Proposed biosynthetic pathway to bacteriohopanoids.

9.1.2 Function of bacteriohopanoids ^[10]

In their molecular dimensions, their amphiphilic character, and their rigid ring structure bacteriohopanoids resemble to sterols (for example,

cholesterol). As surrogates for sterols, bacterial hopanoids regulate membrane fluidity and induce order in the phospholipid matrix of membranes. In many bacteria, hopanoids may play important roles in the adjustment of cell membrane permeability and adaptation to extreme environmental conditions (for example low pH, temperature changes, ethanol accumulation).

However, hopanoid function may not be restricted to membrane-related activities in the light of recent perspectives. Interesting observations indicate an extracellular function for hopanoids in *Frankia* bacteria, which form symbiotic nodules on several shrubby plants. These bacteria express the oxygen-sensitive nitrogenase in vesicles composed by extra-cellular layers of lipids largely consisting of hopanoids, suggesting an oxygen-protection mechanism of the nitrogenase for the hopanoid layers.

In addition, the many structural variants of hopanoids suggest that they may have other interesting but as yet unknown functions.

9.1.3 Biohopanoids and geohopanoids

The pentacyclic ring moiety of hopanoids is very stable and not readily degraded, but after the death of the bacterial cell the functionalized side chain can undergo various abiotic degradative processes, including reduction (up to alkanes), loss of carbon atoms, and epimerization. Degraded hopanoids are ubiquitous in sediments, rocks and crude oil^[11] and, because the characteristic pentacyclic moiety is preserved, may be easily recognized and find valuable use as biomarkers in geochemical studies. Consequently, hopanoids are generally classified into two classes:

biohopanoids, produced in bacteria by biosynthetic processes, and geohopanoids, derived from abiotic degradation of biohopanoids.

There are a few hopanoids that escape this classification. Among them there is 32,35-anhydroBHT (**2**),^[6] which has been first isolated from the Caribbean sponge *Plakortis simplex*, and later shown to be present in the cells of the bacterial symbionts of this sponge,^[12] and should therefore be considered a biohopanoid. Very recently, however, Schaeffer *et al.* demonstrated that anhydroBHT can be formed from BHT and other hopanoids by geochemical processes,^[13] so that the anhydroBHT which is present (sometimes predominantly) in sediments may be of abiotic origin.

Plakohopanoid, isolated from *P. lita*, represents a further example of compounds blurring the boundary between bio- and geohopanoids. It is an ester composed of a C₃₂ hopanoic acid and mannosyl-*myo*-inositol, a building block of phosphatidylinositol mannosides, which are characteristic of Mycobacteria and related species. C₃₂ hopanoic acids are usually classified as geohopanoids, but the presence of plakohopanoid in a marine living organism shows that there is a biosynthetic pathway to C₃₂ hopanoic acids, which therefore should not be considered anymore as sure geohopanoids.

9.2 Isolation and structural determination of plakohopanoid

The MeOH extract of an Indonesian specimen of *Plakortis cf. lita* was partitioned between BuOH and H₂O, and the organic layer was purified by

reversed-phase chromatography followed by normal-phase chromatography, yielding a crude fraction mainly composed of glycolipids, but also containing hopanoids, which was acetylated with Ac_2O in pyridine.

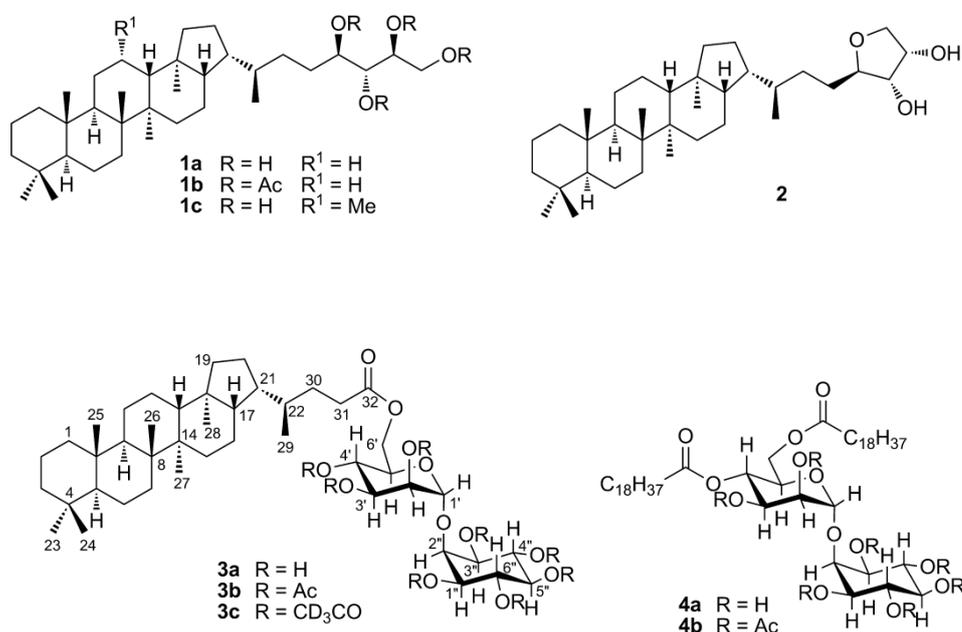


Figure 9.4 Bacteriohopanetetrol (**1a**); bacteriohopanetetrol tetraacetate (**1b**); 12-methylbacteriohopanetetrol (**1c**). 32,35-anhydro-bacteriohopanetetrol (**2**). Plakohopanoid (**3a**); Plakohopanoid peracetate (**3b**); Plakohopanoid peracetylated with $(\text{CD}_3\text{CO})_2\text{O}$ (**3c**). Discoside (**4a**); Discoside peracetate (**4b**).

The composition of the peracetylated glycolipid fraction was analyzed using normal-phase HPLC, and turned out to be very similar to the glycolipid fraction of *P. simplex* from the Caribbean Sea, which has been deeply studied by my research group. ¹H NMR analysis of the collected HPLC fractions showed the presence of the peculiar metabolites previously reported from *P. simplex*, i.e. plakoside A and B,^[1] simplexides,^[2] 12-methylbacteriohopanetetrol,^[5] and crasserides.^[3]

In addition, the HPLC separation provided the new hopanoid derivative, plakohopanoid **3a** as its peracetyl derivative **3b**. Compound **3b** could not be deacetylated to give the natural hopanoid **3a** because **3b** is an ester itself,

nor it was possible to isolate pure **3a** from the crude glycolipid fraction because of the well-known low solubility in most organic solvents of biohopanoids.^[14] However, the presence of the natural compound **3a** in the glycolipid fraction could be demonstrated by an LC-HRESIMS experiment.

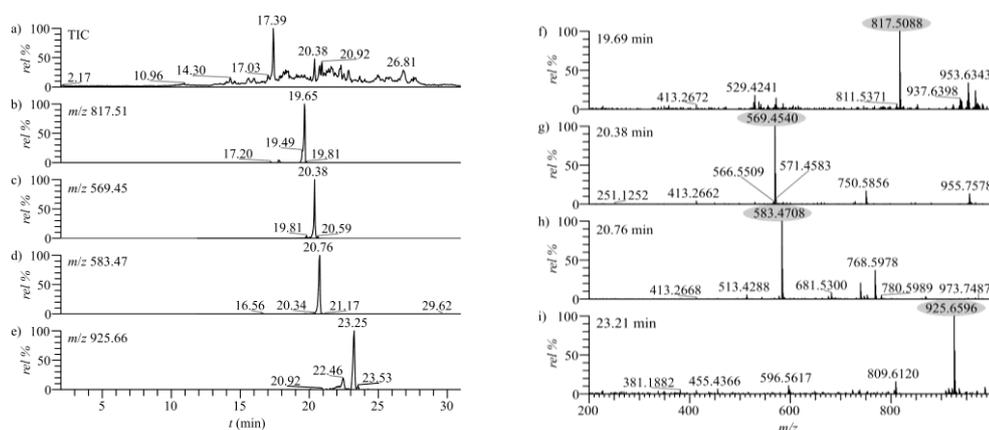


Figure 9.5 HPLC-ESIMS of the crude glycolipid fraction of *P. cf. lita* and extracted-ion chromatograms of the $[M+Na]^+$ ions of several metabolites of the sponge. From top to bottom: (a) total ion chromatogram; (b) extracted ion chromatogram of mass m/z 817.51 for plakohopanoid (**3a**); (c) extracted ion chromatogram of mass m/z 569.45 for bacteriohopanetetrol (**1a**); (d) extracted ion chromatogram of mass m/z 583.47 for 12-methylbacteriohopanetetrol (**1c**); (e) extracted-ion chromatogram of mass m/z 925.66 for the major homologue of discoside (**4a**); (f-i) mass spectra at retention times 19.69, 20.38, 20.76, and 23.21 min, showing peaks for **3a** (calcd. 817.5072), **1a** (calcd. 569.4540), **1c** (calcd. 583.4697), and **4a** (calcd. 925.6587), respectively.

