Biosynthesis of Phenylnannolone A, a MDR Reversal Agent from *Nannocystis pusilla*

DISSERTATION

zur

Erlangung des Doktorgrades (Dr. rer. nat)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhems-Universität Bonn

vorgelegt von

SARAH MUSLIMA BOUHIRED

aus

Donaueschingen

Bonn, Oktober 2012

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrichs-Wilhems-Universität Bonn

1.Gutachterin	:	Prof. Dr. Gabriele M. König
2.Gutachter	:	Prof. Dr. Werner Knöss
Tag der Promotion	:	18.02.2013
Erscheinungsjahr	:	2013

بسم الله الرحمن الرحيم

And over every lord of knowledge there is one more knowing.

Und über jedem, der Wissen besitzt, steht Einer, der (noch mehr) weiß.

Quran 12:76

Publications

Almeida C, Part N, Bouhired S, Kehraus S, and König GM, (2010) **Stachylines A−D from the Sponge-Derived Fungus *Stachylidium* sp.**, *Journal of Natural Products,* 74(1):21-25

Almeida C, Hemberger Y, Schmitt SM, Bouhired S, Natesan L, Kehraus S, Dimas K, Gütschow M, Bringmann G, König GM**,** (2012) **Marilines A-C, Novel Phtalimidines from the Sponge-derived Fungus *Stachylidium* sp**., *Chem. Eur. J.,* 18(28):8827-8834

Research presentation

Opening Symposium NRW International Graduate School Biotech-Pharma; February 2009, Poster presentation**: *Phenylnannolone A: The Biosynthesis of an MDR Reversal Agent***

6th ECMNP (European Conference on Marine Natural Products), 19th-23rd July 2009 in Porto, Portugal: Poster presentation: **Phenylnannolone A: The Biosynthesis of an MDR Reversal Agent,** abstract published in program & abstract book 6th ECMNP

International VAAM (Vereinigung für Angewandte und Allgemeine Mikrobiologie)-Workshop 2010 "Biology of Bacteria Producing Natural Products", 26th -28th September 2010 in Tübingen, Germany, Poster presentation and oral communication (3min): **Phenylnannolone A, a polyketide from *Nannocystis exedens,*** abstract published in program & abstract book International VAAM-Workshop 2010 "Biology of Bacteria Producing Natural Products"

International VAAM (Vereinigung für Angewandte und Allgemeine Mikrobiologie)-Workshop 2011 "Biology of Bacteria Producing Natural Products", 28th -30th September 2011 in Bonn, Germany, Poster presentation and oral communication (3min): **New Insights into the Pathway of Phenylnannolone A**, abstract published in program & abstract book International VAAM-Workshop 2011 "Biology of Bacteria Producing Natural Products"

International Symposium Biotech-Pharma, September 2011 in Bonn, Germany, Poster presentation: **New Insights into the Pathway of Phenylnannolone A**

Directing Biosynthesis III, 19th-21st September 2012 in Nottingham, England: Poster presentation: **Identification of the phenylnannolone A biosynthetic gene cluster,** abstract published in program & abstract book Directing Biosynthesis III

Oral communications

NRW International Graduate School Biotech-Pharma-Scientific colloquium winter 2009/2010
Biosynthesis of phenylnannolone A, 11.12.2009, Bonn

NRW International Graduate School Biotech-Pharma-Annual Retreat 2010
Biosynthesis of phenylnannolone A, 01.03.2010, Maria Laach

NRW International Graduate School Biotech-Pharma-Annual Retreat 2011
Investigations on the biosynthetic gene cluster of Phenylnannolone, 17.3.2011, Attendorn

NRW International Graduate School Biotech-Pharma-Scientific colloquium summer 2011
Assembly of biosynthetic genes from Nannocystis exedens, 29.03.2011, Bonn

Danksagung /Acknowledgments

Mein ganz besonderer Dank gilt meiner Doktormutter Frau Prof. Dr. Gabriele M. König für die herzliche Aufnahme in ihre Arbeitsgruppe und die Überlassung dieses interessanten Forschungsthemas, sowie für ihre Unterstützung und ihre jederzeits offene Tür.

Herrn Prof. Dr. Werner Knöss möchte ich herzlich dafür danken, dass er sich, trotz seiner zahlreichen Verpflichtungen im BfArM, die Zeit für mich genommen hat, das Zweitgutachten für diese Arbeit zu erstellen.

Den weiteren Mitgliedern der Prüfungskommission, Herrn Prof. Dr. Gerd Bendas und Frau Prof. Dr. Gabriele Bierbaum, danke ich für deren Teilnahme, sowie ihrem Interesse an meiner Arbeit.

Aus dem Arbeitskreis Piel möchte ich Dr. Christian Gurgui für seine guten Ratschläge bezüglich der Erstellung der Fosmidbank, und Dr. Max Crüsemann für seine Hilfe beim A domain Assay danken.

Für die Bereitstellung des Programms CLUSEAN und Hilfestellung in bioinformatischen Fragen möchte ich Dr. Tillman Weber sehr danken.

Frau Dr. E. Mies-Klomfass, Frau Dr. M. Koch und Frau E. Gassen danke ich für ihre Hilfe in administrativen Fragen. Herrn Thomas Koegler möchte ich für die immerbereite Unterstützung in computertechnischen Fragen danken.

Ich danke allen Mitgliedern des Arbeitskreises König für die angenehme Arbeitsatmosphäre und freundliche Aufnahme, im Besonderen möchte ich Edith Neu und Ekaterina Eguevera für Ihre ständige Hilfsbereitschaft in organisatorischen, technischen und zwischenmenschlichen Fragen danken. Mila Goralski danke ich für die technische Unterstützung im S1 Labor und ihre Hilfsbereitschaft. Den S1'lern sei an dieser Stelle für das nette Arbeitsklima und die gute Zusammenarbeit gedankt. Dr. Özlem Erol danke ich für die Einführung ins S1-Labor und ihre Freundschaft. Dr. Till Schäberle sei für die Betreuung im S1-Labor gedankt. Desweitern danke ich Dr. Stefan Kehraus für die nette und entspannte Atmosphäre, die er als Kursleiter

während der Betreuung der Praktikumskurse schaffte. Ebenso für sein offenes Ohr und seine Hilfsbereitschaft in jeglicher Angelegenheit.

Meinen Tischnachbarn in unserer „Denkzelle“ (Büro), Dr. Mustafa El Omari, Alexander Schmitz und Kirsten Knapp, möchte ich für die zahlreichen wissenschaftlichen und nicht-wissenschaftlichen Diskussionen, sowie für die tolle Arbeitsatmosphäre danken. Ganz besonders möchte ich Kirsten Knapp, die mir in allen Lebenslagen stets zur Seite stand, für ihre Freundschaft danken.

Einigen möchte ich an dieser Stelle noch in der ihnen jeweils verständlichen Sprache danken:

I’m very grateful for the help and advice concerning the protein expression, fruitful discussions and friendship, which I received from Lavanya Natesan. She was not only a good companion over all these years, but also became a good friend. I’m also very thankful for the friendship of Somia El Saedi, Mamoona Nazir and Fayrouz El Maddah and for the nice moments we shared. Anne-Robin Laaredj-Campell is an old friend of me, whom I thank for proof-reading parts of my thesis.

I’m very grateful for the financial support during the first three years of my thesis by the NRW International Research School Biotech Pharma. I appreciated a lot the possibility to attend diverse seminars and workshops provided by other work groups of the University of Bonn and the mutual exchange between these groups.

My aunts, Louisa und Houria Bouhired, I’d like to thank for their financial and moral support: „Merci beaucoup a votre générosité et pour votre soutien pendant mon doctorat, je vous remercie aussi pour votre amour et encouragement!“.

Meiner Schwester Sabira und meinen Eltern, Gabriele und Amine Bouhired, möchte ich für Ihre Unterstützung und Liebe danken:

شاكرة حسن تربيتكم لي على الاسلام وحب العلم فجزاكم الله عني خيراً

وكل الشكر لله رب العالمين الذي لا حول ولا قوة لي إلا به وأساله أن يتقبل مني هذا العمل

Table of Content

1 INTRODUCTION 1

1.1 Natural products in drug discovery 1

1.2 Myxobacteria: a unique source of natural products 2

1.3 Biosynthesis of myxobacterial secondary metabolites 9

1.3.1 Polyketide Synthases 10

1.3.2 Unusual starter units in PKSs 17

1.3.3 Non-Ribosomal Peptide Synthetases 20

1.3.4 PKS/NRPS hybrids 22

1.4 Phenylnannolone A: a polyketide from *Nannocystis pusilla* 24

2 SCOPE OF THE STUDY 27

3 MATERIALS AND METHODS 28

3.1 Materials 28

3.1.1 Chemicals and other materials 28

3.1.2 Enzymes 31

3.1.3 Molecular weight marker 32

3.1.4 Molecular biological kits 32

3.1.5 Bacterial strains 33

3.1.6 Vectors 34

3.1.7 Fosmids 34

3.1.8 Phages 35

3.1.9 Oligonucleotides 35

3.1.10 Water 36

3.1.11 Culture Media 37

3.1.12 Antibiotics 38

3.1.13 Buffers and solutions 39

3.1.14 Software and Databases 42

3.2 Molecular biological methods 45

3.2.1 Sterilization of solutions and equipment 45

3.2.2 Antibiotic selectivity test 45

3.2.3 Isolation of chromosomal DNA 46

3.2.4 16S rDNA analysis 46

3.2.5 Agarose gel electrophoresis 47

3.2.6 DNA recovery form agarose gels 47

3.2.7 Polymerase chain reaction 47
3.2.7.1 PCR parameters 48
3.2.7.2 PCR protocol 49
3.2.8 Ligation of PCR Products 49
3.2.9 Restriction digestion 50
3.2.10 Preparation of cells competent for DNA-transformation 50
3.2.11 Transformation of host strains 51
3.2.12 Plasmid isolation from transformed *E. coli* 51
3.2.13 Construction of a genomic library 52
3.2.13.1 LMP agarose gel electrophoresis 52
3.2.13.2 DNA extraction from LMP agarose 53
3.2.13.3 End-repair of size selected DNA 53
3.2.13.4 Ligation of end-repaired DNA into the pCC1Fos vector 54
3.2.13.5 Preparation of EPI300-T1R competent *E. coli* cells 55
3.2.13.6 Packaging of fosmid clones 55
3.2.13.7 Plating the genomic library 56
3.2.14 Screening of the fosmid gene library 57
3.2.15 Induction and isolation of identified positive clones 58
3.2.16 Genome sequencing using the 454 sequencing procedure 58
3.2.17 Isolation, cultivation and long term storage of myxobacteria 58
3.2.18 Cultivation and long term storage of recombinant *E. coli* strains 59
3.2.19 Determination of bacterial cell density 59
3.2.20 Determination of nucleic acid concentration and purity of DNA 59
3.2.21 Heterologous expression of proteins 60
3.2.22 Cell lysis by sonication 61
3.2.23 Purification of recombinant proteins by Ni-NTA-columns 61
3.2.24 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) 61
3.2.24.1 Coomassie-staining 62
3.2.25 Concentration of purified proteins and buffer exchange 63
3.2.26 ATP- PP_i exchange assay 63

4 RESULTS 64

4.1 Isolation and cultivation of the myxobacterial strain 150 64

4.2 Isolation of genomic DNA from myxobacterial strain 150 64

4.3 Strain identification of myxobacterial strain 150 65

4.4 Fosmid library production 65

4.5 Genome sequencing 68

4.6 Screening for NRPS/PKS encoding genes in the fosmid library 70

4.7 Fosmid sequencing 72

4.7.1 End sequencing of fosmids 72

4.7.1.1 Comparison of fosmids 11A3 and 21H12 73

4.7.2 Complete sequencing of fosmid 21H12 and 12A9 74

4.8 Elucidation of the putative phenylnannolone A biosynthetic gene cluster 74

4.8.1 Sequencing results for fosmid 12A9 74

4.8.2 Phn1, a Carboxyl Transferase 76

4.8.3 A Polyketide Synthase assembly line encoded by *phn2* 77

4.8.3.1 AMP-Ligase 79

4.8.3.2 Acyl Carrier Proteins (ACPs) 81

4.8.3.3 Ketosynthase domains (KS) 83

4.8.3.4 Acyltransferase domains 85

4.8.3.5 Ketoreductase domains 90

4.8.3.6 Dehydratase domains 93

4.8.3.7 Thioesterase domain 95

4.9 The biosynthetic NRPS/PKS-type gene *sb1* 96

4.9.1 Sequencing result for the fosmid 21H12 96

4.9.2 Analysis of *sb1*, a NRPS/PKS mixed biosynthetic gene 97

4.9.2.1 Condensation domains 99

4.9.2.2 Adenylation domains from Sb1 101

4.9.2.3 PCP domains 104

4.9.2.4 Ketosynthase domain (KS) 106

4.9.2.5 Acyltransferase (AT) 107

4.9.2.6 Thioesterase domain 109

4.10 Antibiotic selectivity test for *N. pusilla* B150 111

4.11 Functional proof of the AMP-ligase from the PKS gene, *phn2* 112

4.11.1 Protein expression of the AMP-ligase from Phn2 112

4.11.2 ATP exchange assay 113

5 DISCUSSION 115

5.1 Construction of the genomic library 115

5.2 Sequencing of DNA from *N. pusilla* B150 117

5.3 The biosynthetic gene cluster for the phenylnannolones 119
5.3.1 Precursor supply - ethylmalonyl-CoA formation 120
5.3.2 AMP-ligase activating cinnamic acid (starter unit) 126
5.3.3 AT_1 domain loading ethylmalonyl-CoA 131
5.3.4 Stereochemistry of double bonds 133
5.3.5 Lactone ring formation through TE 137

5.4 A mixed NRPS/PKS biosynthetic gene cluster 139

6 SUMMARY 141

7 REFERENCES 144

8 APPENDIX 154

8.1 Amino acid sequences for the phenylnannolone A gene cluster 154
8.1.1 Amino acid sequence of Phn1, the putative butyryl-CoA Carboxylase 154
8.1.2 Amino acid sequence of the Phn2, the putative PKS assembly line 154

8.2 Amino acid sequences for the mixed NRPS/PKS protein Sb1 163

8.3 Sequences obtained from end-sequencing of fosmid clones 165

8.4 16S rDNA sequence 168

8.5 Mass spectrometry sprectra of the A domain assay 169

8.6 Substrate predictions of some domains by bioinformatical online tools 170
8.6.1 Prediction for the substrate specificity of the AMP-ligase from Phn2 170
8.6.2 Substrate prediction for the AT1 domain from Phn2 170
8.6.3 Substrate prediction for the A domains of Sb1 the NRPS/PKS hybrid 171

Abbreviations

ACP	Acyl carrier protein
A domain	Adenylation domain
AT	Acyl transferase
BLAST	Basic local alignment search tool
bp	Base pairs
C domain	Condensation domain
CoA	Coenzyme A
Da	Dalton
DH	Dehydratase
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	N,N,N´,N´-Ethylendiamintetraacetat
ER	Enoylreductase
IPTG	Isopropyl-β-D-thiogalactosid
kb	Kilo base pairs (= kbp = 1,000 bp)
kDa	Kilo Dalton
KR	Ketoreductase
KS	Ketosynthase
MALDI	Matrix-assisted laser desorption/ionisation
NADH	Nicotinamide adenine dinucleotide
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PK	Polyketide
PKS	Polyketide Synthase
PPi	Inorganic pyrophosphate
NRPS	Non-ribosomal Peptide Synthetase
rDNA	Ribosomal DNA
RNase	Ribonuclease
SDS	Sodium dodecylsulfate
sp./spp.	Species /species (plural)
T domain	Thiolation domain
TE	Thioesterase
TEMED	N,N,N’,N'-Tetramethylethylendiamin

Abrreviations for amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Gln	E
Glutamine	Glu	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tyrosine	Tyr	Y
Tryptophan	Trp	W
Valine	Val	V

1 Introduction

1.1 Natural products in drug discovery

Numerous natural products, mostly of plant origin, are applied in medicine, such as morphine (*Papaver somniferum*), digitoxine (*Digitalis lanata*), salicin derivatives (*Salix alba*) and taxol (*Taxus brevifolia*). Additionally, the development of new drugs often relies on natural products as leading structures. This is corroborated by the vast number (over 50%) of clinically used drugs that are of natural origin (Newman *et al*., 2003, Butler *et al*., 2004).

Plants and microorganisms provide an immense reservoir of chemically diverse natural products with potent biological activity. Some of these are of utmost medical importance, e.g. lovastatin (*Aspergillus nidulans*), penicillin (*Penicillum chrysogenum*) and daunorubicin (*Streptomyces peucetius*) showing blood pressure lowering, antibiotic and anticancer activity. Biosynthetically, prominent classes of microbial secondary metabolites are the polyketides (PKs) and non-ribosomal peptides (NRPs).

With more than 8000 compounds characterized to date, the members of the Actinomycetales are the declared star producers of secondary metabolites. However, in the last two decades the myxobacteria attracted much attention as an alternative source, as they are promising producers of compounds with both, unique structures and bioactivities (Weissman and Müller, 2009).

1.2 Myxobacteria: a unique source of natural products

In 1892, Roland Thaxter discovered myxobacteria, a fascinating group of microorganisms with remarkable features. Social behavior, gliding on solid surfaces and formation of fruiting bodies under starvation are some of their noticeable features. Eponymous was the immense amount of polysaccharides they excrete, which is retrieved in “myxo”, a word derived from the Greek “myxis” and means slime or mucus (Dworkin, 2007). They are also known to posses the largest genomes amongst bacteria (9-13 Mbp) with a DNA of high GC content of 66-72 mol% (Reichenbach, 1999).

