

**Elucidation of the Putative Gene Cluster for the
Biosynthesis of Siphonazole – a Secondary Metabolite
from *Herpetosiphon***

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Diplom-Biologe (Dipl.-Biol.)

Thomas Christian Höver

aus

Bonn

Bonn, 2012

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

- 1. Gutachter: Prof. Dr. G. M. König
- 2. Gutachter: PD Dr. habil. W. Knöss

Tag der Promotion: 11.07.2012

Erscheinungsjahr: 2012

Acknowledgements

I would like to express my sincere gratitude to Prof. Dr. G.M. König for giving me the opportunity to work on this interesting project and for introducing me to the fascinating field of secondary metabolite biosynthesis. I would like to thank her for her kind supervision and proofreading of my thesis. My thanks go to PD. Dr. W. Knöss for accepting the co-examination of my thesis and to Profs. H.-G. Sahl and E.A. Galinski for their participation in my examination committee.

My dear appreciation goes to all members of AG König for sharing good and hard times with me. I would like to thank all of them for the familiar atmosphere, their help and scientific exchange as well as for the non-scientific chat. My special thanks go to:

Dr. T. Schäberle for all his help and supervision and for all of our discussions. I would also like to thank him for the motivation and new ideas that inspired my work, his assistance with the CLUSEAN pipeline and for conducting IMS and the ATP-PP_i exchange assay.

Dr. Ö. Erol for introducing me into the project and for providing the genomic library and fosmid clones.

E. Goralski for her excellent help and unfailing technical assistance in all areas of molecular biological work, especially the yeast recombination.

S. Bouhired for all the discussions and mutual exchange and for organising the genome sequencing project.

Dr. M. El Omari for his help and good advice and the cheerful atmosphere during the plant anatomy course.

F. Lohr for our mutual exchange and for tackling the problems of λ -Red recombination and her assistance with this method.

E. Neu and E. Eguereva for knowing all the answers to organisational problems. Furthermore, I thank E. Eguereva for performing the LC-MS measurements.

G. Rashid for her excellent work on the mRNA expression analysis.

M. Josten for conducting the MALDI-TOF-MS measurements.

Dr. H. Gross for the kind and relaxed atmosphere during the plant anatomy course and for increasing my teaching experience.

The group of Prof. Dr. J. Piel for providing us with bacterial and yeast strains and for giving us the possibility to perform the ATP-PP_i exchange assay.

Prof. Dr. P. Dorrestein for giving us the opportunity to perform the IMS measurements in his lab.

Dr. T. Weber for the implementation of the CLUSEAN pipeline in our group.

My cordial thanks go to my parents, who are always there for me. I would like to thank them for all the love and help I get from them.

Finally, my special gratitude goes to my wife and my son, without whom everything else would be meaningless.

Contents

Acknowledgements.....	3
Contents.....	I
Figures.....	VII
Tables.....	IX
Abbreviations.....	X
Amino acids.....	XII
Nucleobases.....	XII
1 Introduction.....	1
1.1 Polyketides and non-ribosomal peptides.....	1
1.2 Polyketide synthases.....	2
1.2.1 Modular PKS type I.....	3
1.2.2 Classification of PKS types.....	6
1.2.3 Modular <i>trans</i> -AT PKS.....	7
1.3 Non-ribosomal peptide synthetases.....	9
1.3.1 Editing domains and domain subtypes in NRPS.....	11
1.3.2 Classification of NRPS types.....	12
1.3.3 Hybrid PKS-NRPS.....	13
1.4 The phylum Chloroflexi.....	14
1.4.1 Chloroflexales.....	15
1.4.2 <i>Herpetosiphonales</i>	16
1.4.3 <i>Herpetosiphon</i> sp. 060.....	17
1.5 Siphonazole – a secondary metabolite from <i>Herpetosiphon</i>	17
2 Scope of the Study.....	19
3 Materials & Methods.....	21
3.1 Materials.....	21
3.1.1 Chemicals.....	21
3.1.2 Devices.....	22
3.1.3 Antibiotics.....	24
3.1.4 Disposable material.....	24
3.1.5 Buffers and stock solutions.....	24
3.1.6 Media.....	25
3.1.7 Enzymes.....	26

3.1.8	Kits and standards	27
3.1.9	Organisms	27
3.1.10	Vectors	28
3.1.11	Clones & DNA-constructs	29
3.1.12	Primers	29
3.1.13	Software and Databases	30
3.2	General methods concerning micro-organisms	32
3.2.1	Sterilisation.....	32
3.2.2	Storage of organisms.....	32
3.2.3	Cultivation of bacterial organisms	32
3.2.4	Preparation of a crude extract from <i>Herpetosiphon</i> sp. 060	33
3.2.5	Test for antibiotic resistance of <i>Herpetosiphon</i> sp. 060	33
3.2.6	Imaging mass spectrometry (IMS)	33
3.3	Transformation of bacterial cells.....	34
3.3.1	Preparation of chemically competent cells	34
3.3.2	Preparation of electro-competent cells	34
3.3.3	Transformation of chemically competent cells.....	35
3.3.4	Transformation of electro-competent <i>E. coli</i> cells.....	35
3.3.5	Generation of protoplasts from <i>Herpetosiphon</i>	35
3.3.6	PEG-mediated transformation of <i>Herpetosiphon</i> protoplasts	36
3.3.7	Glycine treatment of <i>Herpetosiphon</i> cultures.....	36
3.3.8	Electroporation of <i>Herpetosiphon</i> sp. 060 cells	37
3.3.8.1	Electroporation of <i>Herpetosiphon</i> protoplasts	37
3.3.8.2	Electroporation of <i>Herpetosiphon</i> cells after glycine treatment	37
3.3.9	Sonication	37
3.3.10	Bacterial conjugation	38
3.4	Molecular biological methods concerning nucleic acids	39
3.4.1	PCR	39
3.4.1.1	Whole cell PCR.....	40
3.4.2	Agarose gel electrophoresis	40
3.4.3	Isolation of DNA from agarose gels	41
3.4.4	Isolation of plasmid and fosmid DNA	41
3.4.5	Isolation of genomic DNA	42
3.4.6	DNA precipitation	42

3.4.7	Restriction digestion	42
3.4.8	Ligation of DNA fragments	43
3.4.9	Dephosphorylation of linear DNA	44
3.4.10	Determination of nucleic acid concentration and purity	44
3.4.11	Sequencing of vector constructs and PCR-products	45
3.4.12	454 Genome sequencing	45
3.4.13	Screening of a genomic library	46
3.4.14	RNA-isolation and expression analysis	46
3.5	Recombination and cloning methods	46
3.5.1	λ -Red recombination	46
3.5.1.1	Generation of linear DNA for λ -Red recombination	47
3.5.2	Cloning of PCR-products	48
3.5.3	Yeast recombinational cloning	48
3.5.3.1	Isolation of plasmid DNA from yeast cells	49
3.6	Molecular biological methods concerning proteins	50
3.6.1	Heterologous expression of proteins	50
3.6.2	Cell lysis by sonication	50
3.6.3	Purification of recombinant proteins by Ni-NTA-columns	51
3.6.4	Denaturing SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)	51
3.6.5	Coomassie-staining	52
3.6.6	Concentration of purified proteins and buffer exchange	53
3.6.7	Determination of protein concentrations after Lambert-Beer	53
3.6.8	O-methyltransferase activity assay	54
3.6.8.1	High performance liquid chromatography (HPLC)	54
3.6.9	Mass spectrometry	55
3.6.10	ATP - PP _i exchange assay	55
4	Results	57
4.1	Knock-out experiments with <i>Herpetosiphon</i> sp. 060	57
4.1.1	Antibiotic resistance tests	57
4.1.2	Transformation of vegetative <i>Herpetosiphon</i> cells	58
4.1.3	Transformation of <i>Herpetosiphon</i> protoplasts	59
4.2	Genome sequencing of <i>Herpetosiphon</i> sp. 060	60
4.2.1	Genome search for the putative siphonazole gene cluster	61
4.2.2	Annotation of the contigs from genome sequencing	62

4.2.3	Other gene clusters discovered from genome sequences	63
4.3	Complete elucidation of the putative siphonazole gene cluster	64
4.3.1	Gap closure and alignment of the contigs	65
4.3.1.1	Positioning of contig 1462	65
4.3.1.2	Gap closure between contigs 0668 and 1280	65
4.3.1.3	Positioning of contig 1357	65
4.3.1.4	Gap closure between contigs 1518 and 1462	66
4.3.1.5	Gap closure between contigs 1518 and 0668	67
4.3.2	Organisation of the putative siphonazole gene cluster	68
4.4	NRPS domains and modules	69
4.4.1	Adenylation domains	70
4.4.2	Peptidyl carrier proteins	71
4.4.3	The condensation domain of sphF	72
4.4.4	Heterocyclisation modules	74
4.4.4.1	Cyclisation tandem domains	75
4.4.4.2	Oxidation domains	77
4.5	PKS domains and modules	79
4.5.1	The <i>trans</i> -AT and oxidoreductase SphA	79
4.5.2	Ketosynthases	81
4.5.3	Acyl carrier proteins	82
4.5.4	Dehydratases	83
4.5.5	Ketoreductases	84
4.5.6	The thioesterase SphJ	85
4.6	Non-PKS/NRPS cluster parts	86
4.6.1	SphB - an O-methyltransferase	86
4.6.2	The α/β -hydrolase encoded by sphH	87
4.6.3	SphI – an aldolase	88
4.7	Domain structure of the putative siphonazole biosynthetic enzymes	89
4.7.1	Classification of carrier proteins in view of the domain structure	91
4.8	Construction of the complete cluster from the fosmid library	92
4.8.1	Screening of the genomic library	92
4.8.2	Yeast recombinational cloning	93
4.8.3	Λ -Red mediated recombination	96
4.9	Protein expression and purification	99

4.9.1	Heterologous expression of the O-methyltransferase SphB	99
4.9.1.1	<i>In vitro</i> methylation of protocatechuic acid by SphB	100
4.9.2	Heterologous expression of the first A domain of SphC	101
4.9.2.1	ATP – PP _i exchange assay	103
4.9.3	Heterologous expression of the DAHP synthase SphI	103
4.10	Expression profile of siphonazole in <i>Herpetosiphon</i> sp. 060	104
4.10.1	mRNA expression of sph genes	104
4.10.2	Imaging mass spectrometry (IMS)	106
4.11	Heterologous expression of the complete gene cluster in <i>E. coli</i>	107
5	Discussion	109
5.1	Genetic accessibility of <i>Herpetosiphon</i> sp. 060	109
5.2	Elucidation of the putative siphonazole biosynthetic pathway	111
5.2.1	NRPS modules	111
5.2.1.1	The cyclisation modules	111
5.2.1.2	The initiation module	112
5.2.1.3	The glycine incorporating module	113
5.2.2	NRPS-PKS interfaces	113
5.2.3	<i>trans</i> -AT PKS modules	114
5.2.3.1	The oxidoreductase of SphA	114
5.2.3.2	Elongating and non-elongating modules	116
5.2.3.3	The cryptic module 3	116
5.2.4	The mosaic structure of the siphonazole gene cluster	117
5.3	Postulated biosynthesis of siphonazole	118
5.3.1	Nature and origin of the starter unit	119
5.3.1.1	The O-methyltransferase SphB	120
5.3.1.2	The DAHP synthase SphI	121
5.3.2	Substrate specificity of the A domain of the initiation module	122
5.3.2.1	ATP-PP _i exchange assay with the initiation module A domain	123
5.3.3	Decarboxylation of the terminal acetyl moiety	124
5.3.4	Transcriptional organisation of the gene cluster	126
5.4	Spatiotemporal biosynthesis of siphonazole	126
5.4.1	Biological role of siphonazole	127
5.5	Complete assembly of the putative siphonazole cluster	128
5.5.1	<i>In silico</i> and genomic library screening	128

5.5.2	Assembly by λ -Red recombination	129
5.5.3	Assembly by yeast recombinational cloning	130
5.6	Heterologous expression of the siphonazole gene cluster	131
5.6.1	Heterologous protein expression	134
5.7	Metabolic potential of Herpetosiphon	135
6	Summary	137
7	References	141
8	Appendix.....	157
8.1	Primary sequences of the proteins SphA – SphJ	157
8.1.1	SphA.....	157
8.1.2	SphB.....	157
8.1.3	SphC.....	157
8.1.4	SphD.....	159
8.1.5	SphE.....	160
8.1.6	SphF	161
8.1.7	SphG	162
8.1.8	SphH.....	162
8.1.9	SphI	163
8.1.10	SphJ	163
8.2	Search results from the NaPDos Database	164
8.2.1	C domains in contigs from the genome sequencing	164
8.2.2	KS domains in contigs from the genome sequencing	165
8.3	Primer sequences	166
8.3.1	Gap closure	166
8.3.2	Protein expression	167
8.3.3	Fosmid library screening	167
8.3.4	Yeast recombinational cloning	168
8.3.5	λ -Red mediated recombination	168
8.3.6	Sequencing primers.....	168
8.3.7	mRNA screening primers	169
8.4	Complete sequence alignment of the DAHP synthase SphI	170
8.5	HPLC UV-spectra.....	171
8.6	Chromatograms from ATP-PP _i exchange assay	172
	Publications.....	173

Figures

Figure 1-1: Examples of natural products	2
Figure 1-2: Post-translational priming of carrier proteins	3
Figure 1-3: Central catalytic processes during PK biosynthesis	4
Figure 1-4: optional reduction steps during PK biosynthesis	4
Figure 1-5: 6-Deoxyerythronolide B (DEB) biosynthesis.....	5
Figure 1-6: structure and mechanism PKS types	7
Figure 1-7: Leinamycin biosynthetic pathway.....	8
Figure 1-8: Basic reactions in NRP biosynthesis	10
Figure 1-9: Reactions performed by optional domains in NRP biosynthesis	12
Figure 1-10: Tyrocidine biosynthetic pathway	13
Figure 1-11: Phylogenetic tree of the major prokaryotic groups	15
Figure 1-12: Images of <i>Herpetosiphon</i> sp. 060	17
Figure 1-13: Molecular structure of siphonazole and phenoxan.....	18
Figure 4-1: Effect of glycine on the growth of <i>Herpetosiphon</i> sp. 060.....	59
Figure 4-2: phase-contrast micrographs of <i>Herpetosiphon</i> sp. 060	60
Figure 4-3: predicted domain structure encoded on selected contigs	63
Figure 4-4: PCR-reactions from genomic DNA bridging the contigs	66
Figure 4-5: Organisation of the putative siphonazole gene cluster	68
Figure 4-6: Modular organisation of the putative biosynthetic enzymes	70
Figure 4-7: Multiple sequence alignment of PCP domains	72
Figure 4-8: Multiple sequence alignment of the C domain from SphF	73
Figure 4-9: GC-content of the putative siphonazole biosynthetic genes	74
Figure 4-10: Multiple sequence alignment of the Cyc domains	77
Figure 4-11: Multiple sequence alignment of the Ox domains.....	78
Figure 4-12: Multiple sequence alignment of the acyltransferase SphA	80
Figure 4-13: Multiple sequence alignment of the oxidoreductase of SphA	81
Figure 4-14: Multiple sequence alignment of the KS domains	82
Figure 4-15: Multiple sequence alignment of the ACP domains	82
Figure 4-16: Multiple sequence alignment of the DH domains	83
Figure 4-17: Multiple sequence alignment of the KR domains	84
Figure 4-18: Multiple sequence alignment of the thioesterase SphJ.....	86
Figure 4-19: Multiple sequence alignment of the O-methyltransferase SphB.....	87
Figure 4-20: Multiple sequence alignment of the α/β -hydrolase of SphH	88

Figure 4-21: Multiple sequence alignment of the DAHP synthase SphI.....	89
Figure 4-22: Hypothetical pathway for the biosynthesis of siphonazole.....	90
Figure 4-23: PCR-screening of the final cluster part	93
Figure 4-24: Linear representation of fosmid clone inserts	94
Figure 4-25: PCR reactions from yeast recombination	95
Figure 4-26: Completion of the putative siphonazole cluster by λ -Red	97
Figure 4-27: Reconstitution of the putative siphonazole cluster	98
Figure 4-28: purification of SphB by affinity-chromatography	100
Figure 4-29: UV absorption spectra from HPLC.....	101
Figure 4-30: Purification of SphC-A1 by affinity-chromatography.....	102
Figure 4-31: Purification of SphI by affinity-chromatography	104
Figure 4-32: Test PCRs for RNA expression analysis	105
Figure 4-33: RT-PCR from RNA samples.....	106
Figure 4-34: False-colour images from IMS	107
Figure 5-1: PKS-NRPS interfaces.....	114
Figure 5-2: Natural products produced by pathways with putative OR domain	115
Figure 6-1: Organisation of the putative siphonazole PKS/NRPS system	138
Figure 8-1: Multiple sequence alignment of SphI	170
Figure 8-2: UV-absorption spectra from methylation assay.....	171
Figure 8-3: Chromatograms from MALDI-TOF-MS measurements	172

Tables

Table 4-1: Antibiotic resistance tests with <i>Herpetosiphon</i> sp. 060	57
Table 4-2: Contigs identified by alignment with fosmid subclones	61
Table 4-3: Hybrid PKS/NRPS cluster on contig 805	64
Table 4-4: BLASTp search results for the translated amino acid sequences from the ORFs surrounding the sph genes	68
Table 4-5: BLASTp search results for the translated amino acid sequences from sph genes.....	69
Table 4-6: Results of the NRPSpredictor for the A domains	70
Table 4-7: signature motifs for condensation domains	72
Table 4-8: Signature motifs for cyclisation-domains.....	75
Table 4-9: Signature motifs for oxidation domains	78
Table 4-10: Signature motifs for methyltransferases.....	86
Table 4-11: Classification of carrier domains as predicted by CLUSEAN.....	91
Table 5-1: Codon adaptation index of the sph genes	133
Table 8-1: C domains encoded on contigs from 454 sequencing.....	164
Table 8-2: C and Cyc domains from the putative siphonazole cluster	165
Table 8-3: KS domains encoded on contigs from 454 sequencing	165
Table 8-4: KS domains from the putative siphonazole cluster	166

Abbreviations

aa	Amino acid
ACP	Acyl carrier protein
A domain	Adenylation domain
AT	Acyl transferase
bp	Base pair
C domain	Condensation domain
Cyc domain	Cyclisation domain
CoA	Coenzyme A
Da	Dalton
DAHP	3-deoxy-D-arabino-heptulosonate 7-phosphate
DH	Dehydratase
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dsDNA	double stranded DNA
DTT	Dithiothreitol
E4P	Erythrose 4-phosphate
EDTA	N,N,N',N'-Ethylenediaminetetraacetat
ER	Enoylreductase
FAS	Fatty acid synthesis
HPLC	High Performance Liquid Chromatography
IMS	Imaging Mass Spectrometry
IPTG	Isopropyl- β -D-thiogalactosid
kb	Kilobases
kDa	Kilodalton
KR	Ketoreductase
KS	Ketosynthase
MALDI	Matrix-assisted laser desorption/ionisation
mRNA	Messenger RNA
NADH	Nicotinamide adenine dinucleotide
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
O-MT	O-methyltransferase

OR	Oxidoreductase
ORF	Open reading frame
oriT	Origin of transfer
PAGE	Polyacrylamide gel electrophoresis
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PEP	phosphoenolpyruvate
pI	Isoelectric point
PK	Polyketide
PKS	Polyketide Synthase
PP _i	Inorganic pyrophosphate
NRPS	Non-ribosomal Peptide Synthetase
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
RT-PCT	Reversed transcription-PCR
SAM	S-adenosylmethionine
SDS	Sodium dodecylsulfate
sp.	Species
ssDNA	Single stranded DNA
T domain	Thiolation domain
TE	Thioesterase
TEMED	N,N,N',N'-Tetramethylethylenediamin

Amino acids

Amino acid	3-letter code	1-letter code	Amino acid	3-letter code	1-letter code
Glycine	gly	G	Threonine	thr	T
Alanine	ala	A	Cysteine	cys	C
Valine	val	V	Tyrosine	tyr	Y
Leucine	leu	L	Asparagine	asn	N
Isoleucine	ile	I	Glutamine	gln	Q
Methionine	met	M	Aspartate	asp	D
Phenylalanine	phe	F	Glutamate	glu	E
Tryptophan	trp	W	Lysine	lys	K
Proline	pro	P	Arginine	arg	R
Serin	ser	S	Histidine	his	H

Nucleobases

Base	IUPAC code
Adenine	A
Thymine	T
Guanine	G
Cytosine	C
Any	N

1 Introduction

1.1 *Polyketides and non-ribosomal peptides*

For decades natural products have played a major role as biological probes, as inspiration for organic chemistry and, most prominent, as an important source for therapeutics (Walsh & Fischbach, 2010). Of all small-molecule new chemical entities between 1981 and 2006, natural products and derived semisynthetic compounds accounted for 68% of antibacterial and 54% of anticancer compounds (Newman & Cragg, 2007), and natural product structures have a great impact on the anti-infective area. Connecting the products to the genes encoding the biosynthetic machinery has stimulated new interest into natural product research. The growing number of available genomic data and the application of genome mining technologies contribute to the discovery of ever new structures, thus illustrating the lasting importance of natural products for drug discovery (Challis, 2008).

Polyketides (PKs) and non-ribosomal peptides (NRPs) represent two of the most prominent classes of natural products with an enormous structural and functional variety (Meier & Burkhardt, 2009), which is reflected by the diversity of producers, ranging from bacteria and protozoa to fungi and plants. Although their physiological role for the producer is often not clear, it is supposed that they carry out functions as pigments, virulence factors, messengers or toxins (Hertweck, 2009). The broad spectrum of biological activity of these metabolites makes PKs and NRPs highly interesting research fields for new drugs and pharmaceutically relevant compounds.

A number of natural products and their derivatives have found their way into clinical use. Amongst them are antibiotics like erythromycin A, tetracyclines or the last resort antibiotic daptomycin (Strieker et al., 2010) and anti-tumor agents like bleomycin and the epothilones. Bryostatin 1 is another potent cytotoxin against different cancer cell lines and in clinical trial phase II and III in combination with other agents (Piel, 2010). Anticholesterol compounds like lovastatin or mevastatin and immunosuppressants like rapamycin or cyclosporin A further exemplify the outstanding activity spectrum of these classes of natural products.

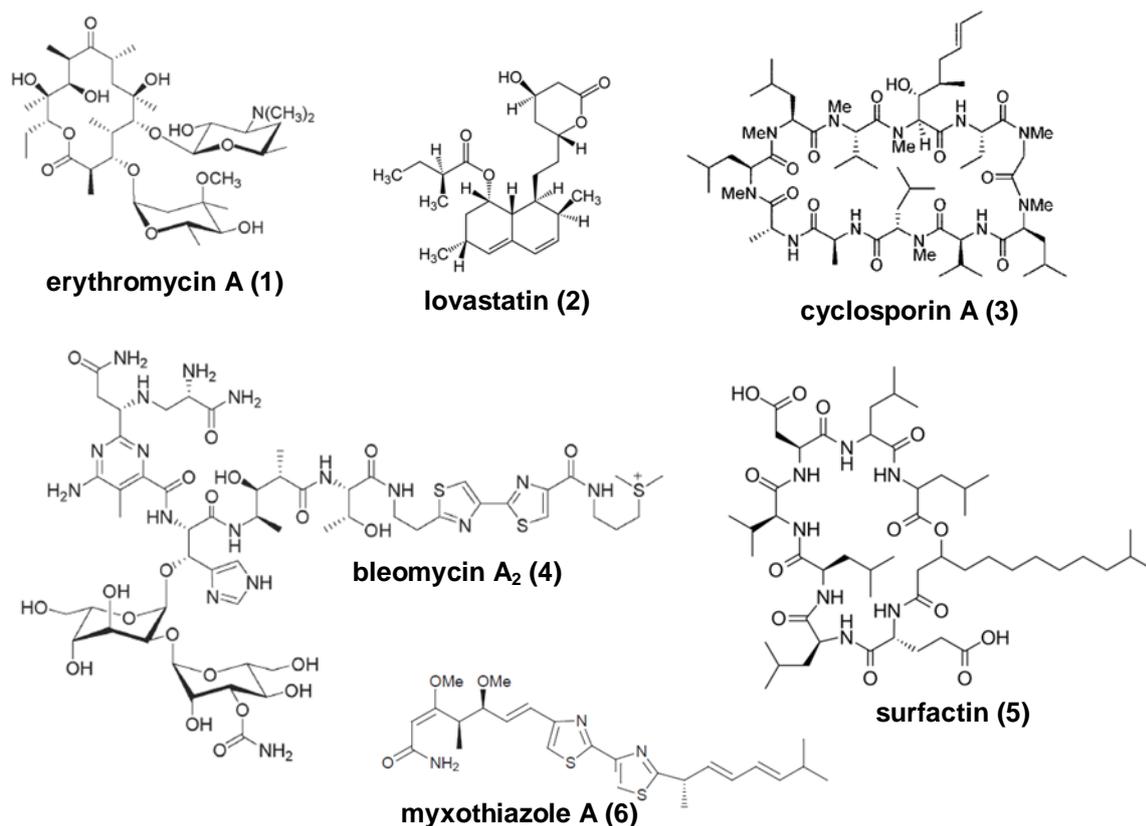


Figure 1-1: Examples of natural products

Derived from PKS (1 & 2), NRPS (3 & 4) or mixed PKS-NRPS (5 & 6) pathways

PKs and NRPs are the product of huge multienzymes or enzyme complexes termed polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS), respectively, which belong to the largest proteins known so far and which will be described in the following.

1.2 Polyketide synthases

Polyketide synthases are multifunctional megaenzymes (Piel et al., 2005) that synthesize complex structures from simple acyl building blocks by a Claisen-type condensation reaction. This biochemistry is shared with fatty acid synthases (FAS) to which PKSs are evolutionary related (Smith & Tsai, 2007). In fatty acid synthesis an activated starter acetyl unit is extended by malonyl moieties through decarboxylative thioester condensation, followed by reduction and dehydration steps to reliably produce saturated fatty acids (Staunton & Weissman, 2001). PKS systems, however, comprise a greater diversity in their building blocks as well as greater flexibility in the

reduction state, which is responsible for the great structural diversity of the produced compounds.

Since Collie's first experiments on the synthesis of orcinol (Collie & Myers, 1893), the last century has seen a huge progress in the unravelling of the polyketide biosynthesis. Today, three types of PKSs have been discovered, each containing different subclasses (Shen, 2003). For an exemplary description of the underlying logic, participating subunits and enzymatic steps the family of non-iterative PKS type I will be described in more detail.

1.2.1 Modular PKS type I

This PKS type is composed of giant multifunctional enzymes, which consist of one or more functional units called module. In the non-iterative subclass each module is responsible for the elongation of the growing polyketide chain by one unit and can be further subdivided into distinct enzyme domains. The minimal module consists of a β -ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP), which are usually organised in a KS-AT-ACP arrangement.

Carrier proteins, also termed thiolation domains (T), are responsible for the covalent attachment of the building blocks as thioesters. The substrates are not directly bound to the protein itself, but to a 5'-phosphopantetheinyl prosthetic group, which is covalently bound to a conserved serine residue. This group is transferred to the ACP by the action of specific 5'-phosphopantetheinyl transferases (PPTases) prior to the biosynthesis of the polyketide, thus turning the *apo*-ACP into its active *holo*-form (Lambalot et al., 1996).

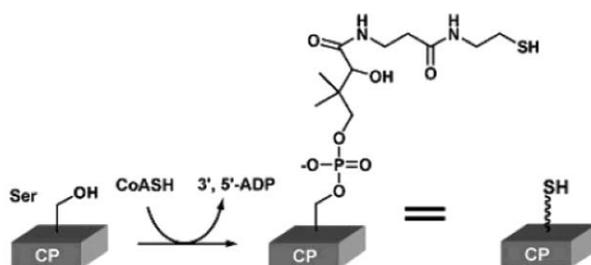


Figure 1-2: Post-translational priming of carrier proteins

Transfer of the phosphopantetheinyl arm from coenzyme A (CoA) to the carrier protein (CP) by PPTase (from Lai et al., 2006)

At the initiation of the biosynthesis, the AT domain loads the primer substrate to the first carrier protein, which passes it to a conserved cystein moiety of the KS domain of the next module. The KS catalyses a decarboxylative condensation with the elongation substrate on the next ACP. Thus, the growing polyketide chain is passed from module to module, while always staying covalently bound to the enzymes. After the final elongation step, the product is released from the protein by the thioesterase (TE) in the termination module. The molecule is either hydrolysed or undergoes intramolecular macrocyclisation (Weissman et al., 1998).

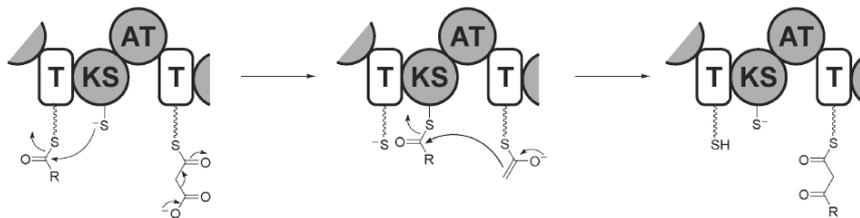


Figure 1-3: Central catalytic processes during PK biosynthesis

Chain elongation by decarboxylative Claisen condensation; T, thiolation domain; KS, ketosynthase; AT, acyltransferase (from Walsh & Fischbach, 2006)

Additionally to this minimal set of domains, a module can contain modifying domains, such as ketoreductase (KR), dehydratase (DH) and enoylreductase (ER). In fatty acid synthesis, the growing chain always undergoes the complete set of reduction steps. First, the keto group is reduced to a hydroxyl group by the KR, followed by the elimination of water by the DH and the reduction of the resulting double bond by the ER. Differently, PKS modules can contain none, one or more of these enzymes, resulting in different grades of reduction.

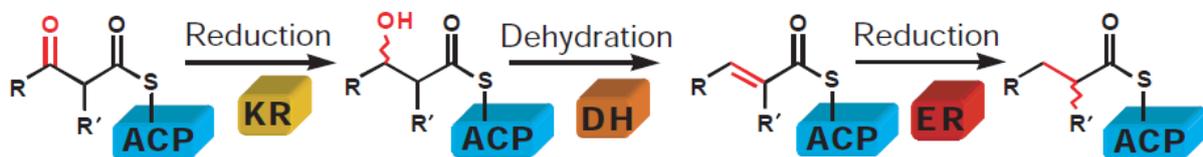


Figure 1-4: optional reduction steps during PK biosynthesis

Catalytic reactions performed on the growing chain; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; ER, enoylreductase (from Cane & Walsh, 1999)

The biosynthetic logic for non-iterative modular PKSs was unravelled through the identification of the 6-deoxyerythronolide B synthase (DEBS) gene cluster in *Saccharopolyspora erythraea* in the pioneer work by the groups of Peter Leadly

(Cortés et al., 1990) and Leonard Katz (Tuan et al., 1990). Consequently, DEBS became the archetype for this PKS class. The discoveries were facilitated by the fact that in microbes all genes for the biosynthesis and ancillary enzymes of a metabolite tend to cluster physically in the genome (Walsh, 2008). It was revealed that the number and order of the modules and their domains in a PKS system mirror the structure of the produced compound, thus allowing a prediction of the molecule structure. This relation is referred to as colinearity rule.

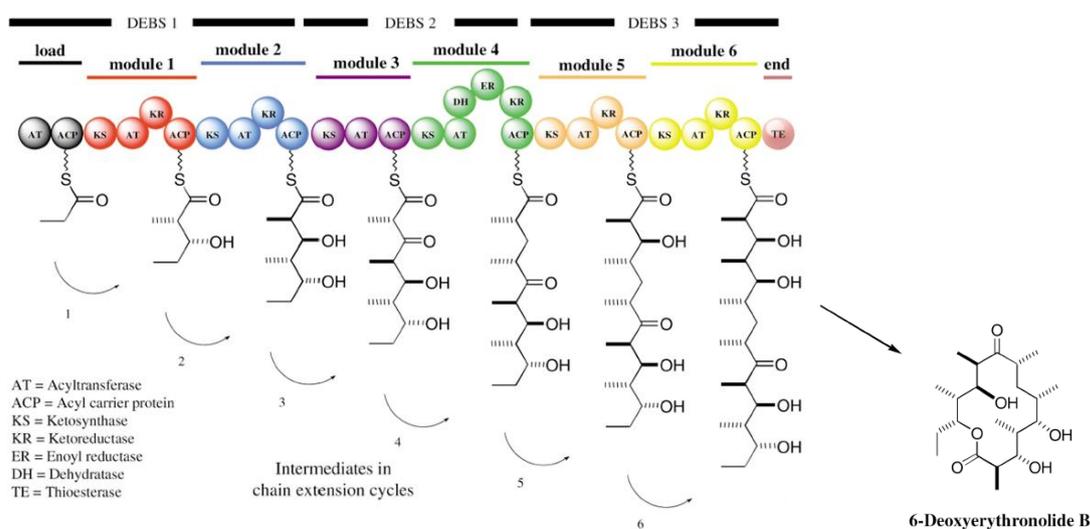


Figure 1-5: 6-Deoxyerythronolide B (DEB) biosynthesis

Organisation of a modular type I PKS system; formation of the erythromycin precursor (mod. after Staunton & Weissman, 2001)

While in FAS a starter acetyl unit is usually elongated by malonyl-CoA, PKS loading modules employ a broader spectrum of incorporated building blocks (Moore & Hertweck, 2002), including acetyl-CoA, malonyl-CoA, propionyl-CoA and methylmalonyl-CoA, isobutyryl-CoA, ethylmalonyl-CoA, propylmalonyl-CoA and hydroxymalonyl-CoA (Pfeifer & Khosla, 2001). Further chain extension is performed by the use of either malonyl- or methylmalonyl-CoA. In this connection, the acyltransferases act as gatekeepers that confer the substrate specificity. In contrast to the generally fastidious ATs of the extender modules (Liou et al., 2003), the AT domain of the initiation module often shows a more relaxed specificity, giving rise to the possibility of incorporating alternate primer units to create derivatives of a natural product (Khosla et al., 1999).

This versatile use of building blocks and the varying degree of reduction both greatly contribute to the vast structural complexity of PKs. Gonzáles-Lergier and co-workers calculated that a PKS with six elongation modules could produce a theoretical number of 100,000 different structures (Gonzáles-Lergier et al., 2005). A further contribution to the diversity is made by so-called post-PKS tailoring enzymes. The bare PK backbone is often modified by hydroxylation, methylation, oxidation or the addition of deoxysugars (Rawlings, 2001) and thus turned into its bioactive form, e.g. as in the case of erythromycin.

1.2.2 Classification of PKS types

Polyketide synthases are generally classified according to their structural organisation into three types (Shen, 2003; Rawlings, 2001). As mentioned before, type I PKSs consist of multifunctional megaenzymes, in which each protein harbours a linear sequence of modules and their catalytic domains (Hertweck, 2009) and are therefore also termed modular PKS. From a structural point of view, this type resembles the animal FAS (Staunton & Weisman, 2001). A further subdivision can be made according to the non-iterative or iterative use of the single modules. In the first case, each module is only used once, so that the colinearity rule applies. This type is widely spread among eubacteria and has already been described before (1.2.1). The group of *trans*-AT type PKS will be described later.

The iterative type I is typical for fungi and was found responsible for the production of compounds like 6-methylsalicylic acid lovastatin. During biosynthesis, modules are used repeatedly for elongation. Even so, the degree of reduction and the substitution pattern can differ with each cycle, as KR, DH, ER and methylation domains (MT) are employed variably. The factors that regulate this alternating use are hitherto unknown. Further subdivision into non-reducing, partially-reducing or highly-reducing PKSs can be made depending on the presence of reductive domains.

Type II PKS consist of discrete, monofunctional enzymes that work iteratively and thus resemble bacterial FAS. A minimal set comprises an ACP and two ketosynthases named KS_{α} and KS_{β} . This type seems to be restricted to bacteria, where it is paradigmatic for the biosynthesis of aromatic compounds, a task that is taken over by iterative type I PKSs in fungi. Chain elongation in type II PKS is similar

to modular systems. The building blocks are transferred from the ACP to KS_{α} , which catalyses the condensation to the next unit tethered to the ACP. In KS_{β} a catalytic histidine is replaced by a glutamine residue, which turns the condensing enzyme into a decarboxylase. This domain, also termed chain initiation factor (Bisang et al., 1999) or KSQ , provides the acetyl starter unit by decarboxylation of malonyl-CoA. $KSQs$ can also be found in the loading module of some type I PKSs. Further modification steps are performed by accessory enzymes. The prototype for type II PKS was the actinorhodin synthase discovered by Hopwood and co-workers (Malpartida & Hopwood, 1984).

Type III PKS are known as chalcone synthases and are distinct from the other two classes. They consist of homodimeric, iteratively acting condensing enzymes that select the starter units and perform the chain elongation in an ACP-independent manner by directly acting on the CoA-bound building blocks. Until the discovery of the RppA synthase in *Streptomyces griseus* (Funa et al., 1999) type III PKS were thought to be distinct for plants, but have meanwhile also been found in fungi (Seshime et al., 2005).

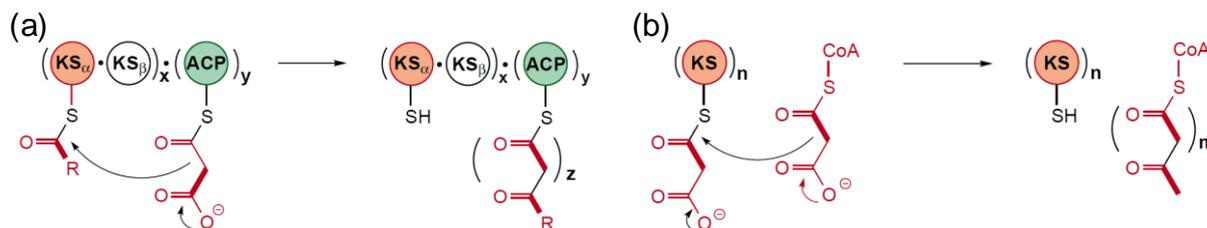


Figure 1-6: structure and mechanism PKS types

(a) iterative type II PKS; condensation reaction is catalysed by KS_{α} ; chain elongation factor KS_{β} acts as decarboxylase; (b) ACP independent, iterative type III PKS; KS acts directly on the CoA-bound substrate (from Shen, 2003)

1.2.3 Modular *trans*-AT PKS

The analyses of PKS clusters from different bacterial phyla led to the discovery of a further family of modular polyketide synthases. In deviation from the textbook type I PKS, the modules are devoid of acyltransferases. Instead, the ACPs are loaded by one to three free standing ATs (Piel, 2010). In consequence, this class of PKSs was named *trans*-AT PKS to distinguish it from the conventional *cis*-type modular synthases. The first *trans*-AT cluster was the bacillaene PKS from *Bacillus subtilis*

(Scotti et al., 1993), although it could not be assigned to a product at that time. The first cluster attributed to a metabolite was the pederin synthase discovered by Piel and co-workers (Piel et al., 2002). The free standing ATs act iteratively and are responsible for the acylation of all PKS modules, as was first shown for the biosynthesis of leinamycin (Cheng et al., 2003). *Trans*-ATs are either encoded as single genes or are fused to an oxidoreductase (OR), which has the function of a *trans*-acting enoylreductase (Bumpus et al., 2008).

Phylogenetic studies on the KS domains suggest a different evolutionary origin for *cis*- and *trans*-PKS enzymes (Piel et al., 2004). KS domains from *cis*-AT PKS usually mirror the phylogenetic group of their organisms, while the clade of *trans*-AT KS includes bacteria of different phyla. It was suggested that *cis*-AT PKS mainly evolved through repeated gene duplication (Jenke-Kodama et al., 2006) and fusion of module-encoding genes within evolving pathways. For *trans*-AT PKSs extensive horizontal gene transfer from various sources and recombination events seem to be responsible for a mosaic-like evolution of these gene clusters. Furthermore, the analysis of numerous KS domains from *trans*-AT clusters revealed that the domain architecture allows conclusions about the substrate specificity of these domains (Nguyen et al, 2008). It was shown that the domains can be grouped corresponding to their substrate types, like e.g. saturated, α,β -unsaturated, α -methyl-branched or β -hydroxylated, which assigns the role of a gatekeeper rather to the KS than the AT.

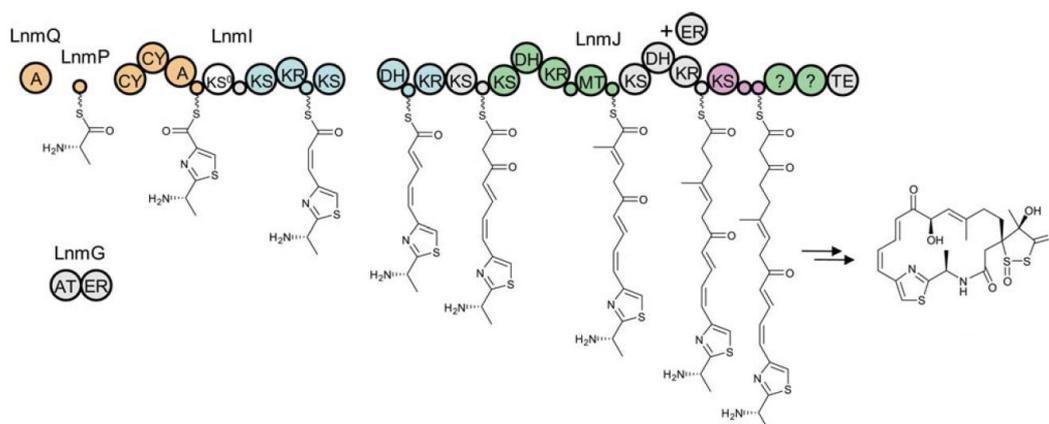


Figure 1-7: Leinamycin biosynthetic pathway

Example of a mixed NRPS-PKS system with a *trans*-AT type PKS part; AT LnmG is responsible for the loading of all ACPs during biosynthesis; A, adenylation; AT, acyltransferase; Cy, cyclisation; ER, enoylreductase; KS, ketosynthase; DH, dehydratase; KR, ketoreductase; TE, thioesterase (from Piel, 2010)

Trans-AT PKSs frequently show a number of peculiarities, which are uncommon in *cis*-type synthases (Hertweck, 2009). As many as 50 variants of module architecture are known so far in contrast to only eight different arrangements in *cis*-AT PKSs (Piel, 2010). These include duplication of domains, unusual orders, division on different genes as well as obviously redundant domains or modules, which apparently do not correspond with the structure of the product. Especially the occurrence of multiple ACPs in one module is frequently observed. These modules are often involved in the introduction of β -branches (Calderone, 2008). Another peculiarity are KSs which lack a catalytic histidine residue (KS₀) and are therefore unable to catalyse chain elongation, but merely pass on the polyketide intermediate.

1.3 Non-ribosomal peptide synthetases

Non-ribosomally derived peptides represent another class of natural products that is of equal structural diversity and pharmaceutical relevance as the already discussed polyketides. Among them are immunosuppressive agents as cyclosporin A, important antibiotics like gramicidin S and vancomycin as well as siderophores like myxochelin A or vibriobactin (Schwarzer et al., 2003). Similar to polyketides they are produced from simple monomers by modular multifunctional mega-enzymes, called non-ribosomal peptide synthetases (NRPS). In general, they follow the same precepts as modular type I PKSs (Fischbach & Walsh, 2006), as was first exemplified at the tyrocidin biosynthetic cluster (Mootz & Marahiel, 1997). Analogous to PKSs, the minimal NRPS module comprises two catalytic domains and a carrier protein for the selection and activation of the substrate and the formation of the peptide bond. These are the adenylation (A) and condensation (C) domain and the peptidyl carrier protein (PCP), canonically arranged as C-A-PCP, whereas the initiation module usually lacks the C domain (Finking & Marahiel, 2004), though exceptions are known (see below).

The A domain is responsible for the selection and activation of the substrate. Herein, it is not restricted to the 20 proteinogenic amino acids like in ribosomal peptide synthesis, but can choose from a diverse pool of building blocks, including a number of non-proteinogenic amino acids like ornithine, hydroxy- or dihydroxyphenylglycine as well as carboxylic acids like aryl acids. Today, more than 500 monomers have

been identified as parts of NRPs (Strieker et al., 2010). The carboxyl group of the cognate substrate is activated as aminoacyl-adenylate at the expense of ATP, which is then tethered to the 4'-phosphopantetheinyl arm of the PCP as thioester. Like in PKS, this requires the priming of the *apo*-PCP with its prosthetic group by the action of PPTases to turn it into its active *holo*-form (Koglin & Walsh, 2009). The basis for the specificity of A domains was investigated by analysis of the substrate binding pocket in the crystal structure of the phenylalanine activating domain PheA from the first module of the gramicidin synthetase (Conti et al., 1997). Sequence comparison with other A domains led to the identification of a set of ten (Stachelhaus et al., 1999) or eight (Townsend et al., 2000) residues that confer substrate specificity. The discovery of this nonribosomal code gave rise to the possibility of altering the specificity by the introduction of point mutations (Eppelmann et al., 2002).

The C domain mediates the chain elongation by catalysing the formation of a peptide bond between the upstream peptidyl-S-PCP and the downstream aminoacyl-S-PCP concomitantly with a translocation of the chain intermediate to the downstream PCP.

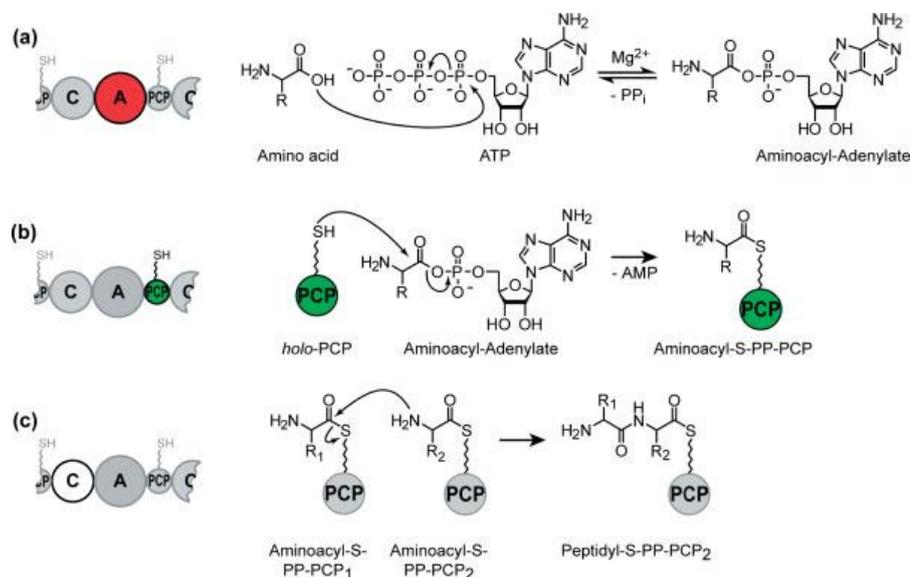


Figure 1-8: Basic reactions in NRP biosynthesis

(a) Substrate activation by the adenylation domain (A) ; (b) loading of the carrier protein (PCP); (c) amide bond formation between two building blocks by the condensation domain (C) (from Marahiel, 2009)

Like in PKS, the peptide chain is released from the synthetase by the action of a TE domain in the termination module. Intramolecular nucleophilic attack of the acyl-O-TE intermediate leads to the formation of a macrocyclic product, as observed e.g. in tyrocidin and surfactin biosynthesis, while simple hydrolysis results in the release of a linear peptide, as observed in the formation of pyochelin and yersiniabactin. Alternative mechanisms have been observed in some fungal PKs like saframycin and cyclosporine, in which chain release occurs through reduction or specialised C domains (Keating et al., 2001).

1.3.1 Editing domains and domain subtypes in NRPS

To the present date, a total of 10 different NRPS domains are known (Sieber & Marahiel, 2005). Besides the already described minimal set of domains, they include enzymes functionalities that introduce modifications to the peptide backbone. Furthermore, a number of functional subtypes of C domains have been identified (Rausch et al., 2007).

Some clusters like the surfactin, lichenysin or fengycin encode a specialised starter C domain in their first module which acylates the following amino acid with a β -hydroxy-carboxylic acid. The C domain of a module can be replaced by a specialised heterocyclisation (Cyc) domain, which not only catalyses peptide bond formation, but also cyclisation of threonine, serine or cysteine residues, resulting in the formation of thiazoline or oxazoline rings, respectively. In NRPs like bleomycin or myxothiazol additional oxidation (Ox) domains convert these heterocycles into more stable thiazole and oxazole rings. These heterocycles are important for the interaction with proteins or nucleic acids and the chelating of metal ions. Reduction (R) domains catalyse the opposite reaction and can also be involved in peptide release. Further functional groups can be introduced by the action of formylation (F) and N-methylation (Mt) domains.

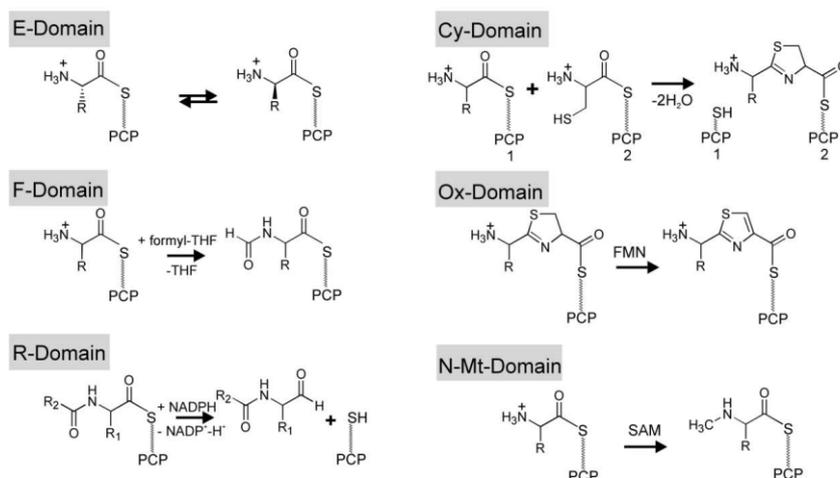


Figure 1-9: Reactions performed by optional domains in NRP biosynthesis

E, epimerisation; **Cy**, cyclisation; **F**, formylation; **Ox**, oxidation; **R**, racemisation; **N-Mt**, N-methyltransferase (from Schwarzer et al., 2003)

A distinct feature of NRPs is the frequent use of D-amino acids. One way of incorporation is the selection by specific A domains. More commonly, L-amino acids are racemised by epimerisation (E) domains. The selectivity for the correct enantiomer is conferred by the C domain that follows directly after the E domain. These ^DC_L domains that catalyse the peptide formation of an upstream D-amino acid with a downstream L-amino acid can be differentiated by phylogeny and by sequence motifs from the subgroup of ^LC_L domains, which connect two L-amino acids. A class of dual E/C domains is able to perform both the racemisation and the condensation reaction. Like in PKS systems, post-assembly tailoring enzymes can introduce further modifications to the non-ribosomal peptide backbone including oxidation, halogenation, methylation, hydroxylation or glycosylation (Koglin & Walsh, 2009).

1.3.2 Classification of NRPS types

By chance, all of the first NRPSs that were discovered and analysed belonged to the class of linear NRPS, which are analogous to the non-iterative type I PKS and whose biosynthetic logic follows the colinearity rule. The synthetases of tyrocidine, actinomycin, balhimycin and bacitracin belong to this group. What was first regarded as exceptions from this paradigm eventually emerged as new classes of NRPS. The three classes known today are denominated linear, iterative and nonlinear NRPS or types A, B and C (Mootz & Marahiel, 2002).

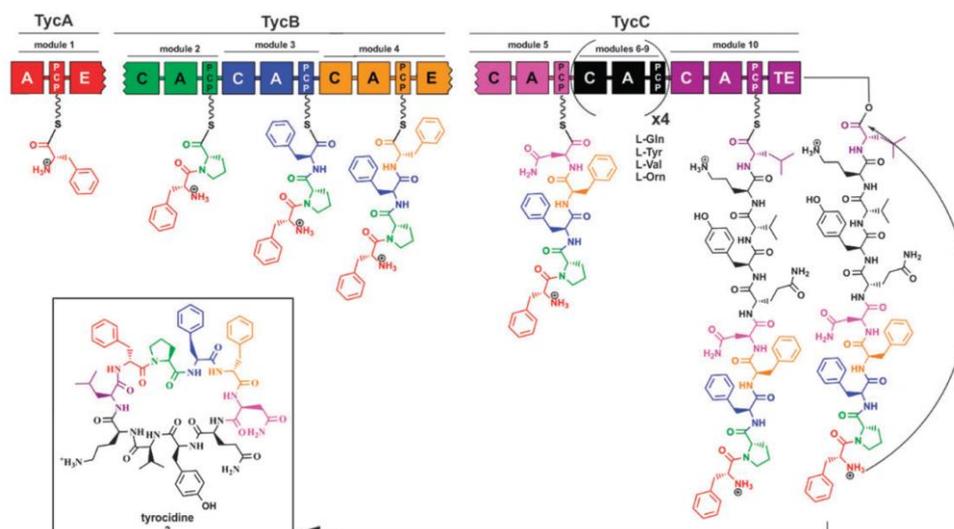


Figure 1-10: Tyrocidine biosynthetic pathway

Example of a linear, non-iterative NRPS system (from Meier & Burkhardt, 2009); A, adenylation; C, condensation; E, epimerization, PCP, peptidyl carrier protein

In the iterative type, the modules are used more than once to produce multimeric compounds. The monomer produced by one round of synthesis is stalled on the C-terminal TE domain, thus regenerating the NRPS for the next chain assembly. The monomers are oligomerised on the TE and finally released usually as cyclised product. This class of NRPS was found responsible for the biosynthesis of e.g. gramicidin S and enterobactin.