The HRESIMS spectrum of compound **3b** showed a pseudomolecular $[M+Na]^+$ ion peak at m/z 1153.5885, indicating the molecular formula $C_{60}H_{90}O_{20}$. Analysis of the 1H NMR spectrum of **3b** ($CDCl_3$) showed the molecule to be composed of a terpenoid moiety (six methyl singlets at $\delta = 0.94, 0.93, 0.84, 0.80, 0.78,$ and 0.69 , and a methyl doublet at $\delta = 0.92$) and of an acetylated carbohydrate moiety (13 carbinol protons between $\delta = 5.55$ and 4.05 and eight acetyl singlets between $\delta = 2.14$ and 2.00).

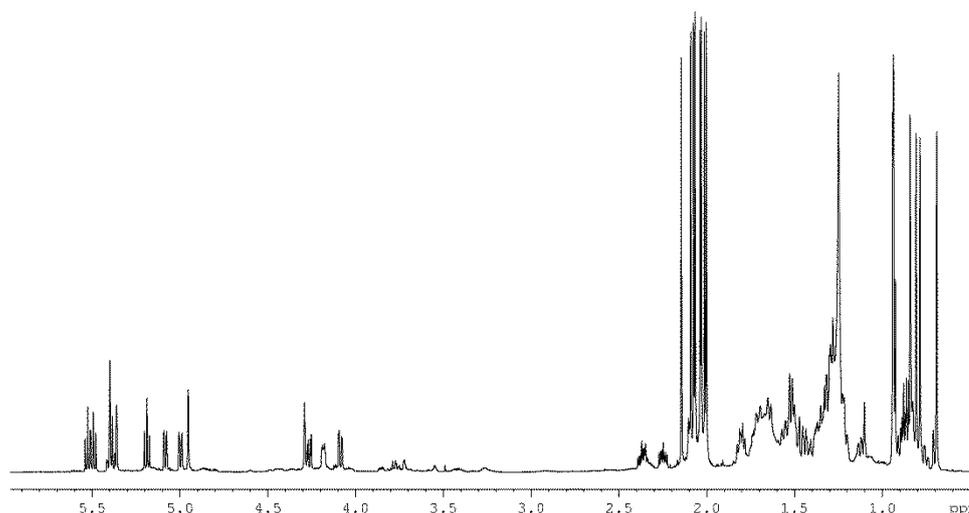


Figure 9.6 ^1H -NMR spectrum of plakohopanoid peracetate (CDCl_3).

Examination of the correlation peaks of the methyl protons in the HMBC spectrum allowed us to assign most carbon atoms in the terpene skeleton of **3b** and to sketch the partial structure depicted in Figure 9.7.

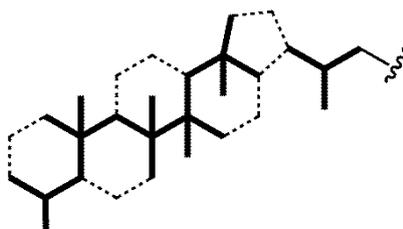


Figure 9.7 Partial structure of plakohopanoid peracetate **3b** as determined by long-range proton-carbon couplings of methyl protons (bold lines).

This fit well in a hopane skeleton, which according to the ^{13}C chemical shifts should be devoid of any functional group. Indeed, a comparison of the ^{13}C spectra of **3b** and bacteriohopanetetrol tetraacetate (**1b**) showed that for all the signals of C-1/C-30 in the ^{13}C spectrum of bacteriohopanetetrol tetraacetate there was a corresponding signal within 0.2 ppm in the spectrum of **3b**. This implied that the terpenoid skeleton of **1b** and **3b** have the same structure and relative configuration, i.e. that of an extended hopane.

Configuration at C-22, which does not affect significantly ^{13}C chemical shifts, was determined on the basis of the ^1H chemical shift of CH_3 -29 ($\delta = 0.92$) which is diagnostic for the $22R$ configuration.^[15]

The chemical shifts of the protons in the hopane part of the molecule were then identified through an HSQC experiment (see Section NMR data - Table 9.1); all the correlation peaks that were present in the COSY, TOCSY, and HMBC spectra were consistent with a hopane skeleton and with the assignment reported in Table 9.1. The structure of the C_2 side chain that completed the structure of the hopanoid skeleton was defined thanks to the correlations of the methylene protons at C-31 ($\delta = 2.37$ and 2.25) with the protons H_2 -30 in the COSY spectrum, and with the ester carbonyl carbon atom at $\delta = 173.9$ in the HMBC spectrum.

For structure elucidation of the carbohydrate moiety of **3b**, a second set of one- and two-dimensional NMR experiments was recorded using C_6D_6 as solvent, because using this solvent a better signal dispersion was observed in the mid-field region of the ^1H NMR spectrum. The ^{13}C NMR spectrum showed the presence of only one anomeric carbon ($\delta = 100.0$), which was correlated with the ^1H doublet at $\delta = 4.98$ ($J = 2.0$ Hz) in the HSQC spectrum. Using this resonance as the starting point, analysis of COSY correlation peaks allowed us to identify a sequence of four methine protons and a couple of methylene protons at $\delta = 5.65$ (H-2'), $\delta = 5.77$ (H-3'), $\delta = 5.85$ (H-4'), $\delta = 4.52$ (H-5'), and $\delta = 4.57/4.44$ (H_2 -6'), accounting for a hexose. This was in the pyranose form because of the relatively shielded chemical shift of H-5' and the coupling between H-5' and C-1' in the HMBC spectrum. The large couplings (10.0 Hz) of H-4' with H-3' and H-5' were

indicative of the axial orientation of these three protons, while the small (3.3 Hz) coupling of H-3' with H-2' showed that the latter proton is equatorial. Therefore, the hexopyranose unit was a mannopyranose unit. The α anomeric configuration of the mannose unit was deduced by the 173 Hz coupling constant between H-1' and C-1' (we could measure this coupling from the residual H-1'/C-1' one-bond correlation observed in the HMBC spectrum), because it is known from the literature^[16] that $^1J_{\text{CH}}$ coupling constants are around 158-162 Hz for axial anomeric protons, and around 169-171 Hz for equatorial anomeric protons. The remaining six oxymethine protons observed in the mid-field region of the proton NMR spectrum were shown to be cyclically arranged from COSY data, and therefore were indicative of an inositol unit.

The *myo* configuration of the inositol was deduced by coupling constant analysis (Table 9.1), which showed all the inositol protons to be axial except for H-2" ($\delta = 4.20$). The shielded chemical shift of H-2" and the HMBC correlation between H-2" and C-1' showed that the *myo*-inositol is glycosylated at O-2' by the α -mannopyranose unit.

Finally, the connection between the terpene and carbohydrate moieties of plakohopanoid was identified as an ester bond between the carbonyl C-32 and the mannose 6'-oxymethylene group on the basis of the correlation peaks between both protons at C-6' and C-32 in the HMBC spectrum.

The carbohydrate moiety of plakohopanoid peracetate (**3b**) is very similar to that of discoside peracetate (**4b**), previously isolated by my group from *Discodermia dissoluta*,^[4] but also present in Caribbean specimens of *Plakortis* (unpublished data) as well as in the Indonesian specimen *P. lita*.

In **4a**, there are two fatty acid residues ester-linked to the mannosylinositol unit at O-4' and O-6'; they are replaced by a single C₃₂ hopanoyl linked at O-6' in plakohopanoid. A comparison between the proton NMR spectra of **3b** and **4b** revealed that all the carbohydrate signals show almost identical chemical shifts and coupling constants in the two compounds, further confirming structure **3b**.

The absolute configuration of plakohopanoid was established by micro-scale chemical degradation (Figure 9.8). Acidic methanolysis of **3b** with 1 M HCl in 92% MeOH produced a mixture of hopanoic acid methyl ester (**5**) and methyl mannopyranoside (**6**). After partitioning between CHCl₃ and H₂O these compounds were found, respectively, in the CHCl₃ and H₂O layers.