Myxobacteria belong to the δ-subgroup of the gram-negative proteobacteria and are united in the order of *Myxococcales*, which is further divided into three suborders: *Cystobacterineae*, *Sorangiineae* and *Nannocystineae.* Soil, the bark of trees, decaying plant materials and the dung of herbivores are all habitats in which these bacteria occur. But beside the terrestial myxobacteria, there are also a few of marine origin mostly living in costal areas (Reichenbach, 1999; Iizuka et al., 2006).

Almost all discovered natural products of myxobacterial origin have unique structures that were not found before in other bacteria. The Helmholtz Centre for Infection Research (Braunschweig, Germany) has isolated more than 7500 strains from the order of *Myxococcales.* Although there are more actinomycetes known to date than myxobacteria, the number of myxobacterial natural products is enormous as they yielded at least 100 natural product core structures and over 500 derivatives (Garcia *et al*., 2009). The biosynthetic capabilities of these microorganisms are tremendous. Moreover, many of these natural products have unique modes of action. The therapeutic spectra of myxobacterial natural products include antibacterial, antifungal, antiplasmodial and antitumor activity.

The most prominent myxobacterial compounds are the epothilones from *Sorangium cellulosum* which possess antitumor activity. Epothilones’ mode of action is comparable to that of paclitaxel (Taxol®), where ß-tubulin is bound promoting tubulin polimerisation and hence, freezes microtubule dynamics in dividing cells. As cancer cells highly depend on cell division, which requires a functioning microtubule, the tubulin system represents an attractive target to defeat cancer cells (Bollag *et al*., 1995, Goodin *et al*., 2004, Reichenbach and Höfle, 2008). The semisynthetic

epothilone B derivative ixabepilone (Ixempra®) is since 2007 available in the United States as a drug for breast cancer treatment (Conlin *et al*., 2007). Other myxobacterial secondary metabolites that interact with the cytoskeleton are tubulysin and disorazol, destabilizing the tubulin target, and rhizopodin that inhibits actin polymerization (Khalil *et al*., 2006, Elnakady *et al*., 2004, Hagelueken *et al*., 2009).

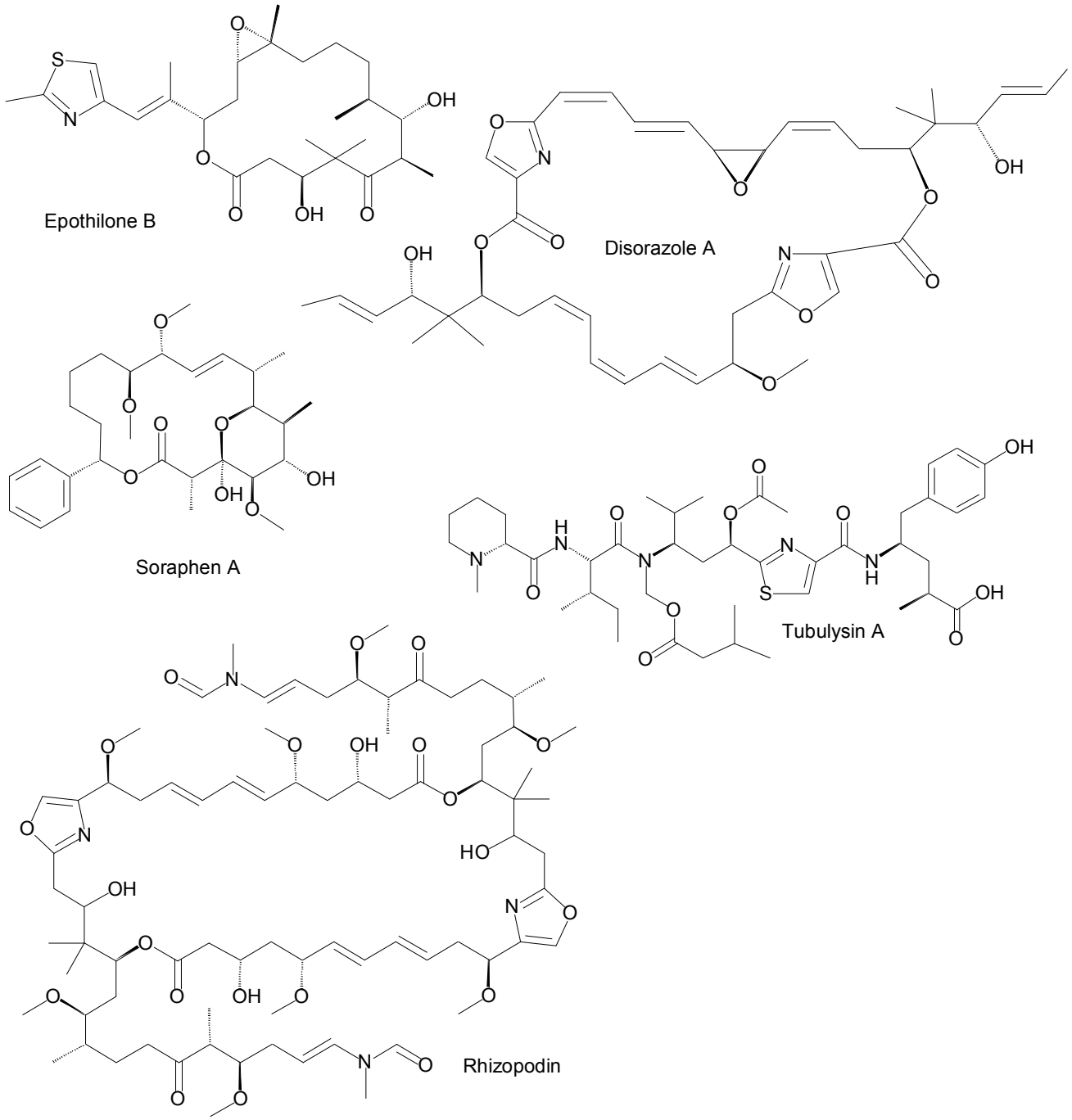


Figure 1.1 Myxobacterial compounds with antitumor activity: Rhizopodin from *Stigmatella aurantiaca*, tubulysin from *Archangium geophyra,* soraphen A, disorazole A and eopothilone B from *Sorangium cellulosum.*

Around 30 % of myxobacterial compounds show antibacterial activity, such as corallopyronin, etnangien, myxopyronin, ripostatin and sorangicin, all inhibiting RNA-

polymerase (Gerth *et al*., 2003; Weissman and Müller, 2010, Erol *et al*., 2010). Corallopyronin was recently tested *in vivo* in mice and has shown antibacterial activity against *Wolbachia*, endobacteria in filarial nematodes that cause lymphatic filariasis. It is a promising alternative to the standard treatment with doxycycline or rifampcin (Schiefer *et al.*, 2012).

Thuggacin interferes with the cellular electron-transport chain in the respiraton of *Mycobacterium tuberculosis*, the bacterium that causes tuberculosis (Steinmetz et al., 2007). Inhibitors of the protein biosynthesis are the compounds angiolam, althiomycin and myxovalargin. Althiomycin interacts within the peptidyltransfer, whereas myxovalargin disrupts the binding of the aminoacly-tRNA. An additional effect of myxovalarginis is that it causes damages on the cell membrane. The cell wall biosynthesis is the target of the antiobiotic myxovirescin, where it inhibits the incorporation of *N*-acetyl-glucosamine (Weissman and Müller, 2010; Gerth et al., 1982).

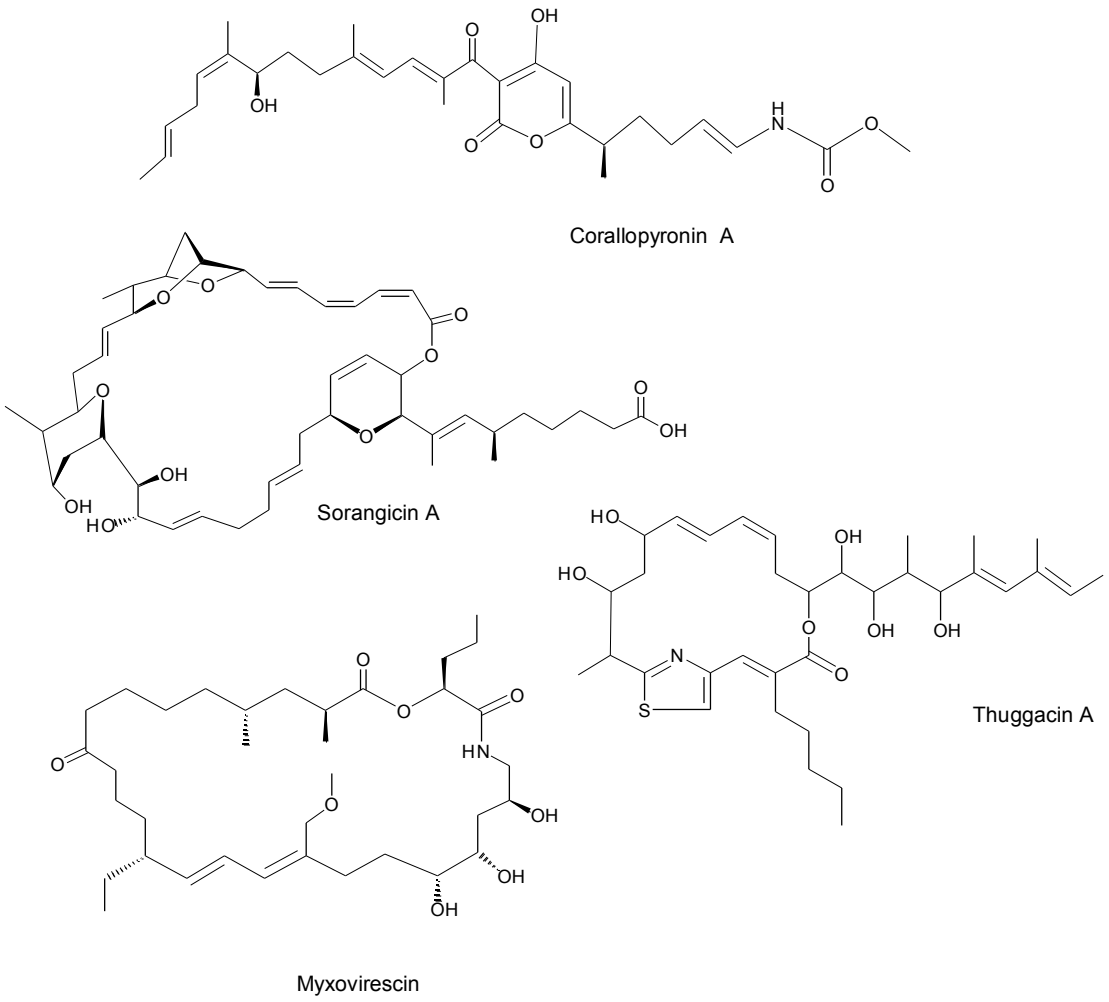


Figure 1.2 Myxobacterial compounds with antibacterial activity: Corallopyronin from *Corallococcus coralloides*, myxovirescin from *Myxococcus virescens*, thuggacin A and sornagicin A from *Sorangium cellulosum.*

Antifungal activity was reported for more than 50% of the myxobacterial metabolites and is the most frequent observed bioactivity of these compounds. The mitochondrial respiratory chain is the main target of most of these natural products, among them are stigmatellin, myxothiazol, haliangicin and myxalamid. Other targets for antifungal NP from myxobacteria are the cell membrane integrity observed for ambruticin, jerangolid and pedein, or the nucleic acid and protein biosynthesis as observed for leupyrrin (Weissman and Müller, 2010; Baker and Alvi, 2004).

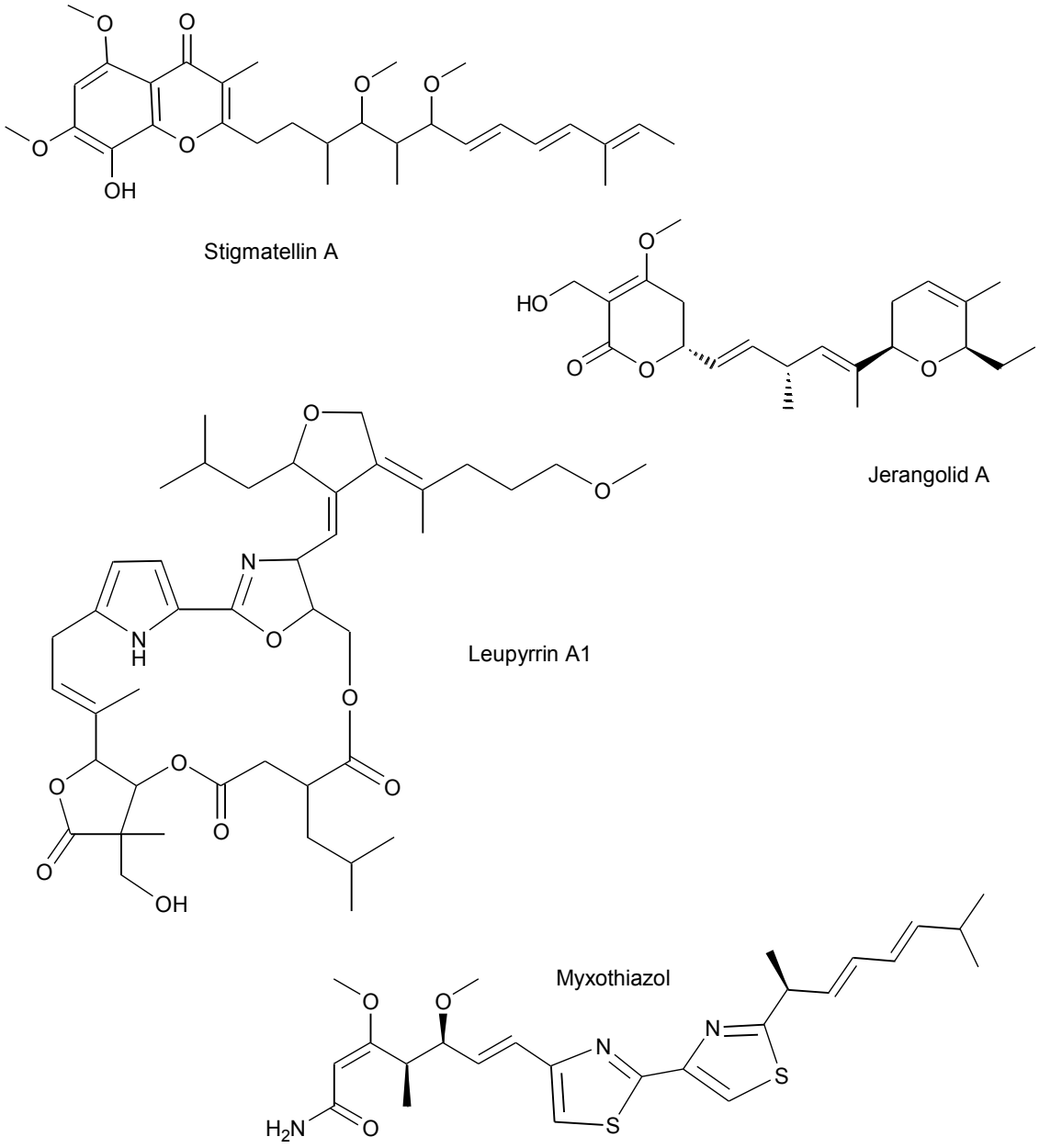


Figure 1.3 Myxobacterial compounds with antifungal activity: Myxothiazol from *Myxococcus Xanthus,* leupyrrin A1 and jerangolid from *Sorangium cellulosum,* stigamtellin A from *Stigmatella aurantiaca.*

The immense diverse spectrum of myxobacterial compounds makes myxobacteria an important source of novel classes of secondary metabolites (Gerth *et al*., 2003). The vast majority of these diverse natural products are PKs and NRPs, whereas more than half of the isolated myxobacterial compounds contain elements of both, PKs and NRPs, and are therefore denoted as hybrid PK/NRP metabolites. Other bacterial producers of secondary metabolites, like the actinomycetes, synthesize prevalently pure PK or NRP compounds (Weissman and Müller, 2009).

Table 1.1 Myxobacterial secondary metabolites and their corresponding type of gene cluster

Compound	Type of gene cluster
Ajudazol	PKS/NRPS
Althiomycin	PKS/NRPS
Ambruticin	PKS/NRPS
Aurachin	PKS
Aurafuron	PKS
Chivosazol	PKS/NRPS
Chondramid	PKS/NRPS
Chondrochloren	PKS/NRPS
Corallopyronin	PKS/NRPS
Crocacin	PKS/NRPS
Cystothiazol	PKS/NRPS
Disorazol	PKS/NRPS
DKXanthen	PKS/NRPS
Epothilon	PKS/NRPS
Etnangien	PKS
Jerangolid	PKS
Leupyrrin	PKS/NRPS
Melithiazol	PKS/NRPS
Myxalamid	PKS/NRPS
Myxochelin	NRPS
Myxochromide S	PKS/NRPS
Myxothiazol	PKS/NRPS
Myxovirescin	PKS/NRPS
Phenylnannolone A	PKS
Rhizopodin	PKS/NRPS
Saframycin	NRPS
Sorangicin	PKS
Soraphen	PKS
Spirangien	PKS
Stigmatellin	PKS
Thuggacin	PKS/NRPS
Tubulysin	PKS/NRPS

The structures for over 60 myxobacterial compounds were already published, and for half of these compounds the corresponding gene clusters were elucidated (table 1.1). Among them are eight gene clusters encoding for a polyketide synthase (PKS) and two gene clusters encoding for a non-ribosomal peptide synthetase (NRPS), the vast majority however, encodes for a PKS/NRPS hybrid.