The class of nonlinear NRPSs contains among others the biosynthetic pathways for vibriobactin, mycobactin and yersiniabactin. They are characterised by deviations from the canonical domain organisation in at least one module. The cysteine-specific A domain in yersiniabactin biosynthesis for example is responsible for the loading of three different PCPs (Suo et al., 2001). The vibriobactin cluster encodes an unusual tandem arrangement of Cyc domains, in which one is responsible for the cyclisation and the other for the peptide bond formation (Marshall & Walsh, 2002).

1.3.3 Hybrid PKS-NRPS

Apart from metabolites, which are exclusively the product of either a PKS or a NRPS, the existence of hybrid PKS/NRPS assembly lines has also been discovered (Cane & Walsh, 1999). Compounds like rapamycin, yersiniabactin, bleomycin and epothilone were found to be synthesised by hybrid systems. They seem to be especially

abundant in myxobacteria, from which e.g. compounds like myxothiazole and melithiazole originate, which are potent inhibitors of the respiratory chain (Weinig et al., 2003).

The interaction of NRPS and PKS systems is enabled by their catalytic and structural similarities as well as their analogous modular architecture (Du et al., 2000). PKS/NRPS hybrids follow the same logic as the already described modular systems, so that the same domain prediction systems can be applied (Fischbach & Walsh, 2006). The question how these assembly lines work at the PKS-NRPS interface has already been addressed by several groups (Koglin & Walsh, 2009). It was found that on the protein level PKSs and NRPSs recognise their correct partners through short recognition sequences at the protein termini, which were named communication-mediating domains (COM) for NRPSs (Hahn & Stachelhaus, 2004) and interpolypeptide linkers or docking domains in the context of PKSs (Gokhale & Khosla, 2000; Broadhurst et al., 2003).

The modular nature of PKS and NRPS assembly lines has inspired the development of “unnatural” natural products (Cane & Walsh, 1999), i.e. modification of product structures through combinatorial biosynthesis. By applying a strategy of “mix and match” (Richter et al., 2008) domains and modules can be modified, deleted or replaced by subunits from other biosynthetic clusters, as has been done intensively for the biosynthesis of epothilone (O’Connor et al., 2003) and erythromycin (McDaniel et al., 1999).

1.4 The phylum *Chloroflexi*

The phylum Chloroflexi is one of the deepest branching lineages of the bacteria and the first evolutionary branch with phototrophic members (Garrity & Holt, 2001). Organisms of this group are characterised by filamentous growth and gliding motility. They stain Gram-negatively, but lack a lipopolysaccharide-containing outer membrane and are therefore proposed to be typically monoderm (Sutcliffe, 2010). The peptidoglycan of their cell wall contains L-ornithine instead of meso-diaminopimelic acid, a characteristic usually attributed to Gram-positive bacteria (Jürgens et al., 1989). The phylum comprises the orders *Chloroflexales* and *Herpetosiphonales*.

1.4.2 *Herpetosiphonales*

The order *Herpetosiphonales* is represented by the family of *Herpetosiphonaceae* with the two species *H. aurantiacus* and *H. geysericola*. Three marine species that were originally placed in this group were moved to the genus *Lewinella* in the phylum *Bacteroides* after analysis of 16S rDNA (Sly et al, 1998). *Herpetosiphon* species contain no bacteriochlorophyll and are supposed to have evolved from *Chloroflexaceae* by loss of the photosystem. They are aerobic chemoheterotrophs with O₂ as terminal acceptor in respiration and a moderate G+C content of ~ 48-53%. Their colonies are of yellow to orange colour due to the presence of carotenoids (Kleinig & Reichenbach, 1977). *Herpetosiphonaceae* may be isolated from various environments like soil, freshwater, decaying organic matter and the activated sludge of sewage plants (Trick & Lingens, 1984), where they are suspected to be involved in bulking of the sludge (Björnsson et al, 2002).

Holt & Lewin first described the type strain *Herpetosiphon aurantiacus* (Holt & Lewin, 1968). The cells are rod-shaped with a diameter of 1.0-1.5 x 5-10 µm, which grow as flexible, unbranched filaments reaching lengths from 100 µm to more than 1200 µm. At the ends of these filaments the frequent occurrence of characteristic sleeves has been observed (Reichenbach & Golecki, 1975). There is a controversial discussion, whether these structures are part of a sheath surrounding the whole filament (Skerman et al., 1977) or if they are the remains of necridial cells. The fact that the filaments exhibit gliding motion as a whole and electron microscopy support the absence of a sheath-like structure (Lee & Reichenbach, 2006). Some *Herpetosiphon* strains were shown to exert bacteriolytic activity against some Gram-positive and negative bacteria (Quinn & Skerman, 1980).

The 16S rDNA of *Herpetosiphon* exhibits idiosyncrasies in two highly conserved sequence regions (Oyaizu et al., 1987), which are regarded as an indication for a rapidly evolving line of bacteria (Woese et al., 1985). The complete genome was sequenced by the Joint Genome Institute (GenBank accession: NC_009972) and comprises 6.35 Mbp and is thus larger than the genome of *Chloroflexus aurantiacus* and other phototrophic Chloroflexaceae (Tang et al., 2011; GenBank accession: NC_010175). Genome mining discovered homologues to the lanM gene and suggests *H. aurantiacus* as a potential producer of lantibiotics (Begley et al, 2009).

1.4.3 *Herpetosiphon* sp. 060

The siphonazole producing strain *Herpetosiphon* sp. 060 was isolated from a mud sample of the intertidal region (Nett, 2007). The strain shows the characteristic physiological features of the *Herpetosiphonaceae* including gliding motility and filamentous growth. On a dish culture, swarming filaments are clearly visible at the growing edge as translucent threads. Repeatedly, orange knob-like colonies are formed by the accumulation of intertwined filaments. Cultivation on VY/2 agar is concomitant with a clearing of the medium. The optimal growth temperature lies around 30°C with peptone as preferred carbon source. The strain exhibits salt tolerance only up to 1% NaCl, but strong bacteriolytic activity against *E. coli*. Sequencing of 16S rDNA revealed 99% identity to an uncultured *Herpetosiphon* species, which lead to the classification of strain 060 as a member of the genus *Herpetosiphon*.

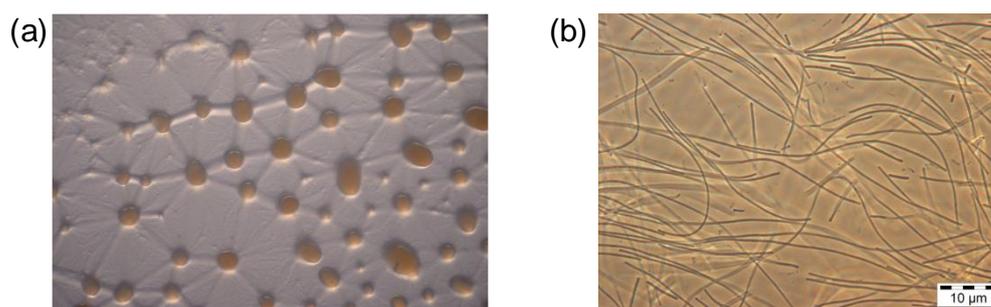


Figure 1-12: Images of *Herpetosiphon* sp. 060

(a) growth on VY/2 agar; orange protuberances and swarming streaks are clearly visible (from Nett, 2007); (b) phase-contrast micrograph (x1000) of *Herpetosiphon* sp. 060 filaments

1.5 Siphonazole – a secondary metabolite from *Herpetosiphon*

Siphonazole was discovered during a metabolite screening of ten different *Herpetosiphon* strains. The complete structure elucidation revealed an unusual composition of a styrene moiety, two oxazole rings connected by a C₂-bridge and a hitherto unprecedented *N*-penta[2,4]dien side chain (Nett et al., 2006). Feeding studies with ¹³C-labeled precursors pointed at hydroxybenzoic acid derived from chorismic acid via the shikimate pathway as precursor. Further labelling studies proposed threonine as origin of the oxazole rings. The amide bond of the pentadien chain involves the incorporation of a glycine moiety, which is then elongated by two acetate units. It was deduced from the labelling pattern that the terminal acetate

moiety undergoes and unprecedented dehydrative decarboxylation. Besides the main compound, a derivative of siphonazole was isolated in small quantities, in which one of the hydroxyl groups of the styrene ring is also methylated.

Due to lack of phenylalanine ammonia lyase in prokaryotes (Bode & Müller, 2003), only few metabolites are described that contain derivatives of cinnamic acid together with polyketide structures and incorporated amino acids. Other examples were isolated from myxobacteria and include crocacins and phenoxan (Jansen et al., 1999; Jansen et al., 1991), of which only the latter exhibits some similarity to siphonazole from a biosynthetic point of view. Though the presence of two oxazole rings is common in natural products, most of them are either linked directly or separated by at least three carbon atoms (Lindner et al, 2008), while the occurrence of a bridge with only one or two carbons is rare.

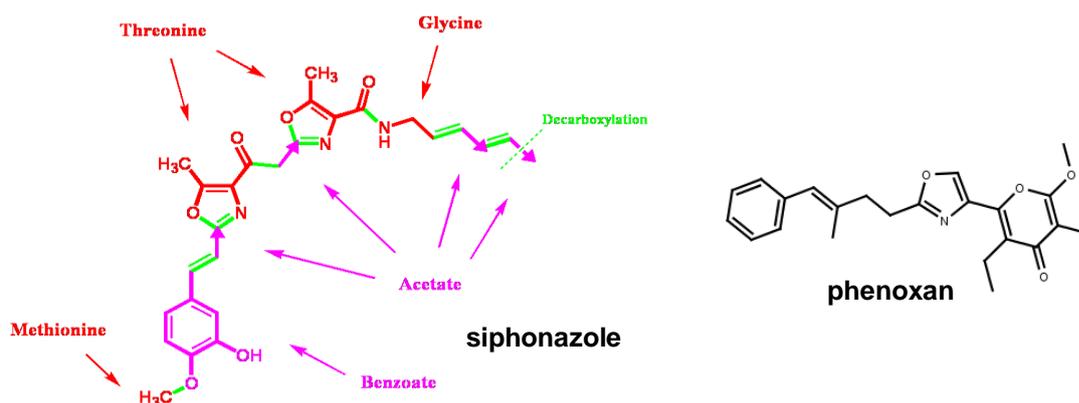


Figure 1-13: Molecular structure of siphonazole and phenoxan

Origin of building blocks are marked according to feeding experiments; amino acid derived parts are coloured red; acetate moieties are shown as purple arrows (from Nett, 2007); phenoxan produced by *Polyangium* sp.

Siphonazole was tested for antibiotic properties, but showed no activity against selected bacteria, fungi and viruses. Also, the cytotoxic effect against 36 cancer cell lines was too weak for a use as anticancer agent. Siphonazole was only shown to have a moderate activity (IC_{50} : 0.59 μ g/ml; 1.27 μ M) against the *Plasmodium falciparum*. A total synthesis of siphonazole and methoxysiphonazole was achieved by the groups of Moody (Lindner et al., 2008) and Ciufolini (Zhang et al., 2009).

2 Scope of the Study

The Herpetosiphonales are an ancient eubacterial group, whose metabolic potential has not been probed so far. To the present date, the compound siphonazole, isolated from *Herpetosiphon* sp. 060, is the only secondary metabolite from this group that has been characterised. The molecule shows several interesting structural features, including a styrene ring, two oxazole rings and an unusual diene terminus. Although the structure hinted at a mixed PKS/NRPS pathway, the biosynthetic enzymes for siphonazole were unknown to the date of the study.

Therefore, the primary intend of this study was the elucidation of the siphonazole biosynthetic gene cluster. In previous works a fosmid library of *Herpetosiphon* sp. 060 has already been established and screened for PKS and NRPS genes. Thereby, several fosmids carrying putative parts of a siphonazole gene cluster have been identified. The obtained sequences were taken as a base for gene knock-out experiments, which aimed at the disruption of the putative cluster. Together with an observation of siphonazole production by LC-MS, a knock-out would allow the unambiguous identification of the biosynthetic enzymes, which can then be subjected to a detailed sequence analysis. This approach necessitated the development of suitable methods for the gene inactivation in *Herpetosiphon* sp. 060 comprising the construction of an integrative vector and first and foremost the establishment of a protocol for the transformation of *Herpetosiphon* cells.

As an alternative approach the genome of *Herpetosiphon* sp. 060 was sequenced to allow the localisation of the identified fosmid inserts and characterisation of the encoded enzymes. Hence, an analysis by bioinformatic methods was included in this study to bring further insights about the encoded enzymes and allow the identification of putative parts of the siphonazole biosynthesis. The obtained data were to be used as a basis for a PCR-guided completion and elucidation of the cluster. Furthermore, the sequences were utilised to probe the genome for further secondary metabolite gene clusters.

After identification of the siphonazole gene cluster, an ensuing part of this study targeted the reconstitution of the complete cluster from fragments identified in the fosmid library. This part comprised the attainment of fosmids covering the complete

cluster range and the application of suitable methods for the recombination of large DNA fragments to a vector carrying the complete functional gene cluster. These working steps aimed at the construction of a vector that can be further modified for an expression in a heterologous host and for ensuing experiments on single domain functions.

The last part of the study addressed a first characterisation of the cluster and individual biosynthetic enzymes. The temporal expression of siphonazole was investigated by analysis of the mRNA expression of the putative genes, while the spatial localisation of the compound was observed by imaging mass spectrometry. Single proteins and domains of the putative biosynthetic enzymes were selected for a heterologous expression in *E. coli*. Purified proteins should be subjected to *in vitro* assays for an investigation of their specific functions. For this purpose, suitable expression vectors, as well as cultivation and purification conditions had to be found.

3 Materials & Methods

3.1 Materials

All chemicals were purchased in research grade or *pro analysi* quality. Ultra pure water was taken from a Milli-Q® Water System and heat sterilised. For media composition deionised water was used.

3.1.1 Chemicals

Reagent	Manufacturer
Acetic acid	KMF Laborchemie Handels GmbH (Darmstadt, Germany)
Aceton-d 99.8% D	Deutero GmbH (Kastellaun, Germany)
Acetonitrile	VWR International GmbH (Darmstadt, Germany)
Adenine hemisulphate	Sigma-Aldrich Co. LLC (St. Louis, MO, USA)
Agar	Fluka Chemie GmbH (Buchs, Switzerland)
Amberlite® XAD16	Sigma-Aldrich Co. LLC (St. Louis, MO, USA)
Ammonium acetate	Roth Chemie GmbH (Karlsruhe, Germany)
Bacto™ Casitone	Becton, Dickinson and Company (Franklin Lakes, NJ, USA)
Bis-acrylamide	Roth Chemie GmbH (Karlsruhe, Germany)
Boric acid	Roth Chemie GmbH (Karlsruhe, Germany)
CaCl ₂ × 2 H ₂ O	Merck KGaA (Darmstadt, Germany)
Chloroform	Roth Chemie GmbH (Karlsruhe, Germany)
Brilliant Blau R 250	Roth Chemie GmbH (Karlsruhe, Germany)
CopyControl™	Epicentre (Madison, USA)
Induction solution	
dNTPs	Promega GmbH (Mannheim, Germany)
Drop-out Mix Synthetic	Biomol GmbH (Hamburg, Germany)
Minus Uracil w/o	
Yeast Nitrogen Base	
Ethanol 99,8% p.a.	Roth Chemie GmbH (Karlsruhe, Germany)
Ethidium bromide	Roth Chemie GmbH (Karlsruhe, Germany)
Ethyl acetate	Julius Hoesch GmbH & Co. KG (Düren, Germany)
Glycerol	Roth Chemie GmbH (Karlsruhe, Germany)
Gel loading dye (6x)	Fermentas GmbH (St. Leon Rot, Germany)

Imidazole	Roth Chemie GmbH (Karlsruhe, Germany)
Isopropanol	Roth Chemie GmbH (Karlsruhe, Germany)
Isovanillic acid	Alfa Aesar (Ward Hill, MA, USA)
Ni-NTA agarose	Qiagen GmbH (Hilden, Germany)
Methanol	Merck KGaA (Darmstadt, Germany)
MgCl ₂ × 6 H ₂ O	Merck KGaA (Darmstadt, Germany)
MgSO ₄ × 7 H ₂ O	Merck KGaA (Darmstadt, Germany)
Na-acetate	Merck KGaA (Darmstadt, Germany)
NaCl	Merck KGaA (Darmstadt, Germany)
Na ₂ -EDTA	Roth Chemie GmbH (Karlsruhe, Germany)
NaOH	Merck KGaA (Darmstadt, Germany)
peqGOLD Agarose	PEQLAB Biotechnologie GMBH (Erlangen, Germany)
Phenol	Merck KGaA (Darmstadt, Germany)
Protocatechuic acid	Roth Chemie GmbH (Karlsruhe, Germany)
SDS	Roth Chemie GmbH (Karlsruhe, Germany)
Sorbitol	Roth Chemie GmbH (Karlsruhe, Germany)
Trifluoroacetic acid	Roth Chemie GmbH (Karlsruhe, Germany)
Tris	Roth Chemie GmbH (Karlsruhe, Germany)
Tris-HCl	Roth Chemie GmbH (Karlsruhe, Germany)
Tryptone	Roth Chemie GmbH (Karlsruhe, Germany)
Yeast extract	Fluka Chemie GmbH (Buchs, Switzerland)
Yeast Nitrogen Base w/o amino acids, carbohydrate & w/AS	Biomol GmbH (Hamburg, Germany)

3.1.2 Devices

Article	Manufacturer
Autoclave	Varioklav®, H+P Labortechnik AG (Oberschleißheim, Germany)
Bandelin Sonorex RK 31	Bandelin electronic GmbH & Co. KG (Berlin, Germany)
Biometra T3000 Thermocycler	Biometra GmbH (Goettingen, Germany)
Bio-Rad Micropulser™	Bio-Rad Laboratories GmbH (Hercules, USA)

BioRad PowerPac™ Basic	Bio-Rad Laboratories GmbH (Hercules, USA)
Branson Sonifier 250	G. Heinemann Ultraschall- und Labortechnik (Schwäbisch Gmünd, Germany)
Centrifuge Heraeus Biofuge <i>fresco</i>	Thermo Fisher Scientific (Waltham, USA)
Centrifuge Heraeus Contifuge Stratos	Thermo Fisher Scientific (Waltham, USA)
Centrifuge Heraeus Fresco 17	Thermo Fisher Scientific (Waltham, USA)
Gel chamber Horizon 58	Life technologies (Karlsruhe, Germany)
Gel chamber Horizon 11.14	Life technologies (Karlsruhe, Germany)
Incubator	Memmert GmbH + Co. KG (Schwalbach, Germany)
Inolab pH meter	WTW GmbH (Weilheim, Germany)
Intas iX Imager	Intas Science Imaging Instruments GmbH (Göttingen, Germany)
Laminar Airflow Clean Bench BSB 4A (Hera Safe, Class II)	Heraeus (Hanau, Germany)
Magnetic stirrer (IKA® RH basic)	IKA® Werke GmbH & Co. KG (Staufen, Germany)
Milli-Q® Water System	Millipore (Eschborn, Germany)
Multitron incubation shaker	Infors AG (Bottmingen, Switzerland)
Olympus BX 51	Olympus America Inc. (Center Valley, USA)
Scale (Sartorius BL 3100)	Sartorius AG (Göttingen, Germany)
Scale (Sartorius BP 221S)	Sartorius AG (Göttingen, Germany)
Thermomixer Eppendorf	Eppendorf (Hamburg, Germany)
UV mini 1240 UV/Vis spectro- photometer	Shimadzu (Kyoto, Japan)
UV cuvettes	Ratiolab GmbH (Dreieich, Germany)
Water bath (Haake DC 10)	Thermo Haake GmbH (Karlsruhe, Germany)

3.1.3 Antibiotics

Reagent	Manufacturer
Ampicillin	Roth Chemie GmbH (Karlsruhe, Germany)
Apramycin	Sigma-Aldrich Co. LLC (St. Louis, MO, USA)
Carbenicillin	Roth Chemie GmbH (Karlsruhe, Germany)
Chloramphenicol	Fluka Chemie GmbH (Buchs, Switzerland)
Gentamycin	Fluka Chemie GmbH (Buchs, Switzerland)
Kanamycin	Sigma-Aldrich Co. LLC (St. Louis, MO, USA)
Nalidixinic acid	Roth Chemie GmbH (Karlsruhe, Germany)
Streptomycin	Sigma-Aldrich Co. LLC (St. Louis, MO, USA)
Tetracyclin	Fluka Chemie GmbH (Buchs, Switzerland)
Zeocin	Invitrogen Life Technologies Co. (Karlsruhe, Germany)

3.1.4 Disposable material

Article	Manufacturer
Amicon Ultra Centrifugal Filters	Millipore GmbH (Schwalbach, Germany)
Centrifuge tubes (15/50 ml)	TPP AG (Trasadingen, Germany)
Eppendorf tubes (0.5, 1.5, 2 ml)	Eppendorf (Hamburg, Germany)
PD-10 Desalting Columns	GE Healthcare (München, Germany)
Polypropylene columns (1ml)	Qiagen GmbH (Hilden, Germany)
Parafilm [®]	Pechiney Plastic Packaging Company (Chicago, USA)
Sterile filter (0.2 µm)	Renner GmbH (Dannstadt, Germany)

3.1.5 Buffers and stock solutions

Buffer	Composition
Buffer P1	50 mM Tris-HCl (pH 8), 10 mM EDTA, 100 µg/mL RNase A
Buffer P2	200 mM NaOH, 1% SDS
Buffer P3	3M potassium acetate (pH 5.5)
TE-buffer	10 mM Tris-HCl (pH 8.0), 1 mM EDTA
10x TBE-buffer	0.89 M Tris base, 0.02 M EDTA, 0.87 M H ₃ BO ₃ , H ₂ O ad 1 l

10x glycine SDS electrophoresis buffer	250 mM Tris, 2M glycine, 1% SDS, pH 8.9
A Domain buffer (Phelan et al., 2009)	20 mM Tris-HCl (pH 7.5), 5% glycerol, 1mM DTT
Trace element solution	20 mg ZnCl ₂ , 100 mg MnCl ₂ x 4 H ₂ O, 10 mg H ₃ BO ₃ , 10 mg CuSO ₄ , 20 mg CoCl ₂ , 5 mg SnCl ₂ x 2 H ₂ O, 5 mg LiCl, 20 mg KBr, 20 mg KI, 10 mg Na ₂ MoO ₄ x 2 H ₂ O, 5.2 g Na ₂ -EDTA x 2 H ₂ O, H ₂ O ad 1 l; filter sterilise
P-buffer (mod. after Okanishi et al., 1974; Hopwood & Wright, 1978)	0.3M sorbitol, 1mM MgCl ₂ x 6H ₂ O, 2 ml trace element solution, H ₂ O ad 800 ml; after autoclaving add the following sterile solutions: 10 ml KH ₂ PO ₄ (0.5%), 100 ml CaCl ₂ x 2 H ₂ O (3.68%), 100 ml TE-buffer (pH 7.2)
T-buffer (mod. after Thompson et al., 1982)	25% PEG 1000, 50 mM Tris (pH 8), 75 mM sorbitol, 100 mM CaCl ₂ x 2 H ₂ O; filter sterilise, do not autoclave
Protein lysis buffer	50 mM NaH ₂ PO ₄ , 300 mM NaCl, 10 mM imidazole, pH 8.0
Protein wash buffer	50 mM NaH ₂ PO ₄ , 300 mM NaCl, 20 mM imidazole, pH 8.0; for more stringent washing 40 mM imidazole were applied
Protein elution buffer	50 mM NaH ₂ PO ₄ , 300 mM NaCl, pH 8.0; for gradual elution buffers with 100, 150, 200 and 300 mM imidazole were used

3.1.6 Media

Medium	Composition
LB medium	10 g tryptone, 5 g yeast extract, 10 g NaCl, H ₂ O ad 1 l, pH 7.5
LB agar	10 g tryptone, 5 g yeast extract, 5 g NaCl, 15 g Agar, H ₂ O ad 1 l, pH 7.5
CY medium	3.0 g casiton, 1.36 g CaCl ₂ x 2 H ₂ O; 1.0 g yeast extract, H ₂ O ad 1 l, pH 7.2; after autoclaving add 1 ml filter sterilised trace element solution and 1 ml vitamin B12 solution (0.5 mg/ml)

VY/2 agar	50 ml sterilised baker's yeast suspension (10%), 1.36 g CaCl ₂ x 2 H ₂ O, 15 g agar, H ₂ O ad 1 l, pH 7.2; after autoclaving add 1 ml filter sterilised vitamin B12 solution (0.5 mg/ml)
Synthetic drop-out medium	6.7 g Yeast Nitrogen Base without amino acids, carbohydrate & with ammonium sulphate; 2 g Drop-out Mix Synthetic Minus Uracil without Yeast Nitrogen Base; 20 g agar; H ₂ O ad 900 ml; pH 5.6 – 5.8; after autoclaving add 100 ml 20% sterile glucose solution
SOB	20 g tryptone, 5g yeast extract; 0.5 g NaCl, 0.186 g KCl, H ₂ O ad 1 l, pH 7.5
SOC	20 g tryptone, 5g yeast extract; 0.5 g NaCl, 0.186 g KCl, H ₂ O ad 1 l, pH 7.5; after autoclaving add 20 ml 1M filter sterilised glucose solution
YPAD medium	10 g yeast extract, 20 g peptone, 20 g glucose, 40 mg adenine hemisulphate; H ₂ O ad 900 ml; after autoclaving add 100 ml 20% sterile glucose solution

3.1.7 Enzymes

Enzyme	Manufacturer
GoTaq® Flexi DNA Polymerase	Promega (Mannheim, Germany)
Lysozyme	Roth (Karlsruhe, Germany)
Proteinase K	Roth (Karlsruhe, Germany)
Restriction enzymes	Fermentas GmbH (St. Leon-Rot, Germany)
RNase (DNase free)	Roth Promega (Mannheim, Germany)
Fast-Link™ DNA Ligase	Epicentre (Madison, USA)
T4 DNA-ligase	Fermentas GmbH (St. Leon-Rot, Germany)
CIAP	Fermentas GmbH (St. Leon-Rot, Germany)
SAP	Fermentas GmbH (St. Leon-Rot, Germany)

3.1.8 Kits and standards

Article	Manufacturer
Gene Ruler™ DNA Ladder Mix	Fermentas GmbH (St. Leon-Rot, Germany)
Gene Ruler™ 1kb plus DNA ladder	Fermentas GmbH (St. Leon-Rot, Germany)
Gene Ruler™ 50bp DNA ladder	Fermentas GmbH (St. Leon-Rot, Germany)
Gene Ruler™ high range DNA ladder	Fermentas GmbH (St. Leon-Rot, Germany)
QIAquick PCR Purification Kit	Qiagen GmbH (Hilden, Germany)
QIAquick Gel Extraction Kit	Qiagen GmbH (Hilden, Germany)
QIAGEN Plasmid Midi Kit	Qiagen GmbH (Hilden, Germany)
QIAGEN Plasmid Maxi Kit	Qiagen GmbH (Hilden, Germany)
QIAprep® Spin Miniprep Kit	Qiagen GmbH (Hilden, Germany)
RevertAid™ H Minus First Strand cDNA Synthesis Kit	Fermentas GmbH (St. Leon-Rot, Germany)
Wizard® SV Gel and PCR Clean-Up System	Promega (Mannheim, Germany)
Wizard® Genomic DNA Purification Kit	Promega (Mannheim, Germany)
PureYield™ Plasmid Miniprep System	Promega (Mannheim, Germany)
DNA Clean & Concentrator™	Zymo Research Corporation (Irvine, USA)
Zymoclean™ Large Fragment DNA Recovery Kit	Zymo Research Corporation (Irvine, USA)
Zymoprep™ Yeast Plasmid Miniprep I Kit	Zymo Research Corporation (Irvine, USA)

3.1.9 Organisms

Organism	Genotype	Provider
<i>Herpetosiphon</i> sp. 060	wild type	own strain collection
XL1-Blue <i>E. coli</i> cells	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI _q ZΔM15 Tn10 (Tetr.)]	Stratagene (La Jolla, CA, USA)
BL21 (DE3) <i>E. coli</i> cells	F ⁻ ompT gal dcm lon hsdS _B (r _B ⁻ m _B ⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Invitrogen Life Technologies Corporation (Karlsruhe, Germany)
BAP1 <i>E. coli</i> cells	n.a.; sfp from <i>B. subtilis</i>	Pfeifer et al., 2001
ET12567 <i>E. coli</i> cells	Dam ⁻ , dcm ⁻ , hsdM ⁻	MacNeil et al., 1992

Phage T1-Resistant TransforMax™ EPI-300™-T1R chemically competent <i>E.coli</i> cells	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) f80Δ <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 endA1 araD139</i> D(<i>ara, leu</i>)7697 <i>galJ galK</i> λ- <i>rpsL nupG trfA tonA dhfr</i>	Epicentre Biotechnologies (Madison, USA)
BW25113 <i>E. coli</i> cells	F ⁻ , Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB-3</i>), λ-, <i>rph-1</i> , Δ(<i>rhaD-rhaB</i>)568, <i>hsdR</i> 514	Gust et al., 2003
BL21 star <i>E. coli</i> cells	F ⁻ <i>ompT hsdS_B</i> (rB ⁻ mB ⁻) <i>gal dcm rne131</i> (DE3)	Invitrogen Life Technologies Corporation (Karlsruhe, Germany)
TOP10 <i>E. coli</i> cells	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80Δ <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ(<i>ara-leu</i>)7697 <i>galJ galK rpsL</i> (Str ^R) <i>endA1 nupG</i>	Invitrogen Life Technologies Corporation (Karlsruhe, Germany)
ArcticExpress (DE3) competent <i>E. coli</i> cells	<i>E. coli</i> B F ⁻ <i>ompT hsdS</i> (rB ⁻ mB ⁻) <i>dcm⁺ Tetr gal λ</i> (DE3) <i>endA Hte</i> [<i>cpn10 cpn60 Gentr</i>] [<i>argU proL Strr</i>]	Agilent Technologies (Santa Clara, CA, USA)
<i>Saccharomyces cerevisiae</i>	n.a.; <i>ura3P</i> deficient	C. Gurgui (Piel group, University of Bonn)

3.1.10 Vectors

Vector	Resistance	Manufacturer
pET28a(+)	kanamycin	Merck KGaA (Darmstadt, Germany)
pIJ773	apramycin	Gust et al., 2003
pIJ778	streptomycin	Gust et al., 2003
pKD46	ampicillin	Datsenko et al., 2000
pCC1FOS™	chloramphenicol	Epicentre Biotechnologies (Madison, USA)
pET151/D-TOPO®	ampicillin	Invitrogen Life Technologies Corporation (Karlsruhe, Germany)
pENTRC1FOS	chloramphenicol	C. Gurgui (Piel group, University of Bonn)
pGEM-T	ampicillin	Promega (Mannheim, Germany)
pUC19	ampicillin	own collection

3.1.11 Clones & DNA-constructs

The following table lists all vectors and DNA-constructs used or created in this study.

Construct	Vector	Insert
EC9	pCC1FOS	Genomic DNA from <i>Herpetosiphon</i> sp. 060 with parts of the putative siphonazole gene cluster
EC10	pCC1FOS	Genomic DNA from <i>Herpetosiphon</i> sp. 060 with parts of the putative siphonazole gene cluster
CF4	pCC1FOS	Genomic DNA from <i>Herpetosiphon</i> sp. 060 with parts of the putative siphonazole gene cluster
IC2	pCC1FOS	Genomic DNA from <i>Herpetosiphon</i> sp. 060 with parts of the putative siphonazole gene cluster
EC9-2, -5, -7	pGEM-T	Subclones with fragments from fosmid EC9
EC10-1, -3, -5, -6, -7, -8, -9, -10	pGEM-T	Subclones with fragments from fosmid EC10
CF4-2, -4, -5, -6, -7, -8, -9, -10	pGEM-T	Subclones with fragments from fosmid CF4
SphB	pET28a(+)	O-MT SphB from <i>Herpetosiphon</i> sp. 060
SphI	pET151	DAHPSynthase SphI from <i>Herpetosiphon</i> sp. 060
SphC-A1	pET151	A domain from <i>Herpetosiphon</i> sp. 060
EC10-aadA	pCC1FOS	Fosmid EC10 with aadA resistance gene
IC2-apra	pCC1FOS	Fosmid IC2 with aac(3)IV resistance gene
EC9+IC2	pCC1FOS	Fosmid with inserts from EC9 and IC2-apra
Siphonazole	pCC1FOS	Fosmid with complete putative siphonazole cluster, recombined from EC9+IC2 and EC10-aadA

3.1.12 Primers

PCR-primers were purchased from Eurofins MWG Operon (Ebersberg, Germany) as desalted, lyophilised powder. The oligonucleotides were reconstituted in sterile water and adjusted to a concentration of 100 pmol/μl. Reconstituted primers were either short-term stored at 4°C or at -20°C for longer periods. A compilation of the nucleotide sequences is given in the appendix (8.3).

3.1.13 Software and Databases

Basic Local Alignment Search Tool (BLAST; <http://blast.ncbi.nlm.nih.gov/>) provided by the National Center for Biotechnology Information (NCBI) was used for multiple sequence alignments of protein primary sequences and alignments of nucleotide sequences. Protein homologies and conserved domains were analysed by application of BLASTp and the Conserved Domain Database (CDD; Marchler-Bauer, 2011).

The **Cluster Sequence Analyzer (CLUSEAN)** pipeline combines database research and specific tools for the identification of PKS and NRPS domains and conserved motifs (Weber et al., 2009). This pipeline was used for the analysis of draft genomic data and the annotation of the putative siphonazole biosynthetic gene cluster.

The NRSPredictor and NRSPredictor2 (Rausch et al., 2005; Röttig et al., 2011) software provided by the University of Tübingen (<http://ab.inf.uni-tuebingen.de/software/NRSPredictor/>) predict the substrate specificity of A domains by analysis of amino acid signature sequences (Stachelhaus et al., 1999). The software was used for the specificity prediction of identified adenylation domains.

ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) of the European Bioinformatics Institute (EBI) was applied to generate multiple alignments of amino acid or nucleotide sequences (Larkin et al., 2007; Goujon et al., 2010). Reference protein sequences were derived from NCBI databases (<http://www.ncbi.nlm.nih.gov/>).

DNASTAR Lasergene® from DNASTAR Inc. (Madison, WI, USA) comprises number of sequence assembly and analysis tools. The SeqMan tool was used for the assembly of data obtained by 454 sequencing.

The **Protparam** tool (<http://web.expasy.org/protparam/>) on the ExPASy server of the Swiss Institute for Bioinformatics (SIB) was employed to calculate the physico-chemical parameters of recombinant proteins like pI, molecular weight and predicted stability and solubility (Gasteiger et al., 2005).

The **Artemis Genome Browser and Annotation Tool** (Rutherford et al., 2000; Carver et al., 2011) by the Sanger Institute was utilised for the annotation of open reading frames, analysis of GC-content and visualisation of bioinformatic data obtained by CLUSEAN. Analysis of codon bias and calculation of the Codon Adaptation Index (CAI) was carried out by application of the **Jcat** tool (Grote et al., 2005; <http://www.jcat.de/>).

Design of PCR-primers and planning of cloning experiments were performed with the **CloneManager 9** software by Sci-Ed Software. Restriction site analysis was accomplished with the **NEBcutter V2.0** (<http://tools.neb.com/NEBcutter2/>) by New England Biolabs (Vincze et al., 2003).

The **NaPDos** tool (<http://napdos.ucsd.edu/>) detects and extracts A and KS domains from nucleotide sequence data. Candidate genes are processed and subjected to database comparisons with known secondary metabolite gene clusters. This tool was utilised to search the draft genome of *Herpetosiphon* sp. 060 for parts of PKS and NRPS gene clusters (Ziemert et al., 2011).

3.2 General methods concerning micro-organisms

3.2.1 Sterilisation

Materials and solutions used for microbiological work were sterilised by autoclaving for 20 min at 121°C and 2 bars pressure in a Varioklav steam steriliser. Heat sensitive solutions were sterilised by filtration through a 0.22 µm disposable filter cap instead.

3.2.2 Storage of organisms

For long term preservation of micro-organisms cryogenic cultures were prepared. 0.75 ml of a fully grown liquid culture was mixed with the same amount of sterile glycerol in a cryogenic vial. The culture was thoroughly mixed by inverting the tube and stored at -80°C.

3.2.3 Cultivation of bacterial organisms

All working steps with bacteria were performed on a laminar air flow clean bench under sterile conditions. All working materials, solutions and tools were either autoclaved, sterile-filtered or heat sterilised in a Bunsen burner flame. For cultivation, always autoclaved media were used.

Small *E. coli* cultures were incubated in either 10 ml glass tubes or 2 ml Eppendorf tubes. Larger cultivation was performed in Erlenmeyer flasks. For the cultivation on agar plates, 100 – 250 µl of a liquid culture were spread on the agar and left under the sterile bench with the lid open, until the liquid was absorbed by the medium. *E. coli* cultures were incubated at 37°C, if not stated otherwise. Liquid cultures of *Herpetosiphon* were always incubated in Erlenmeyer flasks on a horizontal shaker at 30°C and 140 rpm. For cultivation, VY/2 medium supplemented with Vitamin B12 and trace elements was inoculated with blocks from a Petri dish culture of *Herpetosiphon*. Solid cultures were grown on CY-agar at 30°C.

3.2.4 Preparation of a crude extract from *Herpetosiphon* sp. 060

The production of siphonazole was observed by detecting its mass in a crude extract by HPLC-LCMS (3.6.9). Therefore, 1.5L VY/2 medium with 2% (m/v) Amberlite XAD16 adsorber resin were prepared in a 5L flask and inoculated with 100 ml of a pre-culture. The main culture was incubated for seven days to allow extensive production of siphonazole and adsorption of metabolites by the Amberlite. Following, cells and resin were pelleted by centrifugation at 4,000 rpm for 15 min. The liquid was decanted the residue transferred to a suction strainer. Compounds absorbed by the resin and from adherent cells were eluted by extracting three times with 100 ml acetone. The extract was dried at 40°C in a vacuum rotary evaporator and resolved in 250 ml 60% aqueous methanol. Following three times extraction with 100 ml dichloromethane the extract was concentrated and finally solved in 25 ml ethyl acetate. This fraction was dried again and stored until measurement at -20°C.

3.2.5 Test for antibiotic resistance of *Herpetosiphon* sp. 060

Herpetosiphon strain 060 was tested for natural resistance against different antibiotics. VY/2 agar plates each containing one antibiotic agent were inoculated with 50 and 100 µl of a liquid culture and incubated for three days at 30°C. Growth on the plates was observed and the grade of inhibition qualitatively classified.

3.2.6 Imaging mass spectrometry (IMS)

All experiments on IMS with *Herpetosiphon* sp. 060 were conducted by Dr. Till Schäberle in the Dorrestein Lab at the University of San Diego (UCSD, California, USA).

This MALDI-TOF (Matrix Assisted Laser Desorption/Ionisation – Time Of Flight) based mass spectrometry method developed by Esquenazi and co-workers (Esquenazi et al., 2008) allows the detection of secondary metabolites and the visualisation of their spatial distribution. For this purpose, cultures are directly grown on MALDI target plates and sprayed with a matrix, which allows immediate MALDI imaging. The detected masses can then be evaluated with suitable software, where they can be displayed in individual colours and are available for analysis and comparison with known compounds.

3.3 Transformation of bacterial cells

Transformation of bacterial cells describes the introduction of foreign DNA into the cells. Available methods include transformation by heat shock, electroporation and sonication (Yoshida & Sato, 2009). In most cases, the preparation of competent cells is necessary prior to the actual transformation, as most bacteria do not naturally take up foreign DNA. The DNA transfer by conjugation exploits a naturally occurring mechanism of horizontal gene transfer.

3.3.1 Preparation of chemically competent cells

Chemically competent *E. coli* cells were always prepared in the same way independent of the used strain. A 3 ml LB culture was inoculated with a single colony and incubated overnight at 180°C and 180 rpm. With this pre-culture 70 ml of 2xYT-medium were inoculated. The culture was grown to an OD₆₀₀ of 0.3 – 0.4 and the cells harvested by centrifugation for 10 min at 8,000 rpm and 4°C. To confer competence, the cells were resuspended in 10 ml ice cold CaCl₂/MgSO₄-solution (70 mM CaCl₂/20 mM MgSO₄) and incubated on ice for 30 min. The cells were harvested and incubated again in 3.5 ml of the CaCl₂/MgSO₄-solution. After the addition of 875 µl sterilised glycerol, the cells were stored as 100 µl aliquots at -80°C until usage.

3.3.2 Preparation of electro-competent cells

Electro-competent *E.coli* cells were always freshly prepared and used on the same day. The method follows a modified protocol after Gust et al. (2003). All centrifugation steps were carried out at 4°C. The cells were strictly kept on ice all the time.

100 ml SOB medium were inoculated with 1 ml of a 3 ml-overnight culture and grown to an OD₆₀₀ of about 0.5. The culture was divided into two sterile 50 ml falcon-tubes and the cells harvested by 5 min centrifugation at 6,000 rpm. The supernatant was discarded and the cells washed once with 10 ml and twice with 5 ml sterile ice cold 10% glycerol. Each washing step was followed by 5 min centrifugation at 6,000 rpm. After the last step the supernatant was removed and the cells suspended in 200 µl ice cold 10% glycerol. The suspension was aliquotted à 100 µl and kept on ice until use.

3.3.3 Transformation of chemically competent cells

For protein expression and *in vivo* amplification of plasmids, DNA was transformed into chemically competent cells (3.3.1) by heat shock. An aliquot of competent cells was thawed on ice and mixed with 5 μ l of purified plasmid or a ligation reaction (3.4.8). After incubation on ice for 30 min, the sample was subjected to heat-shock at 42°C for 90 s and immediately replaced on ice for 2 min. To allow recovery of the cells they were incubated with 1 ml LB-medium for 1h at constant mixing at 850 rpm in a thermo mixer. 250 μ l of the culture were spread onto an agar plate containing suitable antibiotics for the selection of transformants and incubated overnight at 37°C. Clones were screened for the correct insert by whole cell PCR (3.4.1.1).

3.3.4 Transformation of electro-competent *E. coli* cells

Freshly prepared electro-competent cells (3.3.2) were mixed on ice with 3-15 μ l of the DNA to be transformed. The cell suspension was filled into a pre-chilled electroporation cuvette with 0.2 mm diameter and electroporated at 2.5 kV using a Biorad MicroPulser™ electroporator. The cells were recovered immediately with 1 ml ice cold SOC medium and incubated in 2ml eppendorf tubes at 30°C for 1h. Afterwards, the cells were harvested by 2 min centrifugation at 6,000 rpm and most of the supernatant discarded. The pellet was resuspended in the remaining liquid and spread onto an LB agar plate containing appropriate antibiotics. The plate was incubated over night at 30°C.

As large constructs such as fosmids have a very low transformation rate, high amounts of DNA are needed for successful transformation. In this study amounts >1 μ g DNA were used frequently for electroporation.

3.3.5 Generation of protoplasts from *Herpetosiphon*

Protoplasts of *Herpetosiphon* sp. 060 were generated for transformation experiments following a modified protocol after Thompson et al. (1982). 100 ml CY-medium were inoculated with agar blocks from a *Herpetosiphon* dish culture and grown for 24h. The culture was divided into two 50 ml falcon tubes and the cells harvested by 10 min centrifugation at 4,000 rpm. The pellets were washed in 7.5 ml buffer P and resuspended in 10 ml buffer containing 10 mg/ml lysozyme. The cells were incubated

at 30°C and constant shaking with 180 rpm to allow digestion of the cell wall. Samples were taken after 2, 3 and 4h and progress examined microscopically. At a protoplastation level of approx. 90%, 5 ml buffer P was added and the suspension carefully pipetted up and down with a 5 ml pipette tip to dissolve clumps of cells. To separate the protoplasts from remaining filaments, the suspension was filtered through a syringe with sterile cotton cloth applying gentle pressure only. The filtered cells were pelleted at 3,000 rpm for 10 min, washed with 10 ml buffer P to remove lysozyme and finally resuspended in 1 ml buffer P. The cell suspension was either used directly or stored at -80°C as 100 µl aliquots.

3.3.6 PEG-mediated transformation of *Herpetosiphon* protoplasts

Transformation of protoplasts (3.3.5) was carried out after modified protocols following the procedures as described elsewhere (Bibb et al., 1978; Thompson et al., 1982). 100 µl protoplasts were mixed with 1 µg of the DNA construct to be transformed. Immediately 400 µl of buffer T or buffer P containing 25% – 50% PEG were added and carefully mixed. The mixture was spread on VY/2-agar containing 0.3 M Sorbitol as a protective agent against osmotic stress and incubated at 30°C. The cells were either directly plated on selective medium containing suitable antibiotics or first regenerated for 1-3 days without selection before covering the plate with selective soft agar or a solution of the selective agent. The amount of the used antibiotics was calculated for an agar culture with 25 ml medium.

3.3.7 Glycine treatment of *Herpetosiphon* cultures

Glycine is described as a cell wall-weakening agent (Hammes et al., 1977), which is commonly used prior to the electro-transformation of bacterial cells (Buckley et al., 1999). 100 ml CY-medium were inoculated with agar blocks from a *Herpetosiphon* dish culture and grown for 24h. Glycine was added to the medium at a final concentration of 2% – 5% (w/v) and the culture grown for another 24h. The following steps were all carried out under constant cooling. 2 ml aliquots were pelleted at 5,000 rpm for 10 min and the cells washed two times with 1 ml electroporation buffer. After the second washing step, the pellet was resuspended in 50 µl electroporation buffer and placed on ice until usage (3.3.8.2).

3.3.8 Electroporation of *Herpetosiphon* sp. 060 cells

2 ml from a liquid culture of *Herpetosiphon* were pelleted for 4 min at 13,000 rpm and washed two times with 1 ml electroporation buffer. The pellet was resuspended in 50 µl electroporation buffer, mixed with 5 µl of the DNA to be transformed and incubated for 5 min on ice. Electroporation was performed in 0.1 cm cuvettes at 1.25 and 2.0 kV. The cells were recovered in 1 ml CY-medium and incubated at 30°C for 2h before spreading on selective VY/2 agar plates.

3.3.8.1 Electroporation of *Herpetosiphon* protoplasts

A 100 µl aliquot of *Herpetosiphon* protoplasts was sedimented for 10 min at 3,000 rpm. The supernatant was discarded and the pellet resuspended in 50 µl electroporation buffer. After the addition of 1 µg of the DNA to be transformed, the sample was incubated on ice for 5 min. Electroporation was performed in 0.1 cm cuvettes at 0.35 kV. The cells were recovered in 1 ml CY-medium containing 0.3M sorbitol. One half of the suspension was directly plated on selective VY/2 agar plates containing 0.3M sorbitol. The other half was incubated at 30°C and 500 rpm in a 2 ml-ependorf cap for 24h before plating.

3.3.8.2 Electroporation of *Herpetosiphon* cells after glycine treatment

A 50 µl aliquot of glycine treated cells (3.3.7) was mixed with 1 µg of the DNA to be transformed and incubated on ice for 10 min. Electroporation was performed in 0.1 cm cuvettes at 1.8 kV. The cells were either treated in the same way as protoplasts (3.3.8.1) or alternatively recovered in 5 ml CY-medium containing 0.3M sorbitol and regenerated for 5 days at 30°C and 140 rpm in a 15ml falcon-tube. The cells were harvested by centrifugation for 10 min at 5,000 rpm and resuspended in 1 ml CY medium and spread on selective VY/2 agar plates containing 0.3M sorbitol

3.3.9 Sonication

The use of ultrasound has been described as a method for the delivery of DNA into bacterial cells (Song et al., 2007). A 100 ml CY culture of *Herpetosiphon* sp. 060 was grown for 48h, including the optional addition of 1% glycine after the first 24h. 2 ml of the culture were pelleted by centrifugation, washed two times with electroporation

buffer and finally resuspended in 50 µl electroporation buffer. The cells were mixed with 500 ng of the DNA to be transformed and incubated on ice for 5 min. Sonoporation was performed in a Bandelin Sonorex RK 31 ultrasound-bath. The cells were exposed to a constant ultrasound output for 10s, 30s and 60s. Immediately following the treatment, 1 ml CY medium was added and the cells incubated for 2h at 30°C, before spreading 500 µl on selective VY/2 agar plates.

3.3.10 Bacterial conjugation

Bacterial plasmids can be mobilised and transferred to other cells by horizontal gene transfer (Smillie et al., 2010). During bacterial conjugation plasmids that carry a special origin of replication (origin of transfer, *oriT*) are targeted by a set of proteins, which initiate a rolling circle replication. Through the formation of a pilus, which connects two conjugating cells, the newly synthesised plasmid is transferred to the receptor cell. This process can be exploited for the transformation of bacterial cells. In this study the *E. coli* strain ET12567 was used for this purpose, which harbours the helper plasmid pUZ8002. This plasmid can mediate the mobilisation of other vectors that contain an *oriT*, but is not self-transmissible due to a mutation (Kieser et al., 2000).

A DNA fragment of approx. 1kb was inserted into the *oriT* containing vector pIJ773 and transformed into ET12567 cells. The *E. coli* cells were grown at 37°C in 30 ml LB medium to an OD₆₀₀ of 0.4. 30 ml CY-medium were inoculated with agar blocks from a *Herpetosiphon* sp. 060 dish culture and grown for 48h @ 30°C. The cells of both cultures were harvested by centrifugation and washed with LB medium and finally resuspended in 1.5 ml LB medium. 500 µl of each cell suspension were mixed and pelleted by centrifugation. The supernatant was decanted and the cells resuspended in the remaining liquid, before spreading them on VY/2 agar plates. After incubation at 30°C for 24h, the cultures were covered with 1 ml of an antibiotic solution, which always contained apramycin as selective agent and either nalidixinic acid, carbenicillin or gentamycin to eliminate the *E. coli* cells, but not *Herpetosiphon* (see also 4.1.1). The amounts of antibiotic agents were calculated for an agar plate with 25 ml solid medium. The plates were incubated again at 30°C and observed for several days.

3.4 Molecular biological methods concerning nucleic acids

3.4.1 PCR

The polymerase chain reaction (PCR) allows the rapid amplification of a DNA region of interest between two defined primer regions. First, the DNA is denatured by heating, so that the oligonucleotide primers can anneal to their complement sequences. These primers act as a template for a DNA-polymerase, which elongates the DNA strands. Repetitive cycles of this procedure allow the exponential amplification of the target region. While dNTPs for the elongation are added in high surplus, the amount of primers is usually the limiting factor of the product yield.

Most commonly, the thermostable polymerase from *Thermus aquaticus* (Taq) is used for PCR. The amplification of an open reading frame used for heterologous protein expression was performed with DNA-polymerase from *Pyrococcus furiosus* (Pfu), which is capable of proofreading activity. PCR-reactions were composed as follows:

10x PCR-Buffer	2 µl
MgCl ₂ (25mM)	1 µl
dNTPs (10 mM)	0.4 µl
Forward-Primer (100 µM)	0.5 µl
Reverse-Primer (100 µM)	0.5 µl
DNA-polymerase (5 U/µl)	0.16 µl
Template	1 µl
Water	ad 20 µl

The following thermocycler programme was used for PCR reactions. Elongation times were chosen according to the expected product length, assuming an elongation speed of 1kb/min for the Taq-polymerase. Annealing temperatures were chosen according to the primers used in the reaction and usually were 2-5°C below the melting temperature as calculated by the CloneManager software. As a control reaction for unspecific amplification or contamination with foreign DNA, a reaction without template was used.

1. Initial denaturation	95°C	5 min
2. Denaturation	95°C	30 s
3. Annealing	53-65°C	30 s
4. Elongation	72°C	30-120 s
5. final elongation	72°C	5 min
6. cooling	4°C	hold

Steps 2-4 were repeated 25-30 times.

3.4.1.1 Whole cell PCR

For the rapid screening of bacterial clones, cells were picked from an agar plate with a sterile tooth-pick and stirred directly into the PCR reaction mix. If a clone was to be tested in more than one PCR, 2 µl of a liquid culture were used as template in the reaction. The first step was prolonged to 10 min to ensure complete denaturation of the cells and DNA.

3.4.2 Agarose gel electrophoresis

Non-denaturing agarose gel electrophoresis was employed for the separation of DNA samples for purposes of analysis and preparations like gel extraction (3.4.3). Gels with a concentration of 0.7% – 3% in 1x TBE-buffer were used. Samples were mixed with 6x loading buffer before they were loaded onto the gel. PCR-reactions were directly loaded, as the loading buffer was already included in the reaction mix. An appropriate size marker was filled into one well to enable the estimation of the molecular weight of the samples. Gels were run at 120V and stained afterwards in an ethidium bromide bath for approximately 2 min on a horizontal shaker. Background staining was removed by washing the gel for several minutes in a water bath. Nucleic acids were visualized with the help of an illuminator employing UV-light at a wavelength of 336 nm. Ethidium bromide intercalates in the DNA, which enhances the fluorescence of the dye compared with the unbound reagent. Thus, areas containing nucleic acids can be identified as coloured areas. Gels were documented with the Intas Gel iX Imager.