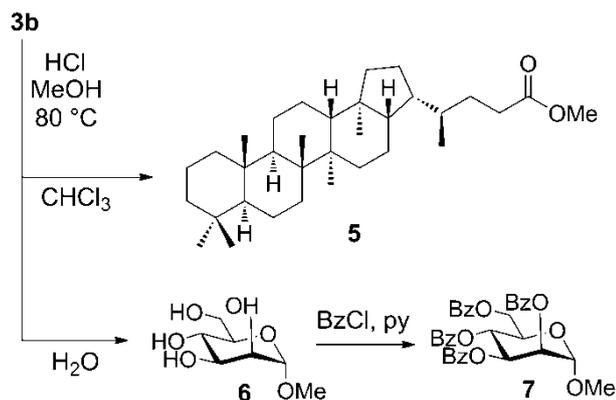


Figure 9.8 Degradation procedure of plakohopanoid peracetate (**3b**).

The hopanoic ester **5** was purified by HPLC (SiO₂, *n*-hexane/EtOAc 95:5), and its ¹H NMR spectrum matched that reported in the literature.^[17] In addition, the positive optical rotation of **5** ([α]_D = +83) confirmed the absolute configuration of the hopane skeleton. Methyl mannopyranoside **6** was benzoylated and purified by HPLC as described in the Experimental Section, yielding the tetrabenzoate **7**. The ¹H NMR and CD spectra of

compound **7** were identical to those reported for synthetic methyl α -d-mannopyranoside perbenzoate.

This completed elucidation of structure and stereochemistry of plakohopanoid peracetate.

9.3 Hopanoids from *Plakortis* specimens

Plakohopanoid represents a novel type of natural hopanoid derivative, composed of a C₃₂ hopanoid acyl ester-linked to a carbohydrate part. Even though **3a** was isolated from a sponge, its moieties, hopanoic acid and mannosyl-*myo*-inositol, are clearly of bacterial origin. Therefore, its bacterial biosynthesis is very likely, especially if one considers that sponge-associated bacteria are extremely abundant in *Plakortis* species (they account for over 50% of the sponge weight) and that there is evidence^[12] that several of the many metabolites isolated from this sponge are actually produced by the bacterial symbionts.

C₃₂ hopanoic acids have been found only in sediments and other geological formations, and are considered diagenetic products which are formed through abiotic degradation of the biohopanoids present in bacteria. Diagenetic transformation from biohopanoids to geohopanoids can take place quickly after the death of bacteria, and is well documented. However, the presence of plakohopanoid in a marine living organism shows that there is biosynthetic pathway to C₃₂ hopanoic acids, which therefore should not be considered anymore as sure geohopanoids. The occurrence of 12-methylbacteriohopanetetrol (**1c**) in the Indonesian specimen of *Plakortis*

also deserves attention. This compound was isolated in 2000 from a Caribbean specimen of *P. simplex*,^[5] and was the first example of a hopanoid methylated at C-12. Since then, no further 12-methylhopanoid has been found, compared to the 19 naturally occurring 2-methylhopanoids and the 25 naturally occurring 3-methylhopanoids described so far. Bacterial communities associated to many species of sponges are highly specific, phylogenetically very different from free-living bacteria in the surrounding water, and consistently conserved in specimens collected in different times and geographical areas,^[18] suggesting vertical transmission within their hosts^[19] and an independent evolutionary history. Therefore, it is well possible that a unique biosynthetic pathway to 12-methylhopanoids has been developed in some symbiotic bacteria of *Plakortis* sponges. This is a further proof of the biosynthetic diversity and biotechnological potential of sponge-associated bacteria.

9.4 Experimental Section

9.4.1 General Methods

Optical rotations were measured at 589 nm on a Jasco P-2000 polarimeter using a 10-cm microcell. High Resolution ESI-MS and ESI-MS/MS spectra were performed on a Thermo LTQ Orbitrap XL mass spectrometer. The spectra were recorded by infusion into the ESI source using MeOH as the solvent. EI-MS spectra were performed using the GC-MS instrument Agilent 6850 II/5973 MSD and a HP-5MS capillary column (Agilent, 5% Phenyl Methyl Siloxane; 30 m, 0.25 mm \varnothing , 0.25 μm). Helium

was used as a carrier gas, injection was in split mode, the program was as follows: hold 150°C for 15 min, heat to 300°C with 5°C/ min, hold 300°C for 10 min. NMR spectra were determined on Varian Unity Inova spectrometers at 500 MHz and 700MHz; chemical shifts were referenced to the residual solvent signals (CDCl₃: $\delta_{\text{H}} = 7.26$, $\delta_{\text{C}} = 77.0$, C₆D₆: $\delta_{\text{H}} = 7.15$, $\delta_{\text{C}} = 128.15$). For an accurate measurement of the coupling constants, the one-dimensional ¹H NMR spectra were transformed at 64K points (digital resolution: 0.09 Hz). Homonuclear ¹H connectivities were determined by a COSY experiment. The single-quantum heteronuclear correlation (HSQC) and multiple-bond heteronuclear correlation (HMBC) experiments were adjusted, respectively, for an average ¹J_{CH} of 142 Hz and a ^{2,3}J_{CH} of 8.3 Hz. High performance liquid chromatography (HPLC) were performed on a Varian Prostar 210 apparatus equipped with a Varian 350 refractive index detector.

9.4.2 Animal Material

A specimen of *Plakortis* cf. *lita* (order Homosclerophorida, family Plakinidae) was collected in January 2008 along the coasts of the Bunaken island in the Bunaken Marine Park of Manado. A voucher sample (no. MAN08-02) has been deposited at the Dipartimento di Farmacia, Università di Napoli Federico II.

9.4.3 Extraction and Isolation

The sponge (380 mL of volume before extraction and 50 g of dry weight after extraction) was cut into pieces and extracted with MeOH (3 × 1.5 L),

MeOH/CHCl₃ (4 × 1.5 L) and CHCl₃ (2 × 1.5 L). The MeOH extracts were partitioned with H₂O and BuOH; the organic layer was added to the CHCl₃ extracts, affording 8.5 g of a dark brown oil, which was chromatographed on a column packed with RP-18 silica-gel. The fraction eluted with CHCl₃ 100% (1.8 g) was subjected to a further chromatography on a SiO₂ column with solvent of increasing polarity. The fraction eluted with AcOEt/MeOH (7:3) was mainly composed of glycolipids. This fraction (0.4 g) was acetylated with Ac₂O in pyridine overnight and subjected to HPLC separation on a SiO₂ column [10 μ, 250 × 10 mm; eluent: *n*-hexane/EtOAc (6:4) flow: 4 mL min⁻¹], affording crasserides (44 mg),^[3] bacteriohopanetetrol **1a** and 12-methylbacteriohopanetetrol **1c** (as a mixture, 112 mg),^[5] plakoside A and B and simplexides (as a mixture, 16 mg),^[1,2] and discoside (5 mg),^[4] all as their respective peracetylated derivatives, which were identified from their ¹H NMR spectra. In addition, the HPLC separation gave pure plakohopanoid as its peracetyl derivative **3b** (2.1 mg).

9.4.4 Plakohopanoid peracetate

Colorless amorphous solid, $[\alpha]_D^{25} = +57$ ($c = 0.1$ in CHCl₃); HRESIMS (positive ion mode, MeOH), $[M+Na]^+$ at m/z 1153.5885, molecular formula C₆₀H₉₀O₂₀Na⁺ (calcd. 1153.5918), ¹H and ¹³C NMR (700 MHz, CDCl₃ and 500MHz, C₆D₆): see Table 9.1.

9.4.5 HPLC-ESIMS Analysis

For LC-MS, the Orbitrap MS spectrometer was coupled to an Agilent model 1100 LC, which included a solvent reservoir, in-line degasser, binary

pump, and refrigerated autosampler. A 2.6 μm Kinetex C18 column (50×2.1 mm), maintained at room temperature, was used. It was eluted at 0.2 mL min^{-1} with H_2O and MeOH, using as a gradient elution 60–100% MeOH over 15 min and hold 15 min. A 2 mg mL^{-1} MeOH solution of the crude glycolipid fraction was prepared, and $5 \mu\text{L}$ of this solution were injected. Results are shown in Figure 9.5.