These gene clusters have been identified either by screening of genomic libraries, transposon-mutagenesis or genome sequencing. Genome sequencing of *Myxococcus xanthus*, *Sorangium cellulosum* and *Stigmatella aurantiaca* revealed

that the number of gene clusters for secondary metabolites exceeds by far the number of identified metabolites. In fact, 8-9% of the genomes of *S. cellulosum* and *M. xanthus* encode secondary metabolites. This is twice the percentage found in the streptomycete *S. coelicolor* (Dworkin, 2007), which indicates that the potential of myxobacteria is enormous and far from utilized.

1.3 Biosynthesis of myxobacterial secondary metabolites

Most of the structurally diverse myxobacterial natural products are synthesized by large multienzymes, i.e. polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) or hybrids of both (Staunton and Weissman, 2001; Wenzel and Müller, 2009). The genes encoding for these biosynthetic machineries are usually clustered in microbial genomes, which simplifies the identification and elucidation of biosynthetic gene clusters. For the majority of myxobacterial PKSs and NRPSs a modular organization exists, where each module is responsible for the incorporation of one building block into the growing product chain. Each module is further subdivided into domains, enzymatic units that are responsible for loading, condensation and further modification of the extender unit. This one-to-one correspondence between the present domains and the biosynthetic transformation that takes place is termed “rule of colinearity” (Staunton and Weissman, 2001, Müller, 2004, Buntin, 2010).

During the assembly process, the substrates and intermediates of the natural product are covalently tethered to the carrier protein (CP) of the respective module through a thioester-linkage. For a functional enzymatic assembly line, the CPs have to be post-translationally activated by a phosphopantetheinyl transferase (PPTase). PPTases transfer the 4’-phosphopantetheine moiety of coenzyme A (CoA) to the conserved serine residue of the inactive *apo*-CP, converting it to the active *holo* form (figure 1.4). Substrates and intermediates are bound to the terminal thiol group of the phosphopantetheine arm, activating them for the condensation reaction. With their long flexible phosphopantetheine arm, *holo*-CPs hand round the intermediates between the individual catalytic domains (Fischbach and Walsh, 2006, Byers and Gong, 2007).

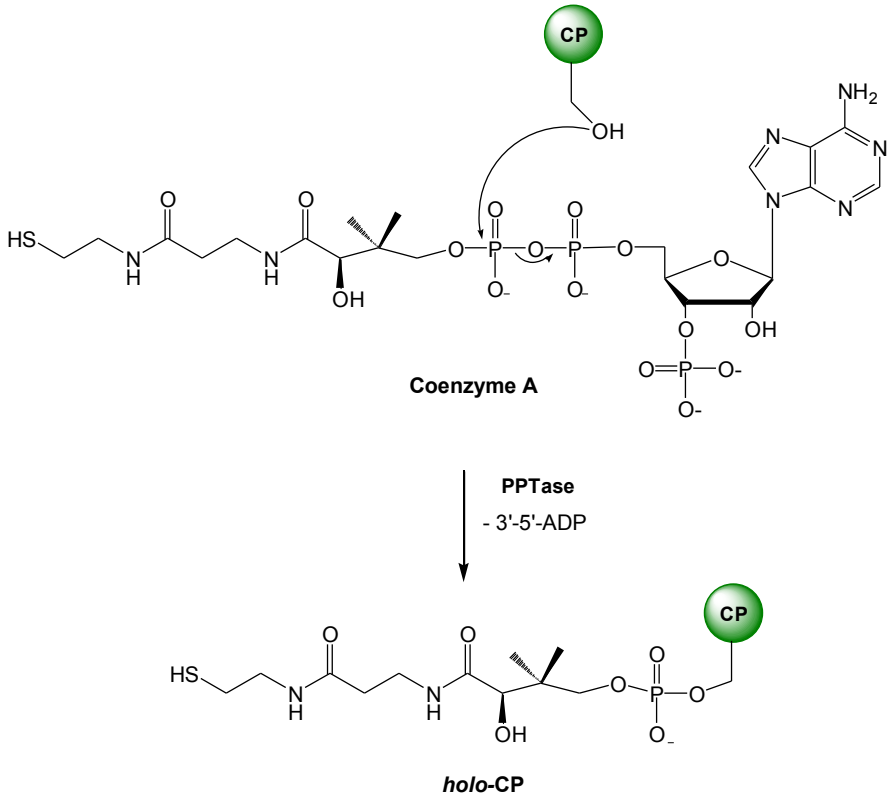


Figure 1.4 Post-translational modification of a CP domain by a PPTase: The PPTase catalyzes the transfer of phosphopantetheine from coenzyme A to a conserved serine in the CP domain, converting the inactive *apo* form to the active *holo*-CP.

1.3.1 Polyketide Synthases

Polyketides are a diverse class of natural products with potent biological activity such as antibacterial, antifungal, anticholesterol, antiparasitic, anticancer and immunosuppressive properties. These structurally diverse natural products are all built by polyketide synthases (PKSs), multienzymes that are found in plants, fungi and bacteria. PKS are classified into various types, based on the architecture and mode of action of their assembly lines (Hertweck, 2009). Initially, they were divided into three major groups: type I, type II and type III PKS. The classification as type I and type II refers to that of the previously classified enzymes of fatty acid biosynthesis (Weissmann, 2009). Type I applies to linearly arranged catalytic domains within large multifunctional enzymes, whereas a dissociable complex of discrete monofunctional enzymes is found in type II PKS. The third group characterizes multifunctional enzymes of the chalcone synthase (CHS) type, mainly found in plants, but sometimes also in bacteria and fungi. Besides the enzyme structures, the mechanism of synthesis, is another characteristic to classify PKS

(table 1.2). Depending on whether a module is used only once or repeatedly, PKS are termed as either modular (non-iterative) or iterative (Hertweck, 2009).

Table 1.2 Overview on the different types of PKSs (adapted from Hertweck, 2009 and Watanabe and Ebizuka, 2004) KS=ketosynthase, AT=acyltransferase, KR=ketoreductase, DH= dehydratase, ER=enoylreductase, ACP=acyl carrier protein, TE=thioesterase

Type	Protein structure	Synthesis mechanism	Found in	Extender units	Domains
PKS I	*Single* protein with ***multiple modules***	**Modular**	Bacteria	Various	**KS, AT, KR, DH, ER, ACP, TE**
PKS I	*Single* protein with ***one module***	**Iterative**	Mainly fungi	Malonyl-CoA	**KS, AT, KR, DH, ER, ACP, TE**
PKS II	*Multiple* proteins each with ***mono-functional*** active site	**Iterative**	Bacteria	Malonyl-CoA	**KS, CLF, ACP, KR, ARO, CYC**
PKS III	*One* protein with ***multiple modules***	**Iterative**	Mainly plants, some bacteria & fungi	Acyl-CoA, Malonyl-CoA	**KS, CHS/ STS**

In bacterial modular type I PKS the principle of colinearity allows the deduction of extension cycles from the number of modules, and can be used for the prediction of the metabolite's basic backbone structure. Conversely, a "retrobiosynthetic analysis" of the compound's structure enables the prediction of the domains involved in its biosynthesis. There are some exceptions, where single modules are used more than once or even are skipped. In *trans*-AT PKS systems, the modules lack the individual AT domain (Piel, 2002, Hertweck, 2009). For the iterative type I PKS, typical for fungi, the number of extension cycles is not predictable from the domain architecture of the protein. Although there is only one module that is used repeatedly, the degree of reduction may alternate as the corresponding domains are used variably.

Biochemistry in modular type I PKS

A minimal elongation module in PKS consists of three essential core domains: a ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP). Initially the AT domains select their specific monomeric units and load them to the corresponding ACP domains. The ACP-bound intermediate from the upstream module is transferred to a conserved serine residue of the KS domain. The KS domain catalyzes the decarboxylative condensation, here the ACP-bound extender unit is decarboxylated and the resulting enolate attacks the upstream KS-bound acyl thioester, forming a C-C bond between the extender unit and the growing polyketide chain (figure 1.5). By this principle the growing chain is passed from module to module.

In respect to the minimal PKS module, the architecture of loading modules can deviate. Most loading modules harbor a KS_Q domain, that decarboxylates the starter unit, but lacks the condensation activity, due to an alteration in the active site. The active cysteine is in this case replaced by a glutamine (Q). Another type of loading modules comprise a didomain, consisting of an AT and an ACP domain, where the AT selects a short chain monocarboxylic acid, as observed for the erythromycin biosynthesis (Bisang et al., 1999, Staunton and Wilkinson, 1997, Hertweck, 2009). For a detailed introduction to diverse starter units and loading modules see also section 1.3.2.

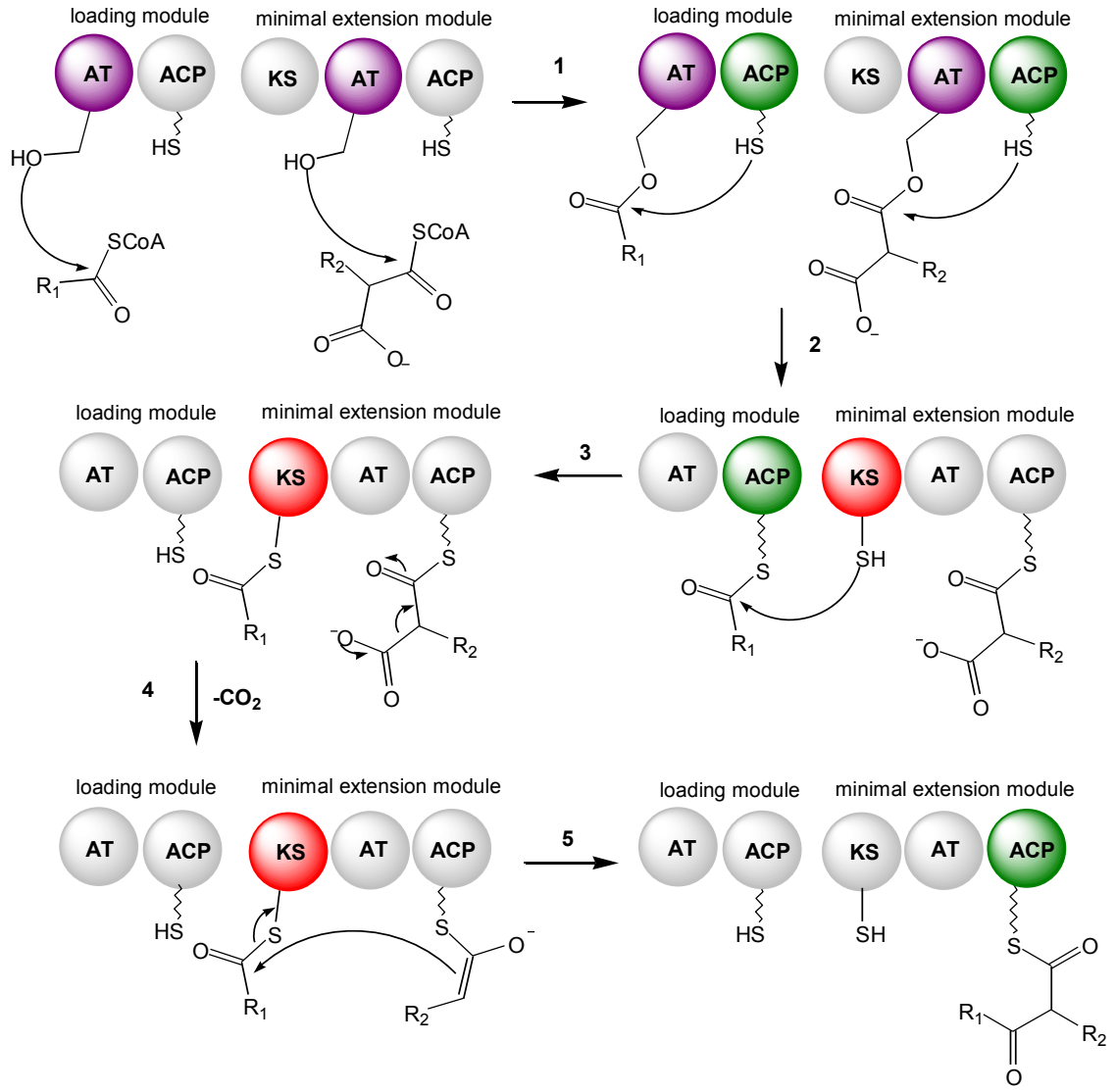


Figure 1.5 Schematic overview of PKS biochemistry: In the first step the AT domains from the loading module and first extension module select their respective acyl-CoA monomers and catalyze in the next step the transthiolation of these substrates to the neighboring ACP domains. The intermediate from the loading module is transferred to the KS of the elongation module (step **3**); the subsequent decarboxylative condensation is catalyzed by the KS domain, forming a C-C bond between the upstream acyl thioester and the downstream enolate (adapted from Buntin, 2010)

Extender units for modular type I PKS may vary; most frequent of all are malonyl-CoA and methylmalonyl-CoA. Less common are extender units such as ethylmalonyl-CoA and chloroethylmalonyl-CoA or the ACP-linked extender units: methoxymalonyl-ACP, hydroxymalonyl-ACP and aminomalonyl-ACP (Chan *et al*., 2009). AT domains in extension modules show high selectivity towards a certain substrate. Sequence analysis of numerous AT domains revealed conserved amino acids motifs that can be correlated to substrate specificity (Yadav *et al*., 2003, Tsai and Smith, 2007).

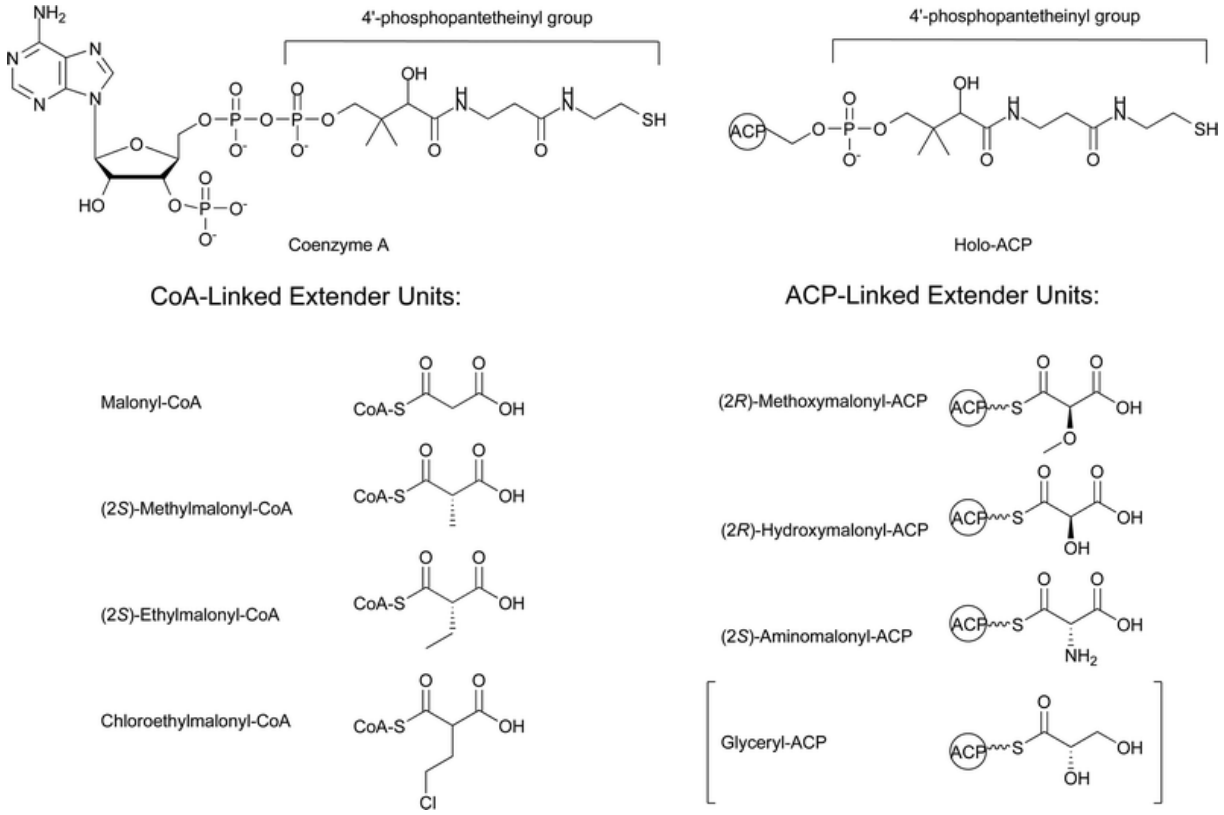


Figure 1.6 Structures of coenzyme A, holo-ACP, and the associated extender units. The brackets around glyceryl-ACP denote that it is not a classic extender unit involved in decarboxylative Claisen condensation reactions (figure origin: Chan *et al*., 2009)

For more structural diversity, a PKS extension module can be further equipped with modifying domains, such as ketoreductase (KR), dehydratase (DH) and enoylreductase (ER). The first reduction step is initiated by the KR domain that catalyzes the stereospecific and NADPH-dependent reduction of the ß-keto function to a ß-hydroxyl-moiety, which is followed by water elimination induced by the DH domain. The so obtained double bond is fully reduced to a saturated acyl chain by the presence of an ER domain. Whereas in fatty acid biosynthesis the full set of reduction steps takes place, in PKS biosynthesis different grades of reduction can occur, depending on the individual domains that are used.