3.4.3 Isolation of DNA from agarose gels

DNA fragments and PCR products which were to be used in cloning experiments were extracted from agarose gels after separation by electrophoresis. The areas of interest were excised with a clean scalpel and subsequently eluted from the gel with the help of the Qiagen gel extraction kit following manufacturer's instructions. The purified DNA was solved in sterile ultrapure water.

3.4.4 Isolation of plasmid and fosmid DNA

Vectors required in this study as well as recombinant plasmids and fosmids generated in cloning experiments were isolated from the corresponding host strain harbouring the construct of interest. Usually 3 ml LB medium were inoculated with a single colony and cultivated overnight.

Plasmids were purified from this culture using either the Promega PureYield Miniprep System or the Qiagen Plasmid Mini Kit following the manufacturer's instructions.

Alternatively, the cells were pelleted by centrifugation at 13,000 rpm for 2 minutes and resuspended in 350 µl buffer P1. Cells were lysed by addition of 350 µl buffer P2 and the mixture neutralised by mixing it with 350 µl buffer P3. Cell debris and proteins were pelleted by 5 minutes centrifugation.

The supernatant containing the DNA was mixed with one volume phenol to remove proteins from the solution. The mixture was centrifuged for 5 min and the aqueous phase recovered. Remaining phenol was removed by the addition of one volume chloroform and a further centrifugation step. Subsequently, plasmid DNA was recovered from the aqueous phase by ethanol precipitation (3.4.6).

Large scale preparations from up to 100 ml culture were either performed as described above or by application of the Qiagen Plasmid Maxi Kit. Fosmids from EPI300 cells were always isolated from 100 ml cultures, which were induced to high copy plasmid production with the CopyControl™ Induction Solution.

3.4.5 Isolation of genomic DNA

Chromosomal DNA of *Herpetosiphon* 060 was isolated using the Promega Genomic Wizard Kit according to the manufacturer's instructions. Genomic DNA was readily extracted from either 2 ml liquid culture or bits of frozen cell pellets from 100 ml liquid cultures. Genomic DNA for the 454 sequencing was isolated with the Qiagen DNeasy Blood & Tissue Kit according to the manufacturer's protocol.

3.4.6 DNA precipitation

For the recovery of DNA from aqueous solutions 0.1 volumes of 3M NaOAc (pH 5.5) were added to acidify the solution and facilitate precipitation of nucleic acids. Three volumes of ice cold ethanol (99.8%) were added and the sample centrifuged at 13,000 and 4°C for at least 30 min. The supernatant was discarded and the pellet washed with 70% ethanol to remove precipitated salts from the pellet. After centrifuging again for 5 min the liquid was removed and the pellet air dried before being solved in an appropriate amount of sterile ultrapure water.

If especially high concentrations of DNA were needed, precipitation was further enhanced by incubating the sample at -20°C for up to 30 min prior to centrifugation.

3.4.7 Restriction digestion

DNA restriction endonucleases catalyse the cleavage of phosphodiester bonds in double stranded DNA in a sequence specific manner. Recognition sites of restriction enzymes are often palindromic sequences with 4 – 8 bp length. While some enzymes produce DNA fragments with blunt ends, most endonucleases generate so called cohesive or sticky ends with short single strand overhangs. Due to the palindromic nature of restriction sites, these fragments can be ligated to DNA cut with either the same enzyme or another that produces compatible ends. Restriction endonucleases originate from the defence mechanism of bacteria against foreign DNA. As in this context the methylation pattern is also important for the discrimination of own and foreign DNA, certain methylations can impair the activity of some enzymes.

Restriction digestion was used to prepare DNA fragments and PCR-products for the ligation into vectors treated in the same way. To ensure the later orientation of the insert in the plasmid, usually two different restriction enzymes were used in cloning experiments. If this was not possible, vectors linearised with only one enzyme were dephosphorylated (3.4.9) to prevent religation.

Reaction buffers for single and double digestions were chosen following the manufacturer's recommendations.

3.4.8 Ligation of DNA fragments

Linear DNA fragments produced by restriction digestion (3.4.7) of PCR-products or vectors were ligated using T4 DNA-ligase. The enzyme catalyses the formation of a phosphodiester bond between the 5'-phosphate and 3'-hydroxyl termini in double-stranded nucleic acids. Thus, DNA fragments with compatible cohesive ends or blunt ends can be linked together, e.g. to form a circular plasmid. This was used to clone PCR-products into suitable vectors. A standard reaction scheme is given below.

10x T4-ligase buffer	2 μ l
T4 DNA ligase	1 μ l
Vector	5 μ l
Insert	8 μ l
Water	ad 20 μ l

The reaction was incubated at 22°C for 2 – 16h, where prolonged incubation increased efficiency. As the enzyme can interfere with downstream processes, it was inactivated at 70°C for 10 min.

The pGEM-T vector system comprises a linearised plasmid with T-overhangs, which are compatible to the A-overhangs produced by the Taq polymerase at the ends of PCR-products. Cloning into pGEM-T was performed according to manufacturer's instructions.

3.4.9 Dephosphorylation of linear DNA

Vectors that are cut with only one restriction enzyme or with endonucleases which produce compatible or blunt ends are able to religate again to circular DNA. This process is unwanted in cloning experiments, where linear DNA shall be inserted into the vector. To avoid self-ligation, the ends of linearised vectors are dephosphorylated using either calf intestine alkaline phosphatase (CIAP) or shrimp alkaline phosphatase (SAP) according to the manufacturer's instructions. Subsequently, the DNA was purified with a Qiagen Miniprep Kit to remove the enzyme, which otherwise could impair following ligation reactions.

3.4.10 Determination of nucleic acid concentration and purity

The concentration of aqueous solutions of nucleic acids was quantified by measurement of the absorption at a wavelength of 260 nm in an UV/VIS spectrometer. Employing the Lambert-Beer law, the concentration can be calculated according to the following formula. The used coefficients describe an average for nucleic acids of unknown sequence or mixtures. As a linear correlation of concentration and absorption is given only for an extinction of approximately 0.1 – 1.0, the samples were diluted accordingly prior to measurement.

$$\text{Formula 1:} \quad \text{concentration} \left[\frac{\text{mmol}}{\text{L}} \right] = \frac{\text{OD} \times k \times d}{\sum \text{nt} \times \text{MW}}$$

OD₂₆₀ = optical density at 260 nm

d = dilution factor

k = coefficient with OD₂₆₀ 1.0 = 50 µg/ml dsDNA

OD₂₆₀ 1.0 = 35 µg/ml ssDNA

OD₂₆₀ 1.0 = 40 µg/ml ssRNA

∑nt = number of nucleotides

MW = molecular weight with DNA = 330 g/mol

RNA = 345 g/mol

Aromatic amino acids show high absorption at 280 nm. Therefore, the contamination of DNA by proteins can be determined by the calculation of the A₂₆₀/A₂₈₀ ratio. Highly pure DNA has a ratio between 1.8 and 2.0.

3.4.11 Sequencing of vector constructs and PCR-products

To determine the sequence of a PCR-product and to exclude mutations in plasmid inserts, their nucleic acid sequence had to be analyzed. Sequencing was performed by GATC Biotech AG (Konstanz, Germany) on an ABI3730xl after the Sanger dideoxy method (Sanger et al., 1977). In the course of this procedure single-stranded DNA is amplified with one primer by PCR in four individual reactions. Each contains, next to dNTPs, also an amount of one of four fluorescently labelled didoxynucleotide triphosphates (ddATP, ddTTP, ddCTP, ddGTP), which are statistically incorporated into the growing DNA chain. The lack of the 3'-OH group leads to chain termination, so that fragments of varying sizes are generated. These can be separated by gel electrophoresis. The visualisation of the different fluorescence tags allows elucidation of the exact sequence.

For sequencing reactions, either PCR-specific primers or general primers (T7 or SP6) for the corresponding plasmid were used. End sequencing of fosmid inserts were performed with either T7 or Epi-RP primers. For the analysis of 16S rDNA, general primers (16S rDNA fwd & rev) were applied.

3.4.12 454 Genome sequencing

A draft genome of *Herpetosiphon* sp. 060 was generated by sequencing of genomic DNA with the next generation Roche 454 technology, which was carried out by GATC Biotech AG (Konstanz, Germany). In this approach, the template is prepared by random fragmentation of the genomic DNA into small sizes and subsequent immobilisation on beads by common adaptors. These beads are encapsulated individually in droplets on a picotiterplate, where amplification is performed by emulsion-PCR on single molecules (Metzker, 2010). The incorporation of specific nucleotides is visualised by a bioluminescence method named pyrosequencing. The produced fluorescence signal allows determination of the assembled DNA sequence. The DNA reads, obtained by these reactions are then assembled into larger fragments called contigs, which represent altogether the draft genomic data. The size of the gaps between the contigs is dependent on the achieved genome coverage by the produced reads.

3.4.13 Screening of a genomic library

The fosmid library of *Herpetosiphon* 060 created by Ö. Erol was screened in a PCR-based approach the final part of the siphonazole cluster. Therefore, specific primers were derived from the DNA sequences obtained by genome sequencing. For rapid screening, always 12 clones from a row of a 96-well plate were pooled and subjected to whole cell PCR (3.4.1.1). PCR-reactions were analysed by gel electrophoresis (3.4.2) to identify the pools yielding a product. The clones of these pools were then analysed individually and the corresponding clone identified.

3.4.14 RNA-isolation and expression analysis

All experiments on RNA and derived cDNA were conducted by G. Rashid. Total RNA was isolated from cultures of *Herpetosiphon* sp. 060 with the help of the NucleoSpin® RNA II kit (Macherey-Nagel) following the manufacturer's protocol. The isolated RNA was used for the synthesis of cDNA from mRNA using oligo (dT)18 primers and the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's instructions.

The expression of open reading frames was probed by PCR-reactions (3.4.1) with standardised amounts of cDNA as template. Specific primer pairs aiming at the intergenic regions and the adjacent regions were derived from the DNA sequences of the targeted gene cluster. Evaluation of the PCR-results was carried out by agarose gel electrophoresis (3.4.2).

3.5 Recombination and cloning methods

3.5.1 Λ -Red recombination

Recombination-mediated genetic engineering is a powerful tool for the site-specific introduction of linear DNA into fosmid or chromosomal DNA. Generally, most bacteria cannot be readily transformed with linear DNA due to intracellular exonucleases, which degrade linear DNA as part of the bacterial defence mechanism. This problem can be alleviated by use of recombination genes from bacteriophage λ , which greatly enhance the recombination rate (Gust et al., 2004). This method allows the replacement of a DNA sequence with a construct of linear DNA containing a

selectable marker and homology arms at both ends. As these homology extensions only need to be as short as 39 bp for successful recombination, they can easily be added by PCR primers.

In order to allow λ -Red recombination steps on fosmids, they had to be transformed into cells of *E. coli* BW25113 harbouring the helper plasmid pKD46, which carries the necessary λ RED (*gam*, *bet*, *exo*) recombination functions.

A single colony of BW25113 was used to inoculate 3 ml LB medium containing ampicillin (100 μ g/ml), which was incubated overnight at 37°C at constant shaking. 100 ml LB medium containing ampicillin were inoculated with 1 ml pre-culture and grown to an OD₆₀₀ of ~0.5. Cells were then prepared for electroporation according to protocol (3.3.2) and transformed with fosmid DNA (3.3.4). The cultures were grown at 30°C and selected on LB-agar containing ampicillin (for pKD46) and chloramphenicol (34 μ g/ml; for pCC1FOS).

Clones containing both pKD46 and the fosmid to be modified were used for λ -Red recombination. Therefore, 100 ml SOB-medium containing ampicillin and chloramphenicol in 500 ml flasks were inoculated with 1 ml of an overnight pre-culture. To induce the expression of recombination genes, 1 ml of a 1M L-arabinose solution was added to the culture, before it was grown at 30°C to an OD₆₀₀ of ~0.5. The cells were prepared for electroporation and transformed with 3 – 15 μ l linear DNA. After incubating the sample in 1 ml SOC-medium at 30°C for 1h, transformants were selected on LB-agar containing ampicillin and appropriate antibiotics.

3.5.1.1 Generation of linear DNA for λ -Red recombination

Streptomycin and apramycin resistance cassettes were amplified from plasmids pIJ778 and pIJ773 respectively by PCR using primers with 40bp homology extensions. PCR products were purified by gel electrophoresis (3.4.2) and extracted according to protocols (3.4.3).

For the recombination of cluster parts already containing selection markers, linear DNA was generated by subjecting the fosmids to restriction digestion (3.4.7) with suitable enzymes. The fragments of interest were obtained by extraction from an agarose gel (3.4.3).

3.5.2 Cloning of PCR-products

The DNA of interest was amplified by PCR (3.4.1) with primers containing suitable restriction sites for insertion into the target vector and was purified by electrophoresis and gel extraction (3.4.2 and 3.4.3). Insert and vector were digested (3.4.7) with the chosen restriction enzymes and ligated (3.4.8), before transforming the generated construct into competent *E. coli* cells (3.3.3). Clones could then be used for further experiments.

For insertion of DNA fragments into the pET151 vector, the Champion Directional TOPO Expression Kit (Invitrogen) was used. Amplification of the insert was performed with Pfu DNA-polymerase to generate blunt end products. The forward primer contained a CACC 5'-extension for directional integration into the topoisomerase-activated linearised vector.

3.5.3 Yeast recombinational cloning

This cloning method exploits the ability of yeast cells to recombine DNA elements with homologous regions. This can be used for the constitution of multiple DNA fragments to one single plasmid construct. This protocol is derived from the LiOAc / ssDNA / PEG mediated transformation method described by Gietz and co-workers (Gietz et al., 2007).

50 ml YPAD medium was inoculated with a 10 ml YPAD overnight culture of *S. cerevisiae* and grown at 28°C and 200 rpm for five hours. Thereafter, the cells were harvested by 5 min centrifugation at 3,000 g and the supernatant discarded. The pellet was resuspended in 1 ml 0.1M LiOAc and transferred to a 1.5 ml tube. The cells were pelleted again for 1 s at top speed and resuspended again in 400 µl 0.1M LiOAc. For the transformation, 50 µl aliquots were prepared.

High molecular single-stranded DNA from salmon testes was used as carrier DNA, which facilitates the delivery of DNA fragments into the cells. 20 mg ssDNA were thoroughly dissolved in 10 ml TE buffer (10 mM Tris/HCl, pH 8.0, 1 mM EDTA) and frozen at -20°C in 250 µl aliquots. Immediately before use, an aliquot was boiled for 5 min and then placed on ice.

The prepared cell aliquots were centrifuged for 15 s at top speed and the supernatant was removed. The cells were mixed with 240 µl 50% PEG (3350) solution, 50 µl 1M LiOAc and 50 µl ssDNA and then vortexed. Equal amounts of the DNA fragments to be transformed were mixed separately and then added to the cell suspension. The samples were first incubated at 30°C for 30 min and then at 42°C for 30 min. The cells were recovered by 15 s centrifugation at 6,000 rpm and resuspended in 1 ml sterile water. A 200 µl aliquot was spread on synthetic drop-out medium (minus uracil) and incubated at 28°C for two days.

3.5.3.1 Isolation of plasmid DNA from yeast cells

Plasmid DNA was purified from yeast cells in small scale using the ZymoPrep™ Yeast Plasmid Miniprep I Kit (Zymo Research) according to the manufacturer's instructions. Plasmids from 50 ml cultures were isolated by the following method.

The cells were harvested by centrifugation and resuspended in 4 ml buffer P1. 5 g glass beads were added and the sample thoroughly vortexed for 3 – 5 min. The culture was frozen at -80°C for 10 minutes and thawed in a 60°C water bath before vortexing again. 4 ml buffer P2 and 4 ml buffer P3 were added and mixed by inversion. The sample was pelleted at 10,000 xg and 4°C for 5 min. 15 ml phenol were added and the sample spun again for 5 min. The upper phase was removed, mixed with 10 ml ice-cold ethanol (99.8%) and placed at -20°C for 30 – 60 min. The sample was centrifuged for 25 min at 10,000 xg and the supernatant removed. The pellet was washed with 70% ethanol and spun again for 5 min. Liquid was removed and the pellet air-dried under a clean bench. DNA was allowed to dissolve in 50 µl H₂O at room temperature for 15 min and by tapping.

3.6 Molecular biological methods concerning proteins

3.6.1 Heterologous expression of proteins

The *in vitro* investigation of protein functions requires the purification of the target protein from the cell. As in most cases the natural protein is only expressed in small amounts, the overexpression in a heterologous host is necessary. Therefore, the DNA sequence of the protein is inserted into an expression vector that usually contains an inducible promoter as well as an affinity tag such as 6x-his (consisting of 6 successive histidines). These tags are expressed at the N- or C-terminus of the protein, depending on the vector, and greatly facilitate the purification via affinity chromatography (3.6.3).

The DNA sequence of the protein of interest was amplified by PCR (3.4.1) and cloned (3.5.2) into the multiple cloning site of the expression vector, so that the start codon of the protein was in-frame with the 6x-his tag on the plasmid. The generated plasmid was first transformed into competent XL1-blue *E. coli* cells (3.3.3), where colonies were screened for correct transformants by whole-cell PCR. To exclude the possibility of mutations in the reading frame, the chosen clones were submitted for sequencing (3.4.11). Plasmid DNA was isolated from confirmed clones and transformed into an *E. coli* expression strain.

3 ml pre-cultures were used to inoculate LB medium containing appropriate antibiotics. Culture sizes ranged from 100 to 600 ml depending on the purpose and expected expression levels. After inoculation, the main culture was grown to an OD₆₀₀ of ~0.4 – 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.1 – 1.0 mM. Again, culture conditions were chosen according to the expression level and solubility of the expressed proteins. Generally, cultivation was carried out for 3h at 30 - 37°C, 4h at 20°C or overnight at 16°C.

3.6.2 Cell lysis by sonication

Cells from protein expression were harvested by centrifugation at 8,500 rpm for 5 min using 50 ml Falcon tubes. The cell pellet was resuspended in lysis buffer in 1/50 volume of the original culture and placed on ice. Cells were lysed with the help of a Branson Sonifier 250, set to output level 4, 50% duty cycle. The samples were

sonified 5 times with ten pulses each. Between the pulses the cells were placed back on ice to avoid overheating of the sample. Cell debris and insoluble parts were pelleted by centrifugation for 15 min at 8,500 rpm and 4°C. The supernatant containing the soluble proteins and the pellet were collected.

3.6.3 Purification of recombinant proteins by Ni-NTA-columns

Proteins harbouring a his-tag can easily be purified by affinity chromatography on a Ni-NTA matrix. The histidine residues are bound by the matrix, while other proteins elute. Unspecifically bound proteins can be eluted by increasing concentrations of imidazole, which competes with the histidines for the binding. The six histidines of the tag ensure that the target protein only elutes at high concentrations.

For the affinity chromatography a gravity-flow column with 1 ml Ni-NTA-agarose was prepared and equilibrated in lysis buffer. The sample was added to the column and allowed to pass the matrix. The column was washed two times with 4 ml washing buffer. Elution was carried out in four steps by adding 0.5 ml of elution buffer with 100, 150, 200, 300 and 300 mM imidazole. All fractions were collected and kept on ice to prevent protein degradation.

3.6.4 Denaturing SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

This method allows the analytical separation of proteins according to their size. In a denaturing approach the proteins are boiled and treated with mercapto-ethanol to remove any disulfide-bonds. During electrophoresis the unfolded proteins are covered with the negatively charged sodium dodecyl sulfate (SDS), which masks the inherent charge of the protein, thus allowing a strict separation by molecular weight.

The separating gel is formed by radical polymerisation of bis-acrylamide to polyacrylamide, which forms a molecular sieve. The concentration of polyacrylamide can be chosen in dependence of the analysed protein size. For a better focussing of protein bands discontinuous gels were used, where the separating gel is covered with a low concentrated stacking gel.

The reaction mix is given below. After polymerisation was initiated by the addition of ammoniumperoxosulfate (APS) and N,N,N',N'-tetraethylendiamin (TEMED), the gel was pipetted between two plates with a spacer distance of 1.0 or 1.5 mm. During polymerisation, the separating gel was covered with isopropanole to ensure a smooth surface. Isopropanole was removed before the addition of the stacking gel.

Glycin SDS-stacking gel (4%)

Tris/HCL pH 6.8 (1M)	1,250 µl
SDS (10%)	25 µl
Bis-acrylamide (30%)	675 µl
Water	3,050 µl
APS (10%)	25 µl
TEMED	2.5 µl

Glycin SDS-separating gel (12%)

Tris/HCL pH 6.8 (1M)	2,500 µl
SDS (10%)	100 µl
Bis-acrylamide (30%)	675 µl
Water	3,050 µl
APS (10%)	50 µl
TEMED	2.5 µl

For each gel run, the reservoirs of the electrophoresis assembly were filled with fresh 1x glycine-running buffer. Protein samples were mixed with 4x denaturing loading buffer and boiled at 99°C for 5 min, before loaded onto the gel. Electrophoresis was performed at 120V until the samples reached the separating gel and voltage increased to 160V afterwards. As a reference, a molecular size marker was loaded in one well. After electrophoresis the gels were analysed by coomassie-staining.

3.6.5 Coomassie-staining

A staining with coomassie brilliant blue was used to analyse the proteins on a polyacrylamide gel. Gels were immersed in the staining solution and shortly heated in a microwave oven for quicker staining and then incubated several minutes on a horizontal shaker. Destaining of the background colour was performed by shaking

with destaining solution for several hours or overnight in the fridge. The gels were documented with the INTAS illuminator.

3.6.6 Concentration of purified proteins and buffer exchange

Samples containing purified protein from heterologous expression (3.6.1) were concentrated with the help of Amicon Ultra centrifugal filters according to the manufacturer's instructions. This method was also used to remove imidazole from the buffer, which could interfere with following assays. Following the manufacturer's manual, the buffer was replaced in several centrifugation steps with 50 mM Tris/HCl, pH8.

In cases, where concentration of a sample led to precipitation of proteins, buffer exchange was carried out with gravity-flow PD-10 desalting columns by GE Healthcare following the manufacturer's protocol.

3.6.7 Determination of protein concentrations after Lambert-Beer

This method allows the determination of a homogenic protein solution based on the absorption at a wavelength of 280 nm. It requires the knowledge of the molar extinction coefficient ϵ , which can be calculated in good approximation for a given amino acid sequence. In this study the program ProtParam provided by the Swiss Institute for Bioinformatics (3.1.13) was used for the calculation of ϵ . The protein concentration can be determined by the following equation.

Formula 2:
$$\text{concentration} \left[\frac{\text{mol}}{\text{L}} \right] = \frac{OD \times d}{\epsilon}$$

Division by molecular weight gives concentration in g / l.

OD₂₈₀ = optical density at $\lambda = 280$ nm

d = dilution factor

ϵ = molar extinction coefficient [$\text{M}^{-1} \text{cm}^{-1}$]

3.6.8 O-methyltransferase activity assay

The activity of a heterologously expressed O-methyltransferase was observed by an *in-vitro* methylation assay similar to the method described by Sherman and co-workers (Shengying et al., 2009). The reaction was performed in 50 mM Tris-HCl buffer with pH 7.5. Reagents and protein solution were always prepared freshly and directly used for the assay. The reaction mixture is given below.

O-methyltransferase (sphB)	25 μ M
protocatechuic acid (substrate)	1 mM
S-adenosylmethionine	500 μ M
MgCl ₂	10 mM
50 mM Tris/HCl (pH 7.5)	ad 100 μ l

The reaction was incubated at 30°C for 2h.

3.6.8.1 High performance liquid chromatography (HPLC)

HPLC analysis of the methylation assay was carried out on a Merck-Hitachi system consisting of a D-6000A interface with an L-6200A Intelligent Pump, a Rheodyne 7725i injection system and a L-4500A diode array detector. Column was a Waters XTerra™ C18 (5 μ m, 4.6 x 250 mm). The complete reaction mix was injected without further preparation and separated using a gradient of 0.1% TFA and acetonitrile as liquid phase. The complete setup is given below.

Solvent A	H ₂ O with 0.1% TFA
Solvent B	acetonitrile
Gradient	0 min: 90% A, 10% B
	20 min: 40% A, 60% B
Flow	1 ml / min

3.6.9 Mass spectrometry

For LC-ESI MS measurements samples were dissolved in methanol at a concentration of 1 mg/ml. All measurements were conducted by E. Eguereva at the institute for pharmaceutical chemistry (University of Bonn, Germany). Experiments were recorded in positive (+Q1) and negative (-Q1) mode and analysed by Applied Biosystems/ MDS Sciex Analyst software.

MALDI-TOF-MS measurements were conducted by M. Josten at the Institut für medizinische Mikrobiologie, Immunologie und Parasitologie (Universitätsklinikum Bonn)

3.6.10 ATP - PP_i exchange assay

Adenylation domains activate their substrates at the expense of ATP. The assay described by Phelan and co-workers (Phelan et al., 2009) measures the consumption of γ -¹⁸O₄-labelled ATP and the formation of ¹⁶O₄-ATP by an excess of unlabelled PP_i. The resulting mass shifts are detected by MALDI-TOF-MS. Thus, the activity of the A domain can be measured. Incubation with different substrates and comparison of the ATP-PP_i exchange rate allows determination of the substrate specificity of the studied domain. The assays were performed by Dr. T. Schäberle in the laboratory of the Piel group at the Kekulé-Institut for organic chemistry and biochemistry (University of Bonn, Germany) according to the described method. In short, 200 nM purified A domain (in A domain buffer, 3.1.5) was incubated with 1 mM γ -¹⁸O₄-ATP, 1mM substrate, 5 mM MgCl₂ and 5 mM PP_i in a reaction volume of 6 μ l for 30 min at RT. Exchange rate by MALDI-TOF-MS was determined by comparison of the ratio of γ -¹⁶O₄-ATP (m/z 506) to the sum of all ATP species, including unlabelled, partially labelled, fully labelled (m/z 514), and monosodium-coordinated ions (m/z 506, 508, 510, 512, 514, 528, 530, 532, 534, 536). Determination of percent exchange was normalised with the following modifier: % exchange = $(100/0.833) \times \frac{^{16}\text{O}}{(^{18}\text{O} + ^{16}\text{O})}$

4 Results

4.1 Knock-out experiments with *Herpetosiphon* sp. 060

In the course of this study intensive research on the genetic modification of the genome of *Herpetosiphon* sp. 060 was carried out. These efforts aimed at a knock-out of the putative siphonazole biosynthetic gene cluster to prove its function. To the present date no protocols for the transformation of *Chloroflexi* or *Herpetosiphon* species are available. The establishment of a working procedure would make this group amenable to genetic modifications. The experiments were based on the assumption that a plasmid carrying a selection marker and a homologous region to the gene cluster would in some cells be integrated into the genome by a single crossover event and thus disrupt the gene cluster. Abolishment of siphonazole production in successful transformants could then be evaluated by LC-MS (3.6.9) of a crude extract.

4.1.1 Antibiotic resistance tests

Herpetosiphon sp. 060 was tested for resistance against several antibiotic agents to select an appropriate selection marker for the knock-out experiments (3.2.5). Table 4-1 summarises the tested antibiotics and the test results.

Table 4-1: Antibiotic resistance tests with *Herpetosiphon* sp. 060

Final antibiotic concentrations in the media are given; (----) indicates no growth, (+) partially inhibited growth, (++) uninhibited growth

Antibiotic	Concentration [µg/ml]	growth
Apramycin	50	----
Tetracycline	50	----
Nalidixinic acid	50	+
Zeocin	50	++
Ciprofloxacin	80	----
Chloramphenicol	170	----
Streptomycin	50	----
Carbenicillin	50	++
Gentamycin	50	++

Previous tests with the *Herpetosiphon* strain showed resistance to kanamycin, ampicillin and hygromycin (Ö. Erol, personal communication). In these earlier experiments, zeocin was selected as a suitable resistance marker, but in this study *Herpetosiphon* sp. 060 exhibited resistance against zeocin to concentrations as high as 150 µg/ml. Nalidixinic acid only led to a partial inhibition of growth, which was possibly the result of an uneven distribution of the agent in the media or caused by chloroform, which is used as solvent for nalidixinic acid. Zeocin, carbenicillin and gentamycin merely slowed the growth of the *Herpetosiphon* strain. Streptomycin, chloramphenicol and apramycin were chosen as selective markers for all further experiments.

4.1.2 Transformation of vegetative *Herpetosiphon* cells

Various methods for delivery of plasmid DNA were applied to *Herpetosiphon* cells. Though extensive research and modifications of available protocols were conducted, none of these procedures succeeded in the transformation of *Herpetosiphon* sp. 060. The filamentous growth of the strain proved to be a problem for all methods, as the filaments tend to form aggregates. Their uneven distribution in a liquid culture impedes the establishment of a growth curve, as a reliable photometric measurement of growth is not possible. Alternative quantification methods like microscopic cell counting, determination of total protein or cell mass were not feasible either for the same reason. Treatment of the cultures with a homogeniser had little effect.

Electroporation (3.3.8) of the cells showed that they readily regenerated even after application of high voltage (up to 25kV / 1cm) in mock transfections, but did not grow on selective plates. The treatment with glycine (3.3.7) resulted in a breakdown of the filaments into smaller fragments and reduced growth rates. An application of 2% glycine (w/v) was sufficient to impair the formation of filaments, while the addition of 5% glycine resulted in severe cell damage. The best time span for electroporation seemed to be within the first 24h of growth. Longer incubation resulted in a thick culture containing clumps of intertwined cell filaments, which favoured arcing during electroporation.

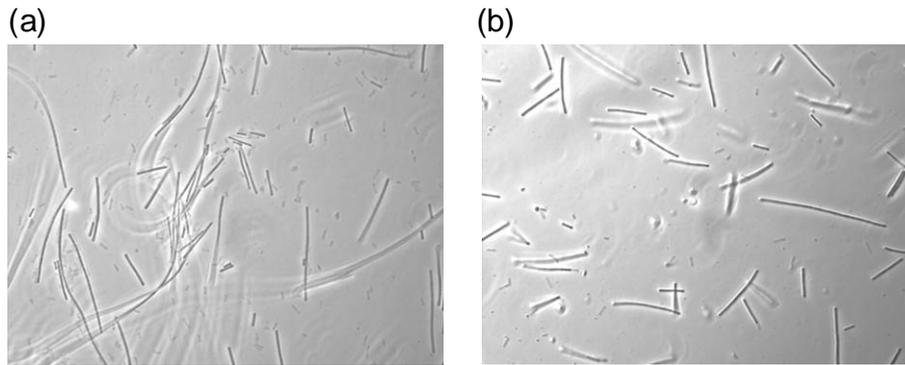


Figure 4-1: Effect of glycine on the growth of *Herpetosiphon* sp. 060

3% (a) and 5% (b) glycine were added to the medium after 24h and the culture was grown for an additional 24h; observation of filaments with phase-contrast microscope (x1000)

In sonication experiments (3.3.9) *Herpetosiphon* cells proved to be quite resistant to ultrasound treatment, as cells in mock transfections readily regenerated even after 60s of sonication.

Similar observations were made for chemical transformation methods (3.3.3) and conjugation experiments (3.3.10). Regardless of varying regeneration periods after treatment, transformants of *Herpetosiphon* were identified neither on selective agar plates nor in liquid culture.

4.1.3 Transformation of *Herpetosiphon* protoplasts

The formation of protoplasts (3.3.5) was conducted as a means to circumvent problems caused by cell filaments or cell wall structures. Protoplast formation was observed over a time of four hours. First protoplasts could be detected after 30 min, but the filaments remained mainly intact even after two hours incubation with lysozyme. Formation of protoplasts was observed at the tips of the filaments, sometimes forming small aggregates. Within the filaments, cells seemed to shrink leaving spaces between them and revealing the sheath-like structure (1.4.2) of the filaments. After four hours the majority of the cells were transformed to protoplasts, though fragments of filaments still remained. Neither by filtration nor by gradient centrifugation it was possible to completely remove the vegetative cells from the protoplasts.

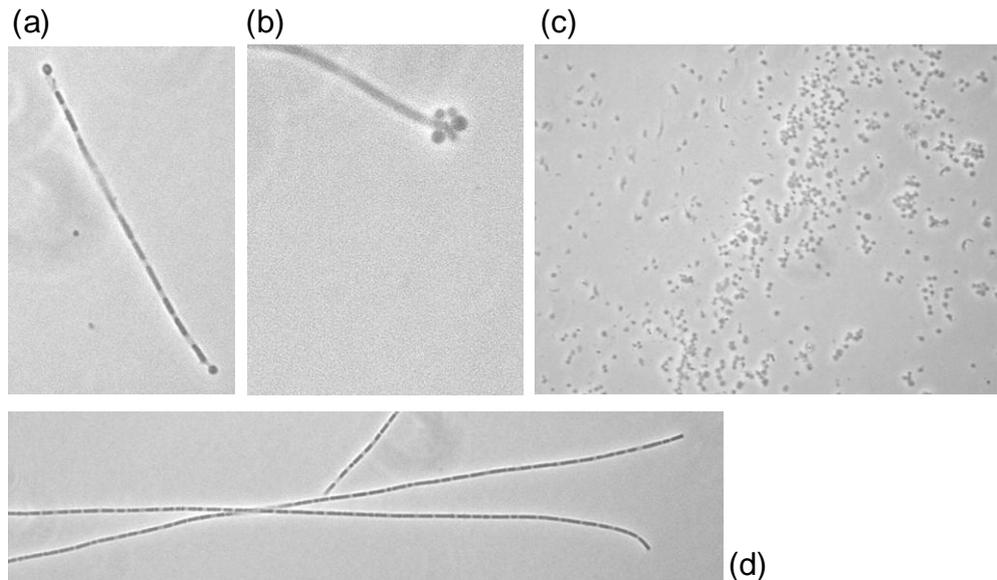


Figure 4-2: phase-contrast micrographs (1000x) of *Herpetosiphon* sp. 060

(a) beginning protoplast formation at the ends of a filament after 2h lysozyme treatment; (b) aggregation of protoplast at the tip of a filament after 2h lysozyme treatment; (c) protoplasts of *Herpetosiphon* after 4h lysozyme treatment; (d) filaments with retracting cells, revealing a sheath-like structure in the clear spacing after 2h lysozyme treatment

Regeneration on non-selective agar plates was readily observed, but due to the mentioned presence of vegetative cells it is uncertain if the regenerated cultures really originated from the protoplasts. Chemical (3.3.6) and electro-physical (3.3.8.1) methods were applied to the protoplasts without yielding successful transformation.

4.2 Genome sequencing of *Herpetosiphon* sp. 060

The elucidation of the complete putative biosynthetic gene cluster of siphonazole necessitated the access to genome DNA sequence information. For this purpose, 454 whole genome sequencing was carried out by GATC Biotech AG (Mannheim, Germany) on a Roche GS FLX Titanium sequencer (3.4.12). For this reaction genomic DNA from *Herpetosiphon* sp. 060 with a concentration of ~400 ng/μl was provided (3.4.5).

In the course of this approach 158,535,259 bp were assembled to 1,663 contigs. Taking the genome size of *H. aurantiacus* with a size of 6,346,587 bp as a reference, 25fold coverage of the genome was achieved by this sequencing. The average contig size was 7,809 bp with the largest contig stretching over 121,725 bp.

4.2.1 Genome search for the putative siphonazole gene cluster

The genome assembly was screened for contigs harbouring parts of the putative siphonazole cluster. For this purpose an .embl formatted file containing the sequences of all contigs was generated and used as database in the CLUSEAN software (Weber et al., 2009). In previous work by Ö. Erol the fosmids EC9, EC10 and CF4 were found to carry NRPS or PKS parts. DNA sequences of subclones from these fosmids as well as end sequences of the inserts were provided by Ö. Erol (personal communication). This set of sequences was subjected to a BLAST search using the genome assembly as reference database. Thus, the subclones were aligned to a total of six contigs, which are listed in table 4-2.

Contigs 1518 and 0668 are covered by all three fosmids, while contig 1280 only appears on fosmids CF4 and EC10. Contig 1357 is unique to fosmid EC9 and contig 1462 only aligns with fosmid CF4. These data indicate that the three fosmids are at least partially overlapping and may carry parts of the same cluster. Subclones EC-5 and CF4-7 both align with contig 0668 as well as with 1518, suggesting that the two contigs lie in close vicinity to each other. The same can be assumed for contigs 1517 and 1518, which both align with subclones CF4-4 and CF4-6.

Table 4-2: Contigs identified by alignment with fosmid subclones

Subclones were sequenced with T7 sequencing primer; subclones homologous to contigs 0668 *and* 1518 are displayed in bold letters; subclones homologous to contigs 1517 *and* 1518 are underlined

contig	length (bp)	homology to subclone
0668	8496	EC9-5; CF4-7; CF4-9 ; EC10-1 ; EC10-3
1280	23571	CF4-2 ; CF4-8 ; CF4-10 ; EC10-6 ; EC10-7 ; EC10-10
1357	503	EC9-5
1462	1762	CF4-5
1517	37170	EC9-2 ; EC9-7 ; <u>CF4-4</u> ; <u>CF4-6</u>
1518	9180	EC9-5; <u>CF4-4</u> ; <u>CF4-6</u> ; CF4-7; EC10-5 ; EC10-8 ; EC10-9

4.2.2 Annotation of the contigs from genome sequencing

To gather further information on possible cluster parts, the genome assembly was analysed using the CLUSEAN pipeline (3.1.13). Through the implementation of different databases all open reading frames (ORFs) with a minimum size of 300 bp were assigned putative protein functions. PKS or NRPS related motifs were identified for all contigs that aligned with the fosmid subclones. Thus, a truncated A domain was identified at one end of contig 0668, followed by parts of a PKS cluster. They are encoded in one single ORF, which stretches over the whole contig. Contig 1280 ends with a fragment of an A domain, preceded by an AT. The other ORFs on this contig have no functions related to a PKS or NRPS pathway. Two ORFs were identified on contig 1518, which both contain parts of a PKS cluster including two Ox domains. The short contig 1462 harbours an A domain and a PCP domain, while contig 1357 contains a partial C domain. Five ORFs encoding for PKS and NRPS modules were found at the end of contig 1517. The modules include two Ox domains and a TE. The adjacent ORFs seem to be unrelated to the PKS/NRPS cluster.

The AT domain on contig 1280 is the only one that was found on these contigs. The arrangement of the predicted domains suggests that contig 1517 encodes the end of a putative biosynthetic gene cluster, while contig 1280 contains the start of a *trans*-AT type PKS cluster.

The domain predictions were refined by applying a BLASTp (3.1.13) search with the translated amino acid sequences of the ORFs in question. An analysis of conserved domain structures revealed the AT to be linked to an oxidoreductase function. Furthermore, an α/β -hydrolase and an aldolase were found in front of the thioesterase on contig 1517.

The specificity of the adenylation domains (see also 4.4.1) was investigated with the help of the NRPSPredictor provided by the University of Tübingen (Rausch et al., 2005). Hence, specificity for threonin was assigned for the A domain on contig 1462 and for glycine for the A domain on contig 1517. The fragments on contigs 1280 and 0668 were insufficient to be analysed. Figure 4-3 summarises the results from this preliminary annotation.

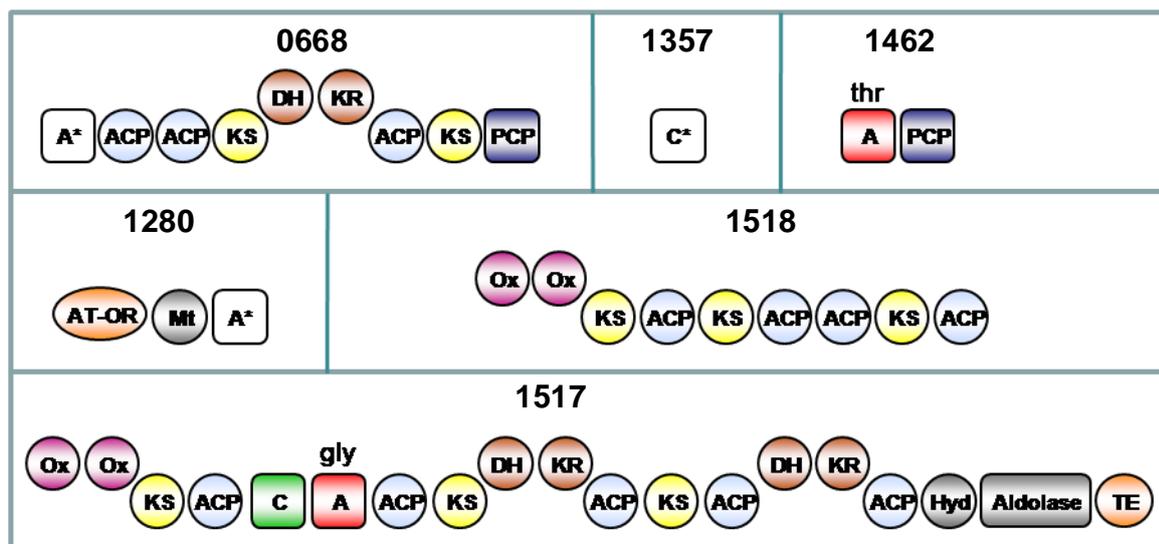


Figure 4-3: predicted domain structure encoded on selected contigs from genome sequencing Preliminary annotation without investigation of domain functionality; contig no. is given in each box; * indicates a truncated domain; predicted A domain specificity is given above the domain; A, adenylation domain; ACP, acyl carrier protein; C, condensation domain; DH, dehydratase; Hyd, hydrolase; KR, ketoreductase; KS, ketosynthase; Mt, methyltransferase; OR, oxidoreductase; Ox, oxidase; PCP, peptidyl carrier protein; TE, thioesterase

4.2.3 Other gene clusters discovered from genome sequences

The contigs from the 454 sequencing were subjected to a search for C domains and KSs using the NaPDos database (Ziemert et al., 2011). This approach detected 38 C domains encoded by ORFs on 12 different contigs. 34 C domains showed between 30% to 50% identity to enzymes from known pathways, while no homology was found for four of the predicted domains. One C domain was detected on contig 1517.

Furthermore, NaPDos predicted 19 KS domains on 13 contigs. 12 of them showed between 35 – 61% identity to enzymes from either secondary metabolite pathways or fatty acid biosynthesis). One KS exhibited no homology to any pathway in the database. The search run found three KSs encoded on contigs 1517 and 1518 each. They were all predicted as *trans*-type PKS and possess the greatest homology to domains from the leinamycin biosynthesis with identities of 43% – 55%. The only exception was one KS encoded on contig 1517 with 58% identity to the kirromycin biosynthetic pathway. A full compilation of all NaPDos- search results is given in the appendix (8.2).

Contigs 1517 and 805 were the only fragments that were identified to contain both a KS and a C domain. Contig 805 contains six ORFs (orfA – orfF) that encode a complete hybrid *cis*-AT type PKS/NRPS cluster. This cluster was further characterised using BLASTp and the NRPS predictor. The NRPS part comprises three epimerisation domains, an ornithine specific A domain and an A domain of unknown specificity. The KS of orfB was annotated as iterative type. Additionally, an acetyl ornithine aminotransferase function was found in this ORF. BLASTp showed that this cluster is homologous to a biosynthetic cluster in *H. aurantiacus*, whose product is yet unknown. OrfA – orfE agree with the order of genes haur1856 – haur1861 and exhibit great sequence identity to their counterparts in *H. aurantiacus*. Furthermore, in both strains genes were found in vicinity to this cluster that are predicted to encode the enzymes for the biosynthesis of a lantibiotic.

Table 4-3: Hybrid PKS/NRPS cluster on contig 805

Identities of the translated amino acid sequences from orfA – orfE to the homologous *H. aurantiacus* enzymes and their respective predicted domains; specificities are indicated in superior and inferior letters as far as known; OAT, acetyl ornithine aminotransferase

ORFs on contig 805	closest homolog (identity on protein level)	predicted domains
orfA	Haur1856 (92%)	Amp-ligase
orfB	Haur1857 (87%)	ACP-KS ^{iter} -AT-ACP-OAT-C-A [?] -Mt-ACP
orfC	Haur1858 (88%)	^L C _L -A ^{asn} -PCP
orfD	Haur1859 (89%)	C ^{epim} -D _{C_L} -A ^{thr} -PCP-C ^{epim} -D _{C_L} -A ^{ser} -PCP
orfE	Haur1860 (84%)	KS-AT-KR-ACP
orfF	Haur1861 (84%)	^L C _L -A ^{orn} -ACP-C ^{epim} -D _{C_L} -A-ACP-TE

4.3 Complete elucidation of the putative siphonazole gene cluster

The comparison of the fosmid subclones with the draft genomic data showed that four of the six identified contigs were covered by more than one fosmid clone, indicating that they may be fragments of one single biosynthetic gene cluster. Additionally, with exception of the AT encoded on contig 1280, none of the predicted modules contained an AT domain. Thus, it was hypothesised that 1280 could be the start of a *trans*-AT type cluster, whose end could be marked by the TE encoded on contig 1517. The presence of two A domains for threonine and one for glycine together with the occurrence of Ox domains encoded on two contigs corroborated the assumption that the putative cluster parts are involved in siphonazole biosynthesis. Therefore, further efforts were taken to obtain the full sequence of the gene cluster.

4.3.1 Gap closure and alignment of the contigs

The coverage of the genome sequences and the number of modules already identified on the contigs gave rise to the assumption that the sequence gaps between the fragments should lie within the range of a few thousand base pairs and thus could be closed by a PCR-based approach (3.4.1). If not stated otherwise, genomic DNA isolated from *Herpetosiphon* sp. 060 was used as template for all reactions.

4.3.1.1 Positioning of contig 1462

Considering the structure of siphonazole, it was proposed that contig 1462 would either precede contig 1517 or 1518. Therefore, PCR-primers were deduced from the sequences of the contigs in question and standard PCR-reactions performed. PCR with forward primer cb1462_f yielded products of similar size for both reverse primers (cb1517_r and cb1518_r). Sequencing of the products revealed homology only to contig 1517, but not to 1518. Therefore the latter product was considered an artefact due to unspecific primer binding caused by the strong similarities between the contigs. These results were confirmed by PCR with the more specific primer pair 1462ktr. The sequencing of both PCR-reactions led to the closure of the gap between contig 1517 and 1462.

4.3.1.2 Gap closure between contigs 0668 and 1280

The analysis of the sequences suggested that contig 0668 may follow 1280 and that the fragments of an A domain encoded on both contigs are actually part of one domain. The gap between the contigs was closed with one PCR step (primers 1280seq_neu & 668seq-1280) and an additional sequencing reaction (3.4.11) from fosmid CF4 to ensure accuracy of the sequence. Both sequencing reactions confirmed that contig 1280 is indeed directly adjacent to 0668 with no bp in between, thus showing that the protein fragments are indeed part of the same A domain.

4.3.1.3 Positioning of contig 1357

The genome assembly was screened with the SeqMan tool from the DNASTAR Lasergene 8 software to search for further reads that align with the small contig 1357. Thus, sequences were found that suggested a connection of this fragment to contig

1518. This was confirmed by sequencing of a PCR-product generated with a primer pair specific to contig 1357 and the end of 1518 (1518-1517 & EC9fwd).

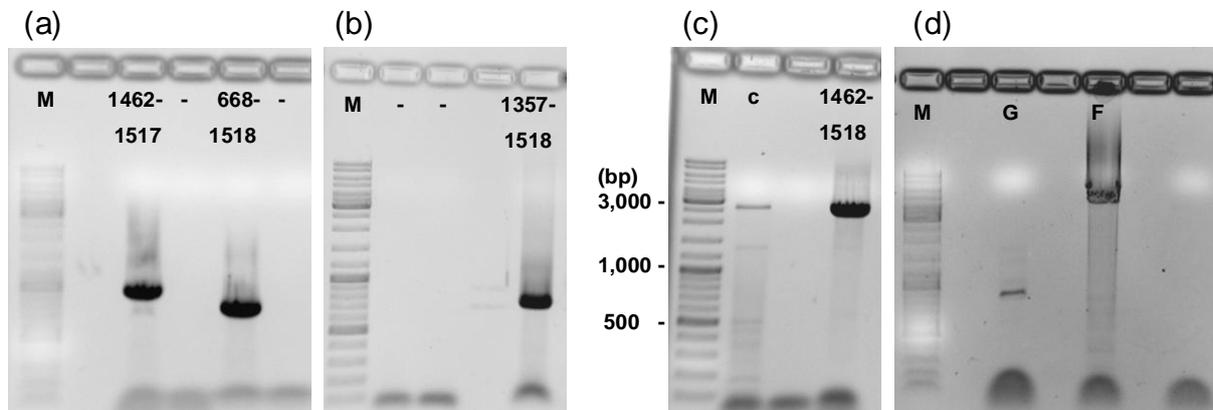


Figure 4-4: PCR-reactions from genomic DNA bridging the contigs

M, size marker; -, negative control; (a) PCR connecting contig 1462 to 1517 and 0668 to 1518; (b) PCR connecting contig 1357 to 1518; (c) PCR connecting contig 1462 to 1518; c, colony-PCR; (d) PCR connecting contig 1518 to 0668; G, PCR from genomic DNA; F, PCR from fosmid clone EC10

4.3.1.4 Gap closure between contigs 1518 and 1462

Contig 1518 was hypothesized to lie in front of contig 1462. This was affirmed by a PCR with primers 1518-1517 and 1462seq between the ends of the contigs, which yielded a product of about 3,000 bp. As a fragment of this size is too long to be completely sequenced under standard conditions both ends were sequenced and new primer pairs deduced from the gained information. These primers were used for a sequencing reaction from fosmid EC9. The obtained sequences were aligned in a preliminary assembly. To close the last gaps and cover regions that had not been sequenced properly, the three primer pairs gap1, gap2 and gap3 were constructed. PCR with these primers resulted in overlapping products that covered the whole gap between the contigs. Together with the already obtained information, the gap could be closed, hereby also confirming the alignment of contigs 1518 and 1357.

Through the additional sequence information a pair of cyclisation domains was discovered. It was also revealed that contig 1357 is actually part of the first of these domains rather than a fragment of a classical C domain.

Concurrently, differences between the sequence of contig 1462 from the genome sequencing and the PCR-products were discovered. Therefore, a thorough PCR-screening was carried out with primer pairs covering contig 1462 and adjacent regions (1462-A fwd & rev) to gather reliable sequence information. The fragments were aligned and compared to the existing contig 1462, which confirmed the discovered differences and revealed more deviations. The new construct was analysed with BLASTp, which yielded the prediction of an A domain containing the complete conserved region. Thus, it was concluded that the PCR originated alignment was correct and the provided sequence for contig 1462 proved to be erroneous. The corresponding sequences were replaced with the new results, thereby also removing two incorrect stop codons. The corrected sequences were used further on in this study.

4.3.1.5 Gap closure between contigs 1518 and 0668

The cluster information obtained up to this point was compared with the chemical structure of siphonazole for a first evaluation of the identified PKS and NRPS modules. A rough assignment of cluster parts to the skeleton of the compound revealed that sequences for one complete NRPS module was still missing to provide the first threonine unit of siphonazole. Considering the average sizes of the DNA sequences encoding the necessary domains, the gap between contigs 1518 and 0668 was estimated to be at least 3,000 bp long. PCR-reactions from genomic DNA yielded only products of less than 1,000 bp in size and were therefore considered artefacts. To minimise the gap, sequencing reactions were carried out from the contig ends and new primers designed with the obtained information (primers 1518-EC10-fwd & EC10seq-rev). A PCR with fosmid EC10 as template (primers 668seq & 1518seq) successfully generated a product between 4,000 and 5,000 bp. In parallel to chapter 4.3.1.4 multiple PCR and sequencing steps (primers EC10seq2 fwd & rev and 1518seq_neu) were carried out and overlapping DNA fragments generated and aligned to the final sequence using BLAST and ClustalW2 alignment software (Larkin et al., 2007; Goujon et al., 2010).

An analysis of the gathered sequence information revealed the presence of a whole NRPS cyclisation module, comprising a pair of cyclisation domains, an A domain and two PCPs.

4.3.2 Organisation of the putative siphonazole gene cluster

The ORFs adjacent to those encoding the AT and the TE were translated into amino acid sequences and analysed with BLASTp to evaluate a possible role in the biosynthesis of siphonazole. The proteins encoded by orf1 and orf2 showed homology to enzymes from the primary metabolism of *H. aurantiacus*. The small protein encoded by orf13 could not be assigned to a specific function, while orf14 aligns with a DNA-polymerase. All four ORFs were considered to be unrelated to PKS or NRPS functions.