9.4.6 Degradation Analysis of plakohopanoid peracetate

Compound **3b** (0.5 mg) was dissolved in 1N HCl in 92% MeOH ($500 \mu\text{L}$), and the solution obtained was kept at $80 \text{ }^\circ\text{C}$ for about 12 h. The reaction mixture was dried under nitrogen and then partitioned between H_2O and CHCl_3 . The H_2O layer contained methyl glycoside **6**, and was benzoylated with benzoyl chloride ($50 \mu\text{L}$) in pyridine ($500 \mu\text{L}$) at 25°C for 16h. The reaction was quenched with MeOH, and after 30 min the mixture was dried under nitrogen. Methyl benzoate was removed by keeping the residue under vacuum with an oil pump for 24 h, and the residue was purified by HPLC [SiO_2 column, 5μ , 250×4.6 mm; eluent: *n*-hexane/*i*-PrOH (99:1); flow: 1 mL min^{-1}]. The chromatogram contained a peak ($t_R = 9.5$ min), which was collected and identified as methyl 2,3,4,6-tetra-*O*-benzoyl- α -d-mannopyranoside (**7**) by comparison of its ^1H NMR and CD spectra with those reported.^[4] The CHCl_3 layer from methanolysis was concentrated and purified by HPLC [SiO_2 column, 5μ , 250×4.6 mm; eluent: *n*-hexane/EtOAc (95:5); flow: 1 mL min^{-1}] to give pure compound **5** (0.2 mg).

9.4.7 Methyl (22R)-33,34,35-trinorbacteriohopan-32-oate (5)

$[\alpha]_D^{25} = +83$ ($c = 0.015$ in CHCl_3), ref. 14 reports +58; EIMS: m/z (relative intensity) 484 (M^+ , 4.6), 469 (9.2), 369 (20.0), 263 (100), 231 (6.4), 191 (32.2); the ^1H NMR (CDCl_3) matched that reported.^[17]

9.5 Mass spectra

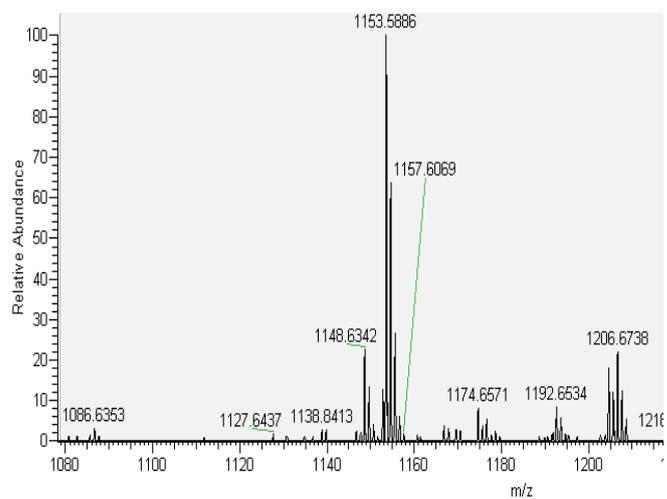


Figure 9.9 High-resolution ESI MS spectrum of plakohopanoid peracetate (**3b**).

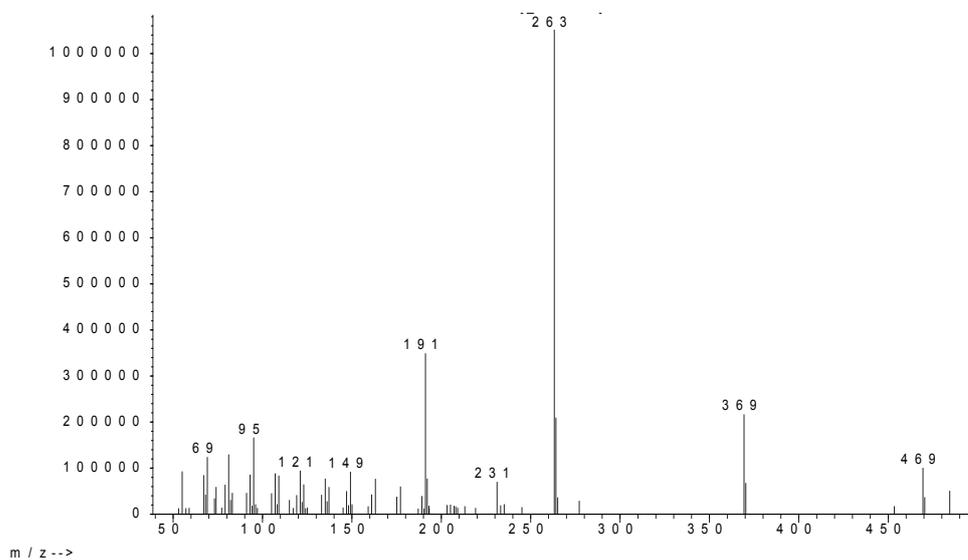


Figure 9.10 EI mass spectrum of hopanoic acid methyl ester (**5**).

9.6 NMR data

Table 9.1 NMR spectroscopic data for plakohopanoid peracetate **3b**.

Position	700 MHz, CDCl ₃		500 MHz, C ₆ D ₆	
	δ_{H} (J in Hz) ^[a]	δ_{C} , type	δ_{H} (J in Hz) ^[a]	δ_{C} , type
1	1.64, 0.76	40.1, CH ₂	1.65, 0.77	40.6, CH ₂
2	1.56, 1.34	18.7, CH ₂	1.60, 1.38	19.1, CH ₂
3	1.34, 1.12	42.1, CH ₂	1.38, 1.16	42.4, CH ₂
4	-	33.4, C	-	33.4, C
5	0.70	56.1, CH	0.75	56.5, CH
6	1.48, 1.31	18.6, CH ₂	1.53, 1.34	19.1, CH ₂
7	1.45, 1.21	33.2, CH ₂	1.49, 1.23	33.7, CH ₂
8	-	41.8, C	-	42.0, C
9	1.24	50.4, CH	1.28	50.8, CH
10	-	37.4, C	-	37.6, C
11	1.52, 1.29	20.9, CH ₂	1.53, 1.27	21.3, CH ₂
12	1.44, 1.39	23.9, CH ₂	1.42, 1.42	24.3, CH ₂
13	1.31	49.2, CH	1.30	49.6, CH
14	-	41.6, C	-	41.9, C
15	1.34, 1.23	33.6, CH ₂	1.31, 1.22	34.0, CH ₂
16	1.71, 1.52	22.8, CH ₂	1.66, 1.50	23.1, CH ₂
17	1.28	54.4, CH	1.18	54.6, CH
18	-	44.3, C	-	44.5, C
19	1.52, 0.90	41.5, CH ₂	1.54, 0.88	41.8, CH ₂
20	1.81, 1.54	27.5, CH ₂	1.81, 1.64	27.8, CH ₂
21	1.73	45.8, CH	1.68	46.3, CH
22	1.51	36.2, CH	1.56	36.6, CH
23	0.84 (s)	33.3, CH ₃	0.912 (s)	33.6, CH ₃
24	0.78 (s)	21.6, CH ₃	0.863 (s)	21.8, CH ₃
25	0.80 (s)	15.9, CH ₃	0.857 (s)	16.2, CH ₃
26	0.93 (s)	16.5, CH ₃	0.976 (s)	16.9, CH ₃
27	0.94 (s)	16.6, CH ₃	0.960 (s)	16.7, CH ₃
28	0.69 (s)	15.8, CH ₃	0.720 (s)	16.0, CH ₃
29	0.92 (d, 7.6)	19.7, CH ₃	0.954 (d, 6.5)	19.9, CH ₃
30	1.81, 1.29	30.5, CH ₂	2.04, 1.43	31.1, CH ₂
31	a 2.37 (ddd, 15.9, 10.6, 5.0)	30.9, CH ₂	2.45 (ddd, 15.9, 9.9, 5.5)	31.3, CH ₂
	b 2.25 (ddd, 15.9, 10.0, 6.0)		2.41 (ddd, 15.9, 9.1, 6.8)	
32	-	173.9, C	-	173.5, C
1'	4.95 (d, 1.7)	99.4, CH	4.98 (d, 2.0)	100.0, CH
2'	5.36 (dd, 3.1, 1.7)	69.6, CH	5.65 (dd, 3.2, 2.0)	70.4, CH
3'	5.39	65.5, CH	5.77 (dd, 10.0, 3.2)	69.3, CH
4'	5.40	68.6, CH	5.85 (t, 10.0)	66.0, CH
5'	4.18 (m)	69.4, CH	4.52 (ddd, 10.0, 3.9, 2.0)	70.5, CH
6'	a 4.26 (dd, 12.5, 4.3)	61.7, CH ₂	4.57 (dd, 12.3, 3.9)	62.0, CH ₂
	b 4.09 (dd, 12.5, 2.2)		4.44 (dd, 12.3, 2.0)	
1''	5.00 (dd, 10.5, 2.5)	69.6, CH	4.88 (dd, 10.6, 2.4)	70.3, CH
2''	4.29 (t, 2.5)	76.4, CH	4.20 (t, 2.4)	77.3, CH
3''	5.08 (dd, 10.5, 2.5)	70.3, CH	5.01 (dd, 10.4, 2.4)	70.7, CH
4''	5.53 (dd, 10.5, 9.9)	69.5, CH	5.902 (dd, 10.4, 9.7)	70.6, CH
5''	5.19 (t, 9.9)	70.6, CH	5.33 (t, 9.7)	71.3, CH
6''	5.49 (dd, 10.5, 9.9)	69.3, CH	5.897 (dd, 10.4, 9.7)	70.0, CH
Ac groups	2.14, 2.09, 2.07, 2.06, 2.04, 2.03, 2.01, 2.00 (8 singlets)	20.9–20.5, CH ₃	1.935, 1.880, 1.672, 1.672, 1.665, 1.625, 1.599, 1.599 (8 singlets)	20.4–20.1, CH ₃
		169.9–169.4, C		169.6–169.1, C