Distinct motifs in KR and ER domains were identified to determine the final stereochemistry of the respective chiral centre, making the prediction of the absolute configuration for a particular metabolite from the related gene sequence possible (Caffrey, 2003, Kwan et al., 2008, Kwan and Schulz, 2011).

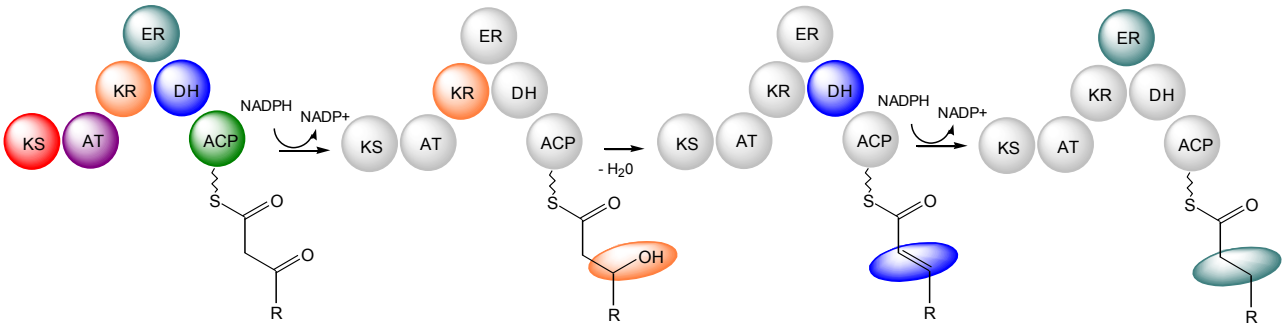


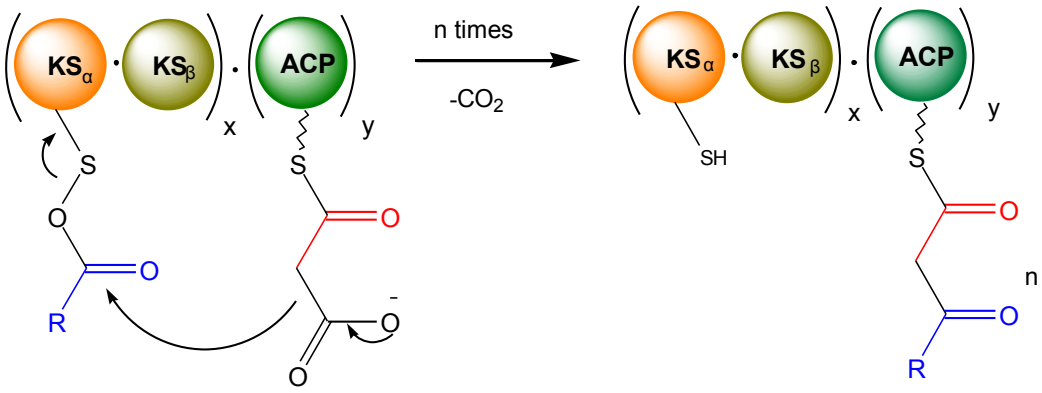
Figure 1.7 Reductive processing of ß-keto-function induced by KR, DH and ER domains. Reductive changes on the keto-group are highlighted, the respective color corresponds to the enzyme, responsible for the functionality. KS=ketosynthase, AT=acyltransferase, KR=ketoreductase, ER=enoylreductase, DH=dehydratase, ACP= acyl carrier protein

PKS biosynthesis is terminated in the last module of the assembly line by a thioesterase (TE) domain that releases the polyketide chain from the last ACP domain either by hydrolysis or cyclization (Du and Lou, 2009).

Iterative type II PKS

The iterative type II PKS consist of a set of discrete and usually monofunctional enzymes, froming a multienzyme complex. PKS II are known to produce aromatic polyketides in bacteria. This system is restricted to prokaryotes, thus fungal aromatic polyketides are generated by iterative type I PKS. A “minimal PKS” in type II PKS comprises a KS (KS_α) a chain length factor (CLF or KS_β) and an ACP. Accessory enzymes such as KR, cyclases (CYC) and aromatases (ARO) are responsible for further modifications defining the molecule’s structure (Hopwood, 1997; Hertweck, 2009; Tsai and Ames, 2009).

Type II PKS



Type III PKS

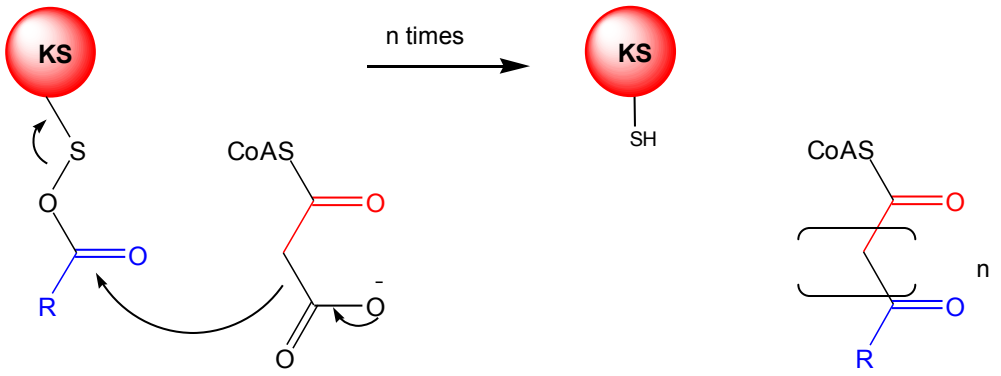


Figure 1.8 Polyketide biosynthesis in type II and type III PKS (adapted from Kira Weissman, 2009). For the iterative type II PKS the condensation reaction is catalyzed by KS_{α}, whereas chain elongation factor KS_{β} acts as a decarboxylase. The ACP independent, iterative type III PKS acts directly on the CoA-bound substrate.

Type III PKS

Type III PKS are remarkable multifunctional enzymes that complete with a single active site the entire biosynthesis, including decarboxylation, condensation and aromatization reactions. The homodimeric enzymes, which are also known as chalcone/stilbene synthases, elongate the PK chain in an ACP-independent way, assembling acyl-CoA bound extender units. This type of PKS was for long time believed to be restricted to plants, but recent studies discovered type III PKS also to be present in some bacteria and fungi (Hertweck, 2009).

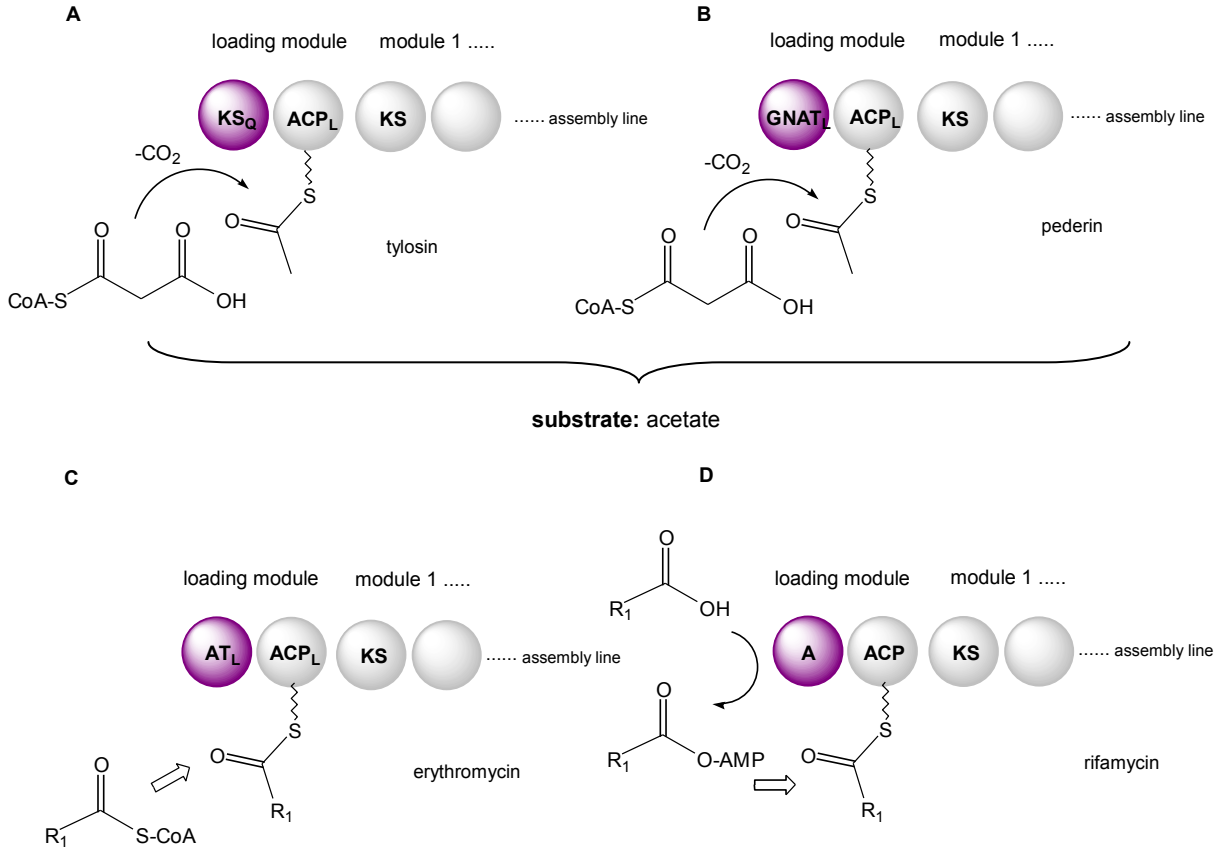
1.3.2 Unusual starter units in PKSs

Loading modules of type I PKS that utilize the common starter units: acetate or propionate, are divided into two major groups. The first type harbors a modified ketosynthase (KS_Q) domain, an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain. This KS_Q domain is responsible for the decarboxylation of the dicarboxylated starter units, e.g. malonyl-CoA, which is loaded by the AT domain to the neighbored ACP domain. However, this modified KS lacks the condensation activity, due to an alteration in the active site, where a glutamine replaces the active cysteine. The second group is characterized by a loading AT (AT_L) domain and a loading ACP (ACP_L) domain. This loading didomain loads short-chain monocarboxylic acids, such as propionyl-CoA, as observed for the erythromycin biosynthesis (Moore and Hertweck, 2001). Recently, a third alternate strategy for the priming of acetate was reported by the Piel group, who observed a GCN5-related *N*-acetyltransferase (GNAT) domain in the context of pederin biosynthesis (Piel *et al.*, 2004).

Beside acetyl-CoA, there are numerous alternate starter units employed by PKS, and various strategies for their activation and loading are employed. Loading modules in type I PKSs, which utilize other starter units than acetyl-CoA, attach their specific primers generally in two different ways, based on the nature of the respective carboxyl group. Either CoA thioesters are used as starters and loaded to their corresponding PKSs by the erythromycin type loading module, the AT_L-ACP_L didomain. Or free carboxylic acids are activated and loaded by a NRPS-like adenylation-thiolation (A-ACP) didomain (figure 1.9) (Moore and Hertweck, 2002).

A broad number of structurally diverse substrates are loaded in the fashion of the erythromycin loading didomain. Among them are substrates such as cylcohexanoyl-CoA, isovaleryl-CoA and benzoyl-CoA, used for the priming of biosynthetic enzymes for molecules such as phoslactomycin, myxothiazole and sorpahen. For the priming of myxothiazole and soraphen the architecture of the loading modules deviates, as they harbor an additional AT domain, which is located adjacent to the other AT domain (Palaniappan *et al.*, 2003, Silakowski *et al.*, 1999, Wilkinson *et al.*, 2001). Another deviation was observed for aureothin PKS from *Streptomyces thioluteus*,

where a single ACP domain is employed for the priming of *p*-nitrobenzoyl-CoA (He and Hertweck, 2003).



substrates: propionyl-CoA, cyclohexanoyl-CoA, isovaleryl-CoA, isobutyryl-CoA, 2-methylbutyryl-CoA, benzoyl-CoA, trans-cyclohexane-1,2-dicarboxyl-CoA

substrates: p-aminobenzoate (PABA), 3-amino-hydroybenzoate(AHBA), dihydroxy-cyclohexene carboxylic acid, 3-5-dihdryoxybenzoate (DHBA), phenylacetate

Figure 1.9 Loading modules in modular type I PKS. A: KS$_Q$-AT-ACP loading domain loads malonyl-CoA, which is decarboxylated, and an acetate unit is attached. This type of loading module is employed in many PKS, i.e. in the formarion of tylosin. B: Priming of acetate by the use of an GNAT domain firstly described for pederin biosynthesis. C: The erythromycin-type loading didomain comprising an AT and an ACP domain loads CoA-thioesters. D: NRPS-like adenylation-thiolation loading didomains load free acids as described for the rifampicin biosynthesis. Possible substrates for the different types of loading modules described for numerous polyketides are listed under each figure.

The rifamycin and candicidin PKSs, each employ a NRPS-like (A-ACP) loading didomain for the priming of 3-amino-5-hydroxybenzoic acid (AHBA) and *p*-aminobenzoate (PABA). A number of polyketides, such as ripostatin A, microcystin-LR, nodularin and cryptophycin 1 load a C_6C_2-phenylalanine derived starter unit, employing an A-ACP didomain in the loading step. The C_6C_2-starter was believed to originate from phenylacetate, but this assumption could not be corroborated by ^{13}C-labeled feeding studies, as it was not incorporated into the structures of microcystin-

LR, nodularin and cryptophycin 1. This indicates that phenylacetate is not a free intermediate and another substrate is probably loaded onto the A-ACP loading didomain. Thus, Hicks and co-workers performed studies on the loading didomain involved in the formation of microcystin-LR, the results of this study are discussed in section 5.3.2. A deviation of the A-ACP loading module was observed for the attachment of dihydrocyclohexene carboxylic acid in the rapamycin PKS, which bears an additional ER domain, reducing the double bond of the starter unit after its attachment to the PKS.

In type II PKS acetate is mostly used as a starter unit, but there are some examples in literature for non-acetate starter units. Besides propionate and malonamate that are employed as starters in the formation of anthracyclines and tetracyclines, a set of short linear branched fatty acids, such as butyryl, valeryl or 4-methylvaleryl may serve as alternate starter units in PKS II. The loading of these short fatty acids is facilitated by a KS III component. Benzoyl-CoA is another possible starter unit, e.g. for the enterocin and wailupemycin biosynthesis (Piel et al., 2000, Kalitzis et al., 2009). In this case, benzoate is activated by a CoA-Ligase to benzoyl-CoA, which can then be loaded to the KS.

PKS of type III use usually starter units such as hydroxyl-substituted and non-substituted cinnamoyl and benzoyl units, as well as activated fatty acids. The substrates are bound through a thioester-linkage to the Cys moiety of the catalytic triad Cys-His-Asn, which is located in the active site cavity and is connected to the substrate binding tunnel of the homodimeric protein. Substrate selection is directed by the spatial constraints of the substrate binding tunnel (Hertweck, 2009).

1.3.3 Non-Ribosomal Peptide Synthetases

NRPSs are similarly to PKSs classified in three different types: linear NRPS, iterative NRPS and non-linear NRPS (Mootz *et al*., 2002). Linear NRPS (type A) shows the same modular character as modular type I PKS that means each module extends the peptide chain by one amino acid. Hence, the number of modules is equal to the number of amino acids in the corresponding peptide. In iterative NRPS (type B), modules or domains are used more than once, as observed for the enterobactin biosynthesis (Gehring *et al*., 1998). The non-linear NRPSs (type C) deviate from the canonical domain arrangement, present in type A and B NRPSs. Examples for type C NRPS are manifested in yersiniabactin, vibriobactin and bleomycin biosynthesis (Suo *et al*., 2001, Marshall *et al*., 2002, Shen *et al*., 2001).

NRPS biochemistry

Peptides that are non-ribosomally synthesized are often assembled in a similar way as polyketides in modular PKS. A minimal NRPS module consists of a condensation (C), an adenylation (A), and a peptidyl carrier protein (PCP) domain, which is also denoted as thiolation (T) domain. NRP biosynthesis starts with the selection and activation of the substrates, which is initiated by the A domain. The substrates that can be selected by A domains are not restricted to the 20 proteinogenic amino acids, but include also non-proteinogenic amino acids and aryl acids. Due to the huge pool of substrates, including more than 500 monomers that were identified to be part of NRPs, a high structural diversity within this class of natural compounds is possible (Strieker et al., 2010).

Investigations on the substrate binding pocket of the phenylalanine-activating domain PheA from the gramicidin synthetase and sequence comparison with other A domains, led to the identification of 8-10 amino acid residues that portray the major determinants of substrate specificity in A domains (Stachelhaus *et al*., 1999, Challis *et al*., 2000). The identification of this “non-ribosomal code” or also called “Stachelhaus code” enabled the prediction of substrate specificity of A domains for unknown NPs. However, in some cases the extracted code can not be correlated to a certain amino acid, especially in the case of unusual substrates. Besides the specific selection of an individual amino acid, the A domain also activates the selected

substrate under ATP consumption to an aminoacyl-adenylate, which is then transferred to the corresponding PCP domain (figure 1.10).