Table 4-4: BLASTp search results for the translated amino acid sequences from the ORFs surrounding the sph genes; length of the sequence is given in column two; identity is given as percentage of aligned amino acids

Gene	size (aa)	Highest Homology (protein level)	Identity (aa)	Accession No.
orf1	863	phosphoenolpyruvate synthase [<i>Herpetosiphon aurantiacus</i> DSM 785]	575/871 (66%)	YP_001544722
orf2	304	GCN5-like N-acetyltransferase [<i>Herpetosiphon aurantiacus</i> DSM 785]	175/299 (59%)	YP_001544430
orf13	205	hypothetical protein OTW25_04915 [<i>Ornithinibacillus</i> sp. TW25]	113/188 (60%)	ZP_08784265
orf14	358	DNA polymerase IV [<i>Clostridium perfringens</i> B str. ATCC 3626]	117/341 (52%)	ZP_02634987

These findings suggest that the putative siphonazole gene cluster comprises ten ORFs (figure 4-5), which extend over 50 kb. The genes were consequently denominated sphA – sphJ. Their translated protein sequences were subjected to a BLASTp alignment and an additional analysis with CLUSEAN. The deduced protein functions and predicted PKS or NRPS domains are summarised in table 4-5.

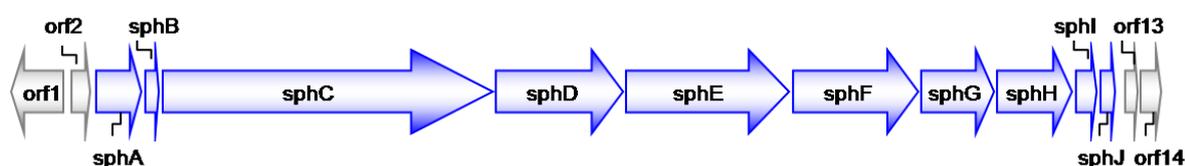


Figure 4-5: Organisation of the putative siphonazole gene cluster

Genes contributing to the biosynthesis of siphonazole are shown in blue; open reading frames without predicted NRPS/PKS functions are shown in gray; representations are drawn to scale according the length of the amino acid sequence

Table 4-5: BLASTp search results for the translated amino acid sequences from sph genes
Length of the amino acid sequence is given in column two; identity is given as percentage of aligned amino acids; column five shows the protein functions and NRPS and PKS domains according to the predictions derived from CLUSEAN and BLASTp search

Gene	Size (aa)	Highest Homology (GenBank accession)	Identity (aa)	Predicted domains
sphA	762	polyketide synthesis protein [<i>Bacillus subtilis</i>] (YP_004207767)	387/768 (50%)	<i>trans</i> -AT-OR
sphB	221	O-methyltransferase, family 3 [<i>Bacillus thuringiensis</i>] (ZP_04094215)	127/220 (58%)	O-methyltransferase
sphC	5451	AMP-dependent synthetase and ligase [<i>Clostridium papyrosolvens</i>] (YP_002505214)	803/2275 (35%)	A-ACP-ACP-KS-DH-KR-ACP-KS-PCP-Cyc-Cyc-A-PCP-PCP-Ox-Ox-KS
sphD	2089	beta-ketoacyl synthase [<i>Clostridium cellulolyticum</i>] (YP_002505214)	719/1951 (37%)	ACP-KS-ACP-ACP-KS-ACP
sphE	2693	Beta-ketoacyl synthase [<i>Clostridium papyrosolvens</i>] (ZP_08192883)	329/730 (45%)	Cyc-Cyc-A-PCP-PCP-Ox-Ox-KS-ACP
sphF	2059	NRPS/PKS [<i>Streptomyces albus</i>] (YP_002505214)	867/2062 (42%)	C-A-PCP-KS-DH
sphG	1186	beta-ketoacyl synthase [<i>Clostridium cellulolyticum</i>] (YP_002505214)	527/1214 (43%)	KR-ACP-KS-ACP
sphH	1228	Polyketide synthase [<i>Sorangium cellulosum</i>] (YP_001614779)	254/780 (33%)	DH-KR-ACP-Hydrolase
sphI	348	phospho-2-dehydro-3-deoxyheptonate aldolase [<i>Sphaerobacter thermophilus</i>] (YP_003320125)	184/285 (65%)	Aldolase
sphJ	243	thioesterase LchAD [<i>Bacillus licheniformis</i>] (YP_077643)	100/231 (43%)	TE

4.4 NRPS domains and modules

The putative siphonazole gene cluster encodes four NRPS and eight *trans*-AT type PKS modules. The functionality of the individual domains was evaluated by an examination of conserved regions within the amino acid sequence. Following a BLASTp search the domain sequences were aligned with homologous proteins derived from the BLAST hits and reference proteins taken from the conserved domain database (Marchler-Bauer et al., 2004, 2009, 2011). Thus, known signature motifs and catalytic residues were identified and assessed for integrity.

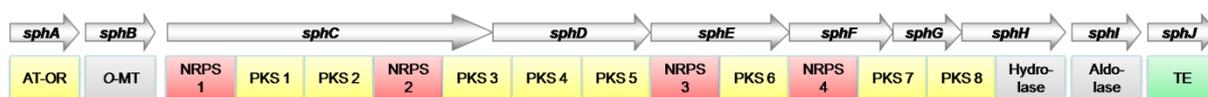


Figure 4-6: Modular organisation of the putative siphonazole biosynthetic enzymes

PKS parts are shaded yellow and NRPS parts red; other parts are shaded grey; thioesterase (TE) is shaded green; AT, acyltransferase; OR, oxidoreductase; O-MT, O-methyltransferase

The predicted domain functions show that NRPS-module 1 represents the initiation module of the cluster and characteristically lacks a C domain. In NRPS-modules 2 and 3 the normal C domain is replaced by tandem Cys domains, while module 4 exhibits a canonical domain organisation. In the following, the domains are either simply numerated (carrier proteins and KSs) or named after their respective proteins and, if necessary, their order within a peptide sequence (e.g. SphF-C, SphC-A1).

4.4.1 Adenylation domains

The A domains are responsible for the selection of cognate amino acids and their activation as aminoacyl adenylate. Stachelhaus and co-workers determined ten amino acids that form the substrate-binding pocket and confer specificity and derived a prediction model for the substrate specificity of an adenylation domain. In this study the prediction was carried out with the NRPSPredictor software (Rausch et al., 2005) for the four A domains that are encoded in the putative siphonazole cluster. As SphC-A2 and SphE-A are 100% identical in their DNA sequence, their prediction results were summarised.

Table 4-6: Results of the NRPSPredictor for the A domains encoded in the putative siphonazole cluster; the first line shows the sequence from which the Stachelhaus-code is deduced; predictions are based on the assumption that either several (large clusters) or only few (small clusters) substrate amino acids have the same properties; specificity alignments are based on comparison with known A domains; a) and c) show individual predictions; b) summarises the predictions for two genetically identical A domains

a) SphC-A1

Residues 8Å around substrate	RWLTFMPSLWEGAAICGGELNDYGATENGVLVNE
Stachelhaus-code	MPWAIGDVLK
Predictions (large clusters)	asp=asn=glu=gln=aad-like specificity Score:0.232448685175
Predictions (small clusters)	val=leu=ile=abu=iva-like specificity Score:0.135284593626
Specificity alignments	none

b) SphC-A2 and SphE-A

Residues 8Å around substrate	LTTHFDFSVWEGNQVFGGEINMYGITETTIVHVTY
Stachelhaus-code	DFWNVGMVHK
Predictions (large clusters)	ser=thr=ser-thr=dht=dhpg=dpg=hpg-like specificity Score:1.3927863256
Predictions (small clusters)	thr=dht-like specificity Score:1.31193117612
Specificity alignments	threonine

c) SphF-A

Residues 8Å around substrate	FAMTFDIAGLELQALCGGEWNLYGPTETTTIWSTA
Stachelhaus-code	DILQLGLIWK
Predictions (large clusters)	gly=ala=val=leu=ile=abu=iva-like specificity Score:1.688107797
Predictions (small clusters)	gly=ala-like specificity Score:1.2144130992
Specificity alignments	glycine

The Stachelhaus-code of SphC-A1 showed no homology to any A domain of known specificity. Also, the small and large cluster predictions have very low score in this case. According to the structure of siphonazole, SphC-A1 supposedly activates a dihydroxy-benzoic acid, which does not agree with the prediction of an aspartate or valine like specificity. SphC-A2 and SphE-A were clearly assigned specificity for threonine and SphF-A was predicted to adenylate glycine. These findings match the arrangement of amino acids in siphonazole.

The A domain sphC-A1 was subjected to a further analysis by the NRSPredictor 2 software (Röttig et al., 2011). Additionally to the already attained results, specificity for a hydrophilic substrate was predicted with a score of 0.118492 and a precision of 0.94. The nearest neighbour based on the Stachelhaus-code has a similarity of 50% and is specific for phenylalanine, which clusters in the hydrophobic-aromatic group of substrates.

4.4.2 Peptidyl carrier proteins

PCP domains belong to the superfamily of PP-binding proteins (GenBank accession: cl0936). They are characterised by a conserved serine residue, which accepts a phosphopantetheine arm as prosthetic group. This arm is essential for the transport of amino acids and peptides within the cluster. Marahiel and co-workers identified the signature motif L-G-G-(DH)-S-L around this residue (Schwarzer et al., 2003). Five

PCP domains were found in the analysis of the putative siphonazole gene cluster and examined by multiple sequence alignment.

PCP1	A	Y	G	L	D	S	I	32
PCP2	S	A	G	A	T	S	L	30
PCP3	S	A	G	A	T	S	L	30
PCP4	S	A	G	A	T	S	L	30
PCP5	S	A	G	A	T	S	L	30

Figure 4-7: Multiple sequence alignment of PCP domains derived from sph genes

Residues critical for the domain function are coloured green; other residues corresponding to the consensus motif are coloured yellow

The alignment (figure 4-7) confirms the presence of the essential serine in all domains. The signature, however, is poorly conserved with the exception of a glycine and a lysine residue.

4.4.3 The condensation domain of sphF

Marahiel and co-workers described seven consensus motifs for condensation domains in NRPS systems (Marahiel et al., 1997; table 4-7). The active site is composed of an H-H-x-x-x-D-G core motif, in which the second histidine and the glycine residue are critical for the condensation activity of the domain (Stachelhaus et al., 1998). The aspartate was found to be of structural importance, while the first histidine participates in the correct folding of the domain. In further works R62, R67 and W202 were identified as additional important residues for the structure and folding of the C domain (Bergendahl et al., 2002; Rausch et al., 2007).

Table 4-7: signature motifs for condensation domains (according to Marahiel et al., 1997)

Catalytic and other important residues are printed in bold letters

C1	S-x-A-Q-R-(LM)-(WY)-x-L
C2	R -H-E-x-L- R -T-x-F
C3	M- H-H -x-I-S- D-G -(WV)-S
C4	Y-x-D-(FY)-A-V-W
C5	(IV)-G-x-F-V-N-T-(QL)-(CA)-x-R
C6	(HN)-Q-D-(YD)-P-F-E
C7	R-D-x-S-R-N-P-L

A sequence alignment of SphF-C was performed (figure 4-8) with C domains from gramicidin (GrsB) and surfactin (SrfA-C) biosynthesis and two similar proteins from *Streptomyces albus* and *Clostridium cellulolyticum* derived from BLASTp analysis.

	Motif C1										Motif C2									
SphF-C	S	E	G	Q	K	A	L	W	11	R	H	P	L	L	A	A	R	L	58	
<i>S.albus</i>	S	A	G	Q	R	A	L	S	11	R	H	P	Q	L	R	A	R	I	58	
<i>C.cel.</i>	S	E	G	Q	K	G	L	W	54	R	H	P	A	L	R	T	R	I	101	
GrsB	S	P	M	Q	E	G	M	L	23	R	Y	D	V	F	R	T	T	F	70	
SrfA-C	S	P	M	Q	E	G	M	L	23	R	Y	D	V	F	R	T	V	F	70	

	Motif C3										Motif C4									
SphF-C	V	H	H	I	M	F	D	G	V	S	136	F	G	D	F	V	A	W	172	
<i>S.albus</i>	F	H	H	I	V	F	D	G	L	S	133	F	A	D	F	V	D	W	164	
<i>C.cel.</i>	F	H	H	I	I	Y	D	G	M	S	180	Y	K	D	Y	V	Y	W	216	
GrsB	F	H	H	I	L	M	D	G	W	C	156	Y	K	Q	F	I	K	W	185	
srfA-C	Y	H	H	I	I	L	D	G	W	C	156	Y	K	D	Y	I	K	W	185	

	Motif C5										Motif C6									
sphF-C	I	G	Y	F	M	N	M	V	V	I	290	H	G	D	Y	P	L	F	323	
<i>S.albus</i>	L	G	Y	F	M	N	M	V	V	L	282	H	S	D	Y	P	L	F	315	
<i>C.cel.</i>	L	G	Y	F	M	N	I	V	V	I	343	H	S	D	Y	P	V	S	367	
GrsB	V	G	L	F	I	N	T	L	P	L	305	H	E	Y	F	P	L	Y	338	
SrfA-C	V	G	L	F	I	N	V	V	P	R	304	H	Q	Y	V	P	L	Y	337	

	Motif C7										Trp important for folding									
SphF-C	F	Y	F	Q	N	W	V	A	346	Y	W	L	190							
<i>S.albus</i>	F	Y	F	Q	N	W	V	E	338	Y	W	L	182							
<i>C.cel.</i>	F	Y	F	Q	N	W	M	D	395	Y	W	L	234							
GrsB	M	V	I	E	N	Y	P	L	362	Y	W	K	203							
SrfA-C	I	V	F	E	N	Y	P	L	358	Y	W	R	203							

Figure 4-8: Multiple sequence alignment of the C domain from SphF

Catalytic residues are coloured green; residues important for structure or folding are coloured blue; other residues corresponding to the consensus motifs are coloured yellow; reference sequences were taken from gramicidin (GrsB, P0C064) and surfactin (SrfA-C, 2VSQ_A) biosynthesis and similar proteins from *Streptomyces albus* (*S.albus*, ABS90470) and *Clostridium cellulolyticum* (*C.cel.*, YP_002505216)

Motifs C1 and C2 only show partly correspondence to the canonical consensus sequences, while the catalytic core moieties in C3 are strictly conserved. In motif C4 the order of the valine and alanine residue are switched in SphF-C. Almost no homology to the consensus sequence could be observed in motif C7, where only the asparagines seems to be conserved, followed by an additional tyrosine in GrsB and SrfA. The sequence alignment identified R50 and W189 in SphF-C as the residues correlating to the structurally important R62 and W202 of the reference protein (TycB1; Bergendahl et al., 2002). Only for R67 (R55 in SphF-C) no direct correspondent was found, though there is an arginine present at position 57, which may take over the same function. Due to the presence of virtually all critical residues, SphF-C was characterised as an active domain.

4.4.4 Heterocyclisation modules

In this module type the usual C domain is replaced by a Cyc domain, which not only catalyses the condensation reaction, but also the formation of heterocycles from serine, threonine or cysteine side chains (Roy & Walsh, 1999). The resulting oxazoline and thiazoline rings can be oxidised to thiazole and oxazole rings by optional Ox domains. The two cyclisation modules in SphC and SphE exhibit several peculiarities. An alignment of the modules revealed a 98% identity of their DNA sequence over a range of 4,750 bp. This area stretches from the Cyc domains to the carrier proteins, but does not include the Ox domains. Through an analysis using the Artemis software (Rutherford et al., 2000; Carver et al., 2011), a deviance in the GC-content of this region was detected (figure 4-9). While the average GC-content of the sph genes lies at 53.1%, the ratio raises to 68% within the cyclisation modules.

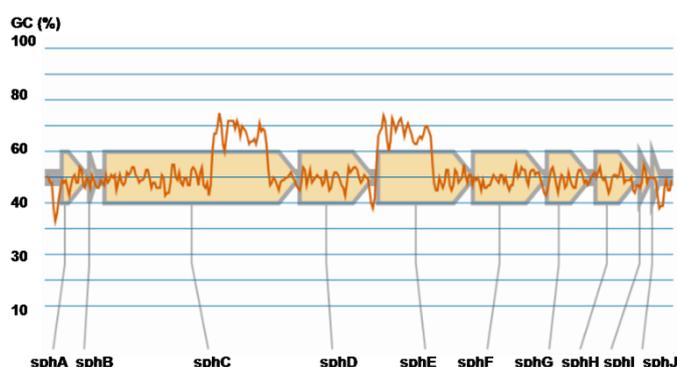


Figure 4-9: GC-content of the putative siphonazole biosynthetic genes

4.4.4.1 Cyclisation tandem domains

In NRPS-modules that form thiazoline or oxazoline rings by intramolecular heterocyclisation the C domain is replaced by a Cyc domain (Keating et al 2002). These domains are capable of catalysing the necessary condensation, cyclisation and dehydration reactions (Marshall et al., 2002). In the putative siphonazole cluster an unusual arrangement of two pairs of tandem Cyc domains was identified in SphC and SphE. As this suggests a redundancy or inactivity within these domains (Marshall et al., 2002), they were subjected to a careful analysis of signature motifs.

The catalytic core motif H-H-x-x-x-D-G (Stachelhaus et al., 1998) of regular C domains is replaced by a D-x-x-x-x-D-x-x-S motif in Cyc domains, in which the two aspartate moieties are essential for the condensation and cyclisation reactions (Keating et al., 2000). Seven further signature sequences are described for Cyc domains (Schwarzer et al., 2003; table 4-8). These conserved regions were analysed by a multiple sequence alignment with Cyc domains from the bleomycin, bacitracin, epothilone and pyocheline biosynthesis. The results are given in figure 4-10.

Table 4-8: Signature motifs for cyclisation-domains (according to Schwarzer et al., 2003)

Catalytic core residues are printed in bold letters

Cy1 **F**-P-L-(TS)-x-x-Q-A-Y-x-x-G-R

Cy2 R-H-x-(IM)-L-(PAL)-x-(ND)-G-x-Q

C3 **D**-x-x-x-x-**D**-x-x-S

Cy3 L-P-x-x-P-x-L-P-L-x-x-x-P

Cy4 (TS)-(PA)-3x-(LAF)-6x-(IVT)-L-x-x-W

Cy5 (GA)-D-F-T-x-L-x-L-L

Cy6 P-V-V-F-T-S-x-L

Cy7 (ST)-(QR)-T-P-Q-V-x-(LI)-D-13x-W-D

	Motif Cy1														
SphC-Cyc1	F	G	L	S	A	V	Q	E	A	F	V	V	G	Q	17
SphC-Cyc2	L	P	L	S	A	V	Q	R	S	Y	L	L	A	R	17
SphE-Cyc1	R	S	V	D	G	G	N	D	I	W	V	V	G	R	21
SphE-Cyc2	L	P	L	S	A	V	Q	R	S	Y	L	L	A	R	17
BlmIV-Cyc	F	P	L	T	D	V	Q	R	A	Y	Y	V	G	R	73
BacA-Cy	F	P	L	T	G	I	Q	L	A	Y	L	V	G	R	49
EposP-Cyc	F	P	L	T	D	I	Q	E	S	Y	W	L	G	R	85
PchE-Cyc	F	E	L	S	S	V	Q	Q	A	Y	W	L	G	R	118

	Motif Cy2								Core Motif C3							
SphC-Cyc1	R	H	P	M	L	T	M	Q	63	D	[4]	D	A	A	S	142
SphC-Cyc2	A	H	P	G	L	R	M	T	58	D	[4]	D	G	R	S	126
SphE-Cyc1	R	H	P	M	L	T	M	Q	74	D	[4]	D	A	A	S	153
SphE-Cyc2	R	P	S	R	P	A	H	D	58	D	[4]	D	G	R	S	127
BlmIV-Cyc	R	H	P	M	L	R	A	V	120	D	[4]	D	A	H	S	198
BacA-Cy	R	H	P	M	L	R	T	I	95	D	[4]	D	D	S	S	174
EposP-Cyc	R	H	D	M	L	R	A	H	131	D	[4]	D	L	G	S	209
PchE-Cyc	R	H	P	M	L	R	A	R	165	D	[4]	D	V	E	S	243

	Motif Cy3											
SphC-Cyc1	I	T	G	G	P	A	L	P	W	[4]	P	209
SphC-Cyc2	R	P	L	G	P	T	L	P	S	[3]	Q	189
SphE-Cyc1	I	T	G	G	P	A	L	P	W	[4]	P	220
SphE-Cyc2	R	P	L	G	P	T	L	P	S	[3]	Q	190
BlmIV-Cyc	L	P	P	G	P	E	L	P	L	[3]	P	266
BacA-Cy	F	P	S	A	P	E	L	P	L	[3]	P	242
EposP-Cyc	L	P	P	P	P	T	L	P	M	[3]	P	277
PchE-Cyc	L	P	D	A	P	A	L	P	L	[3]	P	310

	Motif Cy4										
SphC-Cyc1	A	P	[3]	L	[6]	I	L	R	S	A	266
SphC-Cyc2	S	L	[3]	L	[6]	V	L	A	A	W	229
SphE-Cyc1	A	P	[3]	L	[6]	I	L	R	S	A	277
SphE-Cyc2	S	L	[3]	L	[6]	V	L	A	A	W	230
BlmIV-Cyc	S	P	[3]	L	[6]	V	I	T	A	W	315
BacA-Cy	T	P	[3]	L	[6]	I	L	A	Y	W	291
EposP-Cyc	T	P	[3]	I	[6]	V	I	G	R	W	326
PchE-Cyc	T	L	[3]	F	[6]	V	L	A	R	W	359

	Motif Cy5								Motif Cy6										
SphC-Cyc1	A	P	V	V	S	T	T	L	V	303	P	V	V	F	T	A	M	L	360
SphC-Cyc2	G	D	F	T	A	L	R	W	L	264	P	L	V	F	T	E	L	L	319
SphE-Cyc1	A	P	V	V	S	T	T	L	V	314	P	V	V	F	T	A	M	L	371
SphE-Cyc2	G	D	F	T	A	L	R	W	L	265	P	L	V	F	T	E	L	L	320
BlmIV-Cyc	G	D	F	T	S	L	S	L	L	350	P	V	V	F	T	S	D	L	409
BacA-Cy	G	D	F	T	S	L	M	L	L	326	P	I	V	F	T	S	V	L	386
EposP-Cyc	G	D	F	T	S	M	V	L	L	361	P	V	V	L	T	S	A	L	420
PchE-Cyc	G	D	F	T	T	L	L	L	L	394	P	V	V	F	A	S	N	L	450

	Motif Cy7												
SphC-Cyc1	T	Q	T	P	N	V	W	L	D	[13]	W	D	404
SphC-Cyc2	S	Q	T	P	G	V	A	I	D	[14]	W	D	357
SphE-Cyc1	T	Q	T	P	N	V	W	L	D	[13]	W	D	415
SphE-Cyc2	S	Q	T	P	G	V	A	I	D	[14]	W	D	358
BlmIV-Cyc	S	Q	T	P	Q	V	H	L	D	[13]	W	D	458
BacA-Cy	T	R	T	S	Q	V	Y	I	D	[13]	W	D	431
EposP-Cyc	T	Q	T	P	Q	L	L	L	D	[13]	W	D	464
PchE-Cyc	S	Q	T	P	Q	V	W	L	D	[13]	W	D	495

Figure 4-10: Multiple sequence alignment of the Cyc domains from SphC and SphE

Residues corresponding to the consensus motifs are coloured yellow; C3 shows the alignment for the catalytic core; catalytic residues are coloured green; reference sequences were taken from the biosynthesis of bleomycin (BlmIV-Cyc, AAG02364), bacitracin (BacA-Cy, O68006), epothilone (EposP-Cyc, AAF26925) and pyochelin (PchE-Cyc, AAD55800)

The two Cyc domain pairs of SphC and SphE show great similarity to each other and have identical DNA sequences over long stretches. SphE-Cyc1 differs from SphC-Cyc1 only in a section of 29 bp. Due to this deviation, SphE-Cyc1 shows almost no homology to signature motif Cy1, which is at least partly conserved in the corresponding domain SphC-Cyc1. SphE-Cyc2 deviates from SphC-Cyc2 in a stretch of 69 bp, which results in a different conservation of motif 2. All domains exhibit the essential catalytic aspartate moieties in the core motif C3. Motif Cy5 seems to be only conserved in the second cyclisation domain of each pair. Motif Cy2 generally shows the weakest conservation in all compared sequences. With the exception of the core motif and Cy7, the two domains of each pair in SphC and SphE exhibit differences in the conserved residue of the signature motifs, which complement each other. However, from this bioinformatic data alone it cannot be told, whether each tandem pair catalyses condensation and cyclisation in concerted action or whether one domain of a pair is redundant.

4.4.4.2 Oxidation domains

The ORFs sphC and sphE both encode a pair of Ox domains, which were identified as members of the McbC-like oxidoreductase superfamily (GenBank accession: cl02142). These oxidases are generally associated with cyclisation modules and are responsible for the oxidation of formed thiazoline and oxazoline rings to thiazole and oxazole rings. The domain architecture of the putative siphonazole gene cluster

follows the canonical arrangement, in which the Ox domain follows the corresponding PCP of the module (Schwarzer et al., 2003). Although little biochemical data is available for Ox domains, three signature motifs are described by Schwarzer and co-workers (table 4-9). A multiple sequence alignment was performed with Ox domains from epothilone, tubulysin and rhizoxin biosynthesis.

Table 4-9: Signature motifs for oxidation domains

Ox1	K-Y-x-Y-x-S-x-G-x-x-Y-(PG)-V-Q
Ox2	G-x-x-x-G-(LV)-x-x-G-Y-Y-Y-(HD)-P
Ox3	I-x-x-x-Y-G

Motif Ox1																
SphC-Ox1	K	Y	L	H	A	S	A	G	G	K	Y	A	I	Q	74	
SphE-Ox1	K	Y	L	Y	A	S	A	G	G	K	Y	A	V	Q	54	
SphC-Ox2	L	T	L	E	A	L	E	A	-	-	-	-	I	T	63	
SphE-Ox2	-	-	L	S	S	D	A	G	-	-	-	-	I	E	58	
EpoB	K	F	R	Y	P	S	A	G	S	T	Y	P	V	Q	74	
TubD	R	Y	L	Y	A	S	A	G	G	L	Y	P	V	Q	91	
RhiB	K	Y	L	Y	S	S	A	G	G	L	N	A	I	Q	67	
Motif Ox2																
SphC-Ox1	G	G	I	E	G	L	A	A	G	C	Y	Y	Y	H	P	96
SphE-Ox1	G	A	I	Q	G	L	V	A	G	C	Y	Y	Y	H	P	76
SphC-Ox2	Q	G	V	I	G	V	D	A	G	I	Y	R	Y	D	R	85
SphE-Ox2	E	R	I	A	G	Y	T	G	G	V	Y	R	Y	N	V	80
EpoB	G	R	I	E	G	V	D	E	G	F	Y	Y	Y	H	P	96
TubD	G	R	A	R	G	L	E	P	G	T	W	Y	Y	D	P	113
RhiB	E	R	I	D	G	L	A	G	G	V	Y	Y	Y	Q	P	89
Motif Ox3																
SphC-Ox1	L	R	P	I	Y	Q	132									
SphE-Ox1	L	R	P	V	Y	Q	115									
SphC-Ox2	L	P	S	H	Y	A	133									
SphE-Ox2	L	R	E	R	F	G	129									
EpoB	I	E	S	L	Y	G	143									
TubD	V	E	P	V	Y	A	162									
RhiB	I	A	P	I	Y	N	137									

Figure 4-11: Multiple sequence alignment of the Ox domains from SphC and SphE

Residues corresponding to the consensus motifs are coloured yellow; reference sequences were taken from the biosynthesis of epothilone (EpoB, AAF62881), tubulysin (TubD, CAF05649) and rhizoxin (RhiB, CAL69889)

The alignment (figure 4-11) shows a high correspondence to the Ox1 motif in the first Ox domains of SphC and SphE where only the (PG) residue is not conserved. In SphC-Ox1 also one tyrosine is replaced by a histidine. The second motif completely follows the consensus sequence, while in motif Ox3 the tyrosine moiety is the only residue that is stringently conserved.

The second domains of the cyclisation modules show virtually no conservation of the first motif and also exhibit deviations from Ox2. In SphE-Ox2 the tyrosine in motif Ox3 is replaced by phenylalanine. These results suggest a redundancy or inactivity for both SphC-Ox2 and SphE-Ox2.

4.5 PKS domains and modules

The domain prediction revealed that SphA represents the only AT of the cluster and that none of the eight PKS modules contains an AT of their own. They were therefore all classified as *trans*-AT type PKS modules.

4.5.1 The *trans*-AT and oxidoreductase SphA

A BLASTp analysis of SphA shows that the protein can be divided into two distinct domains. The N-terminal part (amino acids 1 – 295) encodes an AT, belonging to the acyl transferase 1 superfamily (GenBank accession: cl08282). ATs of PKS systems catalyse the transfer of acyl moieties between CoA and ACP using a serine-histidine catalytic diad. They exhibit a characteristic, highly conserved G-H-S-x-G motif surrounding the catalytic serine (Yadav et al., 2003).

Previous studies have shown that the substrate specificity of PKS ATs is influenced by conserved motifs in the peptide sequence. ATs with a specificity for methylmalonyl-CoA exhibit a [RQSED]-V-[DE]-V-V-Q motif 30 amino acids upstream of the catalytic serine, while malonyl-CoA specific ATs have a Z-T-x-\$-[AT]-[QE] sequence instead, with Z being a hydrophilic and \$ an aromatic amino acid (Smith & Tsai, 2007). Furthermore, it was observed that the X in the G-H-S-x-G motif is frequently a branched hydrophobic amino acid like valine or isoleucine. In other ATs usually glutamine or methionine is found in this position. Finally, the specificity is influenced by a further motif 100 amino acids downstream of the active serine. The

[YVW]-A-S-H sequence confers specificity for methylmalonyl-CoA, while a [MILV]-A-F-H motif is found in malonyl-specific ATs. Figure 4-12 depicts an alignment of the conserved regions with similar ATs derived from the BLAST analysis.

Motif 1							
SphA	G	H	S	V	G	90	
BaeE	G	H	S	L	G	87	
PksE	G	H	S	L	G	92	
DfnA	G	H	S	L	G	89	
ChiA	G	H	S	L	G	89	

Motif 2				Motif 3								
SphA	Q	T	Q	Y	T	Q	60	G	A	F	H	193
BaeE	Q	T	Q	F	T	Q	56	G	A	F	H	191
PksE	Q	T	Q	Y	T	Q	61	G	A	F	H	196
DfnA	K	T	Q	F	T	Q	58	G	A	F	H	192
ChiA	Q	T	Q	Y	T	Q	58	G	A	F	H	192

Figure 4-12: Multiple sequence alignment of the acyltransferase SphA

Residues corresponding to the consensus motifs are coloured yellow; reference sequences were taken from the biosynthesis of bacillaene (BaeE, YP_001421287), diffidin (DfnA, YP_001421800) and similar proteins from *Paenibacillus mucilaginosus* (PksE, YP_004640302) and *Pseudomonas fluorescens* (ChiA, AAM12912)

The signature motif around the catalytic serine is completely conserved in SphA and follows, like the reference proteins, the canonical consensus for malonyl-specific ATs. The first motif contains glutamine at the hydrophilic and tyrosine at the aromatic position. The [MILV]-A-F-H motif differs only in the first position. Regarding the alignment this seems to be a frequent deviation. The X position after the active serine is occupied by a valine residue, which supports a prediction for malonyl-CoA specificity.

The C-terminal part of the protein (amino acids 296 – 762) contains an oxidoreductase domain of the TIM phosphate binding superfamily (GenBank accession: cl09108). The primary structure was analysed with regard to the putative catalytic histidine residue (Ha et al., 2006) and subjected to multiple sequence alignment (figure 4-13) with oxidoreductases of analogous proteins from the biosynthesis of other secondary metabolites.

SphA	G	G	T	T	G	518
MmpIII	G	G	H	T	D	904
CorA	G	G	H	T	D	905
PksE	G	G	H	T	D	504
ChiA	A	G	S	S	D	544
DszD	G	G	P	T	D	589

Figure 4-13: Multiple sequence alignment of the oxidoreductase of SphA

The putative catalytic histidine is highlighted in green; deviations in this position are printed in red; reference sequences were taken from the biosynthesis of disorazoles (DszD, AAY32968), chivosazole (ChiA, AAY89048), coralopyronin A (CorA, ADI59523), bacillaene (PksE, YP_004640302) and mupirocin (MmpIII, YP_439869)

It was shown that the catalytic histidine is replaced by a tyrosine moiety in SphA. Likewise, substitution by serine and proline, respectively, was observed in the corresponding genes of chivosazole and disorazole biosynthesis, while the histidine residue is present in genes from the mupirocin, coralopyronin and bacillaene clusters. A function as enoylreductase was suggested for these oxidoreductase domains (Bumpus et al., 2008), but often remained putative as in the case of ChiA (Perlova et al., 2006) and DszD (Carvalho et al., 2005). The lack of the catalytic histidine suggests that the OR of SphA may be non-functional.

4.5.2 Ketosynthases

PKS KSs belong to the superfamily of condensing enzymes (GenBank accession: cl09938) and catalyse a decarboxylating Claisen-condensation. The active centre is formed by cysteine and two histidines, which are distributed between the N- and C-terminus of the peptide chain. The N- and C-terminal domains can be assigned to the beta-ketoacyl synthase superfamily (genebank accession cl09934 and cl08367 respectively). Structurally the beta-ketoacyl synthases are related to the thiolase family (pfam00108).

The active site of KSs consists of a highly conserved C-H-H catalytic triad (Zhang et al., 2006), in which the cysteine residue is part of a G-P-7x-C-S-S signature motif (Cortes et al., 1990). Eight KS domains were found to be encoded in the putative siphonazole gene cluster within the genes sphC to sphG. Multiple sequence alignment was used to screen for the active site residues. The results are given in figure 4-14.

	Motif 1						Motif 2					Motif 3				
KS1	G	P	7x	C	S	S	173	H	G	T	G	T	302	G	H	341
KS2	G	P	7x	C	S	S	175	H	G	T	G	T	303	G	H	344
KS3	G	P	7x	C	S	S	175	A	A	T	G	S	303	G	H	340
KS4	G	P	7x	C	S	S	172	H	G	T	G	T	305	G	H	343
KS5	G	P	7x	C	S	S	172	Q	G	T	G	T	302	G	H	341
KS6	G	P	7x	C	S	S	172	A	A	N	G	T	304	G	H	342
KS7	G	P	7x	C	S	S	170	H	G	T	G	T	300	G	H	344
KS8	G	P	7x	C	S	S	173	H	G	T	G	T	302	G	H	340

Figure 4-14: Multiple sequence alignment of the KS domains

Highly conserved residues are coloured yellow; residues of the catalytic triad are coloured green; deviations in the catalytic core are printed in red

The alignment shows that the cysteine signature box is highly conserved in all KSs. KS 3, 5 and 6 lack the first catalytic histidine, which is supposed to play a role in the activation of a catalytic water molecule during the condensation reaction. KS domains lacking the conserved H-G-T-G-T motif were found to be non-elongating (Nguyen et al., 2008). Therefore KS3, KS5 and KS6 were considered inactive in the cluster.

4.5.3 Acyl carrier proteins

ACP domains show the same G-G-(DH)-S-L signature motif as PCPs. Similarly, the conserved serine is essential for the attachment of the phosphopantetheine prosthetic group.

ACP1	E	L	G	F	D	S	V	32
ACP2	E	L	G	F	D	S	V	32
ACP3	E	F	G	L	D	S	I	32
ACP4	I	Y	G	I	N	S	Q	32
ACP5	E	Y	G	F	D	A	I	32
ACP6	E	Y	G	F	D	S	V	32
ACP7	N	Y	G	I	D	S	V	32
ACP8	R	Y	G	F	N	S	L	32
ACP9	N	Y	G	F	D	S	I	32
ACP10	S	Y	G	V	D	S	V	32
ACP11	E	Y	G	L	D	A	P	32
ACP12	N	Y	G	V	D	S	L	32

Figure 4-15: Multiple sequence alignment of the ACP domains

Residues corresponding to the consensus motifs are coloured yellow; the essential serine moiety is coloured green; deviations in this location are printed in red

In ACPs 5 and 11 the essential serine is replaced by an alanine residue (figure 4-15). This prevents the attachment of the prosthetic group, thus rendering these domains inactive. Like in the PCP domains, the preceding glycine residue is highly conserved in all domains, as well as the aspartate on position 30, which is changed only in ACPs 4 and 8. The remaining amino acids show a partly conserved pattern that differs from the canonical consensus sequence.

4.5.4 Dehydratases

The three DH domains encoded within genes sphC, sphF and sphH were all recognised by BLASTp as members of the PKS DH superfamily (GenBank accession: cl11739). DHs are characterised by a NADPH-binding motif with the conserved sequence H-x-x-x-G-x-x-x-P, which also contains the active histidine and proline residues (Donadio & Katz, 1992).

A sequence alignment showed that the motif is fully conserved in SphC-DH and SphF-DH, while in SphG-DH the active histidine is replaced by a tyrosine and the conserved glycine by aspartate. Alteration of the non-catalytic glycine also occurs in other PKS DHs, while the catalytic histidine and proline moieties are strictly conserved and essential for the function of the enzyme (figure 4-16).

SphC-DH	H	R	W	E	G	Q	A	L	L	P	40
SphF-DH	H	I	V	Q	G	Q	R	V	L	P	40
SphG-DH	Y	Q	V	A	D	S	Q	R	L	P	43
SorB	H	R	V	L	D	M	H	L	L	P	33
Ery chain A	H	V	V	G	G	R	T	L	V	P	42
<i>Beggiatoa</i>	H	V	V	G	S	Q	K	T	L	P	39

Figure 4-16: Multiple sequence alignment of the DH domains of SphC, SphG and SphH
Residues corresponding to the consensus motif are coloured yellow; catalytic residues are coloured green, deviations in these locations are printed in red; reference sequences were taken from the biosynthesis of sorangicin (SorB, ADN68477), erythromycin (Ery chain A, 3EL6_A) and a similar protein from *Beggiatoa* sp. SS (*Beggiatoa*, ZP_01997443)

It can be assumed that the alterations in the signature sequence are responsible for the fact that SphG-DH was not annotated during the CLUSEAN analysis, but was only detected by a manual BLAST search of SphH. DH domains from other gene clusters (e.g. nystatin biosynthesis), in which only the proline is conserved, are known to be non-functional (Pawlik et al., 2007). For this reason, it can be assumed that SphG-DH is also inactive.

4.5.5 Ketoreductases

CLUSEAN search identified three KR domains encoded in sphC, sphG and sphH. Conserved domain analysis by BLASTp assigned them to the class of ketoreductases subclass 2 complex short chain dehydrogenase/reductase (SDR) proteins (GenBank accession: cd08953) and members of the Rossmann fold NADP(+)-binding proteins superfamily (GenBank accession: cd09931). This family contains a characteristic G-x-G-x-x-G-x-x-x-A NADPH-binding motif. The catalytic triad in PKS KR domains consists of a Y-x-x-x-N motif and an active serine upstream of this sequence. A conserved lysine is located ~23 amino acids upstream of the serine moiety. The positions of this lysine and the asparagine seem to be switched in regard to classical SDRs (Reid et al., 2003).

	NADPH-binding motif										
SphC-KR	G	L	G	G	V	G	L	L	C	A	1889
SphG-KR	G	K	G	A	L	G	A	I	F	A	166
SphH-KR	G	A	G	N	V	G	F	K	L	C	504
<i>A.var.</i>	G	T	S	A	V	G	T	E	I	A	1220
PksM	G	A	G	Y	I	G	E	A	W	S	1540
<i>H.che.</i>	G	L	G	K	I	G	L	A	L	A	1125

	Catalytic Core										
SphC-KR	K	[23]	S	[12]	Y	G	Y	A	N	2024	
SphG-KR	K	[23]	S	[12]	Y	A	Y	A	N	297	
SphH-KR	K	[23]	S	[12]	Y	A	A	G	C	639	
<i>A.var.</i>	K	[23]	S	[14]	Y	A	A	A	N	1373	
PksM	K	[23]	S	[12]	Y	A	S	G	C	1675	
<i>H.che.</i>	K	[23]	S	[12]	Y	A	A	A	N	1347	

Figure 4-17: Multiple sequence alignment of the KR domains of SphC, SphG and SphH
Residues corresponding to the consensus motif are highlighted in yellow; catalytic residues are highlighted in green, deviations in these locations are printed in red; reference sequences were taken from similar proteins from *Anabaena variabilis* (*A.var.*, YP_324485), *Bacillus subtilis* (PksM, P40872) and *Hahella chejuensis* (*H.che.*, YP_434161)

The NADPH-binding motifs follow the consensus sequence in all three KR domains with only a single alteration in SphH-KR. The catalytic residues are in agreement with similar reductases encoded by other gene clusters as shown in figure 4-17. In SphH-KR the asparagine is replaced by a cysteine residue, which is also the case in PksM. The catalytic role of the asparagine moiety is still uncertain, but the KRs in PksM have been reported as functional (Butcher et al., 2007). Therefore, the same was assumed for SphH-KR. McDaniel and co-workers identified a diagnostic aspartate residue, which is present in all KRs that catalyse reductions to D-configuration and absent in KRs that reduce to L-configuration. This residue was found in the KR domains of SphC and SphG, but not SphH. Further motifs for the classification of KRs were identified by the Leadly group (O'Hare et al., 2006; Baerga-Ortiz et al., 2006). Neither the signature motifs L-D-D and P-x-x-x-N for type B KRs nor the PQQ signature and the W141 for type A KRs could be identified unambiguously in either of the KRs. A definite categorisation of the KR domains could therefore not be made.

4.5.6 The thioesterase SphJ

A BLASTp search with the amino acid sequence of SphJ confirmed the assignment of the enzyme to the protein family of thioesterases (pfam00975). Like ATs they contain an α/β -hydrolase fold domain and also share the same G-x-S-x-G motif around a catalytic serine. About 140 amino acids downstream of this sequence a G-x-H-F signature was identified surrounding a conserved histidine residue (Schneider & Marahiel, 1998). The BLAST search revealed that SphJ shows the greatest homology to thioesterases from *Bacillus* species (see table 4-5).

A multiple sequence alignment was performed with similar TE domains to identify the signature sequences. SphJ contains a strongly conserved G-H-S-x-G sequence, while the phenylalanine in the second motif is replaced by a methionine in all of the examined sequences (figure 4-18).

SphJ	G	H	S	L	G	89	G	P	H	M	Y	L	220
LchAD	G	H	S	M	G	89	G	G	H	M	Y	L	220
<i>B. cereus</i> AH1134	G	H	S	M	G	90	G	G	H	M	F	L	221
<i>B. thuringiensis</i> IBL 200	G	H	S	M	G	86	G	G	H	M	F	L	217

Figure 4-18: Multiple sequence alignment of the thioesterase SphJ

Residues corresponding to the consensus motif are coloured yellow; catalytic residues are coloured green; reference sequences were taken from the biosynthesis of surfactin (*B. cereus*, EDZ49042), lichenysin (*B. thuringiensis*, ZP_04072876) and a similar protein from *B. licheniformis* (LchAD, AAU22005)

4.6 Non-PKS/NRPS cluster parts

Some of the protein functions encoded within the gene cluster could not be attributed to a PKS or NRPS domain. They include the O-methyltransferase SphB, a hydrolase function in SphH and the aldolase SphI. These enzymes were subjected to a conserved domain analysis according to their BLASTp results and a multiple sequence alignment with related proteins.

4.6.1 SphB - an O-methyltransferase

A homology search with BLASTp annotated SphB as an O-methyltransferase and member of the S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I family (GenBank accession: cd02440). SphB shows strong homology to other O-methyltransferases family 3.

Three conserved regions that are supposed to participate in the binding of S-adenosylmethionine are described (Kagan and Clarke, 1994; table 4-10). Motifs 1 and 3 contain highly conserved glycine residues, and motif 2 includes an invariant aspartate moiety.

Table 4-10: Signature motifs for methyltransferases

Motif1	(VIL)-(LV)-(DE)-(IV)-G-(GC)-G-(TP)-G
Motif2	(PG)-(QT)-(FYA)-D-A-(IVY)-(FI)-(CVL)
Motif3	L-L-(RK)-P-G-G-(RIL)-(LI)-(LFIV)-(IL)

	Motif 1							Motif 2											
SphB	M	L	E	I	G	T	F	T	G	72	N	Y	F	D	F	V	Y	I	142
<i>N. punct.</i>	T	L	E	V	G	V	F	T	G	72	E	T	F	D	F	A	F	I	142
<i>R. brookii</i>	T	L	D	I	G	V	F	T	G	72	E	T	F	D	F	A	F	I	142
<i>B. cereus</i>	V	L	E	V	G	T	F	T	G	76	N	I	F	D	F	I	F	I	146

	Motif 3										
SphB	L	V	R	P	G	G	I	I	G	I	168
<i>N. punct.</i>	L	L	R	P	G	G	L	I	A	I	168
<i>R. brookii</i>	L	I	R	S	G	G	L	I	A	V	168
<i>B. cereus</i>	L	I	R	P	G	G	I	V	A	V	172

Figure 4-19: Multiple sequence alignment of the O-methyltransferase SphB

Residues corresponding to the consensus motifs are coloured yellow; reference sequences were taken from similar proteins from *Nostoc punctiforme* (*N. punct.*, YP_001867016), *Raphidiopsis brookii* D9 (*R. brookii*, ZP_06306391) and *Bacillus cereus* BGSC 6E1 (*B. cereus*, EEK53069)

In the first motif only the glycines at positions 5 and 9 are conserved, but not the residue at position 7 (figure 4-19). The aspartate moiety of the second motif is always present, as well as the two glycines in motif 3.

4.6.2 The α/β -hydrolase encoded by sphH

The C-terminus of SphH contains a domain classified as a member of the α/β -hydrolase 6 family (pfam12697), which contains a wide range of hydrolytic enzymes of diverse specificity. They share the α/β -hydrolase fold as a structural motif, consisting of a central β -sheet surrounded by several α -helices (Ollis et al., 1992). The catalytic triad is formed of a nucleophile, a histidine and an acidic residue (Kourist et al 2010). Most commonly the nucleophile is represented by serine, which is part of highly conserved G-x-S-x-G-G consensus motif, which is shared with the lipase superfamily (cl12031). The catalytic residues were identified by an alignment with similar hydrolases derived from BLASTp (figure 4-20).

SphH-Hyd	G	S	C	F	G	G	51	G	N	L	D	173	A	G	H	201
S. sp.	G	T	S	F	G	G	75	G	G	V	D	184	A	G	H	212
LnMJ	G	A	S	F	G	G	66	G	R	H	D	176	A	G	H	204
PksE	A	A	S	W	G	G	75	G	E	K	E	172	A	G	H	200
MmpIV	G	W	S	M	G	G	76	G	S	D	D	179	A	G	H	207

Figure 4-20: Multiple sequence alignment of the α/β -hydrolase of SphH

Residues corresponding to the consensus motifs are coloured yellow; reference sequences were taken from the biosynthesis of leinamycin (LnMJ, AAN85523) and similar proteins from *Streptomyces* sp. SPB74 (S. sp., ZP_06822181), *Bacillus subtilis* (PksE, ABH03699) and *Cellvibrio japonicus* Ueda107 (MmpIV, YP_001983454)

The alignment shows that in SphH the highly conserved serine in the G-x-S-G-G motif is substituted with a cysteine moiety. The acidic residue follows the most common pattern being either aspartate or glutamate. Additional sequence homologies were observed in the amino acids preceding the acidic moiety and the histidine. The specificity prediction according to the composition of the catalytic elbow established by Kourist and co-workers (Kourist et al., 2010) could not be applied to the SphH hydrolase, as its signature sequences match none of the described classes.

4.6.3 SphI – an aldolase

Database search annotated SphI as a phospho-2-dehydro-3-deoxyheptonate aldolase, which is also known as 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthase. These enzymes catalyse the condensation of phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4P). This reaction is the first committed step of the shikimate pathway, in which chorismate and prephenate are formed as intermediates for the biosynthesis of aromatic amino acids as well as catechols and hydroxybenzoic acids (van Lanen et al 2008). Conserved domain database classified SphI as a member of DAHP synthase I family (pfam0793). A sequence alignment (figure 4-21) with reference DHAPs clearly assigns SphI to the DAHP subfamily AroA_{Iβ} as defined by Jensen and co-workers (Subramaniam et al 1998). This classification corresponds to DAHP class I according to Birck and Woodard (Birck & Woodard, 2001). Figure 4-21 shows a multiple sequence alignment of the putative catalytic regions with enzymes from the same group. A complete alignment for all conserved regions is given in the appendix (8.4).

<i>S. thermophilus</i>	G P C S	104	K P R T S	136	I G A R	187
<i>T. roseum</i>	G P C S	104	K P R T S	136	I G A R	187
SphI	G P C S	104	K P R T S	136	I G A R	187
<i>B. subt. AroA</i>	G P C A	127	K P R T S	159	I G A R	210
<i>S. xyl. AroA</i>	G P C S	32	K P R T S	64	I G A R	115

Figure 4-21: Multiple sequence alignment of the DAHP synthase SphI

Residues corresponding to the catalytic motifs are shaded green; reference sequences are taken from *Sphaerobacter thermophilus* DAHP (*S. thermophilus*; YP_003320125), *Thermomicrobium roseum* DAHP (*T. roseum*; YP_002521366), *Bacillus subtilis* chorismate mutase (*B. subt. AroA*; P39912) and *Staphylococcus xylosus* chorismate mutase (*S. xyl. AroA*; CAA64712)

4.7 Domain structure of the putative siphonazole biosynthetic enzymes

Based on the domain predictions from CLUSEAN and BLASTp and the results of the multiple sequence alignments a model for the modular and domain organisation of the putative siphonazole biosynthetic gene cluster was established (figure 4-22). The initiation module belongs to the NRPS part and is deficient of a C domain. The two cyclisation modules contain all domains in duplicate with the exception of the A domain. The PKS part of the cluster contains three non-elongating modules with mutations in the catalytic core of their C domains (PKS 5, 7 and 9) and one module (PKS 6) with an additional ACP domain.

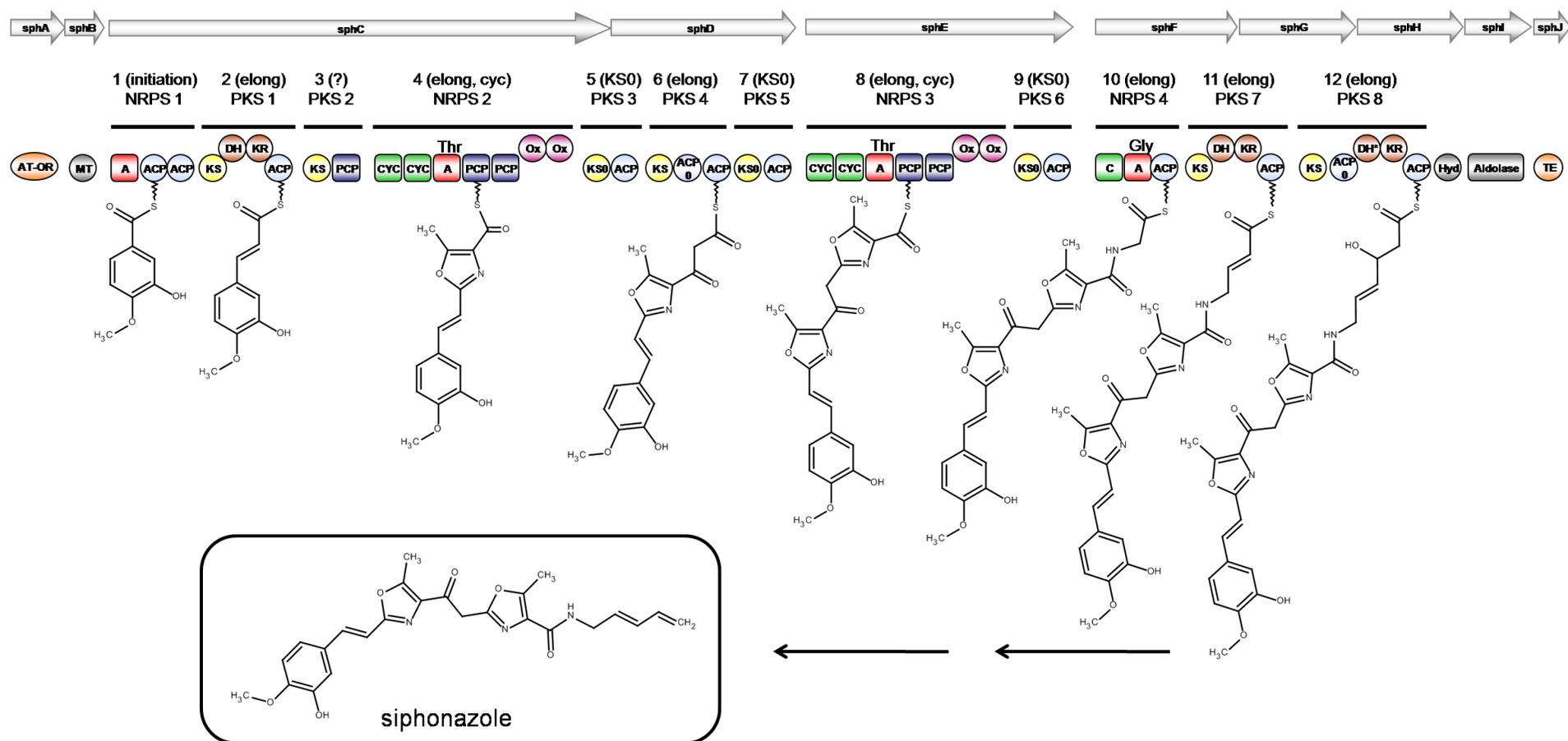


Figure 4-22: Hypothetical pathway for the biosynthesis of siphonazole

Modules are marked by black bars; A, adenylation domain; ACP, acyl carrier protein; Cyc, cyclisation domain; C, condensation domain; DH, dehydratase; Hyd, hydrolase; KR, ketoreductase; KS, ketosynthase; KS0, non-elongating KS; MT, methyltransferase; OR, oxidoreductase; Ox, oxidation domain; PCP, peptidyl carrier protein; TE, thioesterase; domains marked with '0' are non-functional; functionality of DH* in module 12 is uncertain; '?' marks module with uncertain function

4.7.1 Classification of carrier proteins in view of the domain structure

The predicted domain structure exhibits some unusual assignments of carrier proteins. The NRPS-initiation module and the fourth NRPS unit both contain ACP domains instead of PCPs. PKS module 3 in contrast comprises a PCP. Due to these aberrations, the classification by CLUSEAN was re-examined (table 4-11).