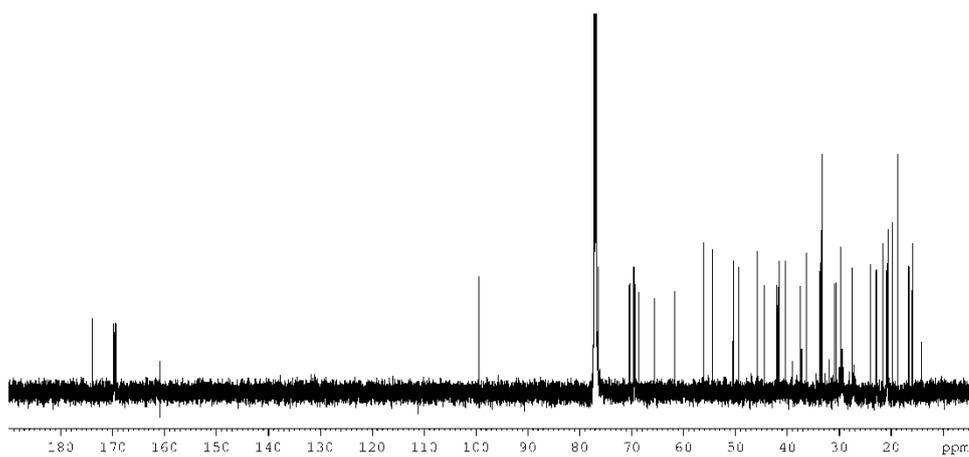


Figure 9.11 ^{13}C -NMR spectrum of plakohopanoid peracetate (**3b**) – CDCl_3 .

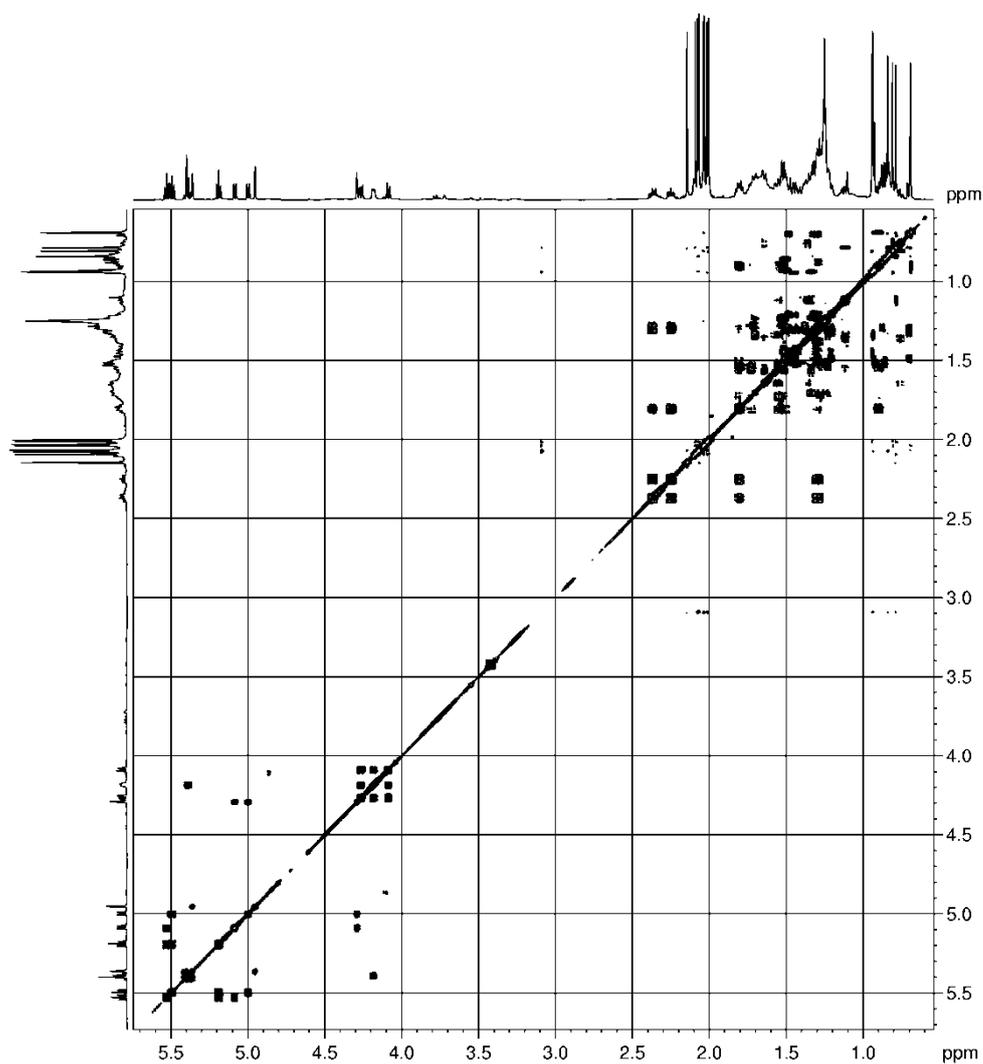


Figure 9.12 COSY spectrum of plakohopanoid peracetate (**3b**) – CDCl_3 .

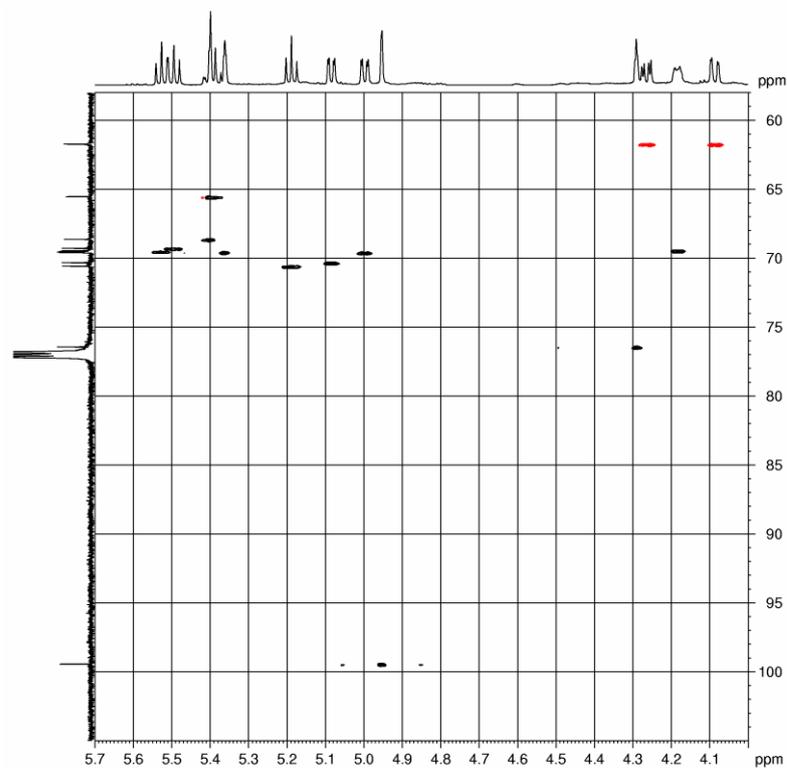
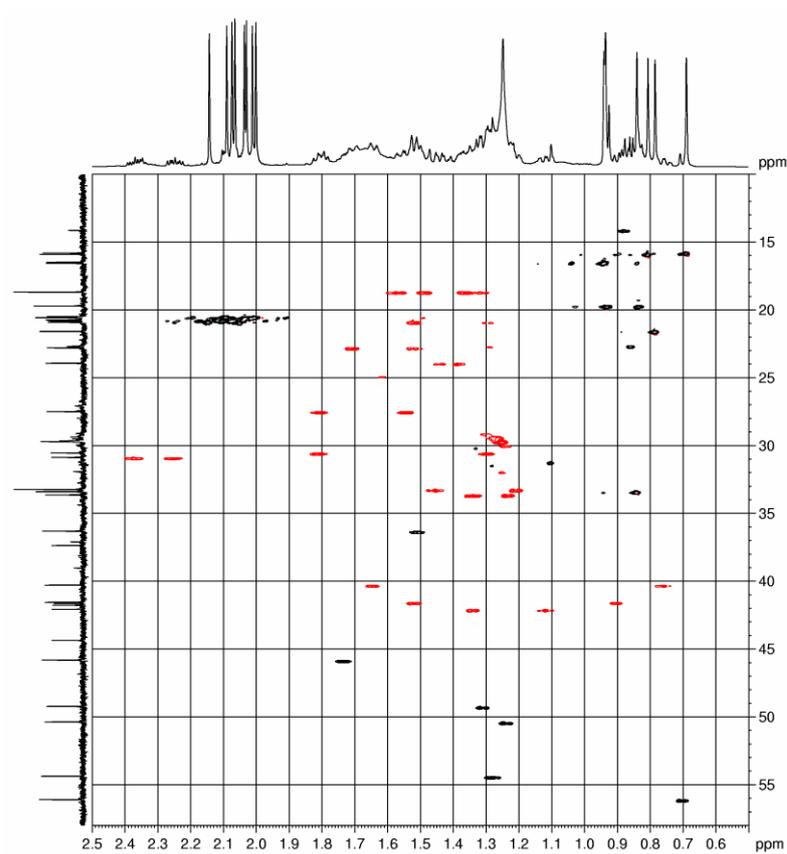