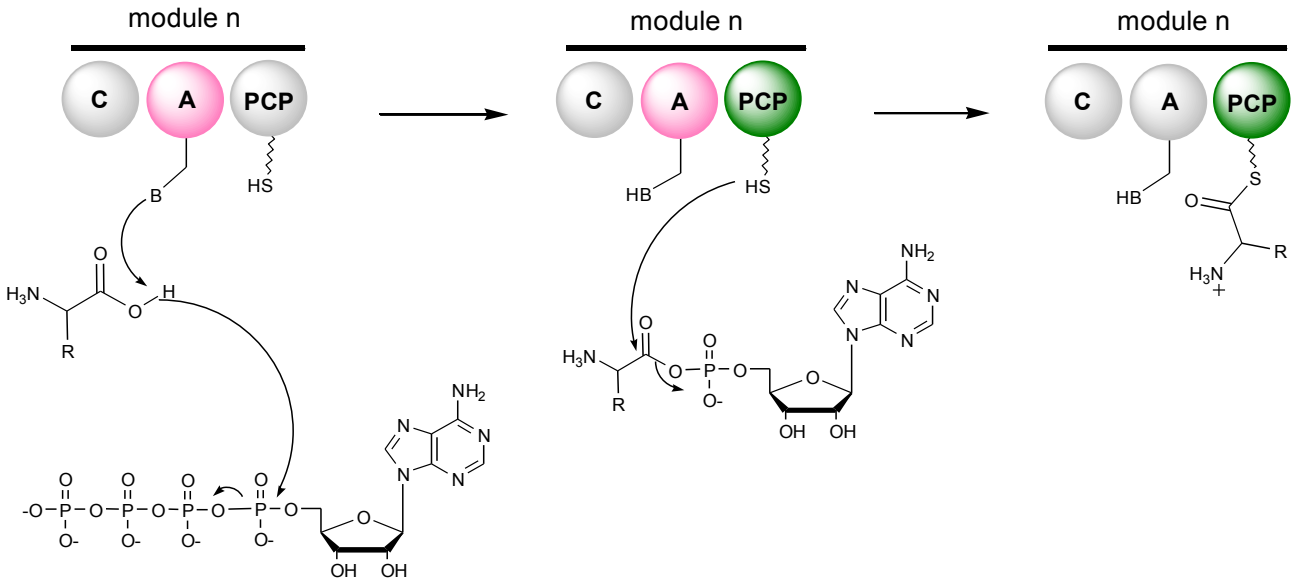


Figure 1.10 Activation of a building block by the A domain in a minimial NRPS module: Selection and activation of a specific amino acid by the A domain (highlighted in pink) under ATP is consumption, with subsequent transfer of the aminoacyl group of the obtained aminoacyl-AMP to the PCP domain (highlighted in green).

Peptide bond formation is catalyzed by the C domain, which possesses two binding pockets: an acceptor site and a donor site. The amino acid that is tethered to the upstream PCP is offered to the acceptor site, whereas the downstream PCP offers its covalently bound amino acid to the donor site of the C domain (Linne and Marahiel, 2000).

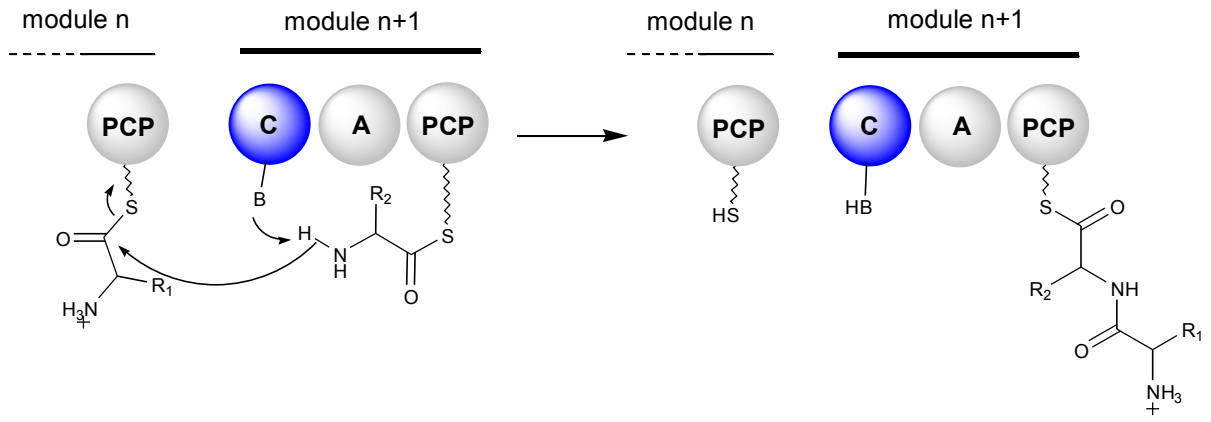


Figure 1.11 Peptide bond formation in a minimal NRPS module initated by the condensation (C) domain (highlighted in blue)**.**

Besides these essential core domains, the NRPS module may be accessorized with other modifying domains, responsible for epimerization, oxidation, methylation or heterocyclization of the respective amino acid, which increases the structural variety of the final products. For the formation of thiazoline or oxazoline rings the C domain is replaced by a heterocyclization (HC or CYC) domain, which catalyzes besides the peptide bond formation also a cyclization of threonine, serine or cysteine residues. These five-membered rings can be further oxidized by oxidation (Ox) domains to stable thiazole or oxazole heterocylces. Some NRP contain D-amino acids, which is mostly introduced by epimerization of an L-amino acid by the presence of an epimerization (E) domain located upstream the PCP domain. Selectivity for the correct enantiomer is controlled by the C domain that follows the E domain. Prevention of premature proteolytic breakdown of the peptides can be achieved through N- or C-methylation of amino acid residues. Methylation is introduced by the so called methyltransferase (MT) domains, transferring the methylgroup from an S-adenosylmethionine (SAM) to the respective nitrogen or carbon atom. These are only some examples for modifications on the peptide backbone. The final domain in NRPS multienzymes is the TE domain, releasing the peptide chain from the assembly line either through hydrolysis or cylcisation, resulting in a linear or cyclic peptide.

1.3.4 PKS/NRPS hybrids

As modular NRPS and PKS system share catalytic and structural similarities, interaction between PKS and NRPS systems is possible (Du *et al*., 2001). Hybrid PKS/NRPS systems that generate mixed PK-NRP compounds are especially common in myxobacterial, firstly described for the myxothiazol gene cluster from *Stigmatella aurantiaca* (figure 1.12). As these hybrid assembly lines are involved in the biosynthesis of clinically valuable natural products, such as rapamycin, epothilone and bleomycin, modifications of these assembly lines through combinatorial biosynthesis was intensively studied for the biosynthesis of epothilone (O' Connor et al., 2003, Richter et al., 2008). This requires deep knowledge on how these assembly lines work coordinate themselves at the PKS/NRPS interface. Two types of NRPS/PKS hybrid systems were described for the biosynthesis of mixed PK-NRP metabolites. In the first class, peptide and polyketide moieties are assembled independently and subsequently coupled by a discrete enzyme, as shown for the

coronatine biosynthesis (Rangaswamy *et al*., 1998). For the other type, there exists a functional interaction between NRPS and PKS modules. Most myxobacterial hybrid systems are of this type, where a PKS-bound ketide chain is elongated by a NRPS module or vice versa (Du and Shen, 2001).

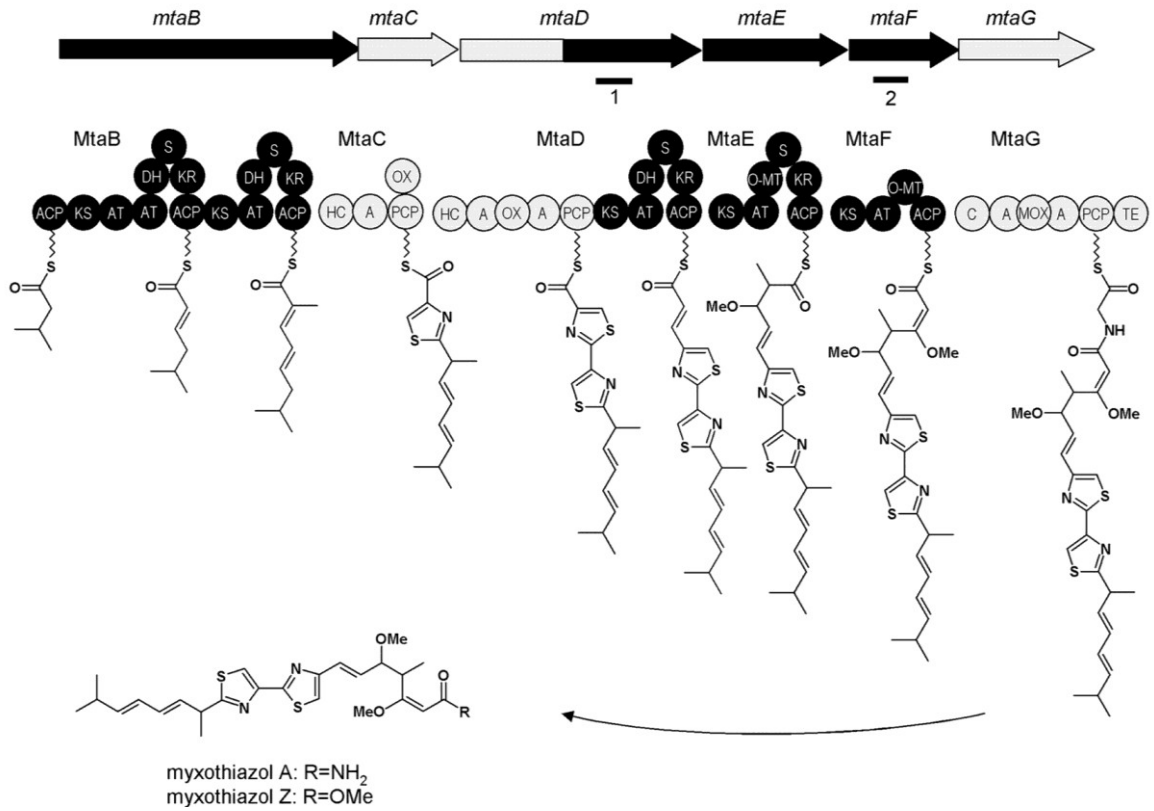


Figure 1.12 The *mta* gene cluster from *Stigmatella aurantiaca*, a NRPS/PKS mixed gene cluster responsible for the formation of myxothiazol. PKS parts of the gene cluster are colored in black, whereas NRPS parts are shown in grey (figure origin: Perlova *et al*., 2006)

The transfer of the intermediates along the PKS-NRPS or NRPS-PKS interfaces is mediated by intermodular communication via “interpolypeptide linkers” or “docking domains” for the case of PKS, or so are called “communication-mediating (COM) domains” in NRPS modules (Gokhale and Khosla, 2000, Broadhurst et *al*., 2003, Hahn and Stachelhaus, 2006).

1.4 Phenylnannolone A: a polyketide from *Nannocystis pusilla*

Three representatives of a new class of compounds were isolated and decribed by our work group in 2008 (Ohlendorf *et al*., 2008): phenylnannolone A-C. They were isolated from *Nannocystis pusilla* B150, a myxobacterium from the intertidal region of Crete. Orange spherical agglomerates of the strain observed in liquid cultures and the nearly strout form of vegetative cells are some characteristic features of this microorganism, as well as agar corodation on solid cultures (figure 1.13).

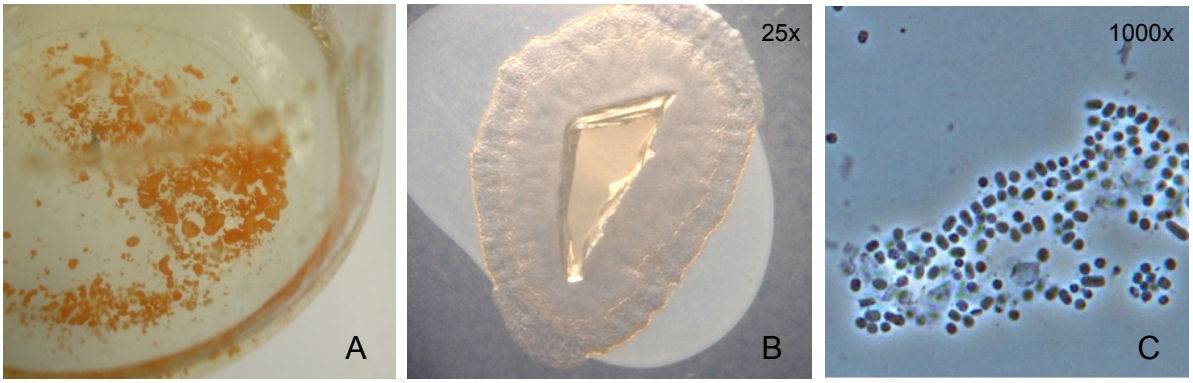
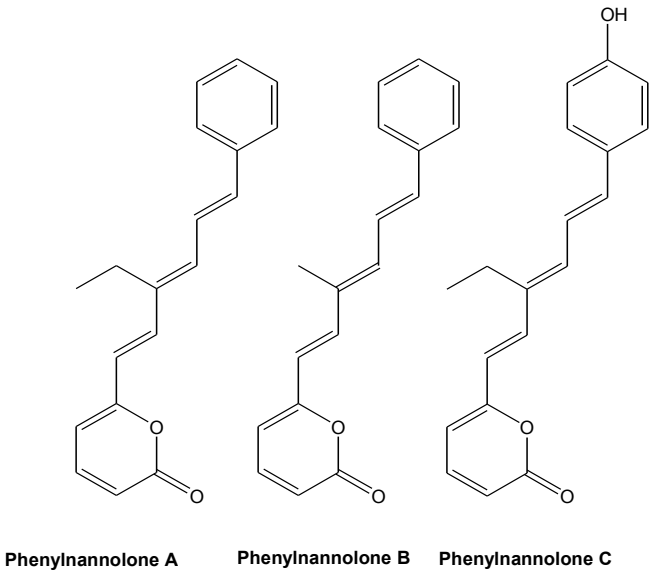


Figure 1.13 *Nannocystis pusilla* B150 in liquid culture (A) and on agar as solid culture (B). C shows the vegetative cells of *N. pusilla* B150



Phenylnannolone A **Phenylnannolone B** **Phenylnannolone C**

Figure 1.14 Molecular structures of phenylnannolone A, B and C isolated from *N. pusilla* B150 by B.Ohlendorf in 2008

The main metabolite of this new class compounds is phenylnannolone A, which had shown inhibitory activity towards p-glycoprotein. The ABCB1 gene product p-glycoprotein leads to the efflux of various drugs from the cell, i.e. antibiotic or

anticancer drugs, and hence results in treatment failure. Therefore, therapy with the anticancer drug daunorubicin is not always successful, as daunorubicin-resistent tumor cells overexpress the ABC-transporter p-glycoprotein. The effect of phenylnannolone A on p-glycoprotein could not be measured with functional assays, as it quenched the fluorescence of daunorubicin and calcein. Therefore it was assayed in an indirect way by an MTT (3-(4,5-Di**m**ethyl**t**hiazol-2-yl)-2,5-diphenyl**t**etrazolium bromide) assay on daunorubicin sensitive and resistant tumor cells for cytotoxicity. The colorimetric MTT assay determines the amount of living cells by measuring the aborbance of a formazan dye, which is only formed in viable cells. When tested alone no toxicity was observed, but together with daunorubicin the compound was able to reverse daunorubicin resistance in the cultured cancer cells. In fact, the combination of daunorubicin with phenylnannolone A resulted in a 10 fold reduction of the resistance factor in the resistant tumor cell line compared to the sole treatment with daunorubicin (figure 1.15). This result is comparable to the effect of p-glycoprotein inhibitors of the 3rd generation, such as tariquidar (Fox *et al*., 2006, Ohlendorf, 2008).

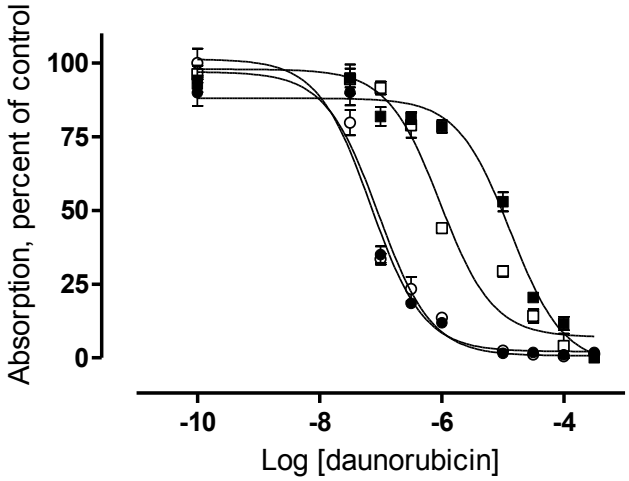
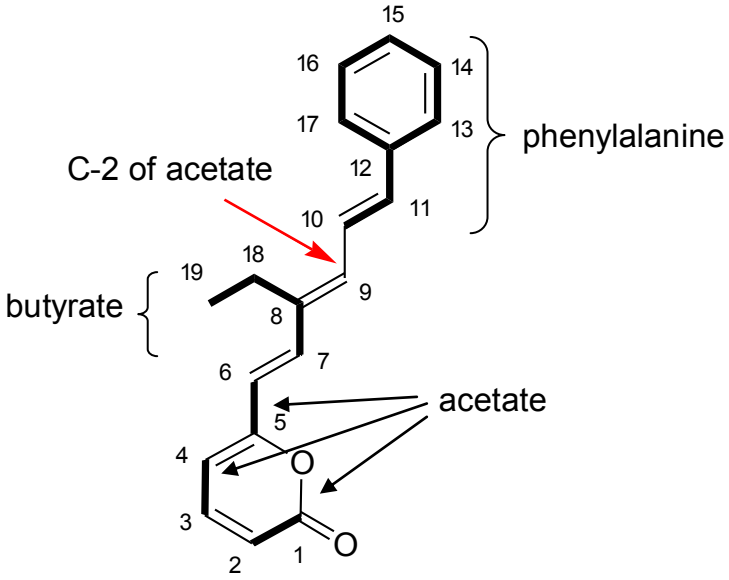


Figure 1.15 Concentration effect curve of the cytostatic drug daunorubicin in resistant (squares) and sensitive (circels) cancer cells in presence (open symbols) or absence (closed symbols) of Phenylnannolone A (Ohlendorf *et al*., 2008)

Phenylnannolone A was also tested for antimicrobial, antiplasmodial, cytotoxic and antiviral activity, but showed no activity in these bioassays, except a weak activity against the Influenza A virus.