Table 4-11: Classification of carrier domains as predicted by CLUSEAN

Scores and E-values from Hidden Markov Model (HMM)-prediction for both domain types are listed; for ACP11 the software provided only results for the ACP prediction

Carrier protein	Score ACP	E-value ACP	Score PCP	E-Value PCP
ACP 1	57.9	8e-17	54.6	7.7e-16
ACP 2	46.8	1.7e-13	46.6	2e-13
ACP 3	70.5	1.3e-20	43.6	1.6e-12
PCP 1	44.2	1e-12	35.4	4.8e-10
PCP 2	37.2	1.3e-10	25.0	2.7e-07
PCP 3	39.4	2.9e-11	25.7	2.2e-07
ACP 4	29.3	3.1e-08	3.4	7.5e-05
ACP 5	43.3	2e-12	0.4	0.00015
ACP 6	34.8	6.9e-10	22.0	1.2e-06
ACP 7	50.9	1e-14	41.3	7.6e-12
PCP 4	39.2	3.2e-11	26.3	1.9e-07
PCP 5	40.1	1.8e-11	26.0	2.1e-07
ACP 8	52.1	4.3e-15	37.5	1.1e-10
ACP 9	41.2	8.2e-12	30.5	1.4e-08
ACP 10	45.4	4.5e-13	8.9	2.2e-05
ACP 11	31.1	8.9e-09	n.a.	n.a.
ACP 12	65.3	4.6e-19	22.7	1e-06

The two carrier proteins of the initiation module (ACPs 1 and 2) were only marginally distinguished as ACPs, while the differentiation of ACP 9 in the fourth NRPS-module was more distinct and comparable to the values of the other thiolation domains. The same applies to the PCP domain in PKS module 3 (PCP 1), which can also be clearly distinguished from an ACP. In most cases however, the values for the refused domain type are close to those of the accepted type. Only in the case of ACPs 4, 5, 10 and 12 the discrimination is definite.

4.8 Construction of the complete cluster from the fosmid library

After the bioinformatic characterisation of the putative siphonazole gene cluster the following work concentrated on the assembly of the complete cluster from partial sequences encoded on fosmid clones from the genomic library. These experiments aimed at providing a basis for further studies. The *in silico* completion of the putative siphonazole biosynthetic gene cluster allowed an exact assessment of the cluster parts that are covered by the genomic library. The ends of fosmids EC9, EC10 and CF4, which were identified to carry parts of the cluster, were sequenced (3.4.11) and aligned with the sph-genes and their adjacent regions. Thus, it was shown that the insert of EC10 starts ~5.5 kb upstream of sphA and ends within the sequences encoding the Cyc domains of sphE. Fosmid CF4 covers a similar area reaching from ~4.8 kb upstream of sphA to the sequence encoding the KS in sphE. The insert of EC9 contains a cluster part encoding the domains from the DH of sphC to the KS of sphG. These results indicate that an area of approx. 7 kb, which contain the final part of the cluster, was not covered by the identified fosmids (figure 4-24).

4.8.1 Screening of the genomic library

To enable further experiments for a full characterisation of the putative siphonazole gene cluster, the fosmid genomic library was screened (3.4.13) for clones that contain the final 7 kb of the gene cluster comprising genes sphH, sphI, sphJ and parts of sphG. For later recombination steps it was also desirable to screen for a fosmid that overlaps with the insert of fosmid EC9 or CF4. Therefore, three specific primer pairs were derived from the DNA sequence to probe the TE domain (TE-end fwd & rev) and the sequence ends of CF4 (CF4-end fwd & rev) and EC9 (EC9-end fwd & rev) within sphE and sphG, respectively. 56 pools of clones from the fosmid library were screened with the CF4 primer pair by whole cell PCR (3.4.1.1), each pool containing the 12 clones of a row on a designated 96-well plate. The screening yielded positive results for pool IC and IIIA, which were subsequently subjected to a further screening with the remaining two primer pairs. Only pool IC proved positive for all three probes. PCR-screening of the individual clones of IC identified the fosmids IC2 and IC3, which both showed positive PCR results for all three fragments. These two fosmids were isolated from 100 ml cultures and the ends of the inserts sequenced. Thus, it was revealed that IC2 and IC3 virtually cover the same

sequence area and only differ in as little as 34 bp. The inserts begin within the Ox domains encoded by sphE and end ~10.8 kb downstream of sphJ. Fosmid IC2 was selected for all further experiments (figure 4-23).

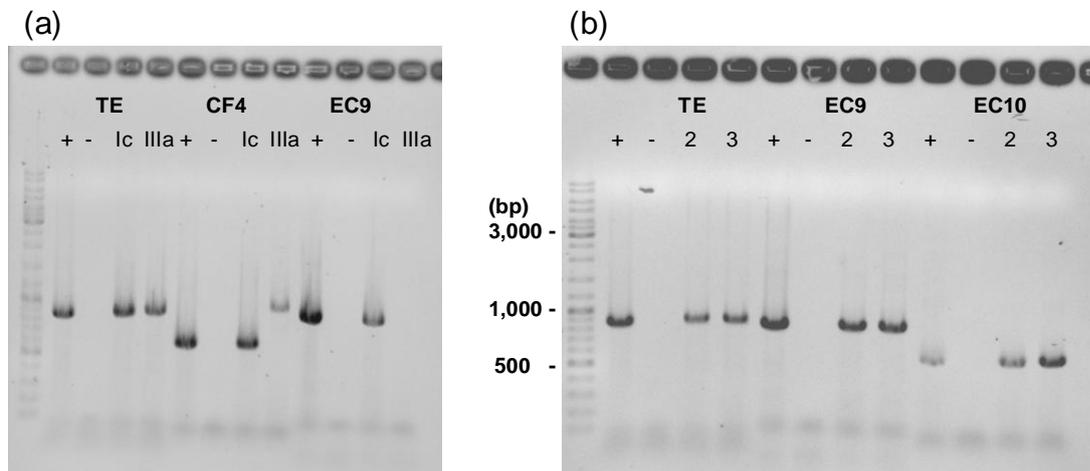


Figure 4-23: PCR-screening of the final cluster part

(a) agarose gel with PCR-products from pool Ic and IIIa, probed for thioesterase (TE) and connection to fosmids CF4 and EC9; **(b)** agarose gel with PCR products from clones IC2 and IC3, probed with the same primer pairs; positive control was always genomic DNA from *Herpetosiphon* sp. 060

4.8.2 Yeast recombinational cloning

As knock-out studies of the putative cluster in *Herpetosiphon* sp. 060 proved unfeasible due to the genetic inaccessibility of the wild type organism, studies in heterologous expression systems must be addressed for a further characterisation of the cluster. This purpose requires a vector construct containing the complete sequence information of the gene cluster. Specific screening of the genomic library (4.8.1) yielded a fosmid (IC2) that comprises the final part of the cluster and overlaps with fosmid CF4, so that the whole sequence was contained on two fosmids only.

The method of yeast recombinational cloning (3.5.3) was applied to combine the inserts into one vector with the complete DNA sequence of the siphonazole gene cluster.

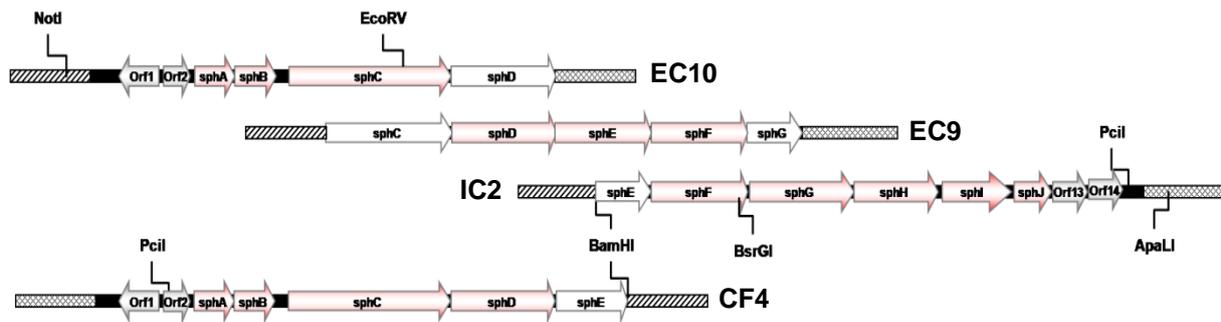


Figure 4-24: Linear representation of fosmid clone inserts

Incomplete reading frames are depicted in white; insert range is shown as black bars; vector backbone is represented by striped (EpiRP primers site) and crossed bars (T7 primer site), respectively; sizes not drawn to scale; restriction sites indicated for yeast recombination and λ -Red experiments

The pCC1FOS vector allows the inserts to be released by a digestion with BamHI (3.4.7). An analysis of the restriction sites in the cluster and the adjacent regions revealed recognition sites for PciI 1,375 bp upstream of the start codon of sphA and 3,387 bp downstream of the stop codon of sphJ. Thus, the desired fragments of the fosmids could readily be cut out by a double digestion with PciI and BamHI. The vector pENTRC1FOS, which is a hybrid of pCC1FOS and yeast specific genes for replication and selection, was chosen for the recombination in yeast. As a selective marker the *ura3P* gene on the plasmid was used, which enables *ura3p* deficient yeast to grow on media lacking uracil. To connect the inserts with the vector, PCR-products of ~1kb in size which cover the ends of the fragments and their adjoining regions were amplified with a primer-generated 20 bp overhang to the vector (primer pairs Hefe-start and -end).

Transformation experiments regularly yielded a number of clones in the range of 10^1 to 10^2 . Transformants were screened for the presence of cluster parts by whole cell PCR (3.4.1.1) with primer pairs used for the fosmid library screening and gap closure (4.8.1 and 4.3.1). The respective primers are given in the appendix and were tested against the empty vector to exclude false positives. By this screening a clone was identified which was PCR-positive for probes in sphA, sphE and sphJ. The PCR reactions were repeated with the same conditions three days later with the original yeast clone, a liquid culture and a plasmid isolated (3.5.3.1) from the liquid culture, which all yielded no product.

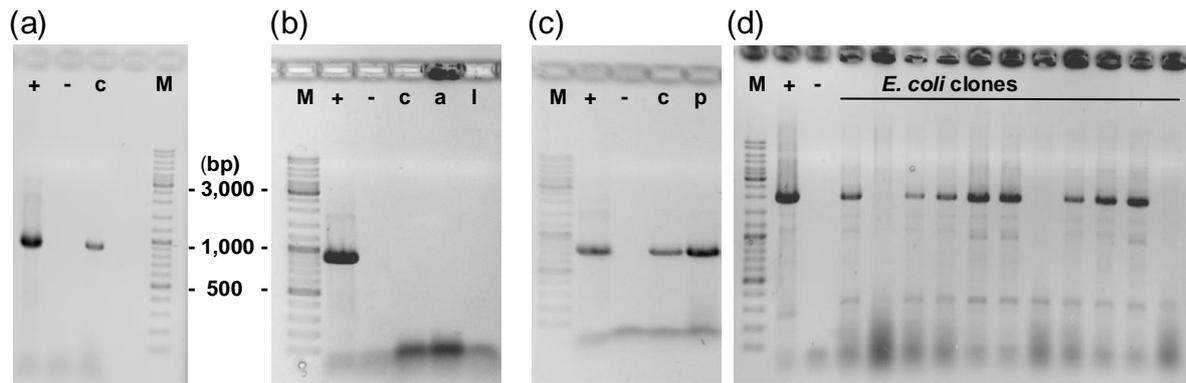


Figure 4-25: PCR reactions from yeast recombination

(a) colony-PCR with TE-primers from yeast clone; +, positive control, -, negative control; **(b)** colony-PCRs from the same clone; c, original colony; a, agar culture inoculated from original colony; l, liquid culture from clone; **(c)** PCR with TE-primers from isolated plasmid of pooled yeast clones; **(d)** colony-PCR probing for empty vector pENTRC1FOS from *E. coli* clones

Following transformations regularly produced yeast clones that showed positive results in PCR screening. Repeatedly, it was observed that PCR-reactions performed with the same clones two or three days later stayed negative. Therefore, yeast clones were pooled and directly subjected to plasmid isolation. Pools and isolated plasmids both showed positive results for all PCR probes. Subsequently, the plasmid was used for the electroporation of electro-competent EPI300 *E. coli* cells (3.3.4). Repetitive transformation experiments yielded about 1 to 30 clones per transformation, which all proved negative in PCR screening. As a control, a primer pair (Ura3P_mid_fwd and Cat_mid_rev) was derived from the vector backbone of pENTRC1FOS, which probes the region adjacent to the integration site of the inserts. A screening of *E. coli* clones with this primer pair discovered that the clones virtually all yielded a PCR-product of 1,900 bp, indicating an empty vector. Clones that did not yield any product were tested again for the presence of cluster parts, but proved to be negative.

4.8.3 λ -Red mediated recombination

Due to the unsuccessful yeast recombinational cloning, another strategy was addressed to recombine the cluster parts to one construct. The end sequencing of the identified fosmids showed that the inserts of EC9, EC10 and IC2 all have the same orientation within the pCC1FOS vector, i.e. the reading frames are all directed from the EpiRP-primer site to the T7-primer site (figure 4-24). The insert of CF4 has the opposite orientation. This arrangement allowed a recombination approach based on a λ -Red mediated strategy (3.5.1). An analysis of the restriction sites on the vector backbone and within the cluster revealed the possibility to generate fragments of EC10 and IC2, which both contain homologous regions to the insert of EC9 and to the vector backbone. Thus, it was possible to recombine the cluster in two subsequent recombination steps by complementation of EC9 with the beginning and the end of the cluster. Therefore, fosmid EC9 was transformed into BW25113 *E. coli* cells, which are capable of λ -Red recombination.

To enable the selection of successful transformants, a streptomycin (primers Str-FRT-fw-EC10 & Str-FRT-re+EC10) and an apramycin resistance cassette (primers Apr-FRT-fwd+IC2 and Apr-FRT-rev+IC2) were amplified together with an adjoining oriT from the plasmids pIJ778 and pIJ773 and cloned upstream of sphA and downstream of sphJ, respectively. Flanking FRT-sites were included to enable later deletion of the cassettes. Using primers with 40 bp homology extensions, the streptomycin cassette was inserted into fosmid EC10, thus deleting 2,884 bp of the upstream region. By application of the same procedure, the apramycin cassette was inserted into IC2, concurrently with the deletion of 3,316 bp.

By digestion with ApaI and BsrGI an 18,842 bp fragment was generated from fosmid IC2 (3.5.1.1), which harbours a 266 bp homology region to the vector and 2,706 bp homology to the insert of EC9. Fosmid EC10 was digested with EcoRV and NotI to obtain a 15,619 bp fragment with 278 bp homologous to the vector and 3,016 bp overhang to EC9. In two subsequent steps of electroporation (3.3.4) and selection these two fragments were transformed and recombined with EC9, thus obtaining a pCC1FOS construct of 71,275 bp, which harbours the complete putative siphonazole gene cluster.

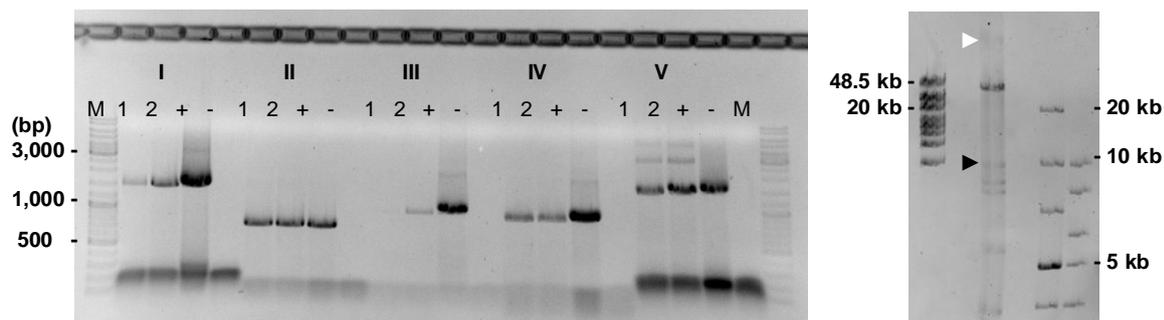


Figure 4-26: Completion of the putative siphonazole cluster by λ -Red recombination

(a) Test-PCRs with two clones (1, 2) after final λ -Red recombination; PCRs for: I, *aadA*; II, front cluster part; III, middle part; IV, final part; V, *aac(3)IV*; (b) restriction digestion with *Hind*III of the isolated fosmid bearing the complete cluster; white arrow indicates undigested fosmid; black arrow denotes unaligned band

Clones obtained from this procedure were screened by PCR (3.4.1) for the presence of the resistance cassettes (corresponding primers see above) and the front (primers 1280seq_neu & 668seq-1280), middle (primers gap1) and final part of the cluster (primers TE-end). The reactions confirmed incorporation of all parts. The generated construct was readily isolated from liquid culture. A further analysis was carried out by a diagnostic restriction digestion with *Hind*III (3.4.7). The obtained restriction pattern agreed with the predictions made with the CloneManager 9 and NEBCutter2 software with the exception of one additional band at 10 kb, which could not be assigned to any restriction fragment. As there were also traces of undigested fosmid detectable, this band was attributed to a product of incomplete digestion.

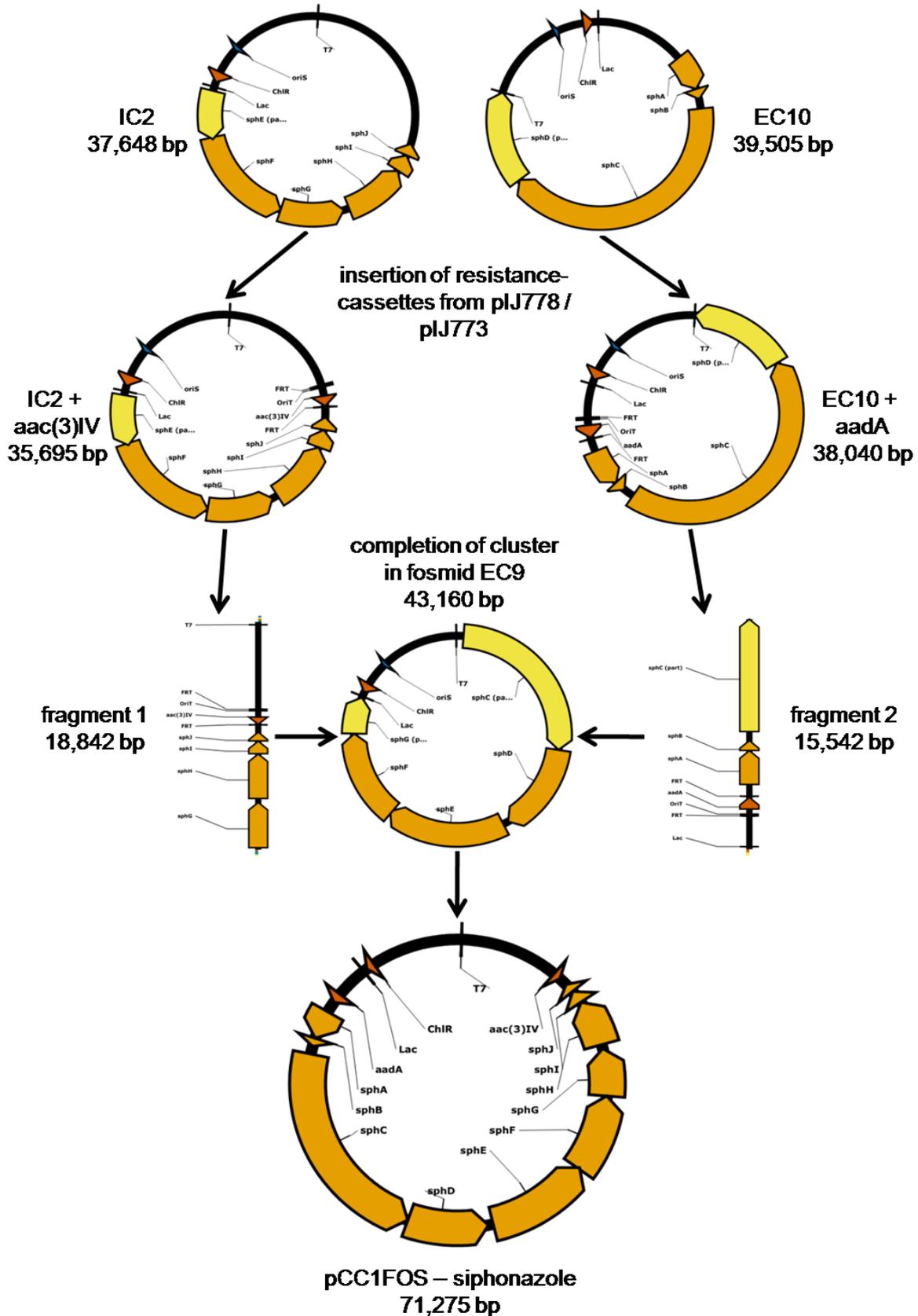


Figure 4-27: Reconstitution of the putative siphonazole cluster

Λ -Red based cloning of resistance markers and cluster fragments into fosmid clones; fragments from EC10 and IC2 were inserted into EC9 to complement the cluster; cluster genes are depicted in orange, resistance genes in red and gene fragments in yellow

4.9 Protein expression and purification

Several proteins from the putative siphonazole cluster were expressed heterologously (3.6.1) and purified from *E. coli* cells (3.6.3) for ensuing enzymatic assays. If not stated otherwise, the corresponding coding sequences were amplified by PCR with specific primers (8.3). The primers were extended by suitable restriction sites that enabled insertion of the DNA fragment into the pET28a(+) vector, which links the proteins to a N-terminal 6x his-tag upon expression. Cloning and transformation reactions were performed according to the described methods (3.5.2 and 3.3.3). After screening and verification of the insert (3.4.11), the recombinant plasmids were transformed into the *E. coli* strain BL21 for expression of the proteins. Purification was performed by affinity chromatography on Ni-NTA columns according to the described protocol (3.6.3) in five elution steps with increasing concentrations of imidazole.

4.9.1 Heterologous expression of the O-methyltransferase SphB

The O-methyltransferase encoded by sphB was amplified (primer pair OMT) and cloned into the pET28a(+)-vector using BamHI and HindIII restriction sites. The recombinant protein was successfully expressed with an N-terminal his-tag and purified from BL21 *E. coli* cells. The analysis by denaturing SDS-PAGE demonstrated very high expression levels of a protein with a molecular weight around 25 kDa. The eluted protein thus seems to be smaller than the calculated mass of 28.95 kDa for the recombinant protein SphB, but it is possible that the chromatography is distorted by the large amount of protein. Through affinity chromatography an estimated purity of 95-99% was achieved in the final elution steps. Elution fractions 3 – 5 were pooled and used for subsequent assays. Though the protein was always present in the soluble protein fraction of the cell lysate, it precipitated rapidly when exposed to low temperatures or during concentration procedures. Therefore, purification had to be carried out at room temperature and assays performed directly afterwards to avoid loss of activity. According to the protein concentrations determined by photometric measurement (3.6.7), an average total of 3 mg protein was purified from elution fractions 3 – 5 of a 100 ml culture.

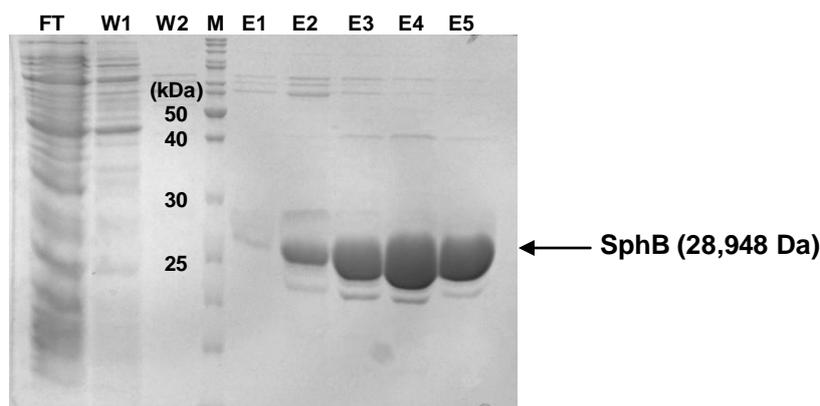


Figure 4-28: purification of SphB by affinity-chromatography on Ni-NTA column

Protein gel shows fraction from the purification; Ft, flow through; W1, wash1 (20 mM imidazole); W2, wash2 (40 mM imidazole); E1 – E5, elution fractions with 100 to 300 mM imidazole; M, size marker

4.9.1.1 *In vitro* methylation of protocatechuic acid by SphB

The *O*-methyltransferase SphB was hypothesized to catalyse the transfer of a methyl group to the 4-OH group of the hydroxybenzoic acid precursor of the starter unit. Feeding experiments identified *S*-adenosylmethionine (SAM) as donor of the methyl group (1.5). To support this hypothesis, the activity of the protein was investigated in an *in vitro* methylation assay (3.6.8). In consideration of the structure of siphonazole 3,4-dihydroxybenzoic acid (protocatechuic acid) was chosen as substrate in this assay, which is a major metabolite in the degradation of aromatic compounds (Buchan et al., 2004). Methylation of the 4-OH group as in siphonazole would yield isovanillic acid as product.

Subsequent to incubation of the enzyme with substrate and SAM, the samples were separated by HPLC on a C₁₈ column (3.6.8.1). Thus, two peaks at 8.29 min and 11.41 min were identified, which exhibit the characteristic absorption spectrum of protocatechuic acid with maxima at wavelengths of 216 nm, 257 nm and 293 nm (Zazza & Sanna, 2010). The corresponding spectra are given in the appendix (8.5). By comparison with the spectra and retention times of protocatechuic and isovanillic acid as reference substances the peak at eight minutes could be assigned to the substrate and the peak at eleven minutes to the product (figure 4-29). This is in accordance with the finding that the product peak does not appear in a control reaction without SAM. This control also lacked a peak at three minutes, which was therefore attributed to SAM.

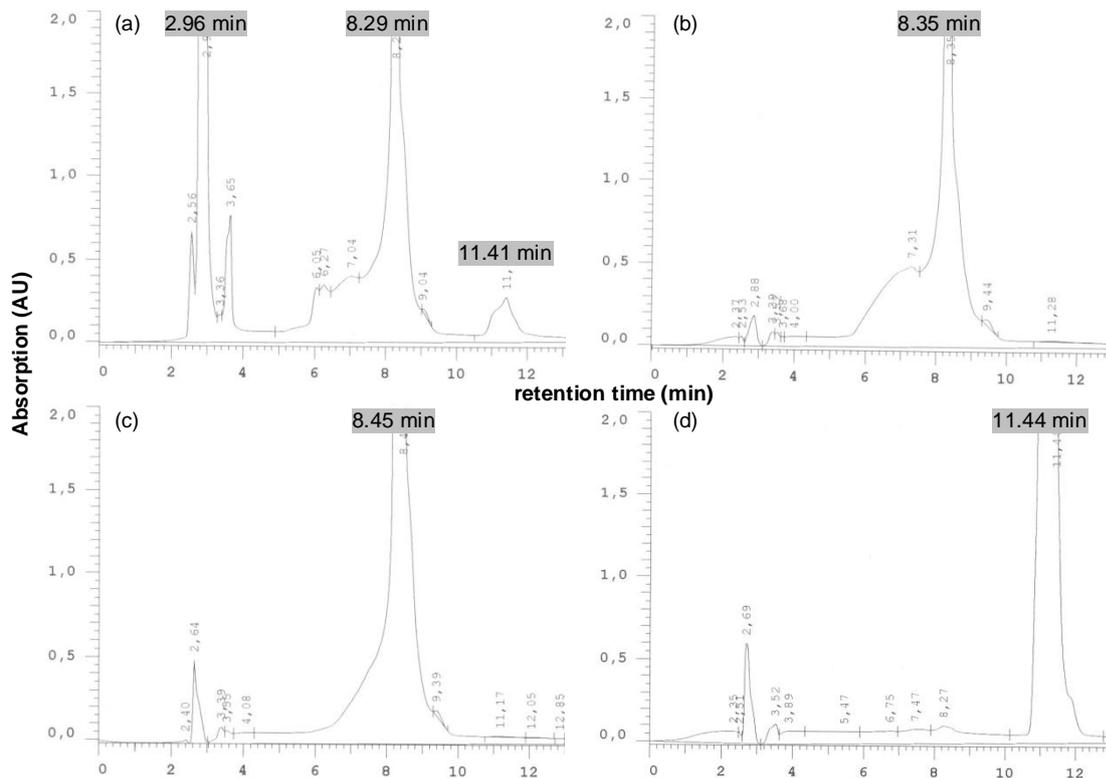


Figure 4-29: UV absorption spectra from HPLC

(a) Methylation assay reaction; SAM at 2.96 min; educt at 8.29 min; product at 11.41 min; (b) control reaction without SAM; (c) protocatechuic acid standard; (d) isovanillic acid standard

4.9.2 Heterologous expression of the first A domain of SphC

The A domain of the initiation module (SphC-A1) is located at the N-terminus of SphC. The forward primers for amplification of the insert were positioned either at the natural start codon of the ORF (primer A-dom_orfstart) or in proximity to the sequence encoding the domain. As the ORF continues downstream, a stop codon was always included in the reverse primers. The boundaries of the A domain sequence were determined according to the predictions with CLUSEAN and the NRPSPredictor, as well as comparative sequence analysis with BLASTp. Integration into the pET28a(+) vector was accomplished by utilisation of BamHI and HindIII restriction sites (3.5.2). A comprehensive list of the tested primer pairs is given in the appendix (8.3.2).

In the first expression experiments no recombinant protein could be detected after purification in the analysis by SDS-PAGE (3.6.4). Variation of IPTG concentration and growth temperature stayed fruitless. By moving the primer sites different constructs were created (primer pairs A-fwd-BamHI & A-rev-HindIII and A_domain

fwd & rev). One primer pair was located directly at the boundaries of the domain (primer pair DHB_A-Dom), while a longer fragment spanned the A domain and the downstream carrier protein (A-PCP-rev). As none of these constructs produced any detectable amount of the target protein, the expression performed in the cold-adapted ArcticExpress (DE3) cells, but without success.

Consequently, the expression was carried out with the Champion™ pET expression system by Invitrogen, using the pET151 vector and the BL21 star™ expression cells. The primer pair sphC-A, employed for generation of the insert, used the same primer sites as before (A-dom_orfstart and DHB_A-dom_rev, respectively). An overnight expression of a 300 ml culture at 16°C yielded a protein with a size just below 60 kDa, which corresponds to the expected size of 58.16 kDa for the recombinant A domain. Most of the protein was located in the insoluble fraction of the pellet, but smaller amounts eluted in the fractions 2 – 4 (figure 4-30). Fraction 2 contained a large amount of impurities and was therefore not used in further assays. An estimated purity of 85 – 90% was achieved in fraction three and more than 95% purity in the fourth elution step. These fractions were pooled and used for further processing.

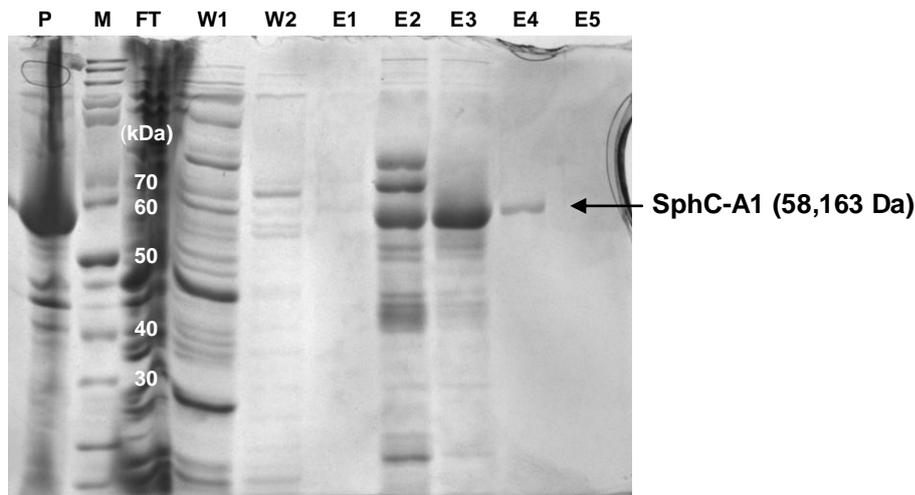


Figure 4-30: Purification of SphC-A1 by affinity-chromatography on Ni-NTA column

Protein gel shows fraction from the purification; Ft, flow through; W1, wash1 (20 mM imidazole); W2, wash2 (40 mM imidazole); E1 – E5, elution fractions with 100 to 300 mM imidazole; M, marker

4.9.2.1 ATP – PP_i exchange assay

The substrate specificity could not be predicted by analysis of the Stachelhaus-code. Therefore, the activation of various substrates by the A domain was investigated by an ATP-PP_i exchange assay (Phelan et al., 2009). The measurement of the conversion of γ -¹⁸O₄-labelled ATP to ¹⁶O₄-ATP allows conclusions about the substrate acceptance of the investigated A domain (3.6.10).

Purified SphC-A1 was incubated with protocatechuic acid, isovanillic acid, glycine, tyrosine, phenylalanine and aspartate as substrates. The exchange rate of ¹⁸O₄-ATP to ¹⁶O₄-ATP was quantified by MALDI-TOF-MS. However, the chromatograms did not show any detectable formation of unlabelled ATP above background level for any of the assayed substrates. The chromatograms are given in the appendix (8.6).

4.9.3 Heterologous expression of the DAHP synthase SphI

The expression of the recombinant protein SphI was accomplished in an analogous way to the expression of the A domain (4.9.2). The primers sphI-fwd and sphI-rev were designed to contain start and stop codon of the gene and used for the amplification of the DNA fragment. The insert was cloned into the pET151 vector and the protein expressed in BL21 star™ cells. For cultivation and purification the same conditions as for the A domain were applied. The purification yielded a protein between 40 and 50 kDa, which mainly resided in the insoluble fraction (figure 4-31). The protein size is in agreement with the calculated weight for SphI of 41.27 kDa. Smaller amounts eluted in fractions 2 – 4. Fraction two still contained several impurities, of which a protein of ~60 kDa was the most prominent. This protein was also present in smaller amounts in fraction three, while the target protein could be obtained with great purity in the fourth elution step.

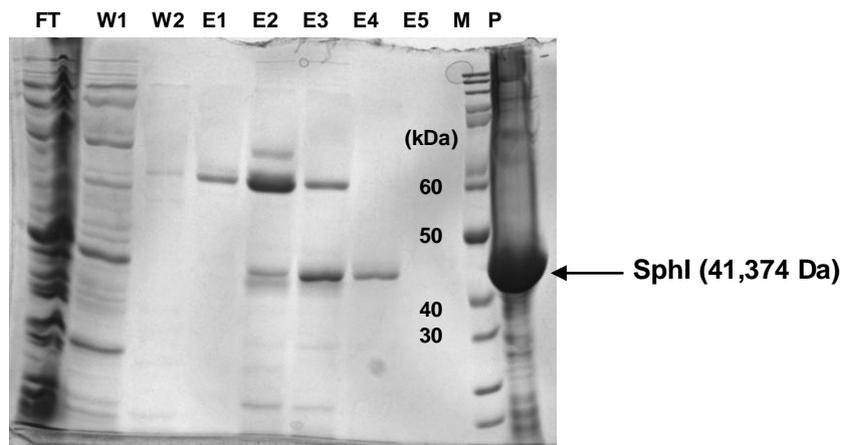


Figure 4-31: Purification of SphI by affinity-chromatography on Ni-NTA column

Protein gel shows fraction from the purification; FT, flow through; W1, wash1 (20 mM imidazole); W2, wash2 (40 mM imidazole); E1 – E5, elution fractions with 100 to 300 mM imidazole; M, marker; P, pellet (insoluble fraction)

4.10 Expression profile of siphonazole in *Herpetosiphon sp. 060*

To the date of the present study, there are no investigations on the time point of siphonazole biosynthesis and the localisation of the compound within a *Herpetosiphon* colony. Therefore, two different series of experiments were conducted to elucidate the temporal expression profile of the gene cluster and the spatial distribution of siphonazole within a solid culture of *Herpetosiphon sp. 060*.

4.10.1 mRNA expression of sph genes

The presence of mRNA indicates the expression of the investigated gene at the time of investigation. In prokaryotes genes are commonly transcribed into polycistronic mRNA carrying the information of several ORFs, which are regulated together in one operon (Zhou & Yang, 2006). PCR-probing of the intergenic regions allows conclusions on the expression pattern of the gene cluster.

Total RNA was isolated (3.4.14) from *Herpetosiphon* cultures after 1, 2, 3, 4 and 7 days and tested for contamination with genomic DNA by a PCR reaction (3.4.1) using primers that bind within the coding region of sphE. RNA isolates that yielded no PCR product were considered pure. Employing oligo dT-primers, a reverse transcription reaction was performed on these isolates to synthesize cDNA of messenger RNAs.

Specific primer pairs (A – J) were derived from the sequence of the putative siphonazole gene cluster to probe the intergenic regions between the sph genes. The pairs always covered the end of an open reading frame and the start of the following gene. An additional primer pair was designed for the upstream region (approx. 200 bp) of sphA. All primers were tested on fosmid DNA containing the whole gene cluster, before performing PCR reactions on the cDNA (figure 4-32).

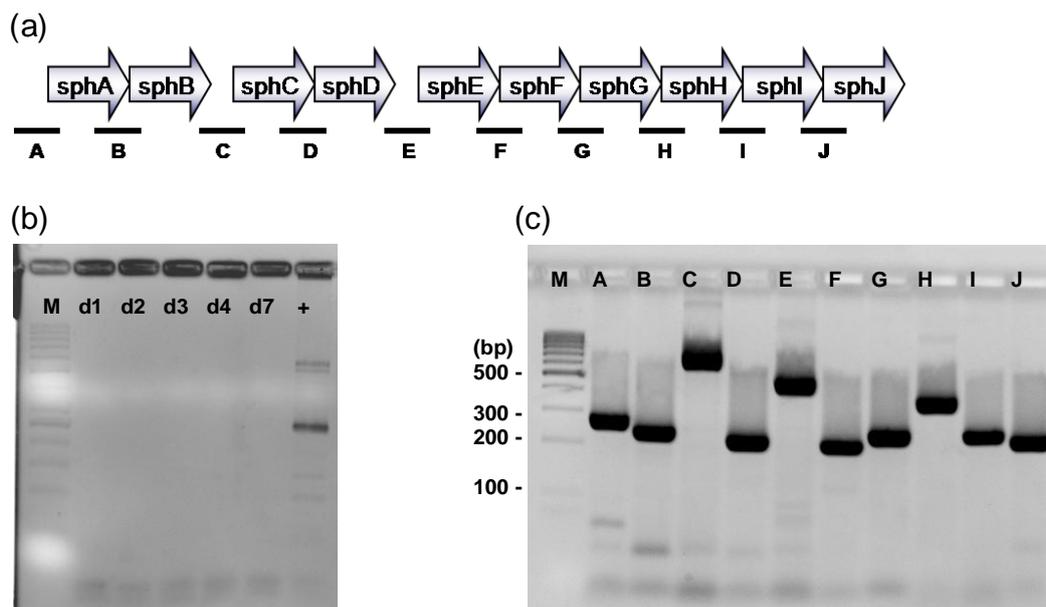


Figure 4-32: Test PCRs for RNA expression analysis

(a) Schematic representation of PCR probes; (b) PCR from isolated mRNA from days 1, 2, 3, 4 and 7 with primer pair gap1; M, marker; +, positive control from fosmid DNA; (c), PCRs from fosmid DNA with primer pairs A – J

PCR with standardised amounts of cDNA as template yielded products only for the first three days of cultivation (figure 4-33). Evaluation of the agarose gels (3.4.2) indicated that the expression is strongest on the second day and already very weak on the third day. PCR products were detected for all probes with the exception of the intergenic region between sphB and sphC. This finding suggests that the two distinct genes sphA and sphB are not part of the same operon as the reading frames that contain the actual PKS and NRPS modules. Furthermore, the upstream region of sphA also yielded a product for all three days, indicating that sphA and sphB are not transcribed as bicistronic mRNA, but may be part of a larger operon. Finding that no expression could be detected on days four and seven suggests that siphonazole is produced exclusively in an early phase of the culture.

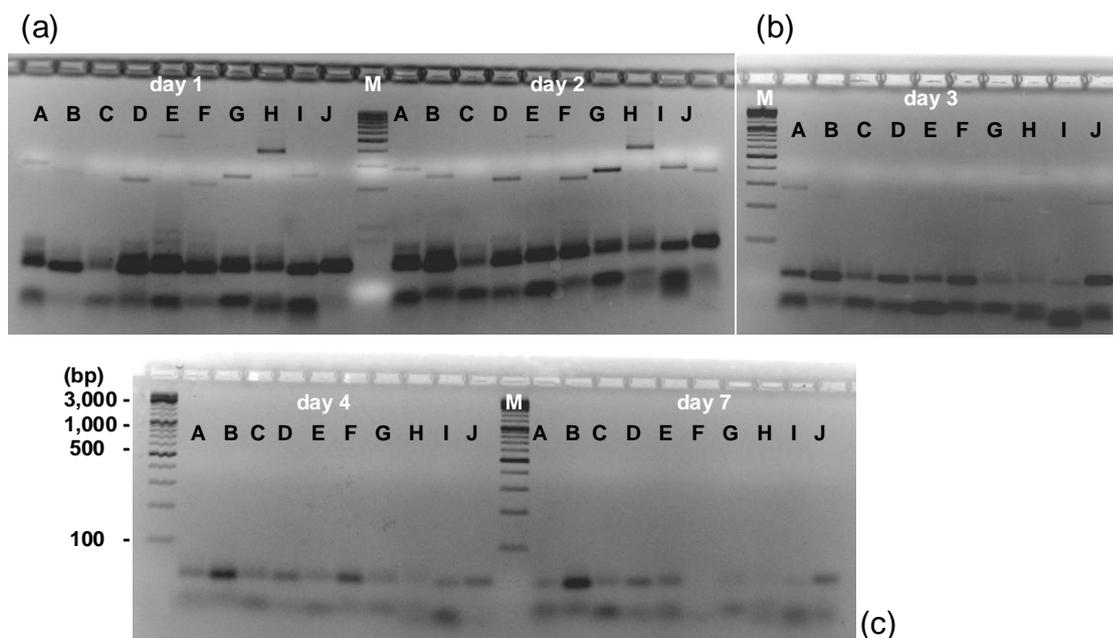


Figure 4-33: RT-PCR from RNA samples

(a) RT-PCR from days 1 and 2; lanes 1 – 10 correspond to primer pairs A – J; (b) day 3; (c) days 4 and 7; M, size marker

4.10.2 Imaging mass spectrometry (IMS)

A *Herpetosiphon* culture was grown on a MALDI plate and subjected to IMS (3.2.6). Thereby, a mass corresponding to siphonazole was detected (m/z 464), which localises around the growing edge of the culture, but not in the centre of the colony. A detected mass with $m/z = 486$ shows the same distribution and was therefore considered to be the sodium adduct of the first compound (figure 4-34).

IMS detected an abundance of other masses, which originate from the colony. A considerable number of them are localised at the edge of the culture. Furthermore, the evaluation of the images revealed two masses ($m/z = 318$ and $m/z = 342$), which did not show even distribution, but were only present at the left and upper edge of the culture. While several products were detected, which exhibit a wider distribution over the colony, one compound ($m/z = 861$) seems to be tightly restricted to the centre of the culture.

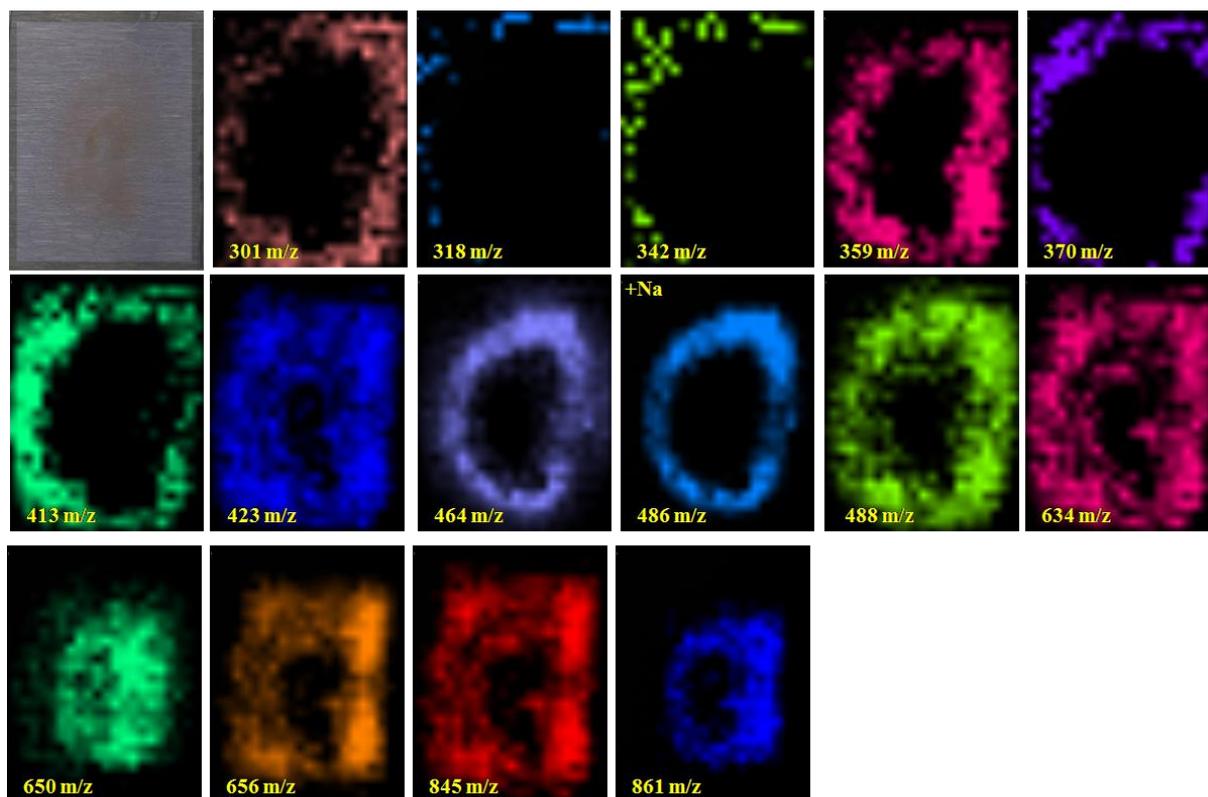


Figure 4-34: False-colour images from IMS

Empty MALDI-plate is depicted in the upper left corner; colours were assigned manually to visualise the spatial distribution of one specific mass; m/z ratios are given in the corresponding pictures; +Na denominates the putative sodium adduct of siphonazole

4.11 Heterologous expression of the complete gene cluster in *E. coli*

The reconstitution of the complete putative siphonazole cluster on one fosmid construct (4.8.3) provided the base for the expression of the cluster in a heterologous host. BAP1 *E. coli* cells were used as expression system in a first trial. These cells contain the promiscuous phosphopantetheinyl transferase Sfp from *B. subtilis*, which is known to prime also heterologous ACPs and PCPs with their cofactor (Quadri et al., 1998). As the fosmid backbone is designed for maintenance in *E. coli*, no further modifications were conducted for the expression in the host cells.

Fosmid pCC1FOS-siphonazole with the completed gene cluster was isolated from an *E. coli* culture by standard methods (3.4.4) and electroporated into freshly prepared electro-competent BAP1 cells as described before (3.3.2 and 3.3.4) using 3 – 15 μ l isolated fosmid. Generated clones were tested by PCR (3.4.1) for the presence of the gene cluster and a positive clone was picked for the inoculation of 30 ml LB pre-

culture (3.2.3). To ensure that the fosmid was maintained in the cells, PCR tests were repeated with the pre-culture. The expression culture was prepared analogous to the cultivation of *Herpetosiphon sp. 060* for subsequent siphonazole extraction (3.2.4). 1.5L LB medium containing 2% (w/v) Amberlite XAD16 adsorber resin were inoculated with 15 ml pre-culture and incubated on a horizontal shaker at 30°C three days. The resin was extracted according to the described protocol (3.2.4), which yielded 178.9 mg crude extract. Unlike the *Herpetosiphon* pellet the *E. coli* cells were not retained in the frit. As it was possible that siphonazole was produced, but not excreted, the cells were harvested by centrifugation and the pellet extracted separately, yielding another 193.8 mg crude extract. Samples of the extracts were subjected to LC-MS analysis (3.6.9) and the spectra screened for the mass of siphonazole. However, a mass corresponding to the compound could not be detected in either of the extracts.

5 Discussion

5.1 Genetic accessibility of *Herpetosiphon* sp. 060

The increasing availability of genome sequences and the development of sequence analysis and prediction tools for NRPS and PKS systems have greatly facilitated the elucidation of new biosynthetic clusters. Nevertheless, these *in silico* studies are no replacement for an experimental proof of concept. Still, assembly lines like the siphonazole cluster rise new questions on biosynthesis that cannot be answered by mere sequence analysis. For a certain alignment of a gene cluster to its corresponding product the genetic knock-out of biosynthetic genes remains an indispensable tool. Equally, detailed studies of specific enzyme functions benefit from the possibility to erase and reconstitute their function by genetic *in vivo* manipulation, which is also a crucial requisite for combinatorial biosynthesis (Galm et al., 2011).

The mentioned experiments all require the introduction of foreign DNA, e.g. plasmid derivatives or linear DNA, into the cell and either the individual maintenance in the host or the integration into the genome. As only few bacteria are naturally competent to take up external DNA, different methods of inducing this competence have been developed over the past decades (Yoshida & Sato, 2009). The physiological characteristics of different bacterial classes and genera have resulted in a great number of protocols, which are available for the transformation of mainly *E. coli*, *Bacillus*, *Pseudomonas* or myxobacterial cells. As to the date of this study no protocols are described for the transformation of *Herpetosiphon* or related Chloroflexi, it was necessary to test different methods and conditions to establish a suitable protocol for the introduction of plasmid DNA into *Herpetosiphon* sp. 060.

Extensive experiments with vegetative *Herpetosiphon* cells (4.1) did, however, not result in any successful transformation neither by electroporation, conjugation, sonication nor chemical treatment. There are several possibilities at which point these transformations were unsuccessful. First, the applied methods failed to deliver the plasmid DNA into the cells. It is likely that features like the unusual cell wall composition, the presence of a sheath-like structure or the filamentous growth contribute to an impaired DNA uptake. The usage of liquid cultures of different age showed that the cell filaments tend to form large clots in fully grown cultures, which

can only be partly separated by pipetting or treatment in a homogeniser. These clots greatly interfere with the electroporation process and lead to arcing at high voltage. Therefore, only cultures that have been grown for one day at maximum seem to be suitable for electrophysical transformations. Additionally, it seems reasonable to assume that cells which are intertwined in a tight knot are less accessible for external DNA in general. The problem of the cell wall was addressed by various methods. As the cell wall structure resembles that of Gram-positive bacteria, protocols for actinomycetes were employed for transformation. The cell wall-weakening agent glycine was shown to prevent the formation of long filaments and thus also reduced clotting, which makes glycine treatment prior to transformation recommendable (4.1.2). It was also shown that protoplasts of *Herpetosiphon* cells can readily be produced by standard methods (4.1.3). Still, a method for a clean separation from remaining vegetative cells has to be found. Observation of lysozyme treated cultures showed that the rod-shaped cells within the filaments lose their shape, revealing a translucent sheath-like structure in the intercellular spaces. This finding supports the presence of such a structure, which would represent an additional barrier for the introduction of DNA. The effect of the glycine on the filaments, however, indicates that this structure should rather be associated with the cell wall. Therefore, the existence of a sheath remains unclear.

A second possibility for the unsuccessful transformation is that the plasmids were not integrated into the genome after uptake into the cell. As the constructs are all based on plasmids designed for the reproduction in *E. coli* cells, they are probably not maintained within *Herpetosiphon*. The integration into the genome by a crossing over event is a statistical process, thus it is possible that this event did not occur due to low transformation efficiency. Further experiments on the transformation of *Herpetosiphon* should therefore include the development of either an integrative plasmid or an *E. coli* - *Herpetosiphon* shuttle-vector. Suitable promoters and origins of replication may be obtained from the genome and natural plasmids of *H. aurantiacus*.