Figure 9.13-14 HSQC spectrum of plakohopanoid peracetate (**3b**) – CDCl₃ low-field region (9.12, above); high field region (9.13, below).



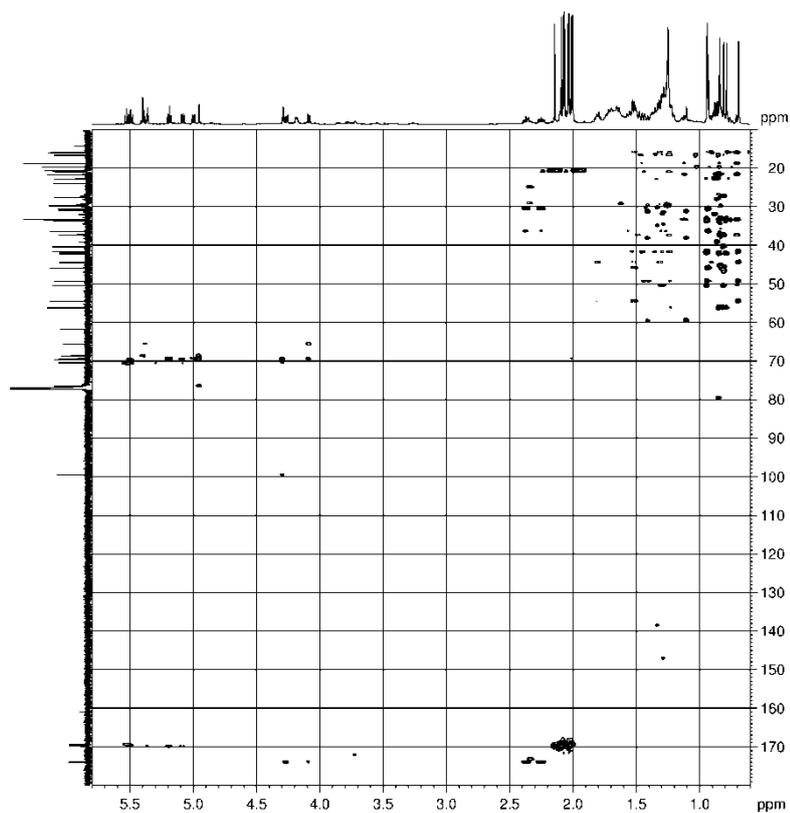


Figure 9.15 HMBC spectrum of plakohopanoid peracetate (**3b**) – CDCl_3 .

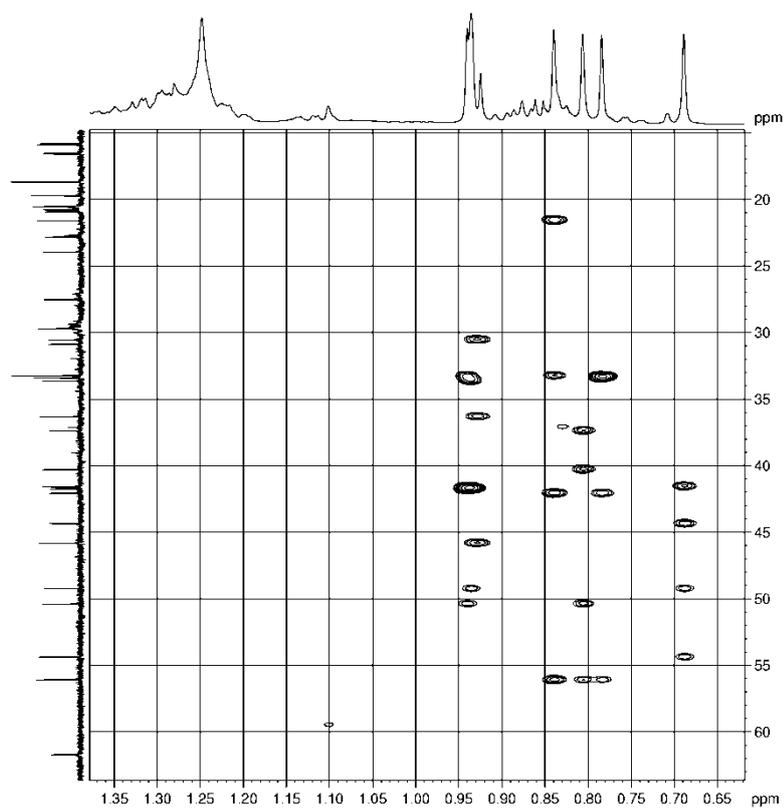


Figure 9.16 HMBC spectrum of plakohopanoid peracetate (**3b**) – CDCl_3 – methyl proton correlations.

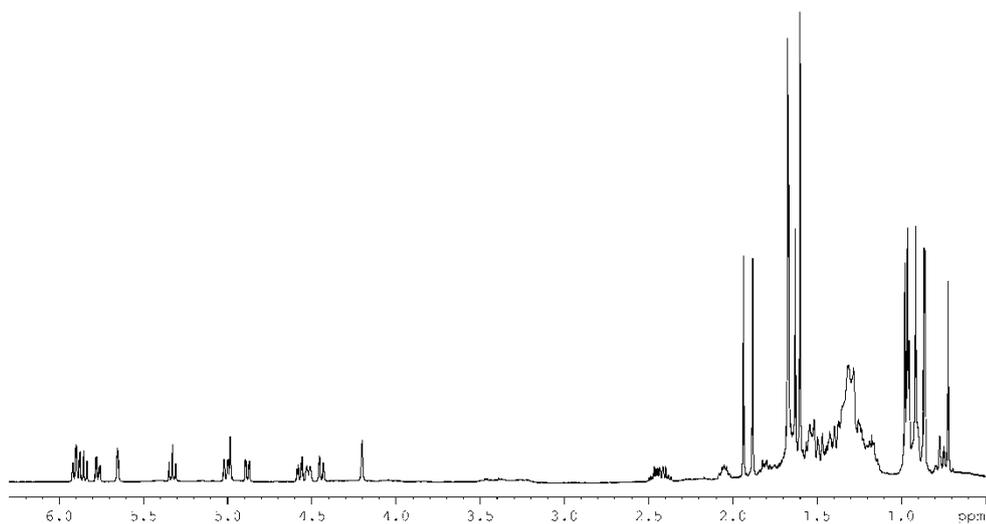


Figure 9.17 ^1H -NMR spectrum of plakohopanoid peracetate (**3b**) – C_6D_6 .