The structure of the polyketide phenylnannolone A shows an unusal architecture. It is composed of an ethyl-substituted polyene chain, which is linked to a pyrone moiety on one side and to a phenyl ring on the other. These structural elements as such are not unusual, but the combination of these components was not described before. First investigations on the biosynthesis of phenylnannolone A with labeled precursors revealed acetate, butyrate and a phenylalanine-derived starter unit as building blocks for phenylnannolone A (figure 1.16). The labeling pattern for the starter unit was somehow remarkable, as it showed an unusual incorporation of a C-2 carbon atom derived from acetate in position 9 and suggested novel biochemical reactions (Ohlendorf *et al*., 2008). The incorporation of the butyrate unit and the three acetate units suggest that the lactone ring and the polyene chain up to the ethyl-side chain are formed by a PKS. However, the biosynthetic origin of the phenylalanine-derived starter unit and its loading can not be explained by the feeding studies.



Phenylnannolone A

Figure 1.16 Molecular structure of phenylnannolone A with biosynthetic building blocks (in bold). Feeding experiments in *N. pusilla* B150 with ^{13}C-labeled building blocks showed incorporation of a phenylalanine derived starter unit, one butyrate and three acetate building blocks. Unprecedented and unusual is the incorporation of a C-2 of acetate in position 9

It may be speculated, that biosynthesis is initiated by a NRPS-like loading module of a PKS system, as described for the rifamycin biosynthesis (see also section 1.3.2, figure 1.9 D). The elucidation of the phenylnannolone A biosynthesis is the focus of the here presented investigations.

2 Scope of the study

This study is focused on bioactive natural products from a *Nannocystis* species. *Nannocystis* spp. are myxobacteria, whose secondary metabolism has hardly been investigated. The recently described phenylnannolone A from *N. pusilla* B150 has been shown to have inhibitory activity towards p-glycoprotein, a membrane protein responsible for multi drug resistance (MDR) (Ohlendorf *et al.*, 2008).

The phenylnannolones represent a new class of compounds with an unusual structure composed of an ethyl-substituted polyene chain linked to a pyrone moiety on one side, and to a phenyl ring on the other. Feeding studies with labeled precursors revealed, besides three acetates, a butyrate and a phenylalanine-derived starter unit, also the incorporation of a C-2 carbon atom derived from acetate. This labeling pattern suggested novel biochemical reactions. From these results we can assume that the *N. pusilla* B150 genome harbors either a mixed NRPS/PKS system or a PKS gene cluster with a NRPS-like loading module to produce the phenylnannolones.

The aim of the presented study was to elucidate the biosynthesis of phenylnannolone A, i.e. identification of the respective gene cluster.

In order to achieve this goal two strategies were envisaged:

1) The genome of *N. pusilla B150* shall be sequenced to get access to the genetic information encoding the biosynthetic enzymes that are involved in the formation of the secondary metabolite phenylnannolone A. Bioinformatic analysis of the corresponding genetic information shall be facilitated by online databases, employing sequence homology searches against already identified genes or their encoded proteins.

2) Construction of a genomic library of *N. pusilla* B150 and screening with suitable primers to identify the complete biosynthetic gene cluster of phenylnannolone A.

In the final part of this study it was intended to characterize the putative biosynthetic enzymes in *in vitro* assays for their functionality and substrate selectivity.

3 Materials and Methods

3.1 Materials

3.1.1 Chemicals and other materials

Table 3.1 Chemical substances and solutions used in this work

Substance	Manufacturer
Acetic acid	KMF Laborchemie Handels GmbH (Darmstadt,Germany)
Aceton-d 99,8%	Deutero (Kastellaun, Germany)
Agar	Fulka Chemie GmbH (Buchs,Switzerland)
Ammonium acetate	Roth Chemie GmbH (Karlsruhe, Germany)
Ammoniumperoxosulfate(APS)	Roth Chemie GmbH (Karlsruhe, Germany)
Ampicillin	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)
Brilliant Blau R 250	Roth Chemie GmbH (Karlsruhe, Germany)
Bromphenol blue	Roth Chemie GmbH (Karlsruhe, Germany)
BSA	Bio-Rad Laboratories GmbH (Munich, Germany)
$CaCl_2$ x 2 H_2O	Merck KGaA (Darmstadt, Germany)
Chloramphenicol	Fulka Chemie GmbH (Buchs,Switzerland)
Chloroform	Roth Chemie GmbH (Karlsruhe, Germany)
D(+)-Maltose x H_2O	Roth Chemie GmbH (Karlsruhe, Germany)
dATP	Fermentas GmbH (St. Leon Rot, Germany)
DMSO	Roth Chemie GmbH (Karlsruhe, Germany)
dNTP	Promega GmbH (Mannheim, Germany)
Ethanol 99,8% p.a.	Roth Chemie GmbH (Karlsruhe, Germany)
Ethidium bromide	Roth Chemie GmbH (Karlsruhe, Germany)
Gel Loading Dye	Fermentas GmbH (St. Leon Rot, Germany)
Glycerol	Roth Chemie GmbH (Karlsruhe, Germany)

GoTaq®Flexi Buffer (10x)	Promega GmbH (Mannheim, Germany)
Imidazole	Roth Chemie GmbH (Karlsruhe, Germany)
Isoamyl alcohol	Roth Chemie GmbH (Karlsruhe, Germany)
Isopropanol	Roth Chemie GmbH (Karlsruhe, Germany)
$MgCl_2$ x 6 H_2O	Merck KGaA (Darmstadt, Germany)
$MgSO_4$ x 7 H_2O	Merck KGaA (Darmstadt, Germany)
N,N,N',N'-tetraethylendiamin (TEMED)	Roth Chemie GmbH (Karlsruhe, Germany)
Na_2-EDTA	Roth Chemie GmbH (Karlsruhe, Germany)
NaCl	Merck KGaA (Darmstadt, Germany)
NaOH	Merck KGaA (Darmstadt, Germany)
Ni-NTA agarose	Qiagen GmbH (Hilden, Germany)
peqGOLD Agarose	PEQLAB Biotechnologie GmbH (Erlangen, Germany)
peqGOLD Low Melt-Agarose	PEQLAB Biotechnologie GmbH (Erlangen, Germany)
Phenol	Merck KGaA (Darmstadt, Germany)
Rotiphorese® Gel 30	Roth Chemie GmbH (Karlsruhe, Germany)
SDS	Roth Chemie GmbH (Karlsruhe, Germany)
Sodium acetate	Merck KGaA (Darmstadt, Germany)
Tris	Roth Chemie GmbH (Karlsruhe, Germany)
Tris-HCl	Roth Chemie GmbH (Karlsruhe, Germany)
Tryptone/Peptone from Caseine	Roth Chemie GmbH (Karlsruhe, Germany)
Yeast extract	Fluka Chemie GmbH (Buchs, Switzerland)

Table 3.2 Technical equipment and other material used in this project

Material	Manufacturer
Amicon Ultra centrifugal filters	Millipore GmbH (Schwalbach, Germany)
Autoclave	Varioklav®, H+P Labortechnik AG (Oberschleißheim, Germany)
Branson Sonifier 250	G. Heinemann Ultraschall- und Labortechnik (Schwäbisch Gmünd, Germany)
Biometra T3000 Thermocycler	Biometra GmbH (Göttingen, Germany)
BioRad PowerPac™ 300	Bio-Rad Laboratories GmbH (Hercules, U.S.A.)

Boekel Replicator	Boekel Scientific (Feasterville, U.S.A.)
Centrifuge Heraeus Biofuge fresco	Thermo Fisher Scientific (Waltham, U.S.A.)
Centrifuge Heraeus Contifuge Stratos	Thermo Fisher Scientific (Waltham, U.S.A.)
Centrifuge Heraeus Fresco 17	Thermo Fisher Scientific (Waltham, U.S.A.)
Centrifuge tubes (15/50 ml)	TPP AG (Trasadingen, Germany)
CopyControl™ Induction solution	Epicentre (Madison, U.S.A.)
Eppendorf Centrifuge 5415 D	Eppendorf (Hamburg, Germany)
Eppendorf tubes 0.5, 1.5, 2 ml	Eppendorf (Hamburg, Germany)
Centrifuge tubes 15/50 ml	TPP AG (Trasadingen, Germany)
Gel chambers Horizon 58 and 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)
Kodak DC290	Kodak GmbH (Stuttgart, 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	IKA® Werke GmbH & Co. KG (Staufen, Germany)
MS2 Minishaker	IKA® Werke GmbH & Co. KG (Staufen, Germany)
Nalgene cryogenic vials	Nalgene Nunc International (Rochester, U.S.A.)
Parafilm®	Pechiney Plastic Packaging Company (Chicago, U.S.A.)
Scale (Satorius BL 3100)	Satorius AG (Göttingen, Germany)
Scale (Satorius BP 221S)	Satorius AG (Göttingen, Germany)
Sterile filter (0.2 µm)	Renner GmbH (Dannstadt, Germany)
Thermomixer Eppendorf	Eppendorf (Hamburg, Germany)
Transferpette®-8	Brand GmbH + Co. KG (Wertheim, Germany)
UV mini 1240 UV/VIS spectrometer	Shimadzu (Kyoto, Japan)

UV cuvettes	Ratiolab GmbH (Dreieich, Germany)
Water bath (Haake DC10)	Thermo Haake GmbH (Karlsruhe, Germany)
XCell SureLock® Mini-Cell	Invitrogen (Karlsruhe, Germany)
1.5 mm cassettes for Mini Cell	Invitrogen (Karlsruhe, Germany)

3.1.2 Enzymes

The enzymes used in the scope of this study are listed in table 1.3. Appropriate enzyme reaction buffers, which provide optimal reaction conditions as for polymerase, restriction and ligation were purchased together with the enzymes.

Table 3.3 Enzymes used in the scope of this study

Enzyme	Manufacturer
Agarase	Fermentas GmbH (St. Leon Rot, Germany)
GoTaq® Flexi DNA Polymerase (5 u/µl)	Promega (Mannheim, Germany)
Lysozyme	Roth (Karlsruhe, Germany)
Proteinase K	Roth (Karlsruhe, Germany)
Restriction enzymes	Fermentas GmbH (St. Leon Rot, Germany), Promega (Mannheim, Germany)
RNase (DNase free)	Roth Promega (Mannheim, Germany)
Fast-Link™ DNA Ligase	Epicentre (Madison, U.S.A.)
T_4-DNA-Ligase	Fermentas GmbH (St. Leon Rot, Germany)

All enzymes were applied following the respective manufacturer's recommendations for use. Lysozyme and Proteinase K were used as stock solutions with concentrations of 100 mg/ml and 20 mg/ml, respectively. Restriction enzymes were purchased together with the appropriate reaction buffers and were used according to the provided company's protocols.

3.1.3 Molecular weight marker

The following DNA standards listed in table 1.4 were used for gel electrophoresis.

Table 3.4 Molecular weight marker applied for size estimation

Molecular weight marker	Manufacturer
Gene Ruler™ DNA Ladder Mix	Fermentas GmbH (St. Leon Rot, Germany)
Lambda Mix Marker	Fermentas GmbH (St. Leon Rot, Germany)
36 kb Fosmid Control DNA (100ng/ul)	Epicentre Biotechnologies (Madison, U.S.A.)
PageRuler Unstained Protein Ladder	Fermentas GmbH (St. Leon Rot, Germany)

3.1.4 Molecular biological kits

Commercial molecular kits that were utilized in this work are listed in table 1.5. The application of the kits was performed according to the manufacturer's instruction manuals.

Table 3.5 Molecular biological kits

Molecular biological kit	Manufacturer
QIAquick PCR Purification Kit	Qiagen GmbH (Hilden, Germany)
pGEM®-T Vector System I	Promega (Mannheim, Germany)
PureYield™Miniprep System	Promega (Mannheim, Germany)
QIAGEN Plasmid Midi Kit	Qiagen GmbH (Hilden, Germany)
QIAquick Gel Extraction Kit	Qiagen GmbH (Hilden, Germany)
QIAprep®Spin Miniprep Kit	Qiagen GmbH (Hilden, Germany)
GeneJET™ Plasmid Miniprep Kit	Fermentas GmbH (St. Leon Rot, Germany)
CopyControl™ Fosmid Library Production Kit	Epicentre Biotechnologies (Madison, U.S.A.)
Fast-Link™ DNA Ligation Kit	Epicentre Biotechnologies (Madison, U.S.A.)
Quick Blunting Kit	New England Biolabs (Frankfurt, Germany)
Wizard® Genomic DNA Purification Kit	Promega (Mannheim, Germany)
innuPREP Bacteria DNA Kit	Analytic Jena (Jena, Germany)
RTP® Bacteria DNA Mini Kit	Invisorb/stratec molecular (Berlin)
NucleoSpin® Tissue	Machery and Nagel(Düren, Germany)

NucleoTrap®	Machery and Nagel(Düren, Germany)
DNA Clean and Concenttator™-5	Zymo Research Europe (Freiburg, Germany)
Wizard® SV Gel and PCR Clean-Up System	Promega (Mannheim, Germany)

3.1.5 Bacterial strains

Within this work, different *E. coli* host strains were utilized for transformation of various insert-vector constructs. The *E. coli* strains, their genotype and manufacturers are listed in table 6.

Table 3.6 *E. coli* host strains applied in the presented study

Strain	Genotype	Manufacturer
XL1-blue *E. coli*	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*qZΔ*M15* Tn*10* (Tetr)]	Agilent Technologies Deutschland GmbH (Böblingen, Germany)
Phage T1-Resistant TransforMax™ EPI-300™-T1R chemically competent *E. coli*	F- *mcr*A D(*mrr-hsd*RMS-*mcr*BC) f80d*lac*ZDM15 D*lac*X74 *rec*A1 *end*A1 *ara*D139 D(*ara*, *leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G *trf*A *ton*A *dhfr*	Epicentre Biotechnologies (Madison, U.S.A.)
One Shot® TOP10 Chemically *E. coli*	F- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen (Karlsruhe, Germany)
BL21 Star™ (DE3) One Shot® Chemically Competent *E. coli* -	F- *ompT hsdS*B (rB-mB-) *gal dcm rne131* (DE3)	Invitrogen (Karlsruhe, Germany)

3.1.6 Vectors

For several intended cloning purpose (construction of a genomic library or protein expression) DNA fragments were cloned into vectors with different properties. The following vectors listed in table 3.7 were used in this work.

Table 3.7 Vectors used for cloning strategies

Vector	Selectable marker	Size	Reference
pet151Topo	ampicillin	5,7 kb	Invitrogen (Karlsruhe, Germany)
pGEM®-T Vector	ampicillin	3 kb	Promega (Mannheim, Germany)
pCC1FOS™ Vector	chloramphenicol	8,1 kb	Epicentre Biotechnologies (Madison, U.S.A.)

Table 3.8 Constructs produced in this work

Construct	Applied vector	Insert size	Insert
pet151AMPACP	pet151Topo	2,2 kb	AMP ligase(PCR product from fosmid 12A9)
Sub12	pGEM®-T Vector	1 kb	DNA fragment produced by sonication
p21H12KS	pGEM®-T Vector	0,7 kb	KS domain (PCR product from fosmid 21H12)
p12A9KS	pGEM®-T Vector	0,7 kb	KS domain (PCR product from fosmid 12A9)

3.1.7 Fosmids

To establish a genomic library the CopyControl™ Fosmid Library Production Kit was used to generate 2290 recombinant fosmid clones. The approximate insert size of recombinant myxobacterial DNA amounts to 36 kb. The fosmid clones 21H12, 11A3 and 12A9 were further investigated during this project.

Table 3.9 Fosmids produced in this work, that were further investigated

Construct	Applied vector	Insert size	Insert
21H12	pCC1FOS™ Vector	36 kb	DNA fragment from genomic DNA of *N.pusilla*
11A3	pCC1FOS™ Vector	36 kb	DNA fragment from genomic DNA of *N.pusilla*
12A9	pCC1FOS™ Vector	36 kb	DNA fragment from genomic DNA of *N.pusilla*

3.1.8 Phages

Phages used for the packaging of concatemer DNA are named in table 3.10.

Table 3.10 Phage extract used for fosmid library production

Phage	Type	Manufacturer
MaxPlax™ Lambda Packaging Extracts	λ-phage	Epicentre Biotechnologies (Madison, U.S.A.)