5.2 Elucidation of the putative siphonazole biosynthetic pathway

The complete sequence of the putative siphonazole cluster was resolved by sequence alignments of subclones from a fosmid library with draft genomic data (4.2). Thus, six contigs were identified carrying the putative genes for the biosynthesis of siphonazole. A subsequent PCR-based approach allowed alignment of the contigs and completion of the cluster sequence (4.3.1).

The putative siphonazole gene cluster comprises ten open reading frames (sphA – sphJ), which cover approx. 50 kB and encode a hybrid PKS/NRPS assembly line. Using BLASTp and the NRPSpredictor software together with detailed studies of conserved signature motifs, the domain structure of the biosynthetic enzymes was elucidated (4.7). They comprise 12 modules, four of which are of modular NRPS type and eight are *trans*-AT type PKS modules.

5.2.1 NRPS modules

The NRPS part of the putative siphonazole gene cluster is represented by four modules, which are all separated from each other by one to three PKS modules (figure 5-1). The NRPS portion comprises the initiation module, the two cyclisation units, and the glycine incorporating module encoded by sphC, sphE and sphF, respectively. All of them are hybrid genes, which encode NRPS and PKS modules side by side in the same ORF.

5.2.1.1 The cyclisation modules

In the two cyclisation modules the C domain is replaced by tandem Cyc domains. A peculiarity of these modules is that they do not only contain two consecutive Cyc domains, but also a double arrangement of PCP and Ox domains (4.4.4). Only the threonine specific A domains are not duplicated. Commonly, a single Cyc domain is responsible for both the peptide bond formation and the cyclisation (Roy et al, 1999), but the unusual arrangement of tandem Cyc domains has its precedence in the biosynthesis of vibriobactin and leinamycin. Marshall and co-workers (Marshall et al., 2002) examined the two domains by site-directed mutagenesis on the catalytic D-x-x-x-x-D-x-x-S core motif and thus found out that the first domain was responsible for

the heterocyclisation, while the second catalysed the peptide bond formation. A similar mechanism was proposed for the Cyc tandem in the leinamycin cluster (Cheng et al., 2002). This provides a rationale for the assumption that heterocyclisation in siphonazole biosynthesis works in a comparable way. Sequence analyses of the Cyc domains revealed an intact catalytic core in all four of them. Deviations only occurred in the other signature motifs (4.4.4.1).

Ox domains are in most cases associated with Cyc domains and are typically found at two characteristic positions (Du et al., 2000; Schwarzer & Marahiel, 2003). The domain can be either inserted into the A domain of the module like in the case of epothilone biosynthesis (Molnar et al., 2000) or positioned downstream of the PCP. The latter case was reported for leinamycin, whereas in the myxothiazole cluster both options are realised (Silakowski et al., 1999). The C-terminal placement of the Ox domains in the siphonazole cluster hence follows this second arrangement. The occurrence of tandem Ox domains, however, has not been reported before. The Ox domains are known to require a FMN cofactor, but only little biochemical data is available. Analysis of the signature motifs (4.4.4.2) revealed that the second Ox domain of both cyclisation modules significantly lacks homology to the K-Y-x-Y-x-S-x-G-x-x-Y-(PG)-V-Q motif. According to the conserved domain database (Marchler-Bauer et al., 2011) the second conserved tyrosine is constituent of a NADPH-binding site (Kobori et al., 2001). This moiety is not conserved in SphC-Ox2 and SphE-Ox2, giving rise to the assumption that these domains may be non-functional due to impaired cofactor binding. This hypothesis would have to be proved by knock-out experiments. Otherwise, it is possible that these additional domains are either redundant or perform hitherto unknown tasks during the siphonazole biosynthesis.

5.2.1.2 The initiation module

The initiation module contains an A domain whose substrate specificity could not be clarified by bioinformatic methods (4.4.1). A further peculiarity is the presence of two ACPs instead of a PCP. Although there are differences between the two types of carrier proteins, which will be discussed later (5.2.3.3), it is imaginable that the domains in this module were confused by the prediction software. The two ACPs were only marginally discriminated from PCPs by CLUSEAN in a hidden Markov model (4.7.1). Thus, these domains could as well be annotated as PCP domains.

5.2.1.3 The glycine incorporating module

The fourth NRPS module contains an A domain with a predicted specificity for glycine (4.4.1). It is the only NRPS module of the putative siphonazole gene cluster that follows the canonical C-A-T domain arrangement. The thiolation domain is, however, annotated as ACP by CLUSEAN. Different from the initiation module, the hidden Markov model clearly distinguishes the ACP with a score of 41.2 (E-value 8.2×10^{-12}) from a PCP with a score of 30.5 (E-value 1.5×10^{-8}).

5.2.2 NRPS-PKS interfaces

Notably, both of the ACP-containing NRPS modules are located at an interface, where the NRPS module is followed by a PKS module, which possesses in each case the same domain architecture KS-DH-KR-ACP. Comparable idiosyncrasies can be observed at the other changeovers from NRPS to PKS and vice versa (figure 5-1). The first cyclisation module (module 4) is preceded by a PKS module, whose thiolation domain is annotated as PCP instead of ACP (5.2.3.3) and followed by a putatively non-elongating PKS module (5.2.3.2). The other two KS⁰ containing modules number 7 and 9 flank the second cyclisation module (module 8), while module 9 also precedes the fourth NRPS module (module 10). Additionally, the three non-elongating modules are all located at the termini of the respective enzymes with module 5 actually being split between SphC and SphD. Regarding this position, it is possible that these modules have their function in providing the docking domains for correct protein-protein recognition of the hybrid subunits (Koglin & Walsh, 2009) than in performing catalytic steps. These small docking elements at the protein termini are known to play an important role in mediating the specific protein interaction (Richter et al. 2008).

Remarkable as these findings are, from the bioinformatic data alone it cannot be told if the arrangement of unusual modules actually contributes to the PKS-NRPS interface and facilitate the switching from one assembly system to the other. It is as well possible that this architecture is just a concurrent result of the mosaic-like structure of the gene cluster (5.2.4). Further insights may be attained by an analysis of docking and communication-mediating domains of the modules in question.

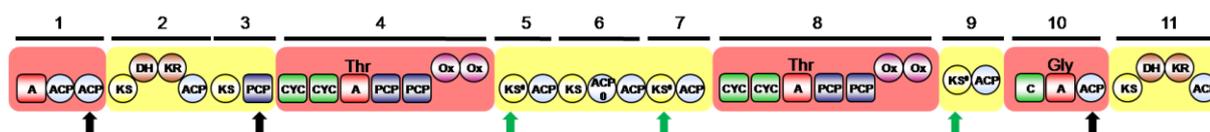


Figure 5-1: PKS-NRPS interfaces

Modules are marked by solid bars and corresponding numbers above; NRPS modules are shaded red; PKS modules are shaded yellow; black arrows indicate idiosyncrasies of carrier proteins; non-elongating KS^0 domains are marked by green arrows; A, adenylation; ACP, acyl carrier protein; C, condensation; Cyc, cyclisation; DH, dehydratase; KR, ketoreductase; Ox, oxidation; PCP, peptidyl carrier protein; ACP₀ is non-functional

5.2.3 *trans*-AT PKS modules

The PKS part comprises eight modules of the *trans*-AT type distributed on sphC to sphH, which are presumably all loaded by the acyltransferase SphA. The AT shows all signature motifs that indicate malonyl-CoA specificity, which is in agreement with the siphonazole structure, as previous feeding studies showed that the polyketide chains were all assembled from acetate units (Nett et al., 2006).

5.2.3.1 The oxidoreductase of SphA

The AT SphA is fused to an oxidoreductase (OR), which is supposed to act as an ER domain (Bumpus et al., 2008). The siphonazole structure, however, indicates no need for the activity of an ER, as complete reduction of the polyketide chain does not occur in the molecule. *Trans*-AT clusters are known to sometimes encode domains which are apparently superfluous and whose function cannot be explained by the product structure (Piel, 2010). This may as well be the case for the OR, as it was already a hypothesis for the before mentioned Ox domains (5.2.1.1).

ERs usually reduce double bonds, formed by the consecutive action of a KR and a DH. Hence, there are theoretically two possible occasions for the OR to act as an ER, i.e. on the double bonds, generated by the KR and DH domains in modules 2 and 11. Yet both double bonds remain in the structure of siphonazole. Consequently, the question arises why complete reduction of the polyketide chain in these positions does not take place, since the fused AT-ER supposedly has the capability to interact with each PKS module. Indeed, the latter ability is a prerequisite for the loading of the ACPs by the AT. Therefore, it seems reasonable to question the catalytic activity of

the OR. Analysis of the primary structure revealed that the histidine moiety, which is putatively essential for the catalytic activity (Ha et al., 2006), is replaced by a tyrosine residue in SphA. A multiple sequence alignment with other AT-OR proteins showed similar deviations in the corresponding proteins of the chivosazole (Perlova et al., 2006) and disorazole (Carvalho et al., 2005) biosynthesis. In both cases, the function of the OR could not be deduced and is still unknown. In contrast, the catalytic histidine is present in the OR domains of the respective AT-ORs encoded in the gene clusters for dihydrobacillaene (Bumpus et al., 2008), mupirocin (Gurney & Thomas, 2011) and corallopyronin A (Erol et al., 2010), for all of which a role as enoylreductase was proposed in agreement with the molecule structures (figure 5-2).

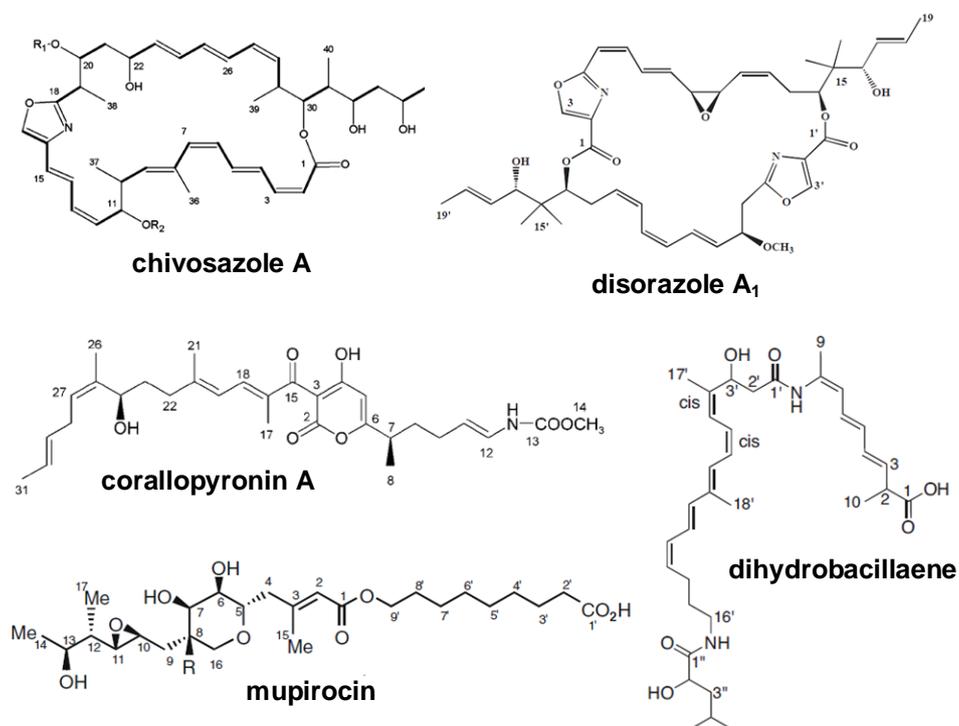


Figure 5-2: Natural products produced by pathways with putative OR domain fused to a *trans*-type AT; chivosazole a $R_1 = \text{Me}$, $R_2 = \text{glycoside side chain}$; mupirocin $R = \text{H}$ in the main product (92%) of *P. fluorescens*

This precedence corroborates the role of the putative catalytic histidine (Ha et al., 2006) and allows the hypothesis that the OR domain of SphA is indeed inactive and has no function in the biosynthesis of siphonazole. Until this assumption is either verified or rejected by experimental data, the possibility cannot be excluded that the domain performs hitherto unidentified steps or has retained its catalytic capabilities,

but is not needed for the assembly of siphonazole, as it was similarly hypothesised for the Ox domains (5.2.1.1).

5.2.3.2 Elongating and non-elongating modules

Modules 5, 7 and 9 all contain KSs, in which an essential histidine moiety of the catalytic triad is replaced by another residue (4.5.2). These KS⁰ termed domains are a common feature of *trans*-AT PKS clusters and are supposed to merely pass on the growing intermediate to the next module instead of performing an elongation step (El-Sayed et al., 2003). As mentioned before, these three modules are all adjacent to NRPS-modules and might therefore be involved in the changeover from PKS to NRPS and vice versa. The presence of only one elongating PKS module between the two NRPS cyclisation modules explains the unusual two-carbon bridge between the oxazole moieties. Modules 6 and 12 both contain two ACP domains. In the first module they are arranged as a tandem, of which the first domain is non-functional due to the lack of the critical serine-moiety. In module 12 another inactive ACP is placed in an unusual position between the KS and an inactive DH. It can be speculated whether the reductive domains were added later to an already existing module by horizontal gene transfer. Only modules 2 and 11 show a canonical domain arrangement. Three of the PKS modules are distributed on different proteins. Module 5 is split between SphC and SphD after the KS⁰. The KS and DH of module 11 are part of SphF, while the rest belongs to SphG, and module 12 spreads over SphG and SphH (figure 4-22).

5.2.3.3 The cryptic module 3

Especially intriguing is the presence of module 3, which consists of a KS and a PCP instead of an ACP. With regard to the siphonazole structure, this module is apparently not needed for the biosynthesis of the compound. Although the KS domain contains an intact catalytic core and all signature motifs (4.5.2), it seems to perform no elongation step. It was already hypothesised that the PCP in this module may be involved in the mediation of a PKS-NRPS interface. This, however, does not explain the obvious lack of catalytic activity in this module. Carrier domains are small flexible peptides that share similar tertiary structure and are usually classified as PCP or ACP according to their accepted substrate. Despite of their shared structure, there

are also distinguishing features. ACPs are generally acidic proteins with a negatively charged surface, whereas PCPs have a more neutral pI value and a non-polar surface (Finking & Marahiel, 2004). The latter ones also cluster according to their substrates and protein partners (von Döhren et al., 1999). Analysis of the protein properties of the PCP in module 3, however, yielded a pI and content of charged amino acids that are very similar to those of the ACPs from the remaining PKS modules. It is nevertheless possible that the carrier domain of module 3 is not recognised by the AT and thus not loaded with an extension unit. Another possibility is that the interaction between the KS and the carrier protein is impaired.

The topic of protein-protein recognition has already been addressed by the Walsh group. By *in vitro* assays it was demonstrated that the iteratively acting cysteine-specific A domain of HMWP2 in yersiniabactin biosynthesis aminoacylates all PCPs, but not an ACP containing fragment (Suo et al., 2001). The same discriminative loading of carrier proteins could also hold true for the AT SphA. They also showed that “hot spots” in the primary sequence of carrier proteins determine the interaction with catalytic partners (Lai et al., 2006). Thus, it is equally possible that the PCP is not recognised by the upstream KS. In both scenarios the intermediate could be passed on to the next module without an elongation taking place.

5.2.4 The mosaic structure of the siphonazole gene cluster

It was suggested that *trans*-AT PKS clusters mainly develop by extensive horizontal gene transfer (Piel, 2010). These events are often reflected by mosaic-like cluster architecture and can be traced back by looking at phyletic clustering, codon usage or GC-content of the gene in question (Lopez, 2003). The latter criterion applies for the two cyclisation module of the siphonazole cluster (4.4.4). Each of them comprises about 5 kb, whose GC-content exceeds 65% in contrast to a ratio of around 50% for the rest of the cluster. Furthermore, the two modules share an almost identical DNA sequence, which adds to the conclusion that they may have their origin in horizontal gene transfer. It is likely that the cyclisation modules are a younger part of the cluster, as the GC-content has not yet been adapted by amelioration (Lawrence & Ochmann, 1997). These hints may indicate that the NRPS parts were inserted into an already existing PKS cluster without loss of function. This event could also explain

some of the carrier protein aberrations and give an answer to the question where the unusual C₂-bridge between the oxazole rings originates from.

It was already noted that the modules encoded by siphonazole biosynthetic gene cluster also comprise non-canonical domain arrangements and KS⁰ domains, which are distributed around the NRPS parts. These idiosyncrasies are typical for *trans*-AT type PKSs and also hint to recombination events by horizontal gene transfer. This kind of genetic rearrangement and the inherent structure of *trans*-AT type PKSs are supposed to favour the generation of new functional clusters, whereas *cis*-AT PKSs generated this way are more likely to be inactive and get lost over time (Nguyen et al., 2008).

5.3 Postulated biosynthesis of siphonazole

With regard to the elucidated domain structure of the PKS/NRPS system, the following way of siphonazole biosynthesis is proposed (figure 4-22). A 3,4-dihydroxybenzoic acid starter unit or its 4-methoxylated derivative is recruited by the A domain of the initiation module and transferred to either of the thiolation domains of the first module. Addition of the methyl group is catalysed by the O-methyltransferase SphB. Following the colinearity rule, the consecutive PKS module elongates the starter by an acetate moiety followed by reduction and dehydration through the corresponding KR and DH domains. As mentioned before, the third module is supposed to merely pass on the intermediate without an elongation step for some unknown reason. The Cyc tandem domains of the fourth module catalyse the condensation with a threonine unit and its consecutive cyclisation. The formed oxazoline ring is converted into an oxazole moiety by either the concerted action of the Ox domain tandem or only one active domain. Of the following three PKS units only module 6 performs an elongation with another acetate moiety. The other two modules contain KS⁰ domains, which do not exert catalytic functions, but may again simply pass on the intermediate. This is followed by the addition of the second threonine derived oxazole ring by the second cyclisation module, which is virtually identical to the first. After another non-elongating PKS unit, NRPS module 11 elongates the intermediate by a glycine moiety. PKS module 12 is congruent with module 2 and adds, reduces and dehydrates a further acetate unit. The last module

condenses the chain with the final acetate unit. As the DH domain of the module is probably inactive, the keto group is only reduced to a hydroxyl before the chain is decarboxylated and dehydrated and the product released by the TE yielding siphonazole.

This reaction scheme follows almost precisely the predicted pathway proposed by Nett and co-workers (Nett et al., 2006), which was based mainly on the results from feeding experiments with labelled precursors. The accuracy of this prediction demonstrates the power and utility of such experiments. The bioinformatic data of the current study support the previous hypothesis and add valuable information on the participating enzymes and additional elements that were not deducible by feeding experiments only.

5.3.1 Nature and origin of the starter unit

Nett and co-workers performed feeding experiment which showed that the styrene residue of the starter unit is derived from the shikimate pathway (Nett et al., 2006). Branching from the key intermediate chorismic acid (Bentley, 1990), this versatile pathway is not only responsible for the production of the aromatic amino acids tyrosine, tryptophan and phenylalanine, but also provides the aromatic building blocks for a number of secondary metabolites including enterobactin, pyochelin and yersiniabactin (van Lanen et al., 2008). Chorismate can be converted to 4-hydroxybenzoic acid by the elimination of pyruvate through chorismate lyase. In vitro experiments with the *O*-methyltransferase SphB showed that the enzyme readily methylates 3,4-dihydroxybenzoic acid (protocatechuic acid). From experiments on the biosynthesis of phenazines it is known that the incubation of chorismic acid with cell-free extracts from *E. coli* clones, which are overexpressing *phz* genes, leads to the complete conversion of the substrate to aromatic acids including predominantly protocatechuic acid (McDonald et al., 2001). Furthermore, protocatechuic acid is a major metabolite in the degradation of aromatic compounds and can be directly generated from 4-hydroxybenzoic acid by *p*-hydroxybenzoate hydroxylase (Buchan et al., 2004). This compound is therefore proposed as the substrate for the adenylation domain of the initiation module of the siphonazole gene cluster. This hypothesis will have to be proved by suitable experiments on the A domain substrate specificity like e.g. an ATP-PP_i exchange assay (Phelan et al., 2009).

5.3.1.1 The O-methyltransferase SphB

The gene *sphB* encodes a SAM-dependent O-methyltransferase, whose apparent function is the methylation of the 4-hydroxy group of the aromatic starter unit. The occurrence of C-, N-, or O-methylations is widespread in the biosynthesis of polyketides and non-ribosomal peptides. In NRPS systems MT domains are often inserted into the A domain of the corresponding module, as this is e.g. the case in pyochelin biosynthesis (Patel et al., 2001). PKSs can also insert α -methyl branches through MT domains, which is usually the case in *trans*-AT type PKS systems, where all modules are loaded with the same extender unit (Piel, 2010), but can also be observed e.g in epothilone biosynthesis (Molnar et al., 2000). Apart from that, methyl groups can also be introduced by *trans*-acting enzymes. The biosynthetic clusters of melithiazole (Weinig et al., 2003) and bryostatin (Sudek et al., 2007) both contain O-MT domains inside the modules, as well as distinct methyltransferases, which are supposed to perform O-methylations.

The *in vitro* assay (4.9.1.1) demonstrated that SphB accepts protocatechuic acid as a substrate for a SAM-dependent methylation. These results support its putative function in the biosynthesis of siphonazole. The exact position of the introduced methyl group was not determined, but in consideration that no siphonazole derivative with a single methylation at the 3-hydroxy group has been found, it is probable that SphB acts quite selectively on the 4-OH. It is, however, known that a dimethoxy-congener of siphonazole is produced in very small quantities by *Herpetosiphon* sp. 060 (Nett et al., 2006). As no other methyltransferases have been found in the vicinity of the putative siphonazole gene cluster, it seems likely that SphB also acts on the 3'-OH at a very small rate. The existence of a 3-methoxy variant of siphonazole can therefore not completely be excluded. More studies on the enzymatic activity and analysis of the methylation products will be needed to fully clarify the specificity of SphB. A comparison with other possible precursors could also help in verifying the identity of the starter unit.

It remains unclear, whether the methylation takes place before or after the starter unit is tethered to the carrier protein of the initiation module. It would seem plausible to first withdraw the starter unit from the shikimate pathway and adjoining metabolism and tether it to the multienzyme complex before adding further modifications. This

way, the assembly could proceed without interruption after securing the first building block. Concomitantly, this would ensure that only those precursors are methylated which have already entered the siphonazole pathway and are thus needed for biosynthesis. Otherwise, the methyltransferase would have to act on the free substrate which would then have to be discriminated and recruited by the first A domain from a pool of predominantly unmethylated units. The assay showed that the methyltransferase itself readily accepts free substrate. Clarification of the point of methylation could arise from comparative experiments with free substrates and carrier protein-bound moieties or from determination of the substrate specificity of the initial A domain. If the starter unit was methylated prior to recruitment by the A domain, the activity of SphB could act as a limiting step for the siphonazole biosynthesis. In this case, addition of a further copy of the gene should lead to increased siphonazole production rates (see also 5.3.1.2).

5.3.1.2 The DAHP synthase SphI

DAHP synthase is the first enzyme of the shikimate pathway and ubiquitously present in bacteria, fungi and plants, but with no known counterpart in animals. The enzyme generates 3-deoxy-D-arabino-2-heptulosonic acid 7-phosphate (DAHP) through the aldol-like condensation of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP). Several classifications for DAHP synthases exist. Birck and Woodard proposed a separation into the metalloenzymes class II and the metal independent class I proteins (Birck & Woodard, 2001). In a homology based approach Jensen and co-workers distinguished between the plant-like AroA_{II}, which is typical for higher plants, and family AroA_I. The latter one can be further subdivided into subfamilies AroA_{I α} , exemplified by the *E. coli* DAHP synthases and AroA_{I β} , represented by the *B. subtilis* DAHP synthase. These subfamilies correspond to Birck's class II and class I, respectively (Jensen et al., 2002).

DAHP synthases are known to be tightly controlled either by transcriptional regulation (Walker et al., 1996) or by feedback regulation through products or intermediates of the shikimate pathway. Members of the AroA_{I α} are usually inhibited by one of the three aromatic amino acids, while AroA_{I β} enzymes are allosterically affected by prephenate or chorismate (Wu & Woodard, 2006). Organisms often contain different paralogs of DAHP synthases. *E. coli* e.g. contains three isoforms of this enzyme,

which are each inhibited by one of the aromatic amino acids, thus allowing a response to the exogenous availability of these products (Bentley, 1990).

Multiple sequence alignment shows excellent consensus of SphB with the conserved motifs of *B. subtilis* AroA and other members of the Aro_{Iβ} subfamily, as defined by the Jensen group (Subramaniam et al., 1998). The presence of shikimate pathway homologues has already been discovered in the biosynthetic gene clusters of secondary metabolites like rifamycin, chloramphenicol and balhimycin (He et al., 2001; Thykaer et al., 2010). Balhimycin contains the non-proteinogenic amino acids hydroxytyrosine and 4-hydroxyphenylglycine, which are both derived from the shikimate pathway. It was shown that overexpression of the *dahp* gene at the border of the cluster led to a threefold increase of balhimycin production, while a *dahp* deletion mutant produced significantly decreased amounts. These additional copies of shikimate pathway genes are hypothesized to ensure precursor supply for the secondary metabolite. The same function can therefore be assumed for SphI in the siphonazole cluster. As the aromatic starter unit of the siphonazole biosynthesis is derived from chorismate, an additional DAHP synthase could attribute to a sufficient production of this building block. Additionally, the PEP synthase Orf1 was found upstream of the cluster, which possibly supplies the DAHP synthase with one of its substrates, thus adding to the precursor formation. These assumptions could be clarified in an analogous way to the balhimycin enzymes, i.e. by insertion of an additional copy or deletion of *orf1* and *sphI* and subsequent observation of siphonazole production rates.

5.3.2 Substrate specificity of the A domain of the initiation module

The substrate specificity of the first A domain could not be reliably predicted by bioinformatic methods. An alignment of the Stachelhaus-code, as derived by the NRPSPredictor software, found no counterpart in the database of known specificities. This lack of significant homology to known A domains underlines the unique position of this domain. The software tool provided two specificity predictions based on the assumption that either several or only few substrates share the same properties (Rausch et al., 2005). These predictions yielded contradicting results and proposed either specificity for polar amino acids like aspartate or glutamine or for hydrophobic amino acids like valine or leucine. Prediction scores, though, were very low in both

cases. The NRPSPredictor2 provided additional information and suggested an affinity for a hydrophilic substrate, based on the separation into three clusters of chemically similar substrates. The deduced Stachelhaus-code shows closest similarity to a sequence specific for phenylalanine, which, however, clusters with hydrophobic aromatic substrates. This would concur with the predictions for a hydrophobic amino acid and is supported by the finding that the starter unit incorporated in the siphonazole biosynthesis is indeed an aromatic acid.

5.3.2.1 ATP-PP_i exchange assay with the initiation module A domain

The generally contradicting results and weak predictions of the software tools demonstrate that the substrate specificity of this domain can only be clarified by experimental analysis. Though the A domain of the initiation module was successfully expressed, the ATP-PP_i exchange assay (Phelan et al., 2009) failed to provide experimental data for a specificity, as no activity could be measured with any of the tested substrates.

It was shown that many NRPS clusters also encode small (~ 70 amino acids) MbtH-like proteins, which sometimes influence amino acid activation (Thomas et al., 2010), though cases are known, where the knock-out of these proteins did not impair metabolite production (Stegmann et al., 2006). An MbtH-like protein was not detected within the siphonazole gene cluster or in its vicinity, but their presence is known in the fully sequenced genome of the closely related bacterium *H. aurantiacus*. It is therefore possible that the siphonazole producer *Herpetosiphon* sp. 060 also contains MbtH-like proteins that interact with the siphonazole biosynthetic enzymes, but are encoded elsewhere. As these proteins are usually encoded within the corresponding gene cluster, it seems more likely that the enzyme activity of the A domain was lost during purification and concentration procedure. It is therefore recommended to test this possibility in further enzyme assays. Furthermore, the gene cluster and its adjoining regions should be carefully scanned again to search for sequences encoding MbtH-like proteins, which might have escaped attention due to their small size. Further experiments could include co-expression with an MbtH-like protein from *H. aurantiacus* to test its influence on the A domain activity. Apart from that, the Challis group observed that in clusters where no MbtH-like protein is encoded nearby, the ATP-exchange assay often yields better results when the

investigated A domain is expressed together with its respective carrier protein (S. Bouhired and G. Challis, personal communication). Hence, the construction of a new expression construct that includes one or both ACPs (figure 4-22) of the first module should be considered. An application of various similar compounds like 3- and 4-methoxybenzoic acid and 3-hydroxybenzoic acid as substrates would be interesting, but the realisation is probably limited by the poor solubility of these substances in aqueous solutions.

5.3.3 Decarboxylation of the terminal acetyl moiety

The siphonazole structure comprises an unusual alkene terminus, which is, according to feeding studies, the product of a decarboxylation of the last acetate unit. The formation of a similar olefinic terminus in the biosynthesis of curacin A was investigated by the Sherman group (Gu et al., 2009). Curacin A is a mixed PKS-NRPS natural product from the cyanobacterium *Lyngbya majuscula* with potent anticancer properties. The curacin gene cluster comprises an unusual termination module with an ACP–sulfotransferase (ST)–TE tridomain. According to the described mechanism, the ST domain first sulfonates the β -hydroxyl of the curacin intermediate with adenosine 3'-phosphate 5'-phosphosulfate (PAPS) as donor group. This turns the hydroxyl into a much more favourable leaving group. The curacin TE represents a new subfamily of thioesterases, which is specifically adapted for β -sulfated substrates. It catalyses not only thioester hydrolysis, but is also responsible for the decarboxylation and concomitant sulfate elimination and thus for the formation of the terminal olefin (Gehret et al., 2011). This ACP – ST – TE tridomain arrangement and the conserved features of the specialised TE were also found in other putative biosynthetic gene clusters. A similar mechanism is therefore proposed for other metabolites like kalkitoxin, also produced by *L. majuscula*, or haliangicin from *Haliangium luteum* (LePage et al., 2005; Fudou et al., 2001).

The putative siphonazole gene cluster, however, does not encode an ST in the termination module or at any other position. Furthermore, the TE SphJ is not related to the TE subfamily determined by the curacin TE and shows none of the specific sequence motifs. A decarboxylation mechanism similar to the curacin A biosynthesis can therefore probably be excluded.

A different mechanism was discovered in the biosynthesis of the immunomodulator tautomycetin from *Streptomyces* sp. CK4412 (Choi et al., 2007). Herein, terminal alkene formation is catalysed by the concerted action of the dehydratase TtnF and the decarboxylase TtnD (Luo et al., 2010), which both work as tailoring enzymes after precursor release by the TE. Gene inactivation studies revealed that TtnF eliminates a β -hydroxyl group, which results in the formation of a diene intermediate. This dehydration step is apparently critical for the decarboxylation by TtnD, which does not occur in the absence of TtnF.

A similar reaction could be proposed for the siphonazole biosynthesis. Analogous to the structures of leinamycin and tautomycetin, the predicted full length precursor of siphonazole harbours a β -hydroxyl group at its terminus. The siphonazole cluster encodes a putative cysteine-hydrolase located in the terminal module. As members of this enzyme class catalyse a variety of hydrolytic reactions, the hydrolase in SphH could very well act as decarboxylating enzyme. Nevertheless, if the reaction proceeds in an analogous way as in tautomycetin biosynthesis, there is still a dehydration step required. This could readily be accomplished if the DH domain of the terminal module is, against the prediction from the analysis of the catalytic residues, active after all. Another possibility is that dehydration and decarboxylation are both catalysed by the hydrolase. Enzymes of the α/β – hydrolase superfamily are known for their catalytic promiscuity and many members can perform more than one chemical reaction (Li et al., 2008). A case of a hydrolase with dehydration activity, however, is not reported so far. Putative hydrolases are also known to be present in other gene clusters for secondary metabolites like mupirocin (El-Sayed et al., 2003) and leinamycin (Tang et al., 2004). In the latter case, the sequence for the putative hydrolase is located at the same position, i.e. after the carrier protein of the termination module. The only difference in the siphonazole cluster is the presence of the DAHP synthase SphI encoded between the hydrolase and the TE SphJ. A function for the hydrolases in the other named gene clusters has not been proposed so far, neither has the involvement of a decarboxylation reaction been reported. An elucidation of the decarboxylation reaction in siphonazole biosynthesis could arise from gene inactivation studies with the SphH-hydrolase and the DH domain of the terminal module. Analogous to the studies on tautomycetin biosynthesis, deletion of the hydrolase gene should result in the production of a siphonazole intermediate with

a carboxyl terminus instead of an olefin and possibly a β -hydroxyl if the enzyme catalyses both reactions. Similarly, a deactivation of the DH gene should yield a hydroxylated derivative, either without or with a carboxyl group if the dehydration is also a prerequisite for decarboxylation. Concomitantly, it would be clarified if the DH domain exerts catalytic activity at all or if dehydration is catalysed otherwise.

5.3.4 Transcriptional organisation of the gene cluster

By PCR-probing with RNA transcripts of the putative biosynthetic genes it was shown that sphC – sphJ are part of one transcriptional unit and are transcribed as a single polycistronic mRNA. The missing PCR fragment between sphB and sphC indicates that the first two genes are part of a different operon. It should be noted that the transcript apparently starts upstream of sphA. This implies that sphA and sphB are either the final part of a longer transcript together with other genes or that they are at least under control of a more distant promoter site. The only ORF in the upstream region that was hypothesized to be possibly related to siphonazole biosynthesis is orf1 encoding a PEP synthase. The assumed involvement in the precursor formation of the starter unit would justify a regulation together with the sph genes, yet orf1 is oriented in the opposite direction and therefore has to be part of a different transcriptional unit than sphA and sphB. Further mRNA studies of the upstream region may help to elucidate the boundaries of this first transcript. The discovery of only two operons is consistent with the fact that all genes of the putative siphonazole cluster are oriented in the same direction. Similar observations were made in the gene clusters for the antibiotics novobiocin and clorobiocin (Eustáquio et al., 2005).

5.4 Spatiotemporal biosynthesis of siphonazole

The gene expression of ORFs sphA – sphJ from the putative siphonazole cluster was examined by RT-PCR from the corresponding mRNA transcripts. Thereby it was demonstrated that mRNA from these genes was only detected in the first three days of cultivation, indicating that siphonazole biosynthesis does only occur in that period. The temporal maximum appears to lie on the second day.

The findings from the mRNA analysis are supported by the IMS results (4.10.2). The spectrometry images depict a mass which matches that of siphonazole and whose

spatial distribution is restricted to the edge of the *Herpetosiphon* culture. Due to the gliding motility of *Herpetosiphon*, the growing cells swarm from the centre of the colony to the rim, where in consequence younger cells are located. Thus, the spatial distribution of siphonazole at the growing edge concurs with its biosynthesis in an early phase of the culture.

The putative siphonazole cluster is flanked by genes whose predicted functions can be assigned to the primary metabolism. It was already mentioned that the first cluster genes belong to a longer operon and might hence be co-expressed with primary metabolism enzymes, which would explain a siphonazole production from the first day on. Likewise, it is possible that siphonazole is produced directly at the transition from growth phase to stationary phase, which may be reached early by subpopulations. Further studies should include monitoring of cell growth to establish a correlation between siphonazole production and growth phase.

5.4.1 Biological role of siphonazole

The analysis of the spatiotemporal distribution of siphonazole allows speculations about its physiological role in *Herpetosiphon* sp. 060. In general, secondary metabolites can take over a variety of functions for their natural producer and may act as pigments, virulence factors, defence mechanisms or messengers (Hertweck, 2009). Nevertheless, the actual role of a metabolite in its natural context is often unclear. Likewise, no biological role could be attributed to siphonazole so far. From a structural point of view, the function of the two heterocycles may be considered. Oxazole or thiazole rings are frequently found in siderophores like vibriobactin or yersiniabactin, where they participate in the chelation of metal ions (Roy & Walsh, 1999). These metabolites are, however, mostly expressed in an iron deficient environment, whereas siphonazole is also produced in the presence of a supply of metal ions. Furthermore, heterocycles create interaction sites for RNA, DNA or proteins. The thiazoles in the antibiotic and antitumor agent bleomycin for example have been shown to intercalate into the DNA double helix (Vanderwall et al., 1997). For siphonazole a function as defensive agent is admittedly rendered unlikely due to the lack of cytotoxicity and antibiotic activity against other microorganisms. An intracellular ligand of siphonazole is not known to the present date.

The results from the RNA expression analysis and the image mass spectrometry (IMS) suggest that siphonazole plays a role in the early growth phase and is expressed at the edge of a growing culture on solid medium. It is also known that siphonazole is excreted into the medium, as it can be readily isolated from the culture supernatant (3.2.4). Therefore, one could imagine a role as messenger or intercellular communication agent. Herein, siphonazole might affect the swarming direction of the growing culture or participate in the interaction with other microorganisms. Small molecules can mediate a great number of responses in interspecies signalling. These are not restricted to growth inhibition, but can also induce changes in metabolism, phenotype or development like sporulation or the formation of fruiting bodies (Shank & Kolter, 2009). A function for siphonazole in intra- or interspecies crosstalk would accord with the absence of an obvious beneficial effect for the producer and missing antibiotic activity. More insights could be gained by further IMS studies, which should address the siphonazole production at the boundary points between a *Herpetosiphon* sp. 060 colony and other bacteria on the same plate (Esquenazi et al., 2008 and 2011).

5.5 Complete assembly of the putative siphonazole cluster

The *in silico* completion of the sequence of the putative siphonazole gene cluster provided the basis for the subsequent assembly of a fosmid construct harbouring the complete sequence of genes sphA – sphJ. By directed re-screening of the genomic library complementary clones were detected and recombined with the already identified fosmids, thus yielding a vector with a size of 71 kb carrying the complete sequence information.

5.5.1 *In silico* and genomic library screening

In a previous study, the fosmid library of *Herpetosiphon* sp. 060 was screened with degenerate primers based on conserved motifs from A, Cys and KS domains (Ö. Erol, personal communication). This screening led to the identification of three cosmids harbouring parts of the putative siphonazole cluster. For the complementation of the whole cluster, the follow-up fosmids IC2 and IC3 were identified by screening the genomic library with specific primers for the TE and its adjacent region. These fosmids had obviously escaped discovery in the original

screening, although they encode not only one A domain, but also three KSs of the putative siphonazole cluster. Derivation of the specific primer sequences was enabled by the sequence data from the genome sequencing. Apart from providing workable DNA constructs for subsequent experiments, the genomic library can in turn complement the information of a draft genome. The two Cys domain tandems encoded in the putative siphonazole cluster were both located in gaps between the contigs from the genome assembly and were first discovered by the library screening. These findings supported the selection of the fosmid clones and facilitated the following identification of the correct cluster in the genome data. This interplay clearly illustrates the usefulness of combining the approaches of a genomic library and an *in silico* screening of genomic data. The incorrect sequences in contig 1462 demonstrated, however, that results from a genome sequencing project cannot always be adopted unaided, but sometimes require revision and should therefore be carefully evaluated.

5.5.2 Assembly by λ -Red recombination

PKS and NRPS assembly lines frequently comprise immense sizes that excel the average insert size of common fosmids or cosmids (Binz et al., 2008). These size limitations can be overcome by the use of BAC libraries with a bigger insert capacity or coexpression of compatible expression plasmids with subsets of the gene cluster (Pfeifer, 2001; Pfeifer et al., 2003). Another possibility is the reconstitution of the gene cluster from several parts on different fosmid or cosmid clones. λ -Red mediated recombination has been established as an appropriate method for rebuilding gene clusters for secondary metabolites (Wenzel et al., 2005). This technique is based on proteins from the bacteriophage λ , which greatly increase the recombination rate of linear DNA by homologous recombination without having to rely on restriction sites (Stewart et al., 1998). The target region in the genome or on a circular DNA construct is defined by short homology extensions at the ends of the linear DNA-fragment. Concomitantly, the same method allows the introduction of modification like the insertion of promoters and homology regions, which are required for the expression in a heterologous host, as well as deletions for gene inactivation studies (Datsenko et al., 2000).

By employing λ -Red recombination techniques, it was possible to stitch the putative siphonazole gene cluster from overlapping parts on three fosmids, thus generating a construct carrying the sequences of sphA – sphJ. The fosmids were based on the pCC1FOS vector backbone and contained their DNA inserts all in the same direction. This allowed, after insertion of selectable markers, a direct isolation of the complementing DNA fragments by restriction digestion. Although the generated fragments had a considerable size of up to ~20 kb, they were readily isolated, electroporated and recombined by standard techniques (Gust et al., 2004). This is in accordance with previous achievements of Müller and co-workers, who utilised a similarly large fragment in the completion of the phenalinolactone cluster (Binz et al., 2008). Recombination of the cluster parts resulted in a fosmid which harbours an insert with a total size of ~63 kb. Even though this exceeds the usual insert size of fosmids, the construct proved to be stable and was readily isolated from an *E. coli* culture. The consistency of the recombined fosmid was proved by restriction and PCR. The straightforward complementation of the putative siphonazole gene cluster clearly demonstrates the effectiveness of λ -Red based recombination methods. The generated construct provides a base for a functional analysis of cluster genes and subunits. Further modifications, to prepare the cluster for an expression in a heterologous host, may be introduced by the described methods.

5.5.3 Assembly by yeast recombinational cloning

Plasmid construction in yeast by homologous recombination is a technique that has been known for over 25 years (Ma et al., 1987) and which has been successfully applied for the assembly of DNA constructs up to artificial genomes from as many as 38 fragments (Gibson et al., 2008; Gibson, 2009). Therefore, recombinational cloning in yeast can be seen as a highly appropriate method for the reconstitution of secondary metabolite gene clusters from several parts. In this study, experiments aimed at the recombination of three overlapping cluster parts with a linearised vector backbone. The PCR screening results suggest that successful recombination in yeast was accomplished repeatedly and that recombinant plasmid constructs could be isolated from yeast cultures, yet the transformation of a plasmid with the complete cluster sequence into *E. coli* cells remained unsuccessful. The occurrence of *E. coli* clones on selective plates and the results from the PCR screening for the vector backbone imply that at least partial constructs were introduced into *E. coli* cells. In

this screening also some clones appeared which did not yield any product. This can either hint to the spontaneous development of antibiotic resistance or to the uptake of plasmids which contained at least fragments of the cluster insert.

Repeated screening of yeast clones on different days point at the possibility that complete plasmid constructs were first generated, but not maintained in the cells. It is known from studies in *E. coli* cells that the “metabolic burden” caused by plasmid maintenance can also affect plasmid stability, but as the utilised pENTRC1FOS is a low copy plasmid, the restraints for the cell should have been kept at a minimum level. Besides, there are also structural factors that can cause plasmid instability like size, polyA sequences or repeated motifs that are prone to deletions, duplications or insertions (Silva et al., 2012). Furthermore, instability can be conferred by environmental stress like antibiotic concentrations or medium composition. The PCR results might therefore indicate that the generated plasmid construct is not stable. This finding would be surprising with regard to the plasmid backbone, which is designed to maintain large inserts, and the fact that a stable fosmid of similar size and the same insert sequence could readily be created by λ -Red recombination. It is therefore questionable if the fragments were actually correctly recombined by the yeast cells. In view of these uncertainties, λ -Red mediated homologous recombination proved to be the more reliable and therefore preferential method for the reconstitution of the gene cluster.

5.6 Heterologous expression of the siphonazole gene cluster

In the previous chapters a number of experiments have been proposed that could help gaining more insights into some of the still unravelled processes in siphonazole biosynthesis. These experiments include the inactivation of certain genes or single domains, which requires several preconditions to be met. Prerequisites include the availability of the gene cluster, genetic systems for *in vivo* modifications of the cluster sequence and the possibility of expressing the altered assembly line in either the native or a suitable heterologous host (Galm et al., 2011). As *Herpetosiphon* sp. 060 seems to be unamenable to genetic manipulation (5.14.1), the heterologous expression of siphonazole remains the only alternative for further studies. For this purpose, the choice of a suitable host strain is critical, as the potential producer must

meet a number of criteria. Apart from genetic accessibility a suitable host needs to provide all necessary precursors and perform the posttranslational modification of the carrier proteins (Pfeifer & Khosla, 2001). While the phosphopantetheinyl transferase (PPTase) Sfp from *B. subtilis* is known to accept virtually all types of ACPs and PCPs, other PPTases are more scrupulous about their substrates. Members of the genus *Streptomyces* have been frequently used as heterologous expression system for natural products. Not only are genomes and metabolism of several species like *S. coelicolor* well characterised, but *Streptomyces* are also prolific producers of secondary metabolites themselves. Strains like *M. xanthus* (Julien & Shah, 2002) or *P. putida* (Wenzel et al., 2005) are also used for similar reasons. Due to its robust expression system and the well developed genetic tools, *E. coli* remains an attractive heterologous host. Several natural products like erythromycin and yersiniabactin (Pfeifer 2001; Pfeifer et al., 2003) have been successfully produced in *E. coli* strains. These heterologous expressions often require extensive metabolic engineering of the host strain (Watanabe et al., 2006) and insertion of additional genes like the mentioned sfp to adapt the organism to the production of the foreign compound. It is long since known that codon usage greatly influences gene expression. Heterologous genes with a codon bias different from the host organism are more likely to be prone to frame shifts, translational errors or decreased expression (Kane, 1995). These difficulties can be overcome by choosing a closely related expression strain, supplementing tRNAs for rare codons or optimising the heterologous genes for the preferred codon bias (Menzella et al., 2006). Gene clusters for natural products are usually explored in genetic libraries based on fosmids, cosmids or BACs, which are designed for *E. coli* and are usually not replicated in other host organisms. The transfer to a heterologous host therefore requires either a specific integrative shuttle vector or the insertion of suitable promoters and homology regions for the integration into the genome. These modifications can readily be performed by the implementation of recombination methods like λ – Red (5.5.2).

The choice of a heterologous host for the expression of the siphonazole cluster is hindered by the isolated genealogical position of *Herpetosiphon*. To the present date, no expression host within the Chloroflexi group has been established and none of the available, well researched expression strains is somehow related to *Herpetosiphon*. On the other hand, siphonazole does not require any unusual building blocks with the

exception of the aromatic precursor derived from the shikimate pathway, which in turn is very common in bacteria. From this point of view, an expression in *E. coli* may well be considered, but taking all the limiting factors mentioned above into account, it is not surprising that the attempted expression in the sfp containing BAP1 cells was not successful (4.11). The phylogenetic distance between *Herpetosiphon* and potential host strains is also reflected in the codon bias. A measure for the similarity of the codon usage of two organisms is the codon adaptation index (CAI), which represents the codon bias of a set of highly expressed genes (Sharp & Li, 1987). The CAI for the siphonazole cluster and several potential host organisms was calculated with the JCat tool of the University of Braunschweig (Grote et al., 2005, Carbone et al., 2003). With a CAI of ~ 0.2 the comparisons revealed only a very low similarity to *E. coli* K12, but showed even lower values for *M. xanthus*, *S. coelicolor* and *P. putida* (table 5-1). It is therefore questionable if these other strains are more suited for a heterologous siphonazole production, as the expression of even single proteins proved to be difficult in *E. coli* (4.9). It can be assumed that the protein expression would deteriorate in a host with an even more different codon usage. The moderate GC-content of *Herpetosiphon* does also not necessitate a high GC-content host like *S. coelicolor*.

Table 5-1: Codon adaptation index of the sph genes

Values were calculated with JCat by comparison of the siphonazole cluster sequence with selected organisms from the software's database

Organism	Codon Adaptation Index
<i>E. coli</i> K12	0.199
<i>Myxococcus xanthus</i> (strain DK 1622)	0.130
<i>Pseudomonas putida</i> (strain KT 2440)	0.174
<i>Streptomyces coelicolor</i> (strain A3(2))	0.060
<i>Bacillus subtilis</i> (strain 168)	0.332
<i>Bacillus licheniformis</i> ATCC 14580	0.488

It was found that *Bacillus licheniformis* has one of the best CAI values with a similarity of over 0.48. In the industry, this organism has been used for decades for the production of extracellular enzymes, lipopeptide surfactants and organic acids and polymers (Waschkau et al., 2008). Some stains are as well producer of the NRPS derived peptide antibiotic bacitracin (Kleinkauf & von Döhren, 1987). The average

GC-content of *B. licheniformis* is with 46.2 % close to that of *Herpetosiphon* sp. 060 (55.7%) (Rey et al., 2004). Regarding these criteria and the fact that transformation and cultivation methods for *B. licheniformis* are well established, this organism should be considered as a possible host for the heterologous expression of the siphonazole gene cluster. Expression and shuttle plasmids for *B. licheniformis* are available (Niu & Wang, 2007), which would facilitate the construction of an appropriate expression vector for the siphonazole production.

5.6.1 Heterologous protein expression

In the course of this study the *O*-methyltransferase SphB, the DAHP synthase SphI and the A domain of the initiation module encoded by SphC have been successfully expressed as heterologous proteins in *E. coli*. The successful expression of functionally active proteins from *Herpetosiphon* species in *E. coli* has already been reported in the early 1990s (Düsterhöft et al., 1993). The performed experiments, however, made several challenges in the heterologous expression of proteins apparent, and only SphB could be expressed as soluble protein in high quantities in a straightforward manner. In contrast, the heterologous expression of the A domain of the initiation module proved to be more difficult.

In contrast to the expression of distinct ORFs, where start and stop codon mark the borders of the gene, the expression of a single domain out of a multimodular protein leaves more freedom for the positioning of start and end. Expression experiments with proteins from a *Pseudomonas* NRPS cluster showed that moving the position of a primer site for as little as five amino acids can have a decisive impact on the production of the heterologous protein construct (T. Höver, unpublished results). Therefore, the primer sites for the A domain were altered repeatedly, without yielding any detectable product. Other significant factors are the choice of vector and expression strain. By changing to the pET151 vector and BL21 star™ *E. coli* cells, the expression of the A domain was accomplished with one of the same DNA insert that failed to produce any protein in pET28a(+) and regular BL21 cells. In the same way expression of SphI was achieved.

It was shown that, amongst other parameters, the mRNA levels and mRNA stability are important factors in protein expression and also affect production of proteins from genes with rare codons (Wu et al., 2004). The problem of diverging codon usage has been discussed before (5.6) and can be answered by the use of optimised expression strains, which either improve RNA stability or contain extra copies of rare codon tRNAs. The BL21 star™ cells are deficient of two protease genes and contain a truncated RNase E gene, which contributes to an increased stability of the produced protein and the mRNA, respectively (Grunberg-Manago et al., 1999; Lopez et al., 1999). The expressed proteins from SphC and SphI turned out to be mainly insoluble. Rosano and Ceccarelli observed in their experiments with codon bias-adjusted *E. coli* strains that overcoming the effects of different codon usage had a detrimental effect on the solubility of the expressed proteins (Rosano & Ceccarelli, 2009). They concluded that the elimination of translation difficulties leads to increased production rates of the heterologous protein, which promotes misfolding and aggregation. The increased mRNA stability in the BL21 star™ cells has most likely similar effects and is probably the cause for the large portion of insoluble protein in the expression of the A domain and SphI. The overproduction of proteins can be sufficient by itself to cause the formation of insoluble aggregates designated as inclusion bodies (Schumann & Ferreira, 2004). Nevertheless, the BL21 star™ cells proved to be the best choice, as both regular BL21 and ArcticExpress cells, which contain special cold-adapted chaperons to promote the correct folding of heterologous proteins (Ferrer et al., 2003), failed to produce a detectable amount of the soluble target proteins. Optimisation of culture and induction condition may help to increase the rate of soluble protein.

5.7 Metabolic potential of *Herpetosiphon*

Members of the Chloroflexi group have been the object of interest with regard to their photosystem for some time (Xin et al., 2011), but apart from the discovery of a diterpenoid from *Chloroflexus aurantiacus* (Hefter et al., 1993) and the characterisation of pigments in *Herpetosiphon* species (Kleinig & Reichenbach, 1977; Reichenbach et al., 1978), a closer examination of their metabolic capabilities has been neglected. The predictions by CLUSEAN and the NaPDos database provided evidence that the genome of *Herpetosiphon* sp. 060 comprises a number of yet

cryptic PKS and NRPS gene clusters. Sequence comparisons showed that the mixed PKS/NRPS cluster found in contig 0805 and the adjoining lantibiotic biosynthetic genes are shared with the closely related *H. aurantiacus*, indicating that the metabolic machinery for the production of secondary metabolites may be more widespread in this bacterial group.

The structure of siphonazole and its putative biosynthetic gene cluster have revealed several peculiarities and raised questions that are still not completely answered. Considering this and the fact that none of the other predicted clusters have been linked to any natural product yet, *Herpetosiphon* and other Chloroflexi may still harbour a lot of untapped potential. An increasing number of these so-called “orphan clusters” (Gross, 2007) from different bacterial classes has been discovered by genome mining approaches, which can and already did provide the basis for the discovery of new natural products. A number of methods have been developed for the awakening of “silent” gene clusters (Chiang et al., 2011) involving, e.g. heterologous expression, gene inactivation and the genomisotopic and OSMAC (one-strain-many-compounds) approach. Refinement of the present bio-informatic data in combination with these techniques may therefore lead to the discovery of further metabolites from *Herpetosiphon*.