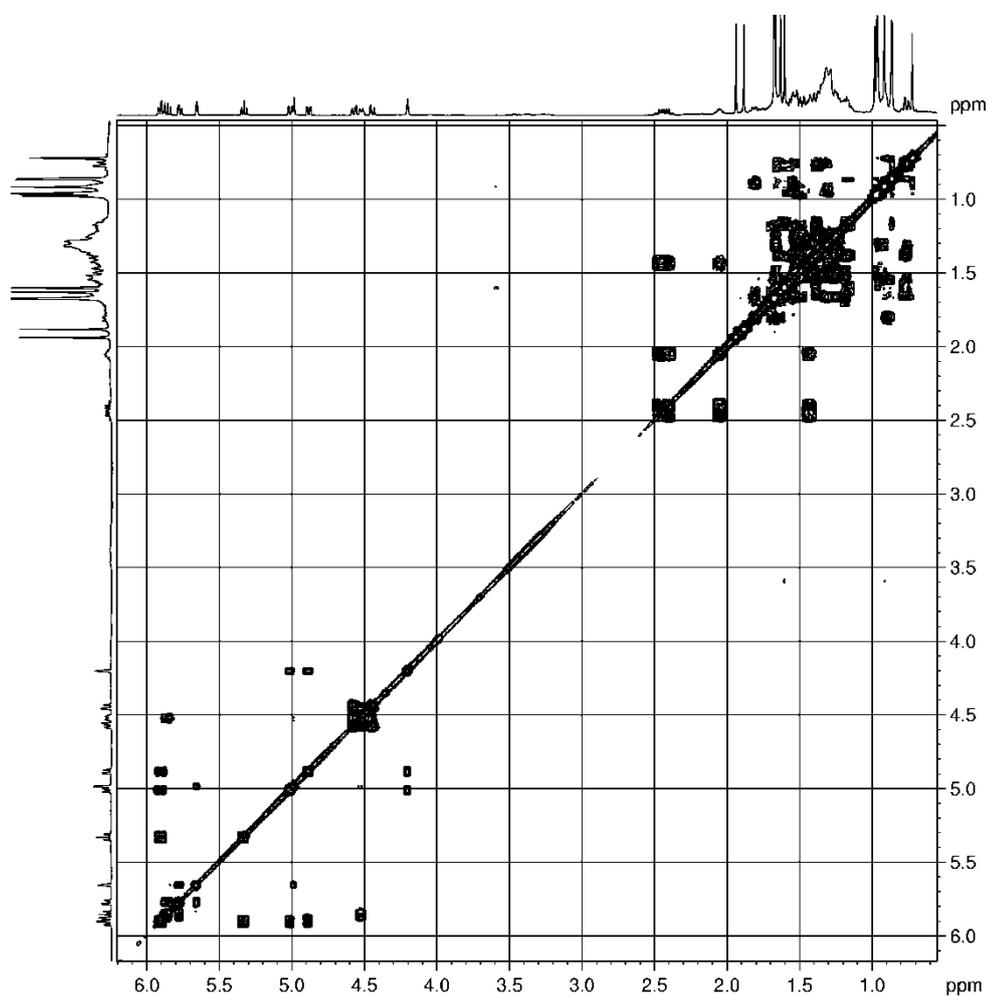


Figure 9.18 COSY spectrum of plakohopanoid peracetate (**3b**) – C_6D_6 .

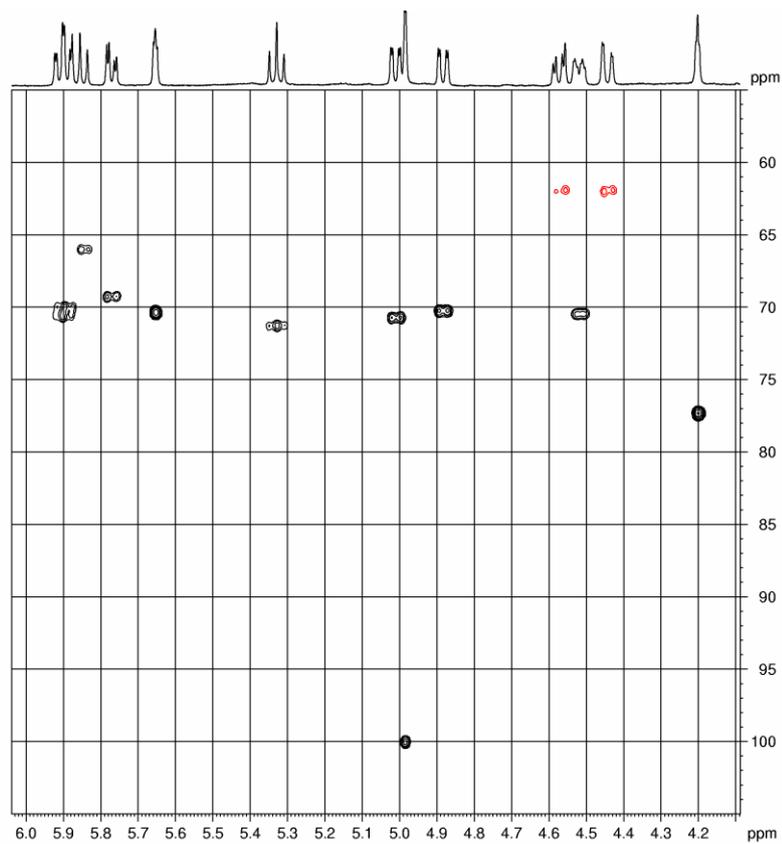
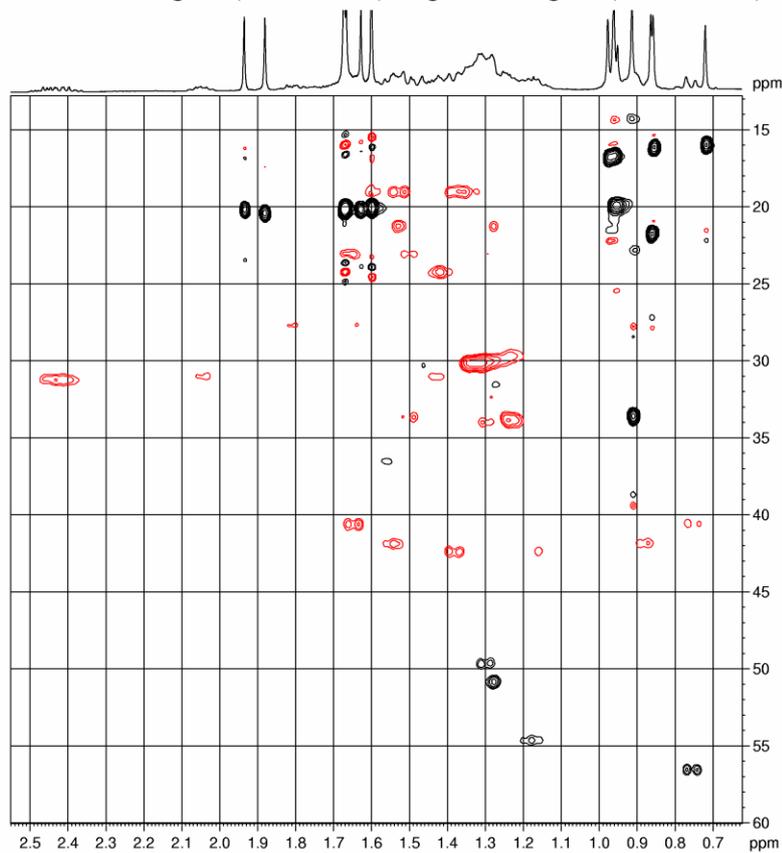


Figure 9.19-20 HSQC spectrum of plakohopanoid peracetate (**3b**) – C₆D₆ – low-field region (9.18, above); high field region (9.19, below).