3.1.9 Oligonucleotides

In the presented work, several oligonucleotides were used as PCR primers, either specific or degenerated. They were designed from multiple sequence alignments. Table 3.12 shows the primers utilized and their corresponding melting temperatures (TM) as stated by the manufacturers. The primers were provided as lyophilized powders by Eurofins MWG Operon (Ebersberg, Germany) and were dissolved in TE buffer, adjusted to a concentration of 100 pmol/µl and stored at -20 °C. In standard PCR reaction mixtures, the primers were used at working concentrations of 10-20 pmol/µl. The base abbreviations were used according to the IUPAC nucleotide code listed in table 3.11. Inosine (I) was integrated as a spacer base to reduce the degradation rate of primers.

Table 3.11 Abbreviations for bases according to IUPAC nucleotide code

IUPAC nucleotide code	Base	IUPAC nucleotide code	Base
A	Adenine	**W**	A or T
C	Cytosin	**K**	G or T
G	Guanine	**M**	A or C
T	Thymine	**B**	C or G or T
U	Uracil	**D**	A or G or T
R	A or G	**H**	A or C or T
S	G or C	**V**	A or C or G
Y	C or T	**N**	any base

Table 3.12 Oligonucleotides used in PCR studies

Primer	Sequence (5'- 3')	T_M	Reference
T7	TAA TAC GAC TCA CTA TA	53	Promega
SP6	TAT TTA GGT GAC ACT ATA G	48	Promega
Epi-RP	CTC GTA TGT TGT GTG GAA TTG TGA GC	63	Epicentre
A3rev.1	CCT CCG GSI CSA CCG GSM IGC CSA AGG	72	Erol et al., 2010
LGDD.S	GCC GCC SAG SIY GAA GAA	46	Erol et al., 2010
KS 1up	MGI GAR GCI HWI SMI ATG GAY CCI CAR CAI MG	56	Beyer et al.
KS d1	GGR TCI CCI ARI SWI GTI CCI GTI CCR TG	58	Beyer et al.
pA	AGA GTT TGA TCC TGG CTC AG	60	Edwards et al. 1989
pH	AAG GAG GTG ATC CAG CCG CA	64	Edwards et al. 1989
1540.F	CTC ACC ATT CAC GGC CAC CTG AG	66	this work
1540.R	CCC GCG GTG ACT GTC GAT TAT TC	64	this work
NRPS-F1	CGC TGA GCT TCG ACG CGT TCG TGT TCG AGC TGC TG	75	this work
NRPS-R1	CGA ACA CCC CGA TCG GCG CCG GGT TGC CGC GG	75	this work
TE-fw1	CCR SCC KMK SGG SGG CAC G	70	this work
TE-rev1	SAR YYG RCG CGC CAY CTC	62	this work
revAMPPCP	CTA GAT CAC CGC GAT CGG CTC GTC	68	this work
forAMPPCP	CAC CGC GTC AAT CCG GCC GCG AGT C	73	this work

3.1.10 Water

For the preparation of culture media, demineralized water was provided by a reverse osmosis system (IMB, Germany). A Milli-Q® Academic Water Purification System (Millipore GmbH, Germany) was used to generate ultra-pure water prepared from demineralized water. Milli-Q® water was used for all applications, if not specified otherwise. For PCR application, autoclaved Milli-Q® water was applied.

3.1.11 Culture Media

Culture media were prepared with demineralized water prior to steam sterilization. The following table shows the composition of the used media.

Table 3.13 List of media and their composition

MD1+Glucose	
3g casiton	
0.7g $CaCl_2$ x 2 H_2O	
2g $MgSO_4$, x 7 H_2O	
2.2 g glucose x H_2O	adjust to pH=7.5
	after sterilization add:
	1ml/L trace element solution
	1ml/L cyanocobalamine solution

VY/2 *Agar:*	
50ml yeast suspension(10%)	
1.36 g $CaCl_2$ x 2 H_2O	
15g agar	
ad 1L H_2O	adjust to pH=7.2
	after sterilization add:
	1ml/L trace element solution

Luria-Bertani (LB)-Medium	
10 g tryptone	
5 g yeast extract	
10 g NaCl	
ad 1 L H_2O	adjust to pH=7.5

LB-Agar	
10 g tryptone	
5 g yeast extract	
10 g NaCl	
15 g Agar	
ad 1 L H_2O	adjust to pH=7.5

Trace element solution	
20 mg $ZnCl_2$	**5 mg LiCl**
100 mg $MnCl_2$ x 4 H_2O	**20 mg KBr**
10 mg H_3BO_3	**20 mg KJ**
10 mg $CnSO_4$	**10 mg Na_2MoO_4 x 2 H_2O**
5 mg $CoCl_2$	**5.2 g EDTA-Na_2 x 2 H_2O**
5 mg $SnCl_2$ x 2 H_2O	ad 1 L H_2O; solution was sterile filtrated

Cyanocobalamine solution	
50 mg cyanocobalamine	
ad 100 ml H_2O	solution was sterile filtrated

3.1.12 Antibiotics

The antibiotics listed in table 3.14 were used as additives in culture media for the purpose of clone selection and for the antibiotic selectivity test described in section 3.2.2. The antibiotics were used as antibiotic stock solutions and were solved either in autoclaved Milli-Q® water (3.1.10) or in ethanol (i.e. chloramphenicol).

Table 3.14 List of antibiotics used as additive in culture media

Antibiotics	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)

3.1.13 Buffers and solutions

All buffers and solutions were prepared with ultra-pure water (3.1.10). Stock solutions not containing any organic solvents or thermo labile components were usually steam sterilized prior to use (3.2.1).

Table 3.15 Buffers and solutions

Buffers for DNA extraction:	
TE buffer	**SET buffer (Sambrook & Russell, 2001)**
10 mM Tris-HCl 1 mM EDTA pH 8.0	75 mM NaCl, 25 mM EDTA 10 mM Tris-HCl pH 7.5
RNase A stock solution (DNase-free)	
50 mg lyophilised RNase A 5 ml Tris-HCl (10 mM, pH 7.5) 15 mM NaCl storage at –20 °C	

Buffers for library construction:	
Phage dilution buffer (PDB)	
10 mM Tris-HCl 100 mM NaCl 10 mM MgCl2 pH 8.3	
Buffers for plasmid and fosmid purification:	
Buffer P1	**Buffer P2**
50 mM Tris-HCl 10 mM EDTA 100 µg/mL RNase A pH 8	200 mM NaOH 1% SDS
Buffer P3	**Buffer QBT (Equilibration buffer)**
3M potassium acetate pH 5.5	750 mM NaCl 50 mM MOPS (pH 7.0) →including 15 % isopropanol and 15 ml of 10% Triton X-100 solution (v/v)
Buffer QC (Wash buffer)	**Buffer QF (Elution Buffer)**
1 M NaCl 50 mM MOPS (pH 7.0) →including 15 % isopropanol	1.25 M NaCl 4M Tris (pH 8.5) →including 15 % isopropanol
Buffers for gel electrophoresis:	
10 x TBE buffer	**Modified 50 x TAE buffer (3.2.13.2)**
0.89 M Tris base 20 mM EDTA 0.87 M H_3BO_3 purified water ad 1000 ml	2 M Tris-acetate 50 mM EDTA (pH 8)
6 x Gel loading buffer (Sambrook & Russell, 2001)	
40 g sucrose 0.25 g bromophenol blue ad 100 ml purified water	

Buffers for protein expression and A domain assay:	
10x glycine SDS electrophoresis buffer	
250 mM Tris, 2M glycine 1% SDS, ad 1000 ml purified water adjust to pH 8.9	
Staining solution	**Destaining solution**
10% acetic acid 50 % ethanol 0.005% coomassie brilliant blue R-250	7.5% acetid acid 30%isopropanol 10% Acetic acid 20% Methanol 50% Methanol 5 % Acetic acid
Protein lysis buffer	**Protein wash buffer**
50 mM NaH2PO4 300 mM NaCl 10 mM imidazole pH 8.0	40 mM NaH2PO4 300 mM NaCl 20 mM imidazole pH 8.0
Protein elution buffer	**A Domain buffer (Phelan et al., 2009)**
50 mM NaH2PO4 300 mM NaCl pH 8.0;	20 mM Tris-HCl (pH 7.5) 5% glycerol 1mM DTT

Buffers QBT, QC and QF were supplied with the QIAGEN Plasmid Midi Kit (3.1.4). They were used for equilibration (QBT) and washing (QC) of the provided DNA purification columns as well as for the elution of purified DNA (QF). The procedure was carried out following the manufacturer's protocol. For gradual elution protein elution buffers with 100, 150, 200 and 300 mM imidazole were used.

3.1.14 Software and Databases

Basic Local Alignment Search Tool (BLAST) is provided by the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/pubmed). This software was applied to analyze nucleotide data for sequence similarities using blastx (nucleotide sequence is translated and compared to the amino acid sequences database). For further analysis of the amino acid sequences blastp (protein database using a protein query) was used. In order to analyze 16S rDNA, the database was searched using blastn (nucleotide query vs. nucleotide databases).

Entrez Nucleotide Database is maintained by the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/gene). This source is linked to numerous databases, e.g. GenBank, NCBI Reference Sequences (RefSeq), and the Protein Data Bank of the Research Collaboratory for Structural Bioinformatics (RCSB PDB)**.** Therefore, it lends itself for detailed analysis of nucleotide and amino acid sequences and putative anticipation on their possible functions in secondary metabolites biosynthesis.

ClustalW2 sequence analysis tool version 2.1 is provided by the European Bioinformatics Institute (EBI), (EMBL) (http://www.ebi.ac.uk/Tools/clustalw). This tool was used to generate multiple sequence alignments based on the Nucleotide Sequence Database, part of the European Molecular Biology Laboratory (EMBL).

DNAtrans (DNA Translator) was developed and provided by Dr. Anke Schiedel (Pharmaceutical Chemistry, Bonn) and Jochen Bosmann. The program helps to clean up DNA sequences, translate them to protein sequence and to create complementary and reverse DNA sequences. With the primer check tool melting temperatures and GC content can be calculated.

Clone Manager 9 is a purchased program with a set of tools for enzyme operations, cloning simulation, graphic map drawing, primer design and analysis and sequence alignments. Clone Manager 9 was mainly used for primer design and the simulation of cloning procedures.

Pfam is a Sanger institute database of protein families that includes their annotations and multiple sequence alignments generated using hidden Markov models.

Artemis is a a free genome browser and annotation tool provided by Sanger institute that allows visualization of sequence features, next generation data and the results of analyses within the context of the sequence, and also its six-frame translation. This program was mainly used to identify open reading frames (ORFs) in the sequence data obtained from genome and fosmid sequencing.

CLUSEAN (CLUster SEquence ANalyzer): This software enables fast access to sequence data from established databases like BLAST and HMMER (Weber *et al.*, 2009). It can be used to identify functional domains and conserved motifs in a given nucleotide or amino acid sequence. Furthermore, the software allowed searching whole genome assembly files as well as assembled contigs of the genome sequencing of *N. pusilla* B150.

AntiSMASH (antibiotics&Secondary Metabolite analysis Shell) is a free online tool that allows the rapid genome-wide identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genomes (http://antismash.secondarymetabolites.org/). It integrates and cross-links with a large number of in silico secondary metabolite analysis tools that have been published earlier. AntiSMASH is powered by several open source tools: NCBI BLAST+,HMMer 3, Muscle 3, Glimmer 3, FastTree, TreeGraph 2, Indigo-depict, PySVG and JQuery SVG.

ASMPKS (Analysis System for Modular Polyketide Synthesis) is an online tool, which provides overall management of information on modular PKS, including polyketide database construction, new PKS assembly, and chain visualization (http://gate.smallsoft.co.kr:8008/~hstae/asmpks/pks_prediction.pl). ASMPKS can predict functional modules for a submitted protein sequence, estimate the chemical composition of a polyketide synthesized from the modules, and display the carbon chain structure on the web interface.

Phylogeny.fr is a free, simple to use web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences (http://www.phylogeny.fr/). It connects various bioinformatics programs to reconstruct a robust phylogenetic tree from a set of sequences (Dereeper et al., 2008, Dereeper et al., 2012).

NaPDoS(Natural Product Domain Seeker) NaPDos is a bioinformatic tool for the rapid detection and analysis of secondary metabolite genes (http://napdos.ucsd.edu/). This tool is designed to detect and extract C- and KS-domains from DNA or amino acid sequence data, individual genes, whole genomes, and metagenomic data sets.

Kodak 1D software version 3.5.4 and **iX Imager software**: These programs were provided together with the respective gel documentation system (Kodak Scientific Imaging Systems and Intas Gel iX imager) and were applied to edit digital photos of electrophoresis gels.

3.2 Molecular biological methods

3.2.1 Sterilization of solutions and equipment

To eliminate potential contaminations with foreign organisms, all heat resistant solutions, buffers and media were autoclaved for 20 min at 121°C and 2 bar atmospheric pressure in a Varioklav® steam sterilizer (table 3.1.1). Heat sensitive components of solutions were sterilized by filtration through membrane filters with a pore size of 0.22 µm (table 3.2). Steam heat sterilization was also applied for decontamination of glass and plastic ware and for inactivation of genetically modified organisms.

3.2.2 Antibiotic selectivity test

To evaluate the resistance of the bacterium towards several antibiotics, *N. pusilla* B150 was cultivated with several antibiotics. For this purpose, a kryo-culture of the strain 150 was cultivated on agar plates with VY/2 medium and grown for 7 days. Small pieces of agar, containing fruiting bodies of *N. pusilla* B150 were transferred to new V/Y_2 agar plates complemented with several antibiotics. The growth of the strain on the agar plates was observed for 14 days with a stereo microscope (Stemi 2000-C, Zeiss, Germany).

For tests in liquid culture a pre-culture of 30 ml MD1-glucose medium was prepared, inoculating 100 µl of a kryo-culture of *N. pusilla* B150. After 7 days of growth under shaking of 140 rpm at 30°C, 5 ml of the pre-culture were inoculated in 100 ml MD1-glucose medium and were cultivated on the same conditions. The growth of *N. pusilla* in liquid was tested respectively for chloramphenicol, streptomycin, tetracycline and carbenicillin. The liquid cultures were checked after 4 and 14 days for bacterial growth.

3.2.3 Isolation of chromosomal DNA

For the isolation of genomic DNA from the myxobacterium *N. pusilla* B150 several molecular biological kits were tested. Amongst them Wizard® Genomic DNA Purification Kit (Promega), innuPREP Bacteria DNA Kit (analytic jena), RTP® Bacteria DNA Mini Kit (Invisorb), NucleoSpin® Tissue and NucleoTrap® (Machery and Nagel) (table 3.5).

But the method described by Neumann *et al*., 1992 was mainly used to isolate genomic DNA from *N. pusilla* B150. Therefore the myxobacterial cells were cultivated for 7 days in 100 ml MD1+G-Medium (table 3.13) at 30°C and 140 rpm and subsequently harvested by centrifugation at 8000 rpm for 10 min at 4°C. The cell pellet was resuspended in 5 ml SET buffer. After addition of 500 µl SDS (10 %), 500 µl lysozyme (40 mg/ml) and 275 µl proteinase K (20mg/ml) the cell solution was incubated at 55°C for 2 h and gently mixed by inverting every 15 min. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to the lysate 1.8 ml 5 M NaCl and and gently mixed by inverting several times. Through centrifugation at 8000 rpm for 10 min phase separation was achieved and the aqueous phase was separated from the organic phase through gentle pipetting with cut tips. The procedure was repeated several times until the interphase remained clear. For removal of phenol residues in the aqueous phase, the same procedure was carried out with chloroform: isoamyl alcohol (24:1).

The resulting aqueous sample containing the DNA was mixed with 1/10 volume of 3 M sodium acetate (pH 5.2) and 2-2.5 volumes of ice-cold ethanol 98% (stored at -20°C). After incubation of this mixture at -20°C overnight the precipitated DNA was centrifuged at 8500 rpm at 4°C for 45 min. The supernatant was decanted off and the DNA pellet dried at room temperature. The DNA was redissolved in TE buffer (pH 8.0) (Table 3.13) containing RNase (20µg/ml) for RNA degradation. It was incubated at 37°C for 30 min and afterwards stored at 4°C.

3.2.4 16S rDNA analysis

A standard PCR reaction (3.2.7) was performed on isolated genomic DNA applying the ‘universal’ eubacterial primer pair pA/pH (3.1.9) to amplify the 16S rDNA (Edwards *et al.* 1989). The obtained PCR product with a size of 1500 bp was excised from an agarose gel and purified using the QIAquick Gel Extraction Kit (3.1.4). Subsequently, the purified amplificate was ligated into the pGEMT®-vector (3.2.8).

The obtained plasmid was sequenced with the T7-primer (3.1.9). With BLAST search a phylogenetic tree was generated and helped to identify the origin of the strain and to prove the purity of the isolated genomic DNA.

3.2.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used to survey DNA manipulations such as preparative DNA isolation, restriction digests or PCR amplification. Standard gels were prepared by dissolving 1% peqGOLD Agarose in 1 x TBE buffer (table 3.1, table 3.15). Small gels were usually run for 40 min at a voltage of 120 V in Life Technologies Horizon® 58 chambers (table 3.2). Nucleic acid probes were mixed with gel loading dye before loading them into the slots. DNA standards were applied to each gel for size estimation. To reduce contamination with ethidium bromide in the laboratory, gels were stained subsequently to the separation run. The dye bath contained 100 ml ethidium bromide solution (10 mg/ml), the gels were stained for 1-3 min, followed by a washing step in water for another 3-5 min. The ethidium bromide solution was saved to be reused several times. Nucleic acids with intercalated ethidium bromide could be visualized by UV illumination at 254 nm. Gels were digitally pictured using the Kodak DC 290 Zoom Digital Camera System, or alternatively the Intas Gel iX Imager (table 3.2). UV light exposure was kept to a minimum, to reduce photochemical damage of the DNA.

3.2.6 DNA recovery form agarose gels

For cloning procedures, stained PCR fragments of interest were directly cut out from electrophoresis-gels and purified using the QIAquick gel extraction kit following the manufacturer's protocol. Purified PCR amplificates were usually cloned in sequencing vectors (pGEMT®) to analyze their nucleotide sequences by BLAST (3.1.14).

3.2.7 Polymerase chain reaction

Polymerase chain reaction was used for the in vitro amplification of DNA sections. Therefore knowledge about the flanking regions from the sequence of interest is a necessity. Oligonucleotide primers are designed complementary to the opposite ends of the desired region. Periodically temperature change causes denaturation, annealing and elongation of the DNA in usually 25-30 cycles. The denaturation of the

double stranded DNA into single stranded DNA enables the primer to anneal and nucleic acid elongation, which is mediated by a polymerase (3.1.2), may proceed. This procedure enables exponential in vitro reproduction of a DNA-section and is therefore the method of choice for rapid amplification of nucleic acid fragments. Within this work, PCR was performed applying a T3 or a T-gradient thermocycler (table 3.2) according to the parameters and protocols described in sections 3.2.7.1and 3.2.7.2.

3.2.7.1 PCR parameters

For PCR amplification of template DNA, PCR-buffers supplied with the GoTaq® Flexi DNA Polymerase (3.1.2) were used to achieve optimal reaction conditions. To ensure hybridization of PCR-primers with their complementary sequence after denaturation, the temperature was adjusted to the calculated annealing temperatures of the primers. Elongation by Taq-polymerase was accomplished at 72 °C, whereby the required elongation time was estimated from the length of the target region (about 1 min/kb). For DNA-extension, the polymerase recognizes the primers annealed to the DNA-template and starts elongation in 3'-direction using dNTPs provided in the reaction mixture.

PCR amplification is depending on several factors as like the concentration of $MgCl_2$, which affects polymerase activity or the choice of the annealing temperature, responsible for the specific hybridization of the primers to the DNA-template. Primer annealing occurs with less specificity at lower temperatures, which was beneficial in the case of homologous, degenerated primer pairs designed from sequence alignments (3.1.14). Optimization of the procedure was achieved for each PCR by adjusting the annealing temperature as well as the elongation time. Self-complementarity of template DNA is usually less likely due to molar excess of oligonucleotides. Despite that, additives like DMSO are useful when working with GC-rich templates avoiding the formation of secondary structures in the polynucleotide molecules (Hung *et al*., 1990).

3.2.7.2 PCR protocol

A standard PCR reaction mixture was composed as follows:

components		volume	final concentration
5x GoTaq® Flexi Buffer		4.0 µl	1x
$MgCl_2$	(25 mM)	1.0 µl	1.25 mM
dNTP	(each 10 mM)	0.4 µl	each 0.2 mM
GoTaq-Polymerase	(5 unit/µl)	0.16 µl	0.08 unit
Forward-Primer	(20 µM)	1.0 µl	0.5 µM
Reverse-Primer	(20 µM)	1.0 µl	0.5 µM
DMSO		1.0 µl	0.5 %
Template		1.0 µl	
H_2O		**ad 20.0 µl**	

The typical PCR thermocycling was performed as described below:

Step	temperature		time
1. initial denaturation step	95 °C		5 min
2. primer annealing	45-60 °C	30x	0.5 min
3. elongation	72 °C		1-1.5 min
4. denaturation	95 °C		0.5 min
5. final elongation	72 °C		4 min
6. end	4 °C		∞

Steps 2 to 4 were repeated in a loop for 30 times, before proceeding with step 5. Annealing temperatures were chosen about five degrees below the lower T_m-value (3.1.9) of one of the used oligonucleotides.

3.2.8 Ligation of PCR Products

PCR fragments produced by Taq-polymerases had a 3'-poly-A nucleotide overhang. Some customary cloning kits, like the pGEM®-T Vector System I (table 3.7), provide a vector, that is compatible to PCR products synthesized by Taq-polymerases. This cloning kit was utilized to directly ligate PCR fragments extracted from electrophoresis gels. A vector to insert ratio of 3:1 was favored in a total volume of 10 µl. Usually, an incubation time of 1h at room temperature was sufficient before

proceeding with the transformation step. The following general pipetting scheme was applied for standard ligations:

Component	pGEM®-T Vector System I
Buffer	5 µl (2 x Rapid Ligation Buffer)
Vector	1 µl (pGEM®-T)
PCR product	3 µl
Ligase	1 µl T4 DNA Ligase
Nuclease-free water	ad 10 µl

3.2.9 Restriction digestion

Chromosomal or plasmid DNA samples were routinely subjected to restriction digestions with restriction endonucleases type II. These enzymes originally belong to the bacterial repertoire of defense mechanisms. They recognize certain DNA sequences, which are mostly palindromes enabling a cleavage of phosphodiester bonds at these specific restriction sites. Most of the restriction endonucleases generate sticky ends, while *SmaI* for example generates blunt ends. However, ligation reactions carried out with sticky ended restriction fragments are more efficient. For optimal conditions, the buffers provided with the restriction enzymes were included in the reaction mixtures. The duration of the restriction digest was depending on the nature and quality of template DNA as well as on the type of restriction enzyme applied. Generally, a reaction mixture included 10 units of enzyme per µg DNA in a total volume of 20 µl and was incubated either for 2 h or overnight at 37 °C.

3.2.10 Preparation of cells competent for DNA-transformation

The *E. coli* strain XL-1 Blue (Stratagene, La Jolla, CA, USA) (3.1.5) with the following genotype was used for general cloning procedures:

recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)]

The preparation of chemically competent cells was followed using the method of Dagert and Ehrlich, 1974. An overnight pre-culture, prepared by inoculating 3 ml LB-medium with a single *E. coli* colony followed by incubation at 37 °C and constant shaking at 180 rpm, was transferred into an Erlenmeyer flask containing 70 ml of 2 x YT-medium. The mixture was again grown at 37°C and shaked at 180 rpm until an

OD_{600} of 0.3-0.4 was reached. At this point the cells were harvested by centrifugation for 10 min at 8,000 rpm and 4°C (Heraeus Contifuge Stratos) (table 3.2). To make the cells transformable, they were treated with ice cold $CaCl_2$/$MgSO_4$-solution (70mM $CaCl_2$/ 20mM $MgSO_4$) twice after resuspension. The first treatment was performed with 10 ml, the second one with 3.5 ml $CaCl_2$/$MgSO_4$-solution. Each time, the cell suspension was kept on ice for 30 min. Storage and deep-freezing of the competent *E. coli* cells was possible adding 875 µl steam sterilized glycerol to the cell suspension, without discernible loss of quality. After the addition of glycerol, the cell suspension was divided into aliquots of 100 µl and stored at -80°C.

3.2.11 Transformation of host strains

In vivo amplification of DNA sequences was achieved via insertion of a vector construct to competent *E. coli* XL1-Blue cells (3.1.5). The exponential growth of those *E. coli* cells leads then to the enrichment of vector DNA. An aliquot of competent cells (3.2.10) was thawed on ice and 5-10µl of the ligation mixture (3.2.8) was added. After incubation on ice for 30 min, the probe was heat-shocked for 90 sec at 42°C to enable DNA uptake. In order to ensure that the delivered plasmid DNA was kept inside the *E. coli* transformants, the cell suspension was again placed on ice for 2 min. To initiate cell growth, 300-600 µl of sterile LB-medium was added and a following incubation was carried out for 1h at 37 °C using the Thermomixer Eppendorf (table 3.2). Depending on the applied vector, an antibiotic selection was performed spreading 100-200 µl of this pre-culture on LB agar plates supplemented with appropriate antibiotics. The plates were incubated overnight at 37 °C using an incubator (3.1.1). Subsequently to plasmid isolation (3.2.12), the transformation of the vector construct was confirmed via PCR (3.2.7).

3.2.12 Plasmid isolation from transformed *E. coli*

The vector construct was isolated from transformed *E. coli* cells (3.2.11) using the GeneJET™ Plasmid Miniprep Kit or PureYield™Miniprep System (3.1.4). For the isolation of plasmid DNA from PCR positive clones, cultures in a scale of 3 ml were sufficient. After inoculation of LB-medium supplemented with an appropriate antibiotic, the culture was grown overnight in a test tube at 37 °C under constant shaking at 180 rpm. Cells were then harvested by centrifugation at 13,000 rpm for 2 min in 1.5 ml tubes (3.1.1). Subsequently, the pellets were resuspended in buffer 1 (3.1.12) by vigorous vortexing applying the MS2 Minishaker (3.1.1). Cell lysis was

achieved by addition of buffer 2 (3.1.12), followed by incubation for not more than 2 min while gently inverting the tube several times. After addition of the neutralization solution (buffer 3) (3.1.12), precipitated proteins and non-circular DNA were pelleted by centrifugation at 13,000 rpm for 15 min. The supernatant was purified with supplied columns following the manufacturer's instructions.

3.2.13 Construction of a genomic library

3.2.13.1 LMP agarose gel electrophoresis

Size selection of sheared DNA fragments, required for the construction of a fosmid library, was achieved via preparative gel extraction. High molecular genomic DNA could not be extracted in using the usually applied QIAquick gel extraction Kit, as this method is only recommended for DNA fragments between 70 bp and 10 kb. Therefore an agarose with a low melting point, in this case the peqGOLD Low Melt Agarose (3.1.1), was used to recover the genomic DNA. According to the manufacturer's manual, gels were prepared with TAE electrophoresis buffer (3.1.12). For separation of DNA fragments in the range of 35-40 kb, a 0.9 % LMP agarose gel was chosen. The 50 x TAE stock solution was diluted to 1 x concentration. 1.8 g of LMP agarose was suspended in 200 ml of 1 x TAE buffer using a magnetic stirrer (3.1.1). The suspension was heated to ebullition using a microwave until the agarose was completely dissolved. The LMP agarose solution was allowed to cool for 20 min to a temperature of about 50°C. In a Horizon® 11.14 chamber, the LMP gel was poured and cooled at 4°C for 1 hour to achieve optimal consistence. The nucleic acid solution was mixed with gel loading dye (3.1.3) and pipetted into the gel slots. For size estimation, molecular weight markers as the GeneRuler™ DNA Ladder Mix, the Lambda Mix Marker and the 36 kb Fosmid Control DNA (3.1.3), supplied with the CopyControl™ Fosmid Library Production Kit (3.1.4) , were used. In a first turn, the gel was run for 10 min at 60 V to assure that the DNA was transferred to the gel matrix. Subsequently, the voltage was lowered to 15-30 V and the gel electrophoresis was carried out overnight. For preparative isolation of the desired band, only the lanes with the molecular weight markers were stained with ethidium bromide, in order to prevent damages to the genomic DNA by intercalation.

3.2.13.2 DNA extraction from LMP agarose

DNA extraction from LMP agarose gels was performed in following the providers' recommends. Guided by the stained band of the Fosmid Control DNA (3.2.13.1), the desired band of about 36 kb had to be excised from the gel using a clean scalpel. To minimize UV light exposure to the band of interest, the latter was masked with a sheet of alu foil. The cut out gel slices were molten at 70 °C for 15 min and transferred to water bath of 42 °C for additional 10 min. According to the manufacturer's protocol, 1 unit of agarase (3.1.2) was added per 100 mg of molten LMP agarose. After repeated pipetting, the mixture was incubated at 42 °C for 2h followed by a heat inactivation of the enzyme for 10 min at 70°C. The mixture was divided into aliquots of 500µl volume, the tubes were then chilled on ice for 5 minutes and subsequently centrifuged for 25 min at 10.000 rpm and 4°C to separate undigested agarose from the sample. 90 % of the supernatant, in this case 1.25 ml, were pipetted in a sterile 15 ml Falcon to precipitate the DNA. This was achieved in adding 125 µl sodium acetate (3M, pH 5,3) and 3.75 ml pure isopropanol to the solution. Incubation for 10 min at room temperature and centrifugation at 6000 rpm resulted in a DNA pellet that was washed with 70% ethanol. The air-dried pellet was then solved in 20µl water or EB-Buffer (3.1.12).

3.2.13.3 End-repair of size selected DNA

Before cloning the size-selected high-molecular DNA fragments into the pCC1FOS vector (0), blunt ends of those DNA fragments were required, as the pCCFOS1 vector itself already provides blunt ends. For this purpose the Quick Blunting Kit (3.1.4) was used according to the manufacturer's protocol. A standard reaction mixture contained the following components:

Purified DNA (up to 5 µg)	1-19 µl
10 x Blunting Buffer	2.5 µl
1 mM dNTP Mix	2.5 µl
Blunt Enzyme Mix	1.0 µl
Sterile distilled water	ad 25 µl

Since the DNA fragments were sheared during DNA extraction, an incubation time of 30 min at room temperature was chosen following the manufacturer's recommendations. Subsequently, the enzyme mix was inactivated by heating at 70°C

for a duration of 10 min. The reaction mixture containing the blunt end DNA could be directly used for ligation into the pCC1FOS fosmid vector.

3.2.13.4 Ligation of end-repaired DNA into the pCC1Fos vector

For the construction of a genomic library, the CopyControl™ Fosmid Library Production Kit (3.1.4) was applied with special regard to the manufacturer's instructions. For blunt end ligation, the molar ratio of the cloning vector to the desired insert DNA has to be taken into account, whereby a 10:1 molar ratio of the CopyControl™ pCC1FOS vector (0) to insert DNA is recommended.

A 10 µl reaction volume was pipetted in a PCR tube according to the following scheme:

Blunt end DNA (~ 250 µg of approx. 36 Kb DNA)	1.5 µl
10 x Fast-Link Ligation Buffer	1 µl
10 mM ATP	1 µl
CopyControl pCC1FOS vector (500µg/µl)	1 µl
Fast-Link DNA Ligase	1 µl
H_2O	4.5 µl

The reaction mixture can be scaled-up, if the ligation effectiveness is limited due to low quality of insert DNA. To calculate the number of clones (N) necessary to cover the whole *N. pusilla* genome by 36 kb fragments with a given probability (P), the following equation was applied:

$$N = \ln(1-P) / \ln(1-f) \qquad P = 99.9\ \%$$

$$N = \ln(1-0.99) / \ln(1-0.0045) = 1021 \qquad f = 36\ kb/8 \times 10^3\ kb = 0.0045$$

$$N = \ln(1-0.99) / \ln(1-0.0040) = 1149 \qquad f = 36\ kb/9 \times 10^3\ kb = 0.0040$$

Figure 3.1 Formula to calculate the number of clones required for fosmid library

In this formula, f is the proportion of the genome present in one clone (Clarke & Carbon, 1976). The size of myxobacterial genomes varies from 8 to 13 Mb (*Myxococcus xanthus*, Goldmann *et al.*, 2006; *Sorangium cellulosum*, Schneiker *et al.*, 2007[1]; *Plesiocystis pacifica*, Shimkets *et al.*, 2010, *Stigmatella aurantiaca*, Huntley *et al.*, 2011[2]). As most myxobacterial genomes have genome sizes between 8 and 9 Mp, we estimate *N. pusilla* B150 to range in this area and thus the number of

required clones was estimated to vary between 1021 and 1149 clones. Based on empirical experiences, this value was doubled, and thus around 2,300 clones should be sufficient to represent the whole genome. According to manufacturer's data, a single ligation reaction is imputed to produce 10^3-10^6 clones. Therefore, only one reaction mixture was set up for ligation. The mixture was incubated overnight at 4 °C. Subsequently, the Fast-Link™ DNA Ligase was inactivated for 10 min at 70 °C. The concatemer formation during ligation is essential for the following lambda packaging step into phage coats (3.2.13.6).

3.2.13.5 Preparation of EPI300-T1R competent *E. coli* cells

The provided *E. coli* strain TransforMax™ EPI-300™, supplied as a glycerol stock solution, was plated on a LB agar plate (without antibiotics). The colonies were grown overnight at 37 °C in an incubator (table 3.2). A pre-culture was prepared in an Erlenmeyer flask containing 5 ml of sterile 100 mM $MgSO_4$ solution and 45 ml of sterile LB medium. After inoculation with a single colony of the plating bacterial strain, the culture was shaken overnight at 160 rpm and 37 °C. A second Erlenmeyer flask contained 40 ml of autoclaved LB medium supplemented with 5 ml of 100 mM $MgSO_4$ solution and 5 ml of sterile filtrated 2 % maltose solution was prepared. The medium was inoculated with 5 ml of the EPI300-T1R overnight pre-culture and shaken at 180 rpm at 37 °C until an OD_{600} of 0.8 was reached. The cells were then centrifuged and resolved in 35 ml of 10mM $MgSO_4$, resulting in an OD_{600} of 0.5. The competent cells could then be stored for up to three days at 4°C.

3.2.13.6 Packaging of fosmid clones

Just like cosmids, fosmids also contain at least one phage lambda cos site, recognized by λ phage particles. Concatemerized DNA, achieved through former ligation, may be then packaged by the λ phages of the MaxPlax™ Lambda Packaging Extracts (3.1.8). The infective phage particles were used to transfect special *E. coli* cells (TransforMax™ EPI-300™; 3.1.5). For the packaging reaction, the manufacturer's protocol was followed. For one ligation reaction, one tube of the MaxPlax™ Lambda Packaging Extracts was thawed on ice. Immediately, 25 µl were transferred to a 1.5 ml Eppendorf tube