The discovery of carotenoids in *Herpetosiphon*, which are common in myxobacteria (Kleinig & Reichenbach, 1977), gave rise to the hypothesis that the two groups may share biosynthetic pathways (Nett, 2007). Database comparisons of the sph genes and predicted A and C domains did not provide further evidence for this assumption. Regarding the highest homology of the corresponding proteins, it was rather illustrated that no relationship to a specific bacterial group emerged, but that similarities were distributed over a wide range of phyla including Bacilli, Streptomyces and Clostridia, Pseudomonads and Cyanobacteria. Sequence identities were generally below 60%, pointing out the evolutionary distance of *Herpetosiphon* to these other groups. Albeit this rather unspecific alignment allows no further conclusions on metabolic relationships, it may be taken as another indicator for the unique position of this phylum. In view of this, unlocking the metabolic potential of *Herpetosiphon* and *Chloroflexus* species still has the capability of coming up with new surprises.

6 Summary

Polyketides (PKs) and non-ribosomal peptides (NRPs) are two major classes of natural products. They comprise a vast variety of structurally and functionally diverse compounds, which include a great number of clinically important substances. These metabolites are the product of giant multienzymes termed polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), respectively.

The present study investigated the putative biosynthetic gene cluster of siphonazole. This compound is produced by the *Herpetosiphon* strain 060, which belongs to the deep rooted phylum of Chloroflexi, and is the first secondary metabolite described from this group. Siphonazole is composed of building blocks derived from acetate, amino acid and shikimate pathways. The gross structure comprises a styrene moiety, two oxazole rings connected by a C₂-bridge and an unusual diene side chain.

Based on a fosmid library from previous work, it was attempted to identify the gene cluster by knock-out experiments. The application of various transformation methods, including heat shock, electroporation, sonoporation, conjugation and chemical transformation of protoplasts, showed that the *Herpetosiphon* strain is unamenable to genetic modification.

In a second approach, genomic DNA of *Herpetosiphon* sp. 060 was subjected to 454 sequencing yielding draft genomic data of the strain. The application of bioinformatic tools led to the identification of several PKS and NRPS cluster parts. Through a comparison with subclones from the fosmid library, six genome fragments were identified carrying partial sequence information of a hybrid PKS/NRPS gene cluster. By employing a combination of PCR and sequencing reactions the gaps between the fragments were closed. The completed sequence information comprises a gene cluster of 50 kb with ten open reading frames (sphA – sphJ), which encode the putative enzymes for the biosynthesis of siphonazole.

Through a detailed bioinformatic analysis of the encoded proteins, it was revealed that they include eight PKS modules, which are assumably all loaded in *trans* by the acyltransferase (AT) SphA, and four NRPS modules (figure 6-1). The AT has a predicted specificity for malonyl-CoA, which is consistent with the siphonazole

structure, and is fused to an oxidoreductase domain with hitherto unknown function. Two of the NRPS units are heterocyclisation modules, in which tandem pairs of cyclisation and oxidation domains catalyse the formation of the oxazole rings from threonine moieties. The first NRPS module initiates the biosynthesis probably by incorporation of a dihydroxybenzoic acid, although ATP-PP_i exchange assays failed to determine the substrate specificity of the corresponding A domain. One module, consisting of the unusual arrangement of a ketosynthase (KS) and a peptidyl carrier protein (PCP), is inconsistent with the molecule structure of siphonazole. It is suggested that no elongation is performed on this module due to unsolved protein-protein recognition issues between its carrier protein and either the KS or the AT. A hydrolase is encoded downstream of the last module, which presumably participates in the formation of the diene terminus of siphonazole. However, the mechanism leading to this unusual terminus is still unsolved.

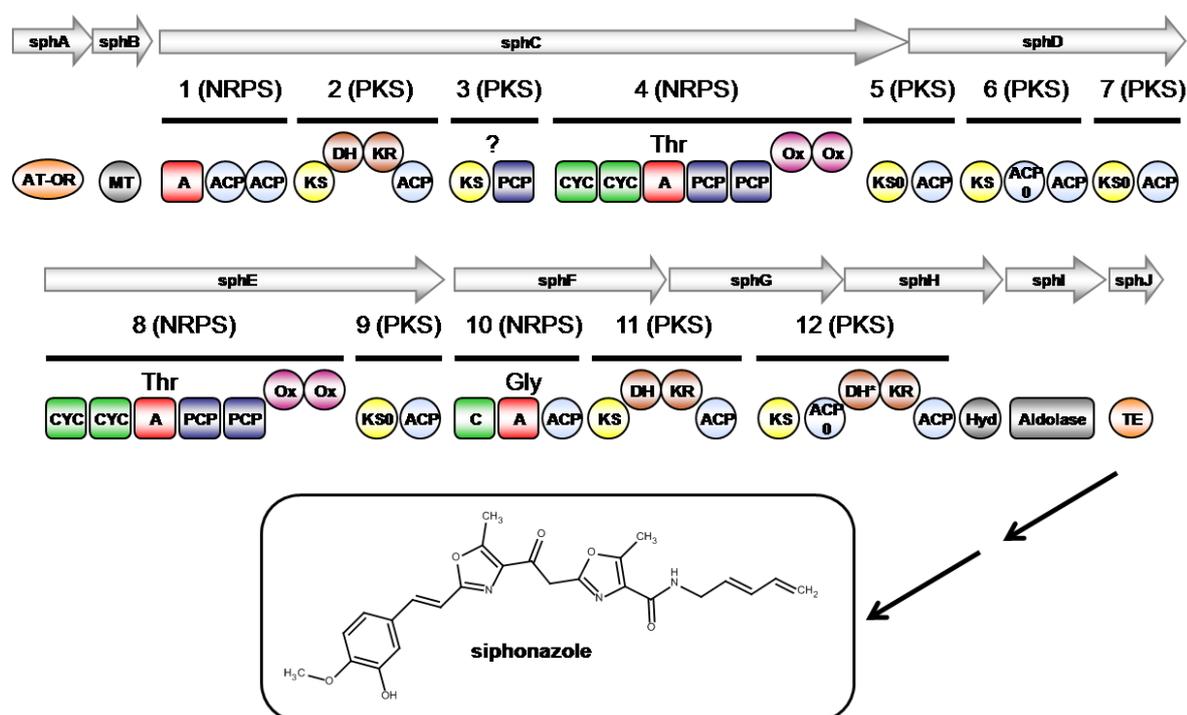


Figure 6-1: Organisation of the putative siphonazole PKS/NRPS system derived by bioinformatic analyses; modules are marked by black bars; A, adenylation domain; ACP, acyl carrier protein; Cyc, cyclisation domain; C, condensation domain; DH, dehydratase; Hyd, hydrolase; KR, ketoreductase; KS, ketosynthase; KS0, non-elongating KS; MT, methyltransferase; OR, oxidoreductase; Ox, oxidation domain; PCP, peptidyl carrier protein; TE, thioesterase; domains marked with '0' are non-functional; functionality of DH* in module 12 is uncertain; '?' marks module with uncertain function

In vitro studies with purified protein showed that the potential precursor molecule protocatechuic acid is readily methylated by the O-methyltransferase SphB. The DAHP synthase SphI, encoded upstream of the thioesterase, is likewise assumed to be involved in the formation of this precursor. This enzyme was successfully expressed and is available for further investigations.

The analysis of total mRNA from *Herpetosiphon* by RT-PCR revealed that sphA – sphJ are only expressed in the first three days of cultivation. The genes are organised in two transcriptional units, one of them containing sphA and sphB and the other one sphC – sphJ. In agreement with this early gene expression, the application of imaging mass spectrometry (IMS) showed that siphonazole is located exclusively at the edge of a growing colony. Hence, it was hypothesised that siphonazole might play a role in intercellular communication.

On the basis of the obtained sequence information, ensuing working steps aimed at the assembly of a vector carrying the complete gene cluster. By λ -Red recombination methods, cluster fragments of three overlapping fosmids from the genomic library were successfully recombined. Thus, a fosmid construct of 71 kb was generated harbouring genes sphA – sphJ. This vector was readily transformed into *E. coli* BAP1 cells, yet heterologous expression of the cluster was not achieved.

In conclusion, this study led to the discovery and elucidation of the putative PKS/NRPS system for the biosynthesis of siphonazole, revealing some unusual structural features. First steps towards a more detailed characterisation of siphonazole biosynthesis were taken by analysis of the spatiotemporal expression profile and purification of several of the biosynthetic enzymes. The reconstitution of the complete cluster on a fosmid laid the groundwork for a subsequent expression in a heterologous host. Thus, the present study provides a basis for further analyses directed at a proof of function of the gene cluster and the unravelling of some still unsolved mechanisms in the biosynthesis of siphonazole. The presence of further gene clusters and likewise the inherent metabolic potential of *Herpetosiphon* were shown by bioinformatic tools. The works presented in this study should therefore encourage a further investigation of the siphonazole pathway and the consideration of *Herpetosiphonales* and *Chloroflexi* as potential producers of new chemical entities.

7 References

- Baerga-Ortiz, A. *et al.* Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase. *Chemistry & Biology* **13**, 277–285 (2006).
- Baril, E., Roy, D. J. & Williams, J. R. Center for Bioethics brings different viewpoints together. *Dimens Health Serv* **63**, 30–31, 48 (1986).
- Begley, M., Cotter, P. D., Hill, C. & Ross, R. P. Identification of a Novel Two-Peptide Lantibiotic, Lichenicidin, following Rational Genome Mining for LanM Proteins. *Applied and Environmental Microbiology* **75**, 5451–5460 (2009).
- Bentley, R. The shikimate pathway--a metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384 (1990).
- Bergendahl, V., Linne, U. & Marahiel, M. A. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry* **269**, 620–629 (2002).
- Berman, F., Gerwick, W. & Murray, T. Antillatoxin and kalkitoxin, ichthyotoxins from the tropical cyanobacterium *Lyngbya majuscula*, induce distinct temporal patterns of NMDA receptor-mediated neurotoxicity. *Toxicon* **37**, 1645–1648 (1999).
- Bibb, M. J., Ward, J. M. & Hopwood, D. A. Transformation of plasmid DNA into *Streptomyces* at high frequency. *Nature* **274**, 398–400 (1978).
- Binz, T. M., Wenzel, S. C., Schnell, H.-J., Bechthold, A. & Müller, R. Heterologous Expression And Genetic Engineering of the Phenalinolactone Biosynthetic Gene Cluster by Using Red/ET Recombineering. *ChemBioChem* **9**, 447–454 (2008).
- Birck, M. R. & Woodard, R. W. *Aquifex aeolicus* 3-deoxy-D-manno-2-octulosonic acid 8-phosphate synthase: a new class of KDO 8-P synthase? *J. Mol. Evol.* **52**, 205–214 (2001).
- Bisang, C. *et al.* A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**, 502–505 (1999).
- Björnsson, L., Hugenholtz, P., Tyson, G. W. & Blackall, L. L. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology (Reading, Engl.)* **148**, 2309–2318 (2002).
- Bode, H. B. & Müller, R. Possibility of bacterial recruitment of plant genes associated with the biosynthesis of secondary metabolites. *Plant Physiol.* **132**, 1153–1161 (2003).
- Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chemistry & biology* **10**, 723–731 (2003).
- Buchan, A., Neidle, E. L. & Moran, M. A. Diverse Organization of Genes of the β -Ketoadipate Pathway in Members of the Marine *Roseobacter* Lineage. *Applied and Environmental Microbiology* **70**, 1658–1668 (2004).

- Buckley, N. D., Vadeboncoeur, C., LeBlanc, D. J., Lee, L. N. & Frenette, M. An Effective Strategy, Applicable to *Streptococcus salivarius* and Related Bacteria, To Enhance or Confer Electroporation Competence. *Applied and Environmental Microbiology* **65**, 3800–3804 (1999).
- Bumpus, S. B., Magarvey, N. A., Kelleher, N. L., Walsh, C. T. & Calderone, C. T. Polyunsaturated fatty-acid-like trans-enoyl reductases utilized in polyketide biosynthesis. *Journal of the American Chemical Society* **130**, 11614–11616 (2008).
- Butcher, R. A. *et al.* The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* **104**, 1506–1509 (2007).
- Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Natural Product Reports* **25**, 845 (2008).
- Cane, D. E. & Walsh, C. T. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *CHEMISTRY AND BIOLOGY-LONDON-* **6**, 319–325 (1999).
- Carbone, A., Zinovyev, A. & Kepes, F. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* **19**, 2005 (2003).
- Carvalho, R., Reid, R., Viswanathan, N., Gramajo, H. & Julien, B. The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. *Gene* **359**, 91–98 (2005).
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: An integrated platform for visualisation and analysis of high-throughput sequence-based experimental data. *Bioinformatics (Oxford, England)* (2011).
- Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & biology* **7**, 211–224 (2000).
- Challis, G. L. Genome Mining for Novel Natural Product Discovery. *Journal of Medicinal Chemistry* **51**, 2618–2628 (2008).
- Cheng, Y. Q., Tang, G. L. & Shen, B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proceedings of the National Academy of Sciences* **100**, 3149 (2003).
- Cheng, Y.-Q., Tang, G.-L. & Shen, B. Identification and Localization of the Gene Cluster Encoding Biosynthesis of the Antitumor Macrolactam Leinamycin in *Streptomyces atroolivaceus* S-140. *Journal of Bacteriology* **184**, 7013–7024 (2002).
- Chiang, Y.-M., Chang, S.-L., Oakley, B. R. & Wang, C. C. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Current Opinion in Chemical Biology* **15**, 137–143 (2011).

- Choi, S. S., Hur, Y. A., Sherman, D. H. & Kim, E. S. Isolation of the biosynthetic gene cluster for tautomycin, a linear polyketide T cell-specific immunomodulator from *Streptomyces* sp. CK4412. *Microbiology* **153**, 1095 (2007).
- Collie, N. & Myers, W. S. VII. - The formation of orcinol and other condensation products from dehydracetic acid. *Journal of the Chemical Society, Transactions* **63**, 122 (1893).
- Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO journal* **16**, 4174–4183 (1997).
- Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178 (1990).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640 (2000).
- Donadio, S. & Katz, L. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111**, 51–60 (1992).
- Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chemistry & biology* **7**, 623–642 (2000).
- Düsterhöft, A., Erdmann, D. & Kröger, M. Isolation and genetic structure of the Avall isoschizomeric restriction-modification system HgiBI from *Herpetosiphon giganteus* Hpg5: M. HgiBI reveals high homology to M. BanI. *Nucleic acids research* **19**, 3207 (1991).
- El-Sayed, A. K. *et al.* Characterization of the Mupirocin Biosynthesis Gene Cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chemistry & Biology* **10**, 419–430 (2003).
- Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the Selectivity-Confering Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics. *Biochemistry* **41**, 9718–9726 (2002).
- Erol, Ö. *et al.* Biosynthesis of the Myxobacterial Antibiotic Coralopyronin A. *ChemBioChem* **11**, 1253–1265 (2010).
- Esquenazi, E., Jones, A. C., Byrum, T., Dorrestein, P. C. & Gerwick, W. H. Temporal dynamics of natural product biosynthesis in marine cyanobacteria. *Proceedings of the National Academy of Sciences* **108**, 5226 (2011).
- Esquenazi, E. *et al.* Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. *Molecular BioSystems* **4**, 562 (2008).

- Eustáquio, A. S. *et al.* Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. *Applied and environmental microbiology* **71**, 2452 (2005).
- Felnagle, E. A. *et al.* MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. *Biochemistry* **49**, 8815–8817 (2010).
- Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N. & Timmis, K. N. Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nature Biotechnology* **21**, 1266–1267 (2003).
- Finking, R. & Marahiel, M. A. BIOSYNTHESIS OF NONRIBOSOMAL PEPTIDES
1. *Annual Review of Microbiology* **58**, 453–488 (2004).
- Fischbach, M. A. & Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chemical Reviews* **106**, 3468–3496 (2006).
- Fudou, R. *et al.* Haliangicin, a novel antifungal metabolite produced by a marine myxobacterium. 2. Isolation and structural elucidation. *J. Antibiot.* **54**, 153–156 (2001).
- Funa, N. *et al.* A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899 (1999).
- Galm, U. *et al.* Comparative Analysis of the Biosynthetic Gene Clusters and Pathways for Three Structurally Related Antitumor Antibiotics: Bleomycin, Tallysomycin, and Zorbamycin. *Journal of Natural Products* **74**, 526–536 (2011).
- Garrity, G. M. & Holt, J. G. Phylum BVI. Chloroflexi ph. nov. *Vol. 1: The Archaea and the Deeply Branching and Phototrophic Bacteria* (2001).
- Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Handbook* 571–707 (2005).
- Gehret, J. J. *et al.* Terminal Alkene Formation by the Thioesterase of Curacin A Biosynthesis: STRUCTURE OF A DECARBOXYLATING THIOESTERASE. *Journal of Biological Chemistry* **286**, 14445–14454 (2011).
- Gibson, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science* **319**, 1215–1220 (2008).
- Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic acids research* **37**, 6984–6990 (2009).
- Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**, 31–34 (2007).
- Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**, 35–37 (2007).
- Gokhale, R. S. & Khosla, C. Role of linkers in communication between protein modules. *Current opinion in chemical biology* **4**, 22–27 (2000).

- Gonzalez-Lergier, J., Broadbelt, L. J. & Hatzimanikatis, V. Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *Journal of the American Chemical Society* **127**, 9930–9938 (2005).
- Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695–699 (2010).
- Gross, H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Applied Microbiology and Biotechnology* **75**, 267–277 (2007).
- Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research* **33**, W526 (2005).
- Grunberg-Manago, M. MESSENGER RNA STABILITY AND ITS ROLE IN CONTROL OF GENE EXPRESSION IN BACTERIA AND PHAGES. *Annual Review of Genetics* **33**, 193–227 (1999).
- Gu, L. *et al.* Polyketide decarboxylative chain termination preceded by O-sulfonation in curacin A biosynthesis. *Journal of the American Chemical Society* **131**, 16033–16035 (2009).
- Gurney, R. & Thomas, C. M. Mupirocin: biosynthesis, special features and applications of an antibiotic from a Gram-negative bacterium. *Applied Microbiology and Biotechnology* **90**, 11–21 (2011).
- Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proceedings of the National Academy of Sciences* **100**, 1541 (2003).
- Gust, B. *et al.* λ Red-Mediated Genetic Manipulation of Antibiotic-Producing *Streptomyces*. *Advances in Applied Microbiology* **Volume 54**, 107–128 (2004).
- Ha, J. Y. *et al.* Crystal Structure of 2-Nitropropane Dioxygenase Complexed with FMN and Substrate: IDENTIFICATION OF THE CATALYTIC BASE. *Journal of Biological Chemistry* **281**, 18660–18667 (2006).
- Hahn, M. & Stachelhaus, T. Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 275 (2006).
- Hammes, W., Schleifer, K. H. & Kandler, O. Mode of Action of Glycine on the Biosynthesis of Peptidoglycan. *Journal of Bacteriology* **116**, 1029–1053 (1973).
- Hanada, S., Takaichi, S., Matsuura, K. & Nakamura, K. *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *International journal of systematic and evolutionary microbiology* **52**, 187–193 (2002).
- He, J., Magarvey, N., Pirae, M. & Vining, L. The gene cluster for chloramphenicol biosynthesis in *Streptomyces venezuelae* ISP5230 includes novel shikimate pathway homologues and a monomodular non-ribosomal peptide synthetase gene. *Microbiology* **147**, 2817 (2001).

- Hefter, J., Richnow, H. H., Fischer, U., Trendel, J. M. & Michaelis, W. (-)-Verrucosan-2 β -ol from the phototrophic bacterium *Chloroflexus aurantiacus*: first report of a verrucosane-type diterpenoid from a prokaryote. *Journal of General Microbiology* **139**, 2757 (1993).
- Heller, M., Edelstein, P. & Mayer, M. Membrane-bound enzymes. III. Protease activity in leucocytes in relation to erythrocyte membranes. *Biochim. Biophys. Acta* **413**, 472–482 (1975).
- Hertweck, C. Die biosynthetische Grundlage der Polyketid- und Vielfalt. *Angewandte Chemie* **121**, 4782–4811 (2009).
- Holt, J. & Lewin, R. *Herpetosiphon aurantiacus* gen. et sp. n., a new filamentous gliding organism. *Journal of Bacteriology* **95**, 2407 (1968).
- Hopwood, D. A. & Wright, H. M. Bacterial protoplast fusion: recombination in fused protoplasts of *Streptomyces coelicolor*. *Mol. Gen. Genet.* **162**, 307–317 (1978).
- Jansen, R. *et al.* Antibiotics from gliding bacteria, XLIII. Phenoxan: A novel inhibitor of HIV-1 infection in cell cultures from *Polyangium* sp., strain PI VO19 (Myxobacteria). *Liebigs Annalen der Chemie* **1991**, 707–708 (1991).
- Jansen, R., Washausen, P., Kunze, B., Reichenbach, H. & Höfle, G. The Crocacins, Novel Antifungal and Cytotoxic Antibiotics from *Chondromyces crocatus* and *Chondromyces pediculatus* (Myxobacteria): Isolation and Structure Elucidation. *European Journal of Organic Chemistry* **1999**, 1085–1089 (1999).
- Jenke-Kodama, H., Börner, T. & Dittmann, E. Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium *Streptomyces avermitilis*. *PLoS Computational Biology* **2**, e132 (2006).
- Jensen, R. A., Xie, G., Calhoun, D. H. & Bonner, C. A. The correct phylogenetic relationship of KdsA (3-deoxy-d-manno-octulosonate 8-phosphate synthase) with one of two independently evolved classes of AroA (3-deoxy-d-arabino-heptulosonate 7-phosphate synthase). *J. Mol. Evol.* **54**, 416–423 (2002).
- Julien, B. & Shah, S. Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*. *Antimicrobial agents and chemotherapy* **46**, 2772–2778 (2002).
- Jürgens, U., Meissner, J., Reichenbach, H. & Weckesser, J. L-ornithine containing peptidoglycan-polysaccharide complex from the cell wall of the gliding bacterium *Herpetosiphon aurantiacus*. *FEMS microbiology letters* **60**, 247–250 (1989).
- Kagan, R. M. & Clarke, S. Widespread Occurrence of Three Sequence Motifs in Diverse S-Adenosylmethionine-Dependent Methyltransferases Suggests a Common Structure for These Enzymes. *Archives of Biochemistry and Biophysics* **310**, 417–427 (1994).
- Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **6**, 494–500 (1995).
- Keating, T. A. *et al.* Chain Termination Steps in Nonribosomal Peptide Synthetase Assembly Lines: Directed Acyl-S-Enzyme Breakdown in Antibiotic and Siderophore Biosynthesis. *ChemBioChem* **2**, 99–107 (2001).

- Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology* (2002).
- Keating, T. A., Miller, D. A. & Walsh, C. T. Expression, Purification, and Characterization of HMWP2, a 229 kDa, Six Domain Protein Subunit of Yersiniabactin Synthetase. *Biochemistry* **39**, 4729–4739 (2000).
- Keppen, O. I., Tourova, T. P., Kuznetsov, B. B., Ivanovsky, R. N. & Gorlenko, V. M. Proposal of Oscillochloridaceae fam. nov. on the basis of a phylogenetic analysis of the filamentous anoxygenic phototrophic bacteria, and emended description of Oscillochloris and Oscillochloris trichoides in comparison with further new isolates. *International journal of systematic and evolutionary microbiology* **50**, 1529–1537 (2000).
- Kieser, T., Binn, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical streptomyces genetics*. (John Innes Foundation: Norwich, 2000).
- Kleinig, H. & Reichenbach, H. Carotenoid glucosides and menaquinones from the gliding bacterium Herpetosiphon giganteus Hp a2. *Arch. Microbiol.* **112**, 307–310 (1977).
- Kleinkauf, H. & Von Dohren, H. Biosynthesis of peptide antibiotics. *Annual Reviews in Microbiology* **41**, 259–289 (1987).
- Kobori, T. Structure and Site-directed Mutagenesis of a Flavoprotein from Escherichia coli That Reduces Nitrocompounds. ALTERATION OF PYRIDINE NUCLEOTIDE BINDING BY A SINGLE AMINO ACID SUBSTITUTION. *Journal of Biological Chemistry* **276**, 2816–2823 (2000).
- Koglin, A. & Walsh, C. T. Structural insights into nonribosomal peptide enzymatic assembly lines. *Natural Product Reports* **26**, 987 (2009).
- Kourist, R. *et al.* The α/β -Hydrolase Fold 3DM Database (ABHDB) as a Tool for Protein Engineering. *ChemBioChem* **11**, 1635–1643 (2010).
- Krieg, N. R. & Garrity, G. M. On Using the Manual. *Vol. 1: The Archaea and the Deeply Branching and Phototrophic Bacteria* (2001).
- Lai, J. R., Koglin, A. & Walsh, C. T. Carrier Protein Structure and Recognition in Polyketide and Nonribosomal Peptide Biosynthesis. *Biochemistry* **45**, 14869–14879 (2006).
- Lambalot, R. H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
- Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and exchange. *Journal of molecular evolution* **44**, 383–397 (1997).
- Lee, N. & Reichenbach, H. The Genus Herpetosiphon. *The Prokaryotes* 854–877 (2006)

- Lerner, R. A. Antibodies of Predetermined Specificity in Biology and Medicine. *Advances in immunology* **36**, 1 (1984).
- Li, C., Hassler, M. & Bugg, T. D. H. Catalytic Promiscuity in the α/β -Hydrolase Superfamily: Hydroxamic Acid Formation, C-C Bond Formation, Ester and Thioester Hydrolysis in the C-C Hydrolase Family. *ChemBioChem* **9**, 71–76 (2008).
- Li, S., Anzai, Y., Kinoshita, K., Kato, F. & Sherman, D. H. Functional Analysis of MycE and MycF, Two O-Methyltransferases Involved in the Biosynthesis of Mycinamicin Macrolide Antibiotics. *ChemBioChem* **10**, 1297–1301 (2009).
- Linder, J., Blake, A. J. & Moody, C. J. Total synthesis of siphonazole and its O-methyl derivative, structurally unusual bis-oxazole natural products. *Organic & Biomolecular Chemistry* **6**, 3908 (2008).
- Liou, G. F., Lau, J., Cane, D. E. & Khosla, C. Quantitative Analysis of Loading and Extender Acyltransferases of Modular Polyketide Synthases. *Biochemistry* **42**, 200–207 (2003).
- Lopez, J. V. Naturally mosaic operons for secondary metabolite biosynthesis: variability and putative horizontal transfer of discrete catalytic domains of the epothilone polyketide synthase locus. *Molecular Genetics and Genomics* **270**, 420–431 (2004).
- Lopez, P. J., Marchand, I., Joyce, S. A. & Dreyfus, M. The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Molecular Microbiology* **33**, 188–199 (1999).
- Luo, Y. *et al.* Functional Characterization of TtnD and TtnF, Unveiling New Insights into Tautomycin Biosynthesis. *Journal of the American Chemical Society* **132**, 6663–6671 (2010).
- Ma, H., Kunes, S., Schatz, P. J. & Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201–216 (1987).
- MacNeil, D. J. *et al.* Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68 (1992).
- Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309**, 462–464 (1984).
- Marahiel, M. A. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *Journal of Peptide Science* **15**, 799–807 (2009).
- Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical Reviews* **97**, 2651–2674 (1997).
- Marchler-Bauer, A. *et al.* CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* **37**, D205–210 (2009).
- Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327–331 (2004).

- Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–229 (2011).
- Marshall, C. G., Hillson, N. J. & Walsh, C. T. Catalytic Mapping of the Vibriobactin Biosynthetic Enzyme VibF. *Biochemistry* **41**, 244–250 (2002).
- McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel ‘unnatural’ natural products. *Proceedings of the National Academy of Sciences* **96**, 1846 (1999).
- McDonald, M., Mavrodi, D. V., Thomashow, L. S. & Floss, H. G. Phenazine Biosynthesis in *Pseudomonas fluorescens*: Branchpoint from the Primary Shikimate Biosynthetic Pathway and Role of Phenazine-1,6-dicarboxylic Acid. *Journal of the American Chemical Society* **123**, 9459–9460 (2001).
- Meier, J. L. & Burkart, M. D. The chemical biology of modular biosynthetic enzymes. *Chemical Society Reviews* **38**, 2012 (2009).
- Menzella, H. G. *et al.* Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster. *Journal of Industrial Microbiology & Biotechnology* **33**, 22–28 (2005).
- Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31–46 (2009).
- Molnár, I. *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chemistry & Biology* **7**, 97–109 (2000).
- Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Natural Product Reports* **19**, 70–99 (2002).
- Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *Journal of bacteriology* **179**, 6843 (1997).
- Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **3**, 490–504 (2002).
- Nett, M. *Structure elucidation and biosynthetic studies of secondary metabolites from gliding heterotrophic bacteria.* (Verl. Dr. Hut: München, 2007).
- Nett, M. *et al.* Siphonazol, ein ungewöhnlicher Naturstoff aus *Herpetosiphon* sp. *Angewandte Chemie* **118**, 3947–3951 (2006).
- Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years. *Journal of Natural Products* **70**, 461–477 (2007).
- Nguyen, T. *et al.* Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nature Biotechnology* **26**, 225–233 (2008).

- Niu, D. & Wang, Z.-X. Development of a pair of bifunctional expression vectors for *Escherichia coli* and *Bacillus licheniformis*. *Journal of Industrial Microbiology & Biotechnology* **34**, 357–362 (2007).
- O'Connor, S. E., Walsh, C. T. & Liu, F. Biosynthesis of Epothilone Intermediates with Alternate Starter Units: Engineering Polyketide–Nonribosomal Interfaces. *Angewandte Chemie International Edition* **42**, 3917–3921 (2003).
- O'Hare, H. M., Baerga-Ortiz, A., Popovic, B., Spencer, J. B. & Leadlay, P. F. High-Throughput Mutagenesis to Evaluate Models of Stereochemical Control in Ketoreductase Domains from the Erythromycin Polyketide Synthase. *Chemistry & Biology* **13**, 287–296 (2006).
- Okanishi, M., Suzuki, K. & Umezawa, H. Formation and reversion of Streptomycete protoplasts: cultural condition and morphological study. *J. Gen. Microbiol.* **80**, 389–400 (1974).
- Ollis, D. L. *et al.* The α / β hydrolase fold. *Protein Engineering Design and Selection* **5**, 197–211 (1992).
- Oyaizu, H., Debrunner-Vossbrinck, B., Mandelco, L., Studier, J. A. & Woese, C. R. The green non-sulfur bacteria: a deep branching in the eubacterial line of descent. *Syst. Appl. Microbiol.* **9**, 47–53 (1987).
- Patel, H. M. & Walsh, C. T. In Vitro Reconstitution of the *Pseudomonas aeruginosa* Nonribosomal Peptide Synthesis of Pyochelin: Characterization of Backbone Tailoring Thiazoline Reductase and N-Methyltransferase Activities. *Biochemistry* **40**, 9023–9031 (2001).
- Pawlik, K., Kotowska, M., Chater, K. F., Kuczek, K. & Takano, E. A cryptic type I polyketide synthase (cpk) gene cluster in *Streptomyces coelicolor* A3(2). *Archives of Microbiology* **187**, 87–99 (2006).
- Perlova, O., Gerth, K., Kaiser, O., Hans, A. & Müller, R. Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *Journal of Biotechnology* **121**, 174–191 (2006).
- Pfeifer, B. A. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
- Pfeifer, B. A. & Khosla, C. Biosynthesis of polyketides in heterologous hosts. *Microbiology and molecular biology reviews* **65**, 106–118 (2001).
- Pfeifer, B. A., Wang, C. C. C., Walsh, C. T. & Khosla, C. Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host. *Applied and environmental microbiology* **69**, 6698–6702 (2003).
- Phelan, V. V., Du, Y., McLean, J. A. & Bachmann, B. O. Adenylation Enzyme Characterization Using γ -¹⁸O₄-ATP Pyrophosphate Exchange. *Chemistry & Biology* **16**, 473–478 (2009).

- Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proceedings of the National Academy of Sciences* **99**, 14002 (2002).
- Piel, J., Hui, D., Fusetani, N. & Matsunaga, S. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environmental microbiology* **6**, 921–927 (2004).
- Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports* **27**, 996 (2010).
- Piel, J. *et al.* Exploring the Chemistry of Uncultivated Bacterial Symbionts: Antitumor Polyketides of the Pederin Family. *Journal of Natural Products* **68**, 472–479 (2005).
- Pierson, B. K. & Castenholz, R. W. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Archives of microbiology* **100**, 5–24 (1974).
- Quadri, L. E. N. *et al.* Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry* **37**, 1585–1595 (1998).
- Quinn, G. R. & Skerman, V. B. D. Herpetosiphon—Nature’s scavenger? *Current Microbiology* **4**, 57–62 (1980).
- Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology* **7**, 78 (2007).
- Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Research* **33**, 5799–5808 (2005).
- Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part A-erythromycin biosynthesis). *Natural product reports* **18**, 190–227 (2001).
- Reichenbach, H. & Golecki, J. R. The fine structure of Herpetosiphon, and a note on the taxonomy of the genus. *Arch. Microbiol.* **102**, 281–291 (1975).
- Reichenbach, H., Beyer, P. & Kleinig, H. The pigments of the gliding bacterium Herpetosiphon giganteus. *FEMS Microbiology Letters* **3**, 155–156 (1978).
- Reid, R. *et al.* A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases. *Biochemistry* **42**, 72–79 (2003).
- Rey, M. W. *et al.* Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome biology* **5**, r77 (2004).
- Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nature Chemical Biology* **4**, 75–81 (2008).

- Rosano, G. L. & Ceccarelli, E. A. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microbial cell factories* **8**, 41 (2009).
- Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research* (2011).
- Roy, R. S. *et al.* Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Natural Product Reports* **16**, 249–263 (1999).
- Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
- Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**, 5463–5467 (1977).
- Schneider, A. & Marahiel, M. A. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Archives of Microbiology* **169**, 404–410 (1998).
- Schumann, W. & Ferreira, L. C. S. Production of recombinant proteins in *Escherichia coli*. *Genetics and Molecular Biology* **27**, 442 (2004).
- Schwarzer, D., Finking, R. & Marahiel, M. A. Nonribosomal peptides: from genes to products. *Natural Product Reports* **20**, 275 (2003).
- Scotti, C. *et al.* A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* **130**, 65–71 (1993).
- Seshime, Y., Juvvadi, P. R., Fujii, I. & Kitamoto, K. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochemical and biophysical research communications* **331**, 253–260 (2005).
- Shank, E. A. & Kolter, R. New developments in microbial interspecies signaling. *Current Opinion in Microbiology* **12**, 205–214 (2009).
- Sharp, P. M. & Li, W. H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research* **15**, 1281 (1987).
- Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology* **7**, 285–295 (2003).
- Sieber, S. A. & Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chemical Reviews* **105**, 715–738 (2005).
- Silakowski, B. *et al.* New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274**, 37391–37399 (1999).
- Silva, F., Queiroz, J. A. & Domingues, F. C. Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnology Advances* (2012).

- Skerman, V. B. D., Quinn, G. R., Sly, L. I. & Hardy, J. V. Sheath Formation by Strains of Herpetosiphon Species. *International Journal of Systematic Bacteriology* **27**, 274–278 (1977).
- Sly, L. I., Taghavi, M. & Fegan, M. Phylogenetic heterogeneity within the genus Herpetosiphon: transfer of the marine species Herpetosiphon cohaerens, Herpetosiphon nigricans and Herpetosiphon persicus to the genus Lewinella gen. nov. in the Flexibacter-Bacteroides-Cytophaga phylum. *Int. J. Syst. Bacteriol.* **48 Pt 3**, 731–737 (1998).
- Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. Mobility of Plasmids. *Microbiology and Molecular Biology Reviews* **74**, 434–452 (2010).
- Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports* **24**, 1041–1072 (2007).
- Song, Y. *et al.* Ultrasound-mediated DNA transfer for bacteria. *Nucleic Acids Research* **35**, e129–e129 (2007).
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & biology* **6**, 493–505 (1999).
- Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *Journal of Biological Chemistry* **273**, 22773–22781 (1998).
- Stackebrandt, E., Rainey, F. A. & Ward-Rainey, N. Anoxygenic phototrophy across the phylogenetic spectrum: current understanding and future perspectives. *Archives of microbiology* **166**, 211–223 (1996).
- Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **18**, 380–416 (2001).
- Stegmann, E., Rausch, C., Stockert, S., Burkert, D. & Wohlleben, W. The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. *FEMS Microbiology Letters* **262**, 85–92 (2006).
- Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Current Opinion in Structural Biology* **20**, 234–240 (2010).
- Subramaniam, P. S., Xie, G., Xia, T. & Jensen, R. A. Substrate Ambiguity of 3-Deoxy-d-manno-Octulosonate 8-Phosphate Synthase from *Neisseria gonorrhoeae* in the Context of Its Membership in a Protein Family Containing a Subset of 3-Deoxy-d-arabino-Heptulosonate 7-Phosphate Synthases. *Journal of Bacteriology* **180**, 119–127 (1998).
- Sudek, S. *et al.* Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from ‘Candidatus Endobugula sertula’, the Uncultivated Microbial Symbiont of the Marine Bryozoan *Bugula neritina*. *Journal of Natural Products* **70**, 67–74 (2007).

- Suo, Z., Tseng, C. C. & Walsh, C. T. Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase. *Proceedings of the National Academy of Sciences* **98**, 99 (2001).
- Sutcliffe, I. C. A phylum level perspective on bacterial cell envelope architecture. *Trends in Microbiology* **18**, 464–470 (2010).
- Tang, G. L., Cheng, Y. Q. & Shen, B. Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chemistry & biology* **11**, 33–45 (2004).
- Tang, K. H. *et al.* Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC genomics* **12**, 334 (2011).
- Thompson, C. J., Ward, J. M. & Hopwood, D. A. Cloning of antibiotic resistance and nutritional genes in streptomycetes. *J. Bacteriol.* **151**, 668–677 (1982).
- Thykaer, J. *et al.* Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster. *Metabolic Engineering* **12**, 455–461 (2010).
- Trick, I. & Lingens, F. Characterization of *Herpetosiphon spec.* - A gliding filamentous bacterium from bulking sludge. *Applied Microbiology and Biotechnology* **19**, 191–198 (1984).
- Tuan, J. S. *et al.* Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene* **90**, 21–29 (1990).
- Van Lanen, S. G., Lin, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism. *Proceedings of the National Academy of Sciences* **105**, 494–499 (2008).
- Vanderwall, D. E. *et al.* A model of the structure of HOO-Co· bleomycin bound to d (CCAGTACTGG): recognition at the d (GpT) site and implications for double-stranded DNA cleavage. *Chemistry & biology* **4**, 373–387 (1997).
- Vincze, T., Posfai, J. & Roberts, R. J. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res* **31**, 3688–3691 (2003).
- von Döhren, H., Dieckmann, R. & Pavela-Vrancic, M. The nonribosomal code. *Chem. Biol.* **6**, R273–279 (1999).
- Walker, G. E., Dunbar, B., Hunter, I. S., Nimmo, H. G. & Coggins, J. R. Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. *Microbiology (Reading, Engl.)* **142** (Pt 8), 1973–1982 (1996).
- Walsh, C. T. The Chemical Versatility of Natural-Product Assembly Lines. *Accounts of Chemical Research* **41**, 4–10 (2008).

- Walsh, C. T. & Fischbach, M. A. Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society* **132**, 2469–2493 (2010).
- Waschkau, B., Waldeck, J., Wieland, S., Eichstädt, R. & Meinhardt, F. Generation of readily transformable *Bacillus licheniformis* mutants. *Applied Microbiology and Biotechnology* **78**, 181–188 (2007).
- Weber, T. *et al.* CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology* **140**, 13–17 (2009).
- Weinig, S., Hecht, H. J., Mahmud, T. & Muller, R. Melithiazol Biosynthesis:: Further Insights into Myxobacterial PKS/NRPS Systems and Evidence for a New Subclass of Methyl Transferases. *Chemistry & biology* **10**, 939–952 (2003).
- Weissman, K. J. *et al.* The Thioesterase of the Erythromycin-Producing Polyketide Synthase: Influence of Acyl Chain Structure on the Mode of Release of Substrate Analogues from the Acyl Enzyme Intermediates. *Angewandte Chemie International Edition* **37**, 1437–1440 (1998).
- Wenzel, S. C. *et al.* Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering. *Chemistry & Biology* **12**, 349–356 (2005).
- Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
- Woese, C. R., Stackebrandt, E. & Ludwig, W. What are mycoplasmas: The relationship of tempo and mode in bacterial evolution. *Journal of Molecular Evolution* **21**, 305–316 (1985).
- Wu, J. Thermotoga maritima 3-Deoxy-D-arabino-heptulosonate 7-Phosphate (DAHP) Synthase: THE ANCESTRAL EUBACTERIAL DAHP SYNTHASE? *Journal of Biological Chemistry* **278**, 27525–27531 (2003).
- Wu, J. & Woodard, R. W. New Insights into the Evolutionary Links Relating to the 3-Deoxy-D-arabino-heptulosonate 7-Phosphate Synthase Subfamilies. *Journal of Biological Chemistry* **281**, 4042–4048 (2006).
- Wu, X., Jörnvall, H., Berndt, K. D. & Oppermann, U. Codon optimization reveals critical factors for high level expression of two rare codon genes in Escherichia coli: RNA stability and secondary structure but not tRNA abundance. *Biochemical and biophysical research communications* **313**, 89–96 (2004).
- Xin, Y., Pan, J., Collins, A. M., Lin, S. & Blankenship, R. E. Excitation energy transfer and trapping dynamics in the core complex of the filamentous photosynthetic bacterium Roseiflexus castenholzii. *Photosynthesis Research* (2011).
- Yadav, G., Gokhale, R. S. & Mohanty, D. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *Journal of Molecular Biology* **328**, 335–363 (2003).

- Yoshida, N. & Sato, M. Plasmid uptake by bacteria: a comparison of methods and efficiencies. *Applied Microbiology and Biotechnology* **83**, 791–798 (2009).
- Zazza, C. & Sanna, N. Photoabsorption spectra of a natural polyphenol compound for therapeutic applications: the protocatechuic acid in dilute water solution at room temperature. *Physical Chemistry Chemical Physics* **12**, 3859 (2010).
- Zhang, J. & Ciufolini, M. A. Total Synthesis of Siphonazoles by the Use of a Conjunctive Oxazole Building Block. *Organic Letters* **11**, 2389–2392 (2009).
- Zhang, Y., Buchholz, F., Muyrers, J. P. P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature genetics* **20**, 123–128 (1998).
- Zhang, Y.-M., Hurlbert, J., White, W. & Rock, C. O. Roles of the Active Site Water, Histidine 303, and Phenylalanine 396 in the Catalytic Mechanism of the Elongation Condensing Enzyme of *Streptococcus pneumoniae*. *Journal of Biological Chemistry* **281**, 17390–17399 (2006).
- Zhou, D. & Yang, R. Global analysis of gene transcription regulation in prokaryotes. *Cell. Mol. Life Sci.* **63**, 2260–2290 (2006).
- Ziemert, N., Podell, S., Penn, K., Allen, E. E. & Jensen, P. NaPDoS, a new genome mining webtool for the discovery of natural products. *manuscript submitted* (2011).

8 Appendix

8.1 Primary sequences of the proteins SphA – SphJ

8.1.1 SphA

MLMRAFIFPGQGSQKQGMGEDLFEAFPELTDQASQILGYSIRELCLRDPQRQLAQT
QYTQPALYVICVLSYLAEVSRGIRPDFVAGHSVGEYSALFAADVDFDFETGLRLVQKR
GELMAKAQNGGMAAVGGLSEAQIRSILNQHQLYSLDIANLNAPTQVVLSPNDIIA
AQAVFEAAQAQLYIKLNVSGAFHSRYMKNRSRDEFGTFIASMQFAEPRIPVIANLDAN
PYTASTIKRNLVEQITHSVRWVESIEYLLHAGVNEFKEIGPGTVLTKLNQRITANQAA
TPTVSTTPTPQPLPNKPATPTPTPTQPTLNGSPTLRRAINPLSLGSEAFRRAYGVKYAY
AAGGMHHGSIAMIERLAQAKIMSFFGTGGLALSEIEQALQQQLQQLGTAPYGVNL
LWADMQNPEHEAQVFACLERYSIHHIEVAGYLDVTPALLRFRAKGLTQHADGSINST
RTIMAKITRPEIAEVFLSPPTTEQIQGLVAAKLLTSQQAQYVASLPIADDICVEADSGG
TTGLGSAYALLPTIIQLRDRAQGAVRIGGAGGIGTPAAAAAFMLGADFILTGSINQC
TVEAKTSNAKELLQDVNPQDTIHIPNIDLLSPGSKTQVLKKGVFFHVVRANRLYDISR
FYENLEQLDQTVRQQIEEKYFKRTLADIWRELCTTTASQIIERAERNPRQKLALTFKW
YFDQSIQWAIHGDLSRKVDFQIACGPALGAFNQWVKGTPLERWQARHVDTIADSLM
HGTADYLNQRMHILHHQ

8.1.2 SphB

MAKASLNLTDLQLYEYLLQHSREPELVRELREETARMLMYAMQTPPEETQFIAWLL
TLTNAKRMLEIGTFTGYTTLWAALTMPDDAKIVACDVADKWVNVGKKYWQQAGVA
HKIDLRIAPALQTLDQLLQNHFANYFDFVYIDADKENYWNYYERALMLVRPGGIIGID
NVLWGGSVVNELNQSADVKAIREFNQRLHQDQRIAAMSLHPIGDGLTLVIRK

8.1.3 SphC

MVQQDLQSQSQNDLYTLFTNQARQTPQAEALRWCIDEQWQSLSFEQLAKHIEQLA
AALAATGIGVGSVGVLAERSPAIVTSILALFKLGAIVVPLDLGYSAANIAQISAQTAM
ALIITRHAYVAKIANVPINYADIEELIVRENPTPASVMPDPARPALLLYTSGTTGQPKG
ALHCQRQLLNRFNWLWQNYPFAGDLMVQRTSINFMPSLWEWLGGLVCGVATLII
DDTTIKNPRRFLKTLAEQRATHLAAVPSLLGMFLGEGNIADLAGLRLICGGEPLLP
LRSQLRANLPTTTILNDYGATEMNGVLFNEYQPQQPATETGMRPIANVEVLILDHDL
NLITDGSSGQIYIGGPCLAYGYVNQANLTAERFIPHPAATLPGSRLYRTGDVGRQLA
DGSIVISGRQDHQVKIRGMRVELAGIEQTLAQHPEVRIAVVAHEAKSGQKSLAGYV
VTRGAITTAELHTYLTDLNLPNYMVAQLVILDEMPRLPNGKIDRMALTRRETVKSLTA
NSVIDQLCQIAATVLETDPPTLEPERKFYELGFDSVNVVSVFVRINQHLGCDIDVSAIY
NYATLADLASALATIVPAQVPAPQPIAIEKHLESTSAVAPIQSTAKPLIVPTQPIAASSVI
DQLCQIAAIVLETDPALLEPERKFYELGFDSVNVVSVFVRLINQHLGCDIDVSAIYNYAT
LADLALALATEHGVLQQTSAAPILKAPVAIPAAPVPTFNPSIATTTTPVESPPQPQI
MTVADTAQGTEPTEVAIIGLAGRFPGAASVAEFWELLAAGQSAIREVPADRWSNQT
YFDPNPQTPQHHTTSKWGGFLDQIDYFDHSFFNISKSEAAGMDPQQRLCLEEGWKA
LEDAGYVEDAVAGKTVGVFVGMVPGDYLERIRSQANELTAYNLLGSDSAILAARLAY
YLNKGPSIAIDTACSSSLVAIHLACQSIRAGESELALAGGVSILSTPSLYITSSKMG
LSATGQCKTFDNQADGFVPGAVGFVVLKALSQAQRDGDRIYGVIKGSGINQDGKT

NGITAPSGQAQTTLERAVYAQAKLSPRSISYIEAHGTGTRLGDPIEIRALTESFRHWT
EDRQFCAIGSVKTNMGHALAAAGIAGLAKVLLSLQQRQIPPSLNFVSPNEHIPFADS
PFFVATQLQPWLTPAGEPRRAAISSFGFSGTNSHLVLEEFAPRVDRPESQWHLIV
LSAKTETALRQQIVALSQALEQPKQTYRLDDIAYTLLVRRAAFVRCAMIVRDYSDLQ
HQLRLLQQPQTQFTKAPKAQPLHTTLTFQRSPDDLASLKECFLSNQVVDWQHLYP
QTEHNLLSLPHYCFERERFWIDTDQPLATNQAPNTAQLHPLIEANCSTLQGIRFNTV
LQPTDFYVADHRWEGQALLPGVVYAEIARVASSLALNQPVSHIEQLLWLRPLIVEQTI
NLQSRUYADEHDVLVEIGVADPADQLTIHAKARIPLAKQNVDSQYMAALDLPAIRAICV
EQIDQAEFYQHFSQLGLNYGSRLQAISISYNAQQALAELELPTELSADFARYSLHP
ALLDAALQTVLVLQAKQAGRDKADLPFELAAIDLIQPLERQCVAYVQRNAEQGRLKR
SQIVIANLQGEPLVVLHGFAQGAPKATTGNEVEALQRGEIAYYSATEYETALVQGSIA
QGTMLVFGSSQVQKLLNSSGANVQVTYGTYSYRQIDRYTYTIDWLNPNDYHQLFATL
EREGLVPSSILQLANLETKQSSSLEQTVEATIYPLIYLYQALGMLKFMPLNRILWAWQ
EFVHNSSSAAAQAALSGVLQTLANEYPALPCKLVSFVQQTDLQLIVASLLNELQTTD
FTFETINYVNQKRLSQGYQRQPASAEPTTTPFRSGGTYLITGGLGGVGLLCAHHLAK
HYRAKLALVGRSQPTPAIEAQLAELANAGAEVCYIRADIGATQQAHAIAQAQAQFG
QIHGIIHSAGVLHDQPLRNARRAEIEAVLHPKVYGTAAALDQATASENLDFLALFSSITT
VVGNGFGQTAYGYANGFLDQFAMQRERLRQAGQRFGRVTSINWPYWQAGGMRMS
AQAEQLMETLIGAMPLETSSAIEALYHALTAPASQIVVFAGDRSTFEAAMGTRLYQP
SIRSEAMSYPSPITESNPFQRQIQADLLRMAADILRARPEDIDPQIELTEFGLDSISM
MEFANHINKQFKLVPTLFLLEHLTLASVANYLAQQARPTDQAIPPMPEVVARTATK
VETPISKPIAPVIAQPLPAPTLSQSINTFVAQPVPQPKNQPIAVIGMSGFRFPQADDL
NEFWELLRSGTSAISEIPSDRWVWQYHGDSDLSSKTSVKWGGFLRDIAGFDAA
FFNISPTAQMLDPQQRLFLETVWQTIEDAGYRPSAFNGSRTAVFVGAFTKDYAELL
QAHGINAAHTTTGTDHSIIANRVSYFFDWRGPSEAITACSSAFVAIHRAVQALRNG
EADLAIAGGVNALLAPQGFRLSTQAGMLSQSGTIRSFQSDGYLREGGAVLLK
PLAQAERDGDIIHGIVIRGSAVNHNRSYSLTAPNPAAQTDVIIRAYANAEIDPASVSYI
EAHGTGTVIGDPAEIQAFQRAFEQLAEQSSLPNSNCGIGSIKPNIGHLEAASGIAAFIK
VLLALRHATIPATLNLETINPQIDLAETPFQLIRTAGPWQAQHGQRRASLHSFGFGGT
NAHIVLDEYPQVRASAAAAPVIVPCSAKHPKSLRAGLAALHQHVNNPQLRLDDLA
YTLQIGREAFDERVAFVVSNAELITLLESYLNNTITPDQAIYVGSQMHNDRTLKTLA
GKLGQSLQELHTPQELSQIALLWTKGVDVAWEQLDQTERRQRLALPTYRFNHQS
YWIPTNPDDHNPSNNGDPEASNPAKPEHAPTSVIGQLQQLFSQVLGISAEQLDPY
GNLEAYGLDSILATKARYLIEQRFGINVPMNIFKQQEHIAALAQQIPADASFAEDLETE
SVEALKVGTTFGLSAVQEAFFVVGQHLLPFDHLGCHFFLTFPLTTTDPQTVAAAWLTL
CQRHPMLTMQLVPPAAQTIRDPRDLPAPTIRDCTSADDATIAACAHEHTDRLSHAAY
PLDAWPRWSLQILALPHHACRVFLSVDLTDLDAASLVILLHEWHTLVLDPATVLPAS
TAFPALQRLQTRAAAAAADLAYWRTRLHGITGGPALPWIPLQPHHATTAPDGRR
FLPRTRLAACLPAWASALQALAAQHAVAPSTLLLAVFAHILRSATADPRAPFALIT
LANRVPLIPASEQLVAPVVSTTLVRLDPADSLATLLPSLQAQLWDTLDHPFTSGIAVL
RDLKHQQLLPATTTIPVVFTAMLNFGNHAQQPSFFPAVDTALTQTPNVWLDHQLL
ERNGALHFNWDILPAAFPRGLAEALFSRYTAALQTLACDPAALHAAALGLPDPTPLP
LSAVQRSYLLARPAAPARFYQEFRLPQHAIPRLAAAWLRVAAAHPGLRMTLHTDHL
VCASASPTPLDLPITPERGPAITATRTAMQHAPLQPIAVRLTHDTDATTVLHCAIDAL
YLDGRSLMTLFWAACAAPFAPFAPPATSPHALRAALAAATSPTAADWATVWA
DRPLGPTLPSGTAQAARYNVPIVQPHALAAQAAAHLGSLNLTLLTSFAAVLAAWLPA
TPWSLTVVSWDRPTTLPDAHRVIGDFTALRWLTLPLPAADWTTAAHAVAALAAADA
ARPGDPIAALRPRLRHASQQTFFPLVFTLLELDYDPSSWAVGAGWSQTPGVAIDCLP
QWTTTDTLLLQWDVAAQISDPQTLQQLLATQAAWLDRLATDPTAWTLLVPTSLPAG
APSTNHAQLLPALATAVNHTPSAAAPAPQVAWTEHRSVPAVIAVAHATPHAIALS
RDAHCSYADLETRATALAHTLIAAGAAPGQLIGLLAHRSLDLPIAILKTAAYLPLD