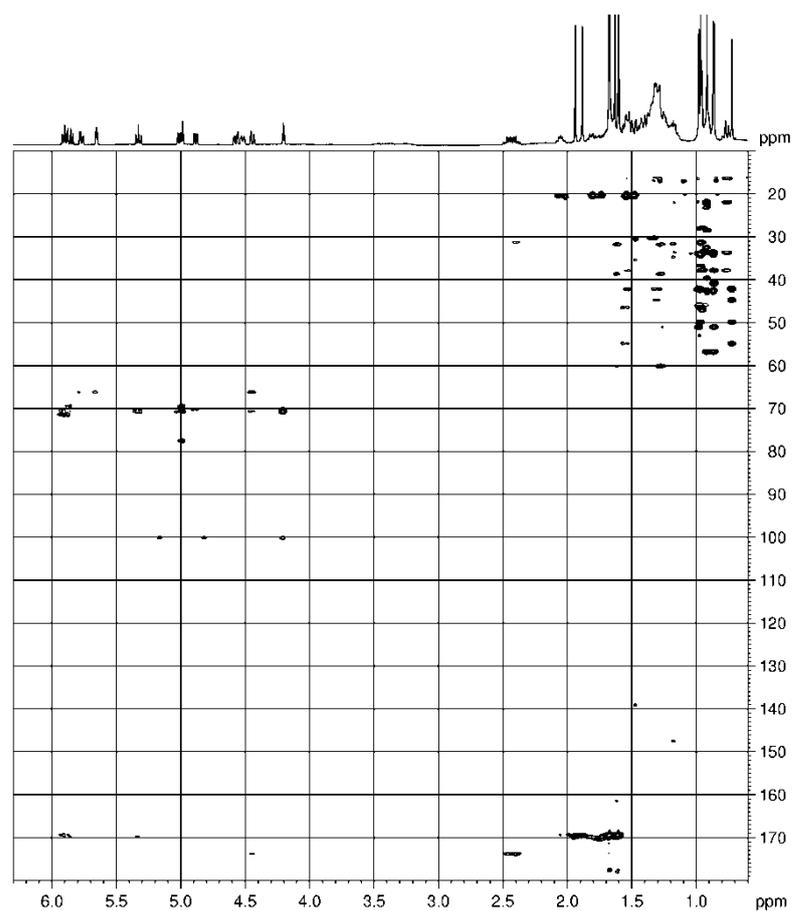


Figure 9.21 HMBC spectrum of plakohopanoid peracetate (**3b**) – C_6D_6 .

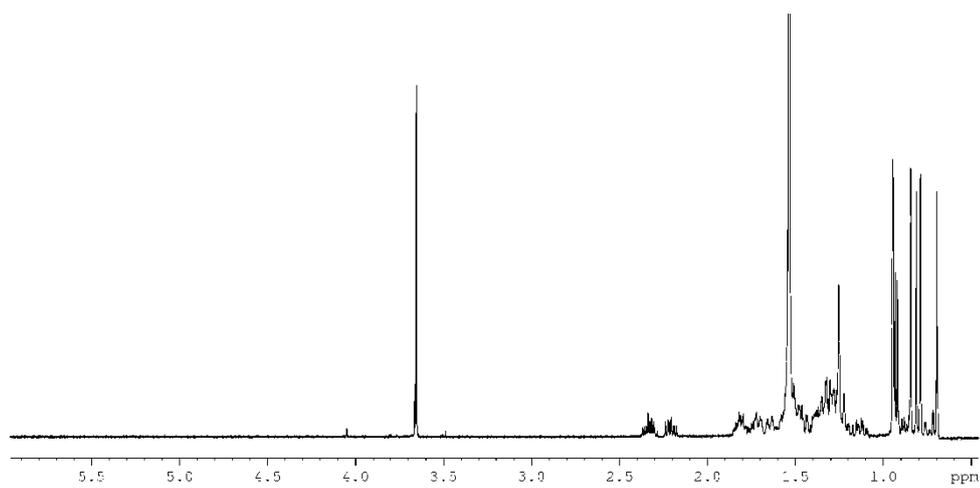


Figure 9.22 1H -NMR spectrum of hopanoic acid methyl ester (**5**).

References

1. Costantino V., Fattorusso E., Mangoni A., Di Rosa M., Ianaro A. (1997). Glycolipids from Sponges. 6. Plakoside A and B, Two Unique Prenylated Glycosphingolipids with Immunosuppressive Activity from the Marine Sponge *Plakortis simplex*. *J. Am. Chem. Soc.* 119, 12465–12470.
2. Costantino V., Fattorusso E., Mangoni A., Di Rosa M., Ianaro A. (1999). Glycolipids from sponges. VII. Simplexides, novel immunosuppressive glycolipids from the caribbean sponge *Plakortis simplex*. *Bioorg. Med. Chem. Lett.* 9, 271–276.
3. Costantino V., Fattorusso E., Mangoni A. (1993). Isolation of five-membered cyclitol glycolipids, crasserides: unique glycerides from the sponge *Pseudoceratina crassa*. *J. Org. Chem.* 58, 186–191.
4. Barbieri L., Costantino V., Fattorusso E., Mangoni A. (2005). Glycolipids from Sponges. Part 16. Discoside, a Rare *myo*-Inositol-Containing Glycolipid from the Caribbean Sponge *Discodermia dissoluta*. *J. Nat. Prod.* 68, 1527–1530.
5. Costantino V., Fattorusso E., Imperatore C., Mangoni A. (2000). The First 12-Methylhopanoid: 12-Methylbacteriohopanetetrol from the Marine Sponge *Plakortis simplex*. *Tetrahedron.* 56, 3781–3784.
6. Costantino V., Fattorusso E., Imperatore C., Mangoni A. (2001). A biosynthetically significant new bacteriohopanoid present in large amounts in the Caribbean sponge *Plakortis simplex*. *Tetrahedron.* 57, 4045–4048.
7. Rohmer M. (1993). The biosynthesis of triterpenoids of the hopane series in the Eubacteria: A mine of new enzymatic reactions. *Pure Appl. Chem.* 65, 1293–1298.
8. Neunlist S., Rohmer M. (1985). Novel hopanoids from the methylotrophic bacteria *Methylococcus capsulatus* and *Methylomonas methanica*. (22S)-35 aminobacteriohopane-30,31,32,33,34-pentol and (22S)-35-amino-3 beta-methylbacteriohopane-30,31,32,33,34-pentol. *Biochem. J.* 228, 769–771.

9. Seemann M., Bisseret P., Tritz J. P., Hooper A. B., Rohmer M. (1999). Novel bacterial triterpenoids of the hopane series from *Nitrosomonas europaea* and their significance for the formation of the C₃₅ bacteriohopane skeleton. *Tetrahedron Lett.* 40, 1681-1684.
10. Kannenberg E. L., Poralla K. (1999). Hopanoid Biosynthesis and Function in Bacteria. *Naturwissenschaften.* 86, 4, 168-176.
11. Rohmer M. (2008). From Molecular Fossils of Bacterial Hopanoids to the Formation of Isoprene Units: Discovery and Elucidation of the Methylerythritol Phosphate Pathway. *Lipids.* 43, 1095–1107.
12. Laroche M., Imperatore C., Grozdanov L., Costantino V., Mangoni A., Fattorusso E. (2007). Cellular Localization of Secondary Metabolites Isolated from the Caribbean Sponge *Plakortis simplex*. *Mar. Biol.* 151:1365–1373.
13. Schaeffer P., Schmit G., Adam P., Rohmer M. (2010). Abiotic formation of 32, 35-anhydrobacteriohopanetetrol: A geomimetic approach. *Org. Geochem.* 41, 1005–1008.
14. Ourisson G., Rohmer M. (1992). Hopanoids. 2. Biohopanoids: a novel class of bacterial lipids. *Acc. Chem. Res.* 25, 403–408.
15. Neunlist S., Bisseret P., Rohmer M. (1988). The hopanoids of the purple non-sulfur bacteria *Rhodospseudomonas palustris* and *Rhodospseudomonas acidophila* and the absolute configuration of bacteriohopanetetrol. *Eur. J. Biochem.* 171, 245–252.
16. Agrawal P. K., Jain D.C., Gupta R. K., R. Thakur S. (1985). Carbon-13 NMR spectroscopy of steroidal saponins and steroidal saponins. *Phytochemistry.* 24, 2479–2496.
17. Peiseler B., Rohmer M. (1991). Prokaryotic triterpenoids. (22*R*,32*R*)-34,35-Dinorbacteriohopane-32,33-diols from *Acetobacter acetii* ssp. *xylinum*: new bacteriohopane derivatives with shortened side-chain *J. Chem. Soc. Perkin Trans. I*, 2449–2453.
18. Hentschel U., Usher K. M., Taylor M. W. (2006). Marine sponges as microbial fermenters. *FEMS Microb. Ecol.* 55, 167-177.

19. Lee O. O., Chui P. Y., Wong Y. H., Pawlik J. R., Qian P. Y. (2009). Evidence for Vertical Transmission of Bacterial Symbionts from Adult to Embryo in the Caribbean Sponge *Svenzea zeai*. *Appl. Environ. Microbiol.* 75, 6147–6156.