PHAPADRLTFILHDAQVALILHDPTIDSTPFAAPQRTFVPIHTTTPSPSSTPLPRIDPTA
 DAYVIYTSGSTGQPKGVPITHRHILRLLSTTHPWFRFQPTDVTWTLFHSVAFDFSVWE
 LFGALCHGARLVIVPAAITRDFPAFYRLVAAEGVTILNQTPTAFAQFAAADAHHPLPL
 ALRTIVFGGEALDLRLLAPWFDRHGDQQPQLINMYGITETT VHVTYRPITTADLAHAR
 SVIGVPLPDLQLAVWQPDGSPTPVGEV GELVVFGAGVSRGYLRRPALTAQKFVTD P
 HRGSGYRSGDLVRVLPDGDLEYLGRNDLQVKVRGFRIELGDIEAALRRDPAVAAV
 ATVVRTGVAEPTLVGYLVPHPGATIDLPRLRAQAATWLPPYMPVPSVLGILPAVPLTV
 NGKLDRRALPWPLPTTPTSAAPSEVAPHPDVAAITATISGLVTTLLHGTRVAPTADLF
 SAGATSLTIVTLVQQQLADLYPVSVP I AQILEFPTISAIAAWLAAQLPAPAVTTTPAMAV
 APQPELAAITATVSGLVTTLLHGTPVAPTADLFSAGATSLTIVTLVQQQLADHYPVSVPI
 AQILEFPTISAIAAWLATQLSAAQPAASPTAPPIPAASIKLAVAPNSELFGATWNTNVK
 ANQPINLTQIAGLLSTIQRLEIAGEAKYLHASAGGKYAIQVYLAVRAGGIEGLAAGCYY
 YHPEEQALILLSTEIPAMPTAPFMLILAGHLNLRPIYQDGSVPFGLLDAGYLSQVLRS
 RSGNHGIGLTPWVDLDQTAIRHACQLDADDQIALCFGGGNLAATAQPTQIIPHQYG
 GATVTTAEIAALSGPMKYHQTDASTALAEPKRMHAIRNFPTEQARQILPTFKYAPEE
 YLVRTCQRRYEAPVAFNDFSRFLSFLQPDSTKLDHLEALEAITIYLSINQQGVIGV
 DAGIYRYDRASHSLIANDQFSYPLQSCHIPFNRQQLSQAFCGYFVIDSHALPSHYA
 ATSALIAAGALSQYLLEQQSKVNIGMVP I GGLHFERVKQDFGLTEQHVLLHSFIAGAF
 THQVQRSESTQFAQQPVILTKPQTSNGLAIVGYSGLYPGASNLEQFWVNLASGTSTI
 SLPPSERWSLEQGFDDQDRAGKIYSWAGGYLENIMQFDHLLFNISPLEARSLDPQE
 RLLLQAVWECLEQAGYTNESSLDDAPRVGVFVGAMWNDYQLQGLEGWQASHKAA
 TAALHSSLANRISHYFNFNGPSMAVNTSCSSAITALHLAAESIQRGECDAAI VGGVNL
 ILHPYHHAALCAMNLVATDGHSRAFSAQGTGLVPGEV GALLIRPVAQAQANNTIH
 GIIRSTAISHYGR TNQFGMANTKSQTQLILQAFERAKLRPQTVNYIEAATGSALADA
 SEIAALSNTFKHYGATTNSLIGSIKPNIGHLEAASGLSQITKVLLQLKHRRLAPSINYE P
 INPLLQLETSPFTIVHDQQDWVTTTRDDHGVIQPRRALINAFGATGSGGGCVIIIEYCDE
 E

8.1.4 SphD

MHLGQLVQAAVSSLKNIATR SKLMFQNLFILSAASSNQLVLTTRQLYDFLTQAEQDA
 PDLTAIAYTLQVGRRALAYRLALMAATQSELLAGLQAF LAGDHSNPQLIIGKADHGV
 SEVAKPQTALEAAHAWVAGSPIAWQQLWNQVSQRPKRIPLPTYPFAEASHSIAIANA
 ASNQAKPQPAQQPKPAQPELVLRVAEYLRGHFCVAE LAESIVDLQAPLEIYGINSQ
 LMTQTLARLEHDLGSLSKTLFFEHQTLFDLATHLSKTHATR WATLLGIESQPEQPPQ
 PVALPQPQTQAPSLNNEPIAIIGLAGRYPQADSIAEFWQLLANGNDCIREIPLERWDY
 RPLFNPNKAARGSIYAKWGGFLNNIAGCDPRFFNIAPREME LTPQERLFLEVAWE
 TFEDAGYTRSRLRQHFNQIGV FVGVMYAEYQLWGAQEALKGHNVALGSSPGSIA
 NRISHFFNLNGPSLAIDTMCSSTLMGLHLAVQSIRNRECQAALVGGVNL SLHPNKYL
 MHSQMNMLSTDGHCRSFGAGGDGFVPG EAVGAVLLKPLSQAERDGDHIYGLIKAS
 AVNNDGKTNGYTVPNPVAQTAVIKQALVQAAIDPRTISYIEAHGTGTALGDPIEIVGL
 NNAFRDFTTDRQFCAIGSVKSNIGHCEAAAGIAGLTKVLLQLKHQQLVPSLHAEQLN
 PNIDFEQSPFFVQQGLTAWQSQTPRRAGISSFGAGGVNTHVIIIEYPAAVAPTSTQP
 EQPYLVVLSAKTPAQLQLYAQRLADFI AQSSGSGAPANLEQQILNSLSDVLRVDPAA
 IDPAEALEEYGFDAIMLAELARNLSIQLEREILPAAII EYGSNLALVSGLNPSKSIQQPS
 TEPLNLRDLAYTLQIGREAFEERLTIIASNLQTLQQQLTDFAIHELVAESMARATIKR
 NHQAVDFRTEEVAGNDWHALAQHWLSGASIDWAALPQNQQARRISLPTYPFAREQ
 HWIEIVPMPQIGSNVQALHPLIDANQSTFERQCFQKSLVANDP LLLAGSQQGGYLHSG
 LLEMALAAARISSQQPSHMITLTIDECEPRLTDQQLWIQLTPTSAGCEYYISLDSVE
 HPAIAWGNVNSSEPQPQATLDLAALERDYRSANAQFLSQQE VYQQQLQTLDIACQP

WLEQVAINGNQAWSSLAVAFTNSDYSDFFVVVPAVWEATCQAVALLSNQAVEATLLV
 NYATIKLVQSLAETKHIVIVQTEAGYTLNFADQQGLILGIAELELRQSPSDVIAEIELNP
 SNRIDNPTQTTIDLPRIAATILKYDPAQLDLHVPSEYGFDSVMLTEFINQINREYGQKL
 VLAQILQLNQPTLAAIGALIPSQTPPPMAAPPARPKPTKVQKSMPQPCKTIQRQRPTM
 SDAAEPIAVIGMSGFRFPQADDLNEFWEVLRSDTSAIGEIPAERWDWQAYAGDPQQ
 GNKTRIKWGGFLRDIAGFDAFFNISPTEARLMDPQQRLFLETVWHTIEDAGYRPSA
 LNGSRTGVFVGVNAHDYERLMLEQGITIEAHSATGNAHAILANRVSYLFDWHGPSE
 AIDTACSSSAVAIHAIQALRNGDADLALAGGVHLLLSPHDFISYNKAGMLSESGRVK
 TFDAAADGYLRGEGVGAVLLKPLAQAERDGDPIYAVILGSAVNHNGRGFALTVPNP
 NGQADVIARAYQNAGVDPASVSYIEAQGTGTTLGDVPEIMAFTEVFKTTSTVMDQP
 CLIGSLKPHIGHLEAASGVAALIKLILALRHSSIPATLNLETINPQIELTETRFQLAQQPQ
 AWQAQPNMRRAGLHSFGVGGSSNAHLVIEEYSSPQHSSNQHEPALIVLSAQKPTSL
 RAMLIRLQNELASQPELNLHELAYTLQVGREAFEHRVAFIAESVATLSALIERYLAGE
 PASEAQSVFEGALRTIDPALKRLLTGKSGELLITAALAERDLGQVAQLWTQGAFVSW
 NSLYPANTVQRLHLPGYHFEHQQFWFDQTPNSTPTTQPISLMPVMSISTELAEQAI
 FGNVLGIPATQIDPSEPLENYGIDSVLMTKVKYLIEEKFAITIPLALLSQQATLETLSQHI
 PTNAQKTIPDLTTLSGTALDDLFIQIQGR

8.1.5 SphE

MLTMQLVPPAAQTIRDPRDLPAPTIRDCTSADDATIAACAHEHTDRLSHAAYPLDAW
 PRWSLQILALPHHACRVFLSVDETLDDAASLVILLHEWHTLVLDPATVLPASTAFPA
 FLQRLQTRAAAAAADLAYWRTRLHGITGGPALPWIPLQPHHATTAPDGRRFLPRT
 RLAACLPAPAWSALQALAQHAPSTLLLAVFAHILRSATADPRAPFALILTLANRV
 PLIPASEQLVAPVVSTTLVRLDPADSLATLLPSLQAQLWDTLDHPFTSGIAVLRDLKH
 QQLLPATTTIPVVFTAMLNFAFGNHAQQPSFFPAVDALTQTPNVWLDHQLLERNGA
 LHFNDILPAAFPRGLAEALFSRYTAALQTLACDPAALHAAALGLPDPTPLPLSAVQ
 RSYLLARPAAPARFYQEFRLPQHAIPRAGGGLAARGRRPSRPAHDPPYRPPRLCQ
 RQPHHGGFADHARAGSGWPLLATGTAKQHAPFQPIAVRLTQDTDATTVLHCAIDAL
 YLDGRSLMTLFREWAACAAAPFAPFAPPATSPHALRAALAAATSPTAADWATVWA
 DRPLGPTLPSGTAQAARYNVPIVQPHALAAQAAAHLGSLNLTLLTSFAAVLAAWLPA
 TPWSLTVVSWDRPTTLPDAHRVIGDFTALRWLTLPLPAADWTTAAHAVAAALAAADA
 ARPGDPIAALRPRLRHASQQTFFPLVFTLELDYDPSSWAVGAGWSQTPGVADCLP
 QWTTTDTLLQWDVAAQISDPQTLQQLLATQAAWLDRLATDPTAWTLLVPTSLPAG
 APSTNHAQLLPALATAVNHTPSAAAPAPQVAWTEHRSPAVIAAVAHATPHAIALSF
 RDAHCSYADLETRATALAHTLIAAGAAPGQLIGLLAHRSLDLPIAILILKTGAAYPLD
 PHAPADRLTFILHDAQVALILHDPTIDSTPFAAPQRTFVPIHTTSPSPSTPLPRIDPTA
 DAYVIYTSGSTGQPKGVPITHRHILRLLSTTHPWFRFQPTDVWTLFHSVAFDFSVWE
 LFGALCHGARLVIVPAAITRDFPAFYRLVAAEGVTILNQTPTAFAQFAAADAHHPLPL
 ALRTIVFGGEALDLRLLAPWFDHRHGDQQPQLINMYGITETT VHVTYRPITTADLAHAR
 SVIGVPLPDLQLAVWQPDGSPTPVGEVGVGELVVFAGVSRGYLRRPALTAQKFVTD
 HRGPGYRSGDLVRVLPDGDLEYLGRNDLQVKVRGFRIELGDIEAALRRDPAAAAV
 ATVVRTGVAEPTLVGYLVPHPGATIDLPRLRAQAATWLPPYMPVPSVLGILPAVPLTV
 NGKLDRRALPWPLPATPTSAAPSVVAPHPDVAAITATVSGVVTTLLHGTPVAPTADL
 FSAGATSLTIVTLVQQLADHYPVSVPIAQILEFPTISAIAAWLAAQLPAPAPVTAPAVA
 VAPHPDVATITTTVSDLVTTLLHGTPVAPTADLFSAGATSLTIVTLVQQLADHYPVSV
 PIAQILEFPTISAIAAWLAAQLSAAQPAAPPTAPPIPAAASLLNLNGLSNLLGLLQPT
 AGKRKYLYASAGGKYAVQVYLAVRTGAIQGLVAGCYYYHPEQHGLVCLSNEEPDFL
 LESAAPFGFLLVGQLAALRPVYQDFSPMFAALDAGYISQLLINYADNQGGINLTPNHTI
 TTQNLNSALQLSPDHRVVQCLLGGTANAGISKQQSGQAIEFVSLES LAEYIGQMTYS

QLSQEELARLDEQKLHLRPIVNPQQIDLPEALIDSDRYLRRGCQRCYEQSPIQTTNF
HGLLKSLSKLLSSDAGIEAYLYLKAERIAGYTGGVYRYNVATQLLEPIVQELSYPLQN
CHTPFNRPHLSSAGFCLYLIGDINRLRERFGPAAFHGGLLSAGAIGQHILLEQQANHK
LGLVPIGGMNERIKQDFGLSEQHVLLHSFLGGSYQHQTNASQSPSPMPTANPVQP
DQPMIAIVVGISGRYPQAATIQEFWQNLLAGKNCISEIPAERWDHTPYFDPKAQPG
KTRSKWGSFLADIDQFDPLLFNIAPEAEELDPQVRLFLETVWRLFEDAGYSREALK
ALNVAGALPVGLFVGS MYQHYQLLANDPQSQALLAIQSYSSIANRTSFFFDLRGPSI
AIDTACSSSLIALHLACESIRKQESGMVVGGINLHLHPAKYLALSELGLVGGDELSR
SFGQQGGFVPGEGVGAVLLKPLAAAERDGDRIYCVIKGGAINHSGHGSYIALPQAS
AYTSLIQNALKQAEVDPRTISYLEAAANGTPLADSIEIAGLTRAFQASTSDQQFCAIGS
VKSNIHLEAASGISQLIKVALQLHHRTLVP TLNAEPLNPLIELHNSPFYLQQTQQPW
QRPTIEVAGRVEQQPLRAALGSYGAGGSSAFVIVEEYAPSPVLAPSGDTP TIIALSAQ
NRQSLVLQAENLRDALATAPLLDVATTLLRGREPLAERLAFIASNRHEAISKLEQFIN
GTLTEKEGSVGHANETHVVSSLFDSADEFQAFIASQIAEQRSAKLARLWTVGAISNQ
VLRVIPAGQIVSLPGYLFHQRCWLPNPAQHQPSPIAAQRSQSAKPTDDLQEI VAD
LLKIPVASLDRKTKLQRYGFNSLLAMRLLHVLS ELYGCEIPTREFLRLESVAAISDWLT
QAGFTRTSLEQAEPNPAHPDSWDADLDDHQLLHLLHAIADGKISPSDALNYEYTLN
GSEN

8.1.6 SphF

MPILMITNYCICYTLSPMAKLAQAMLLIMNIPSMGRRISAMHRSKATIYQQVARGELS
VGMALQALRELASLPAISNQSQQFALSEGQKALWMAYQLDPNQYAYNVPLAYI LAP
NTQPVILLEQALNDLIKRHPLLAARLVETDQGQIEQVWDGATTSISLERQQLSNVDWV
QEVRIAHPFDLLKGPLFKATL FELPNRELLLLLNHHIMFDGVS IQILLGELQAVYA
AYAQQQSPQLAPLTTDFGDFVAWQQTMLASPNGQRQRRYWL EQLAGRVTPLLEP
YDRQPSGQPSFKGAMWFEFELDTQLTAHIRSLGLEHNRSYFSLIMSGFSMLLARYGR
QQSIYLGTPLSGRPEARFDNLIGYFMNMVIGADLSDNPSYLELSRRIYETALDAMD
HGDYPLFAIADELGGNMPFHAAFYFQNWVAPLAEQT NANATLLTKPVLSIHQEGEF
SLTLEVIELGDRYSCYFKYNPDYFDQATIQQMAEHLIQLLGAAINQPEQPIHSLPSLLA
AEQRLLTAWNQTQTTYPRELALPSLVAQQAATPD AIAIGSTLAKPNQTINYRTLMA
AVDQIAAHLQSHGIGPGQRVAVLLNRSIEMVLSLLAIAKTGAA YIPLDPIYPAERLAAM
LDDSQASLILLHGELAVHLPTTSIAQLEIEQLLAKQPAKPLVAVEIDPNSPVYLIY TSGS
TGTPKGVVISHSGLTNFLWAMREQLVFS AKDRILAITTICFDIAGLELYLPLISGGSVEI
LPAEVTRDGYLLKQAIANS DATWLQATPATWTMLLAAEW EQPLQ TILCGGERLSYD
LAQQLVARASTVWNLYGPTETT IWSTASRIQANTPITIGQPISNTTLYILDQAMQAVPI
GVAGELYIGGAGVALEYWQQPALTNARFIDYPDANQQISRIYKTGDLARFRPDGQL
EHLGRLEDEQVKVRGFRIELMEIETTLRRQPEVREAVVVQRAEASQQLIGFVMLEDPR
IGVERPFNSEALRQALAQSLPEYMLPTKIIGLREFPLTLN QKIERKVLKSLPLATIINRY
GYPQTNIEQPKQPVK TQPSAPLLAELAELASTIVAIAPNEIDPLL PFGNYGFDSIRFTQ
FSTLLRKRFSLPIMP NIFYLKPTLQELAQHLQSLLPQQPEPVNQQVIEQPTLTQPTTK
ANAAIAIIGIAGTLPQSRDLAEFWQHLVAGADLITVPANREAWTTPASANLNATERQ
SINWGGFIPAVDTFDAFFGISPREAALMDPQQRLMLETVWKAIEDAGYRASSLQAA
NVGVFIGASGADFLGMTEDSIDGYTLTG IARSVIANRISYLLNFHGPSEPVD TACSSS
LVAVHRAAQAIRNGECE LALAGGVNVM LSDFASAAASKVGM LSPDGRCKTFDASA
NGYVRGEGA AVILLKPLELALRDRDHIYAVIRGS AVNHGGGRANSLTAPNPNAQSALI
YQAYRNAGIDPASVGYIETHGTGTALGDPIEVDGLKSAFQKLYHDADQAWIDGHCA
LGSVKSNIHLEAAAGIAGLIKTVLALKQRYIPATIHLQTLNPIYIEVEASPF SISRFRD
WNTSGVRRAGVSSFGFGGSNAHIILEEAPVQSSSSVSDQPTLIVLSAKTATSLQLAA
HQLATHIQRLQSDQPLWFGATSLRD LAYTLLTGREAFRERLAIVASDFAELAVALNA

YLAGNSAANIYSNTSAQTAAEATMLHSLAQQWVAGATIEWSQLDVTNAPLPQRQPV
 PSYAFMPDHYPARRNQAQAVLHPLVERNSSTFNTTQFSTRLLAEAFYLRDHIVQQG
 RVLPGVTYLEIARAAGSFAAEKPVQRISNLIWASPCVVEQVRELTISLQPQPNAAGFS
 ITTQSATGSKITHAEGTLRYDRPDDQPNMAKIDLAPVRSRLTHQKTAACEYHLFAEH
 QFAYGPSFQVIEQIAYNADESLATIKLPAAQQSQFADFGHPSLLDQALQTVILLSD
 ASDSESRLALPFAIGELVIFGQLAPECLVHAIRNQAQAGEKGLVKYTIQICTPDGQSLI
 QINDYTIRAMPNQASVQPASSNLSTIAYYTPQWEQVGIYETIEGV

8.1.7 SphG

MVPETIVYLWSEQASANQPASASLDLTLYPLFLLSQAFIHHKRAAMRVLYCYSDAPQ
 PIAPFIMGLAGFARSLQMEHQYWIQLLALQSSDVAMNRQVLAIIIRDELADTTSSRE
 FELRYQEGKKFARRFLPHAAPTPTADPQQVLVLRPHGCVIAGGKALGAIFARYFAS
 QAKVQLVLLGRSPIDQPTQKLISELEQTGSQAHYLQVDISDQAALSLALAPFGTIHGII
 QAAGISNDSLLVHKTKASIDQTIAAKVQGTALDAITAKQPLDFFILFGSSSGAFGNLG
 QADYAYANSFLNGFAAYREYLATKALRHGKTIANWPLWADGGMGHNQATQQRLA
 QLGLSPLATEEQGLRAFELSLHTNQLQMIVLGGDRQRIANLVNPMPTPQPIQPNLVVD
 TPSMSETPITPTTTEPTMAETPSIDAALQHYTRDYLRALLADATKLSAQRINPNPFE
 SYGVDSVMILGLNQLGDRFGELPKTLFFEYQTLEQLTGYFVENHRTTLLKDVRLPE
 PASVVAAKPTPQPTSVNPNPPSAPELSQLAPAQPATPKGEVDIAIIGVSGRYPLAAN
 LDEFWQNLQTGRNCVTTIPAERWDYRRYFHPEKGTYSIWGGFLDDIDKFDPL
 FFNISPLEAEMLDPQERIFIQTVWHTLEDAGYTRNALSRYQVGVYVGMWGWYQLY
 GAAEPDDGVILTPASSYASIANRVSYFFNFRGPSIALDTMCSSSLTTIHLACESIKRGE
 IEVALAGGVNLTLPKHFILSQTKFASSEGLCRSFGDGGDGYVPSEGVGAILLKAL
 PQAIQDGDRIYGVIKASAVNHGGKTNQYTVPNPRQQEDLIRQAFRKAQIDPQTISYIE
 AHGTGTALGDPIEITGLTKAFNLPANQRHVAVGVSVKSNIGHCESAAGIAGITKVLQ
 MQHGQLVPSIHAETLSSNINFAESPFIQRNLSAWNSAGLRRAGVSSFGAGGANH
 IVLEEYQTEIQPEPNDQRPQLIVLSARDNERLRANAAALEAYLVKQTGSNQTNRAF
 AGTNLQHELAELVANTLNVASSELDPEPLHEYGLDAPAFQLQKQFEQAYQLTLPS
 NNPIDAFSIDQLAEYLQAEYPQRFTHAHSFAQAPNLAPPQLRLADIAYTLQVGREAM
 EARVAFVSSIEVLIERLHAFAGGATAIANCYLSATPKPSGD

8.1.8 SphH

MDYQVADSQRPAELLAMLHTATQIALPATPIEMQRLQLSAPQLLSEPDQEIVNLT
 PQADGSITAIVGQSTNQPWLQATIQTNAASAEQRVALEPILARCSNQATLSQQSNGV
 SYGPAYAGLERIVWINDSEALGYTLDPFQPKQLGKTAFTPAHGAIMLRLASEFHQR
 HGLASEFRKLGSVQIVADCPAEGLVYLKRAAGSNHLAILDHSGRVYAKFQGLSTHS
 EATPVQALPAPSTSQAPSDAEQILERFFYAPRWQPTLANAQSTPQVGLVIIASAES
 QALVSALCANQTAAINLIWLDSTLKHPTIPTWLVRSDPNALSHALAAIGSIDSIYLLS
 GLEQASTTFDLGHFEQQQERGVLSLFRLLKALQEQQLLDQELSLKIITANTNHVVAN
 DQVQPFASMFGLAKSAREYRRLRIACIDLEQHDLPETVAATAAFIHGEAPLAPL
 QEVAQRQGQRYQRILEPVKLPNSLAFRKGGMVVLGGAGNVGFKLCCYLAEQQCQ
 AKVVITGRRAITPAIEAQLQSIEERGGRAIYIADMSDLDSMRSAVATTKQHFGGLHG
 VFHSAMSFETYPLNGLTEAQLYQDLAPKARGSVVLHHALVDEALDFKLYFSSGEAF
 TGNIGWGSYAAGCTFEDAFHAERAQVNYPVQVINWGFWESERDQYLDALKAKGI
 NPIETAIGMQAVSRVVAAGIPQVMALNVADKVLMLMGVVLPKPNVASATPAISQAAT
 PIPAAATPPVAVVPLVTSQPVAANPNQREQLYGYILGQLAAVLKIDPSRIDVTSE
 LTNYGVDSLVTDIHKRFEQDLGSMPTLLENTTVAAIANFLQQDYADRVAFFSP

MVAATVTSQATSNTHEQLIDTADLFALEVGSNQPSSATQPPTAPASITLLRQIEPAAI
ASELDRYGDQYAQKLFPAWKQNGGSLVNLTELDANPQLLKHLLVNVADKTQAEVW
MIGNGPPLVLIPAIGLTAPVWINQIQQWAADYRVIVIHQQGYGMDLTDISTAAVAK
LFISTLDQLGINRPCHVIGSCFGGVAAQYLTQAYPERVCSLTLCGTFNKNFGLPDIDV
SELTIDQMIEGAKMIGSSINRDFDAVAEGLASDQAQPIVEQARSLLLKSQCVSPLVVM
RYITQILTLNGQAWLPRIQAPTLCLSGNLDTIVAPETSRTISQQIPAGRYIEIPGAGHYP
FLTHVDLFEQAVRPFLREQEAQLMSTLV

8.1.9 SphI

MIVTIEHAAGFEPIEQVLRCCDEYEASLNLTRDKRYTRIRITVDQAYSELFVRLKSIVG
VAKVETTGKPYPLAAREAKSAASVINVRGIEIGGSDLVVMAGPCSVESREQILATAH
AVRAAGARILRGGAYKPRTSPYDFRGLGEEGLQLLAEAREATGLAIVTEVMSIADLE
LGAYYADIIQIGARNMQNFSLHACGSINRPILLKRGPSATIEEWLLAAEYILAAGNEQ
VILCERGIRTYEPFTRNTFDVSAIASIKHLSHLPIIGDPSHGKAEVLPQMARAIIAA
GADGLIIEVHPNPAQAWSDGQQSLTFEHFEQLMNDLAPIANALGRQLNPNVQALVV
S

8.1.10 SphJ

MASLIKALAPDPQLPQLFLFPFAGGTANSYRGLANALKSHFSVYAIDPQGHINRQEP
LLDDLETMVEAYLAALLPLIKPDFTLFGHSLGGAVVYRLTQRLEQLGHAPVTVFISGY
HPPHISDKSTAYLNDDQFLEHIEAMGGVPPEIGQDQSFMRYPFLPIFRADFRATETFIH
SDRTKIKAPVFLNGDKDDDAIKHMQEWSYWLNQVRYRIFSGPHMYLSSQPELIAT
YIIECNQAISIVDE

8.2 Search results from the NaPDos Database

8.2.1 C domains in contigs from the genome sequencing

Table 8-1: C domains encoded on contigs from 454 sequencing

Domains were identified and annotated with NaPDos; contigs with parts of the putative siphonazole cluster are printed in bold letters

Contig	Database match id	Identity (%)	align length	pathway product	domain class
00282	syryn1_C5_LCL	40	300	syringomycin	LCL
00282	micro2_C1_DCL	47	297	microcystin	DCL
00282	syryn1_C7_LCL	45	296	syringomycin	LCL
00282	syryn1_C6_LCL	46	299	syringomycin	LCL
00282	ituri3_C1E	43	295	iturin	epim
00462	Sare2407_1	33	298	pksnrps2	C
00462	syryn1_C9_LCL	39	300	syringomycin	LCL
00462	syryn1_C5_LCL	45	299	syringomycin	LCL
00585	syryn1_C9_LCL	30	274	syringomycin	LCL
00585	fengy2_C1_DCL	30	208	fengycin	DCL
00689	micro1_C2E	41	303	microcystin	epim
00689	fengy2_C1_DCL	30	288	fengycin	DCL
00738	bacil2_C1_start	43	293	bacillibactin	start
00782	syryn1_C7_LCL	50	300	syringomycin	LCL
00782	micro2_C1_DCL	44	301	microcystin	DCL
00782	syryn1_C7_LCL	41	300	syringomycin	LCL
00782	cyclo1_C8_LCL	42	201	cyclosporin	LCL
00782	syryn1_C9_LCL	43	303	syringomycin	LCL
00782	micro1_C2E	38	304	microcystin	epim
00805	micro2_C1_DCL	41	298	microcystin	DCL
00805	mycos3_C1E	36	299	mycosubtilin	epim
00805	micro2_C1_DCL	41	297	microcystin	DCL
00805	micro5_C1	38	292	microcystin	C
00805	syryn1_C9_LCL	41	302	syringomycin	LCL
00805	fengy2_C1_DCL	39	297	fengycin	DCL
00805	syryn1_C6_LCL	43	300	syringomycin	LCL
00805	mycos3_C1E	34	295	mycosubtilin	epim
00874	micro1_C1	44	298	microcystin	C
00874	syryn1_C9_LCL	46	297	syringomycin	LCL
00897	Sare2407_1	33	296	pksnrps2	C
01402	syryn1_C7_LCL	42	299	syringomycin	LCL
01515	micro2_C1_DCL	49	298	microcystin	DCL
01515	micro1_C2E	47	303	microcystin	epim
01517	Sare2407_1	30	297	pksnrps2	C

**Table 8-2: C and Cyc domains from the putative siphonazole cluster
Alignment and classification by NaPDos**

gene	Database match id	Identity (%)	align length	pathway product	domain class
sphC	pyoch3_C1_cyc	30	434	pyochelin	cyc
sphE	bleom9_C3_cyc	32	331	bleomycin	cyc
sphF	bacit3_C5_LCL	29	438	bacitracin	LCL

8.2.2 KS domains in contigs from the genome sequencing

Table 8-3: KS domains encoded on contigs from 454 sequencing

Domains were identified and annotated with NaPDos; contigs with parts of the putative siphonazole cluster are printed in bold letters

Contig	Database match id	Identity (%)	align length	pathway product	Domain class
00180	FabF_Bacillus_FAS	42	405	fatty acid synthesis	FAS
00282	bleom_AAG02357_RH	56	427	bleomycin	RH
00330	FabF_Bacillus_FAS	53	408	fatty acid synthesis	FAS
00433	FabB_Ecoli_FAS	61	403	fatty acid synthesis	FAS
00585	KirAIV_CAN89634_11T	44	426	kirromycin	trans
00668	LnMJ_AF484556_2T	52	423	leinamycin	trans
00668	KirAll_CAN89632_5T	50	435	kirromycin	trans
00805	yersi_YP_070123_RH	48	423	yersiniabactin	RH
00805	CALO5_12183629_i1	49	424	calicheamicin	iterative
00838	FabF_Ecoli_FAS	48	418	fatty acid synthesis	FAS
00863	FabF_Bacillus_FAS	51	417	fatty acid synthesis	FAS
01348	FabF_Bacillus_FAS	35	407	fatty acid synthesis	FAS
01517	LnMJ_AF484556_4T	55	437	leinamycin	trans
01517	KirAIV_CAN89634_7T	58	423	kirromycin	trans
01517	LnMJ_AF484556_3T	49	440	leinamycin	trans
01518	Lnml_AF484556_2T	55	435	leinamycin	trans
01518	LnMJ_AF484556_4T	52	434	leinamycin	trans
01518	Lnml_AF484556_2T	43	435	leinamycin	trans

Table 8-4: KS domains from the putative siphonazole cluster

Alignment and classification by NaPDos

gene	Database match id	Identity (%)	align length	pathway product	domain class
sphC	KirAll_CAN89632_5T	50	435	kirromycin	trans
sphC	KirAll_CAN89632_5T	46	437	kirromycin	trans
sphC	KirAll_CAN89632_5T	36	438	kirromycin	trans
sphD	Lnml_AF484556_2T	56	437	leinamycin	trans
sphD	Lnml_AF484556_2T	46	434	leinamycin	trans
sphE	Lnml_AF484556_3T	49	444	leinamycin	trans
sphF	Lnml_AF484556_4T	54	428	leinamycin	trans
sphG	KirAIV_CAN89634_7T	58	423	kirromycin	trans

8.3 Primer sequences

8.3.1 Gap closure

Primer	Sequence (5' – 3')	Purpose
Cb1462_f	TGTGCCAATTGCCAGATTC	gap closure 1462-1517
Cb1517_r	CGCCAGCAGAAGCATAGAG	gap closure 1462-1517
Cb1518_r	GGCCAGCAATTTGGGTCAG	gap closure 1462-1517
1462ktr-f	CCAAGTGTGCTGGTATC	gap closure 1462-1517
1462ktr-r	TAGGAAATCCGGTTCCTC	
Gap1_f	CGGTGGGAACGACATTTG	gap closure 1462-1518
Gap1_r	TTGGAGGGACGGTAGAAG	
Gap2_f	CCGTCTTTGCCACATTC	gap closure 1462-1518
Gap2_r	TCGAGGAGTTCGGTAAAG	
Gap3_f	TGCTGCTCACCAGCTTTG	gap closure 1462-1518
Gap3_r	AGACGGAGAAGTCGAAG	
1518-1517	GGCTGCCATTTCTTCCTGAC	gap closure 1462-1518
1462seq	GCATAAGGCTTGACGTAAGG	
1462-A-fwd	GGCAGCTCATTGGGTTG	Control primers for
1462-A-rev	GCACCAAGGTGACGATG	contig 1462 A domain
1280seq_neu	GCGCAATTTACGTTCCG	Gap closure 1280-0668
668seq-1280	CCGCATTCCAGTTTCGG	
1518-1517	GGCTGCCATTTCTTCCTGAC	gap closure 1518-1357
1357seq_f	CCGCCTTGCTTGGATTAC	
1518seq	GTTGCGCCAAATAACTCG	gap closure 0668-1518

668seq	CGACGTTTGGGTTATCG	gap closure 0668-1518
1518seq_neu	CCAGCAATTTGGGTCAG	gap closure 0668-1518
1518-EC10-fwd	GCCCAGAAATTCGTGAC	gap closure 0668-1518
EC10seq-rev	GCCATTGACGGTCAAGG	gap closure 0668-1518
EC10seq2-fwd	CTCGACCATCCCTTCAC	gap closure 0668-1518
EC10seq2-rev	TCGGCATAGGAGCAGTG	

8.3.2 Protein expression

Primer	Sequence (5' – 3')	Target
OMt-fwd-BamHI	GCGGATCC ATGG CAAAGGCATCATTGAAC	SphB
OMt-rev-HindIII	CGAAGCTT GTT ATTTGCGAATAACGAGAG	
A-fwd-BamHI	GCGGATCC ATGG TGCAACAAGATTTAC	SphC-A1
A-rev-HindIII	GCAAGCTT TCAT GCAGTTAGGCTCTTCAC	
A-PCP-rev	GCAAGCTT TC AAAGGGCGCTGGCCAAATC	SphC-A1
DHB A-Dom fwd	TAGGATCCAGCCTGAGCTTTGAAC	SphC-A1
DHB A-Dom rev	GCAAGCTT CTA ATCAATTTTGCCATTGGGAAG	
A-dom orfstart	CG CATATG GTGCAACAAGATTTACAATCAC	SphC-A1
A Domain fwd	TAGAATTCGCCCGCCAAACTCCGCAAG	SphC-A1
A Domain rev	TAAAGCTT TC AGAGCGCCATGCGATC	
sphC-a-fwd	CACCATGG TGCAACAAGATTTAC	SphC-A1
sphC-a-rev	CTA ATCAATTTTGCCATTGGGAAG	
sphI-fwd	CACCATG ATCGTAACGATAGAGC	SphI
sphI-rev	CTA ACTGACCACAAGTGC	

8.3.3 Fosmid library screening

Primer	Sequence (5' – 3')
TE-end-fwd	GCGTACCACCAGAAATTG
TE-end-rev	GCGGTAGGTCTTACCAAC
CF4-end-fwd	GGCCTGTGAGAGTATTCG
CF4-end-rev	CACCAACGTGCGATGATG
EC9-end-fwd	CGGTGATGATTCTTGGACTC
EC9-end-rev	AGCCAAGTGGATCGTAG

8.3.4 Yeast recombinational cloning

5'-Homology extensions to the vector backbone are underlined

Primer	Sequence (5' – 3')
Hefe-start-fwd	<u>TGTACAAAAAAGCAGGCTCCGGGACGGCCATCATATAG</u>
Hefe-start-rev	CGCCAAATGGTTGTTCTC
TE-end-fwd	GCGTACCACCAGAAATTG
Hefe-end-rev	<u>AGCTGGGTTCGGCGCGCCCACAAGGCTCCACTTTCTACG</u>
Ura3P mid fwd	GGCGGCAGAAGAAGTAAC
Cat mid rev	TCGCTCTGGAGTGAATACC

8.3.5 λ-Red mediated recombination

5'-Homology extensions to the fosmid inserts are underlined

Primer	Sequence (5' – 3')
Str-FRT-fw-EC10	<u>CAGCGTCGCAGGATTCTCGCTCCAGCGAGTTCTGGTTATG</u> GATCCGTCGACCTGCAGTTC
Str-FRT-re+EC10	<u>TAGAGCAGAATCATCAATGCTTAAGACCGCGAACGGCTCG</u> GCAGGAATTTCGATGTGTAGG
Apr-FRT-fwd+IC2	<u>ATGATTGGCTGGATTGCACGCTGCGATGGCAGTAACCTTC</u> GATCCGTCGACCTGCAGTTC
Apr-FRT-rev+IC2	<u>GCGATGCAGTTGCCGCATGGAACGAAGAAGAAGAACGCAA</u> GTGTAGGCTGGAGCTGCTTC

8.3.6 Sequencing primers

Primer	Sequence (5' – 3')
T7	TAATACGACTCACTATA
Epi-RP	CTCGTATGTTGTGTGGAATTGTGAGC
SP6	ATTTAGGTGACACTATAGAA
16S rDNA fwd	AGAGTTTGATCCTGGCTCAG
16S rDNA rev	AAGGAGGTGATCCAGCCCCA

8.3.7 mRNA screening primers

Primer	Sequence (5' – 3')
A-fwd	GCGAGGCATACCATGAAAGG
A-rev	CCTGCTTCTGTGATCCTTG
B-fwd	CAGGCTCGTCATGTGCGATAC
B-rev	TTCTGGTGGCGTTTGCATGG
C-fwd	TTGGTGATGGCTTGACTCTC
C-rev	CGGGCTTGGTTGGTAAACAG
D-fwd	TGGCGTAATTCAGCCACGTC
D-rev	GCAATCGCGGTTAAATCAGG
E-fwd	TTCAGGAACTGCGCTTGATG
E-rev	GCAGTCGCGAATGGTTGGTG
F-fwd	AATCCCGCCCACCCAGATTC
F-rev	AATGCCATGCCGACGCTAAG
G-fwd	GGAGCAGGTTGGCATCTACG
G-rev	GTTTCAGGCACCATCCCTTG
H-fwd	TTGTTATTTGAGCGCCACCC
H-rev	AACGCTGACTATCAGCAACC
I-fwd	TTCTGCGCGAGCAAGAAGCC
I-rev	CTTTGGCAACGCCCAACAATG
J-fwd	CCGACAACCTCAATCCCAATG
J-rev	TCAAGCAGCGGCTCTTGACG

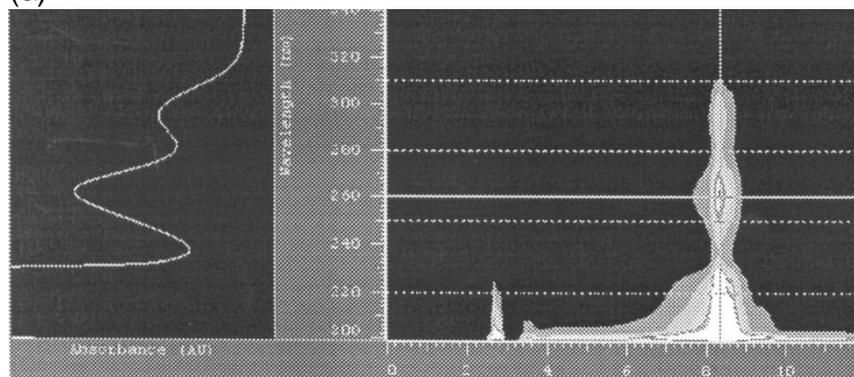
8.4 Complete sequence alignment of the DAHP synthase SphI

<i>S. thermophilus</i>	HPDDTVIHVG-DVAVGGNEVVVIAGPCSVESEAQIIQTAHAVKAAAGASML	126
<i>T. roseum</i>	YPLDTVIQVR-DVTIGGNEVVVIAGPCSVENERQILETARAVKAAAGASML	126
SphI	KSAASVINVR-GIEIIGSDLVVMAGPCSVESREQILATAHAVRAAGARIL	126
<i>B. subtilis_AroA</i>	KPEDTIVDIK-GEKIGDGQORFIVGPCAVESEYQVAEVA AAAAKKQGIKIL	149
<i>S. xylosus_AroA</i>	KPEDTIVQFDNGGIIGDGNKSFVFGPCSVESQEQVDAVAANLQARGEKFI	54
	. : : : . . : * . : . : * * : * * . * : . * : * : :	
<i>S. thermophilus</i>	RGGAFKPRTPYDFRGLGEEGLRLLAKARAETGLPIVTEVLSVADLDMVA	176
<i>T. roseum</i>	RGGAFKPRTPYEFRGLGEQGLKLLAKAREETGLPIVTEVLSAQHVDLVA	176
SphI	RGGAYKPRTPYDFRGLGEEGLQLLAEAREATGLAIVTEVMSIADLELGA	176
<i>B. subtilis_AroA</i>	RGGAFKPRTPYDFQGLGVEGLQILKRVAFEDFLAVISEIVTPAHIEEAL	199
<i>S. xylosus_AroA</i>	RGGAFKPRTPYDFQGLGVEGLKILLKNTKDKYGLNVVSEIVNPADFEVAD	104
	* * * * : * * * * * * : * * * * : * * * * * * . . . * : * * * * : . . .	
<i>S. thermophilus</i>	EYADVLIQIGARNNTQNEILLLEAVGKTKPKVLLKRGMSQIEEWLLAAEYIM	226
<i>T. roseum</i>	EYADLLQIGARNCONYLLLEAVGRAKKPVLLKRGMAVQIEEWLLAAEYIV	226
SphI	YYADIIQIGARNMQNFSLHACGSINRPILLKRGPSATIEEWLLAAEYIL	226
<i>B. subtilis_AroA</i>	DYIDVIQIGARNMQNFELLKAAGAVKKPVLLKRGLAATISEFINAAEYIM	249
<i>S. xylosus_AroA</i>	EYLDVFIQIGARNMQNFELLKEAGRSNKPVLLKRGLSATIEEFYIAAEYIA	154
	* * : * * * * * * * * : * * . * : * * : * * * * * * . . * . * : * * * * *	
<i>S. thermophilus</i>	AQGNQVMLCERGI RTFFETITRNTL DLSAVPVVKRLSHLPIIVDP SHGTG	276
<i>T. roseum</i>	AQGNQVLLCERGI RTFFETITRNTL DLNAVPAVKHLSHLPIIVDP SHGTG	276
SphI	AAGNEQVILCERGI RTYEPFTRNTFDVSAIASIKHLSHLPIIGDP SHGVG	276
<i>B. subtilis_AroA</i>	SQGNQDIILCERGI RTYETATRNTL DISAVPILKQETHLPEVFDVTHSTG	299
<i>S. xylosus_AroA</i>	SQGNENIILCERGI RTYEKATRNTL DISAVPILKQGTHLPEVMDVTHSTG	204
	: * * : * * : * * * * * * * * : * * * * * * : * * : * * * * * : * * * * *	
<i>S. thermophilus</i>	KWYLVPSMMLA AVASGADGLIVEVHPNPDHALSDGSQLTFENFSAAMPK	326
<i>T. roseum</i>	RWFMVPAMMLA AVAAGADGLIVEVHPNPDHALSDGAQSLTPENFAAVMPK	326
SphI	KAELVPQMARAAIAAGADGLIEVHPNPAQAWSDGQSLTFEHFEQLMND	326
<i>B. subtilis_AroA</i>	RRDLLLPTAKAALAIAGADGVMAEVHPDPSVALSDSAQQMAIPEFEKWLNE	349
<i>S. xylosus_AroA</i>	RKDIMIPTAKAGLAVGADGIMAEVHPDPSVALSDSQQMDFNEFDKIFYNE	254
	: : : * . : * * * * : * * * * : * * * . * . : . * . .	

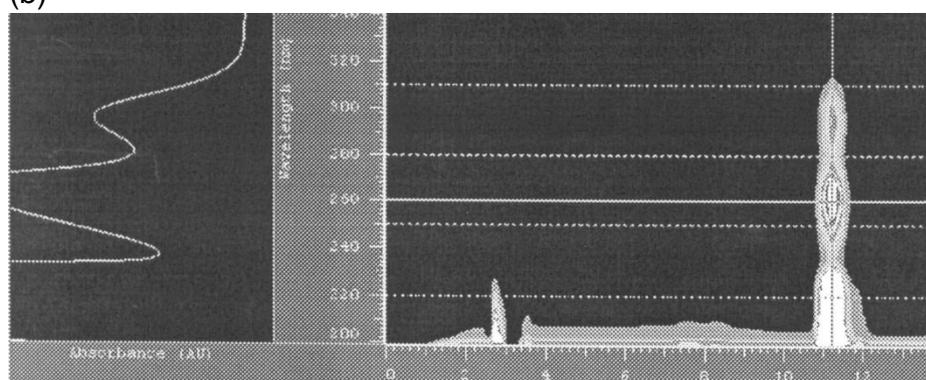
Figure 8-1: Multiple sequence alignment of SphI with other DAHP synthases from subclass AroA_{lg}; conserved residues are shaded yellow; possible motif regions for catalytic activity of DAHP synthases are shaded green; conserved regions were defined as reported by Subramaniam et al., 1998; reference sequences from *Sphaerobacter thermophilus* DAHP (*S. thermophilus*; YP_003320125), *Thermomicrobium roseum* DAHP (*T. roseum*; YP_002521366), *Bacillus subtilis* chorismate mutase (*B. subtilis* AroA; P39912) and *Staphylococcus xylosus* chorismate mutase (*S. xylosus* AroA; CAA64712)

8.5 HPLC UV-spectra

(a)



(b)



(c)

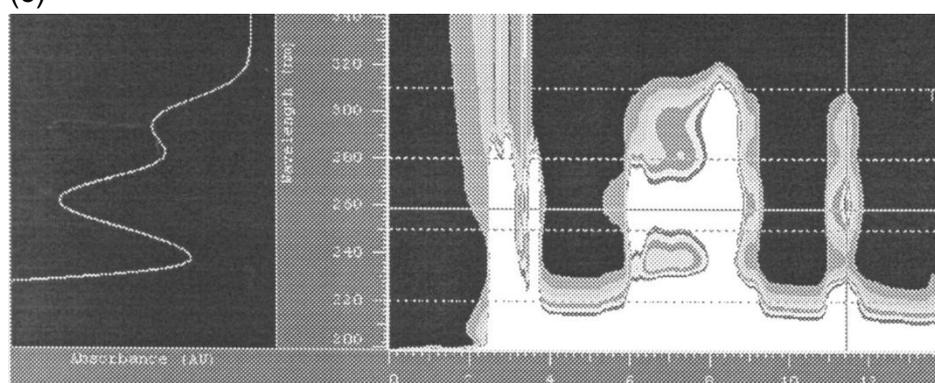


Figure 8-2: UV-absorption spectra from methylation assay;
(a) Protocatechuic acid standard; (b) isovanillic acid standard; (c) methylation reaction;
absorbance on the left side shows UV spectrum of the product peak

8.6 Chromatograms from ATP-PP_i exchange assay

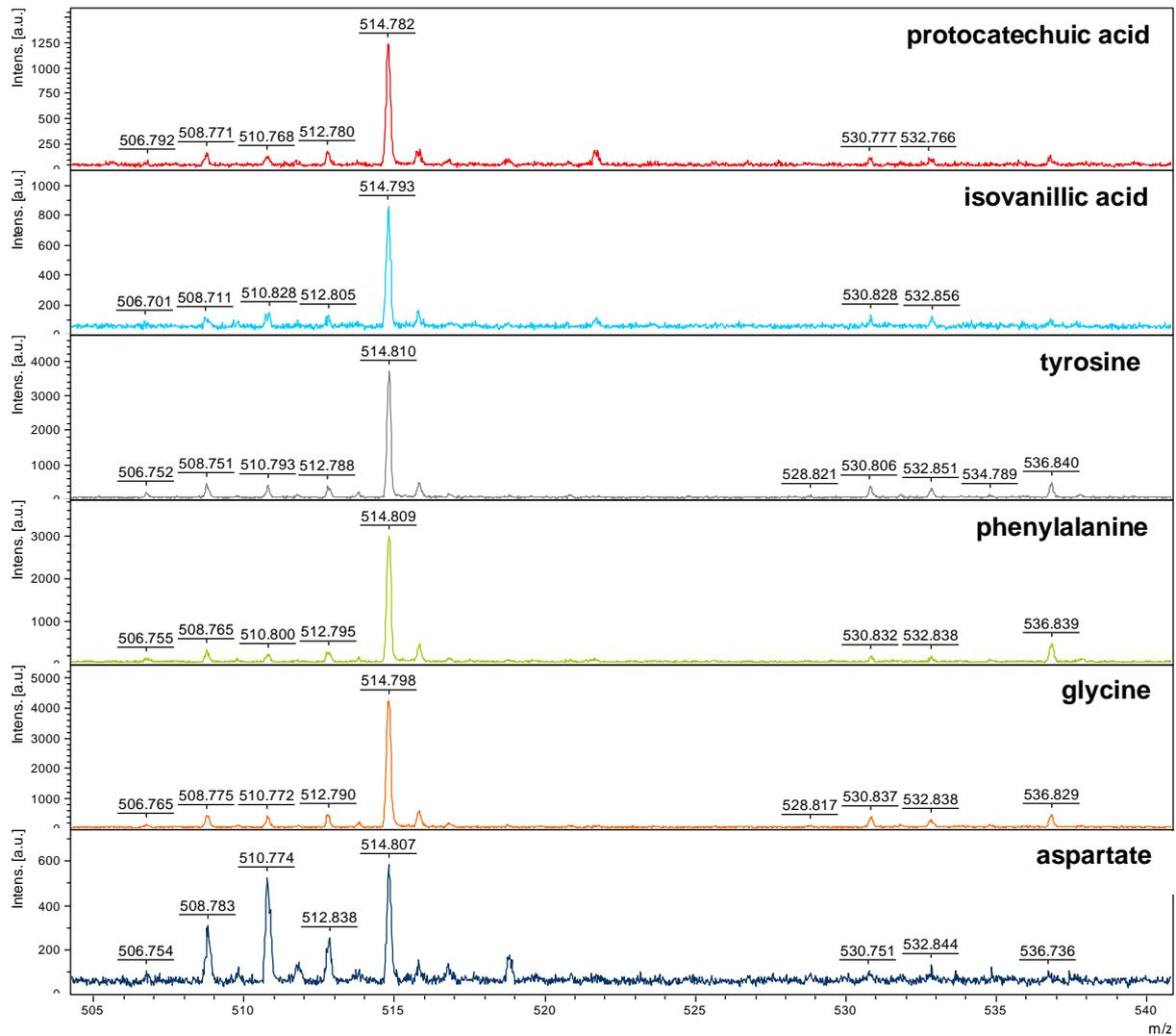


Figure 8-3: Chromatograms from MALDI-TOF-MS measurements

Marked are fully labeled ATP (m/z 514), unlabelled ATP (m/z 506), partially labelled ATP and monosodium-coordinated ions; the tested substrate is indicated at each chromatogram

Publications

Posters

Ö. Erol-Hollmann, T. Höver, M. Nett and G.M. König. „Siphonazole – a secondary metabolite from *Herpetosiphon*”. 6th European Conference on Marine Natural Products, Porto, Portugal, July 19 – 23, 2009

T. Höver, T.F. Schäberle, Ö. Erol-Hollmann, G.M. König. „Siphonazole – a secondary metabolite from *Herpetosiphon*”. International Workshop of the VAAM Section "Biology of Bacteria Producing Natural Compounds", Tübingen, Germany, September 26 – 28, 2010

T. Höver, T.F. Schäberle, Ö. Erol, G.M. König. „Siphonazole – a secondary metabolite from *Herpetosiphon* sp. 060”. International Meeting “NatPharma: Nature Aided Drug Discovery” (NADD), Naples, Italy, June 5 – 8, 2011

T. Höver, T.F. Schäberle, Ö. Erol, G.M. König. „Siphonazole – a secondary metabolite from *Herpetosiphon* sp. 060”. International Workshop of the VAAM Section "Biology of Bacteria Producing Natural Compounds", September 28 – 30, 2011

Oral Presentation

T. Höver, T.F. Schäberle, Ö. Erol, G.M. König. „Siphonazole – a secondary metabolite from *Herpetosiphon* sp. 060”. International Workshop of the VAAM Section "Biology of Bacteria Producing Natural Compounds", September 28 – 30, 2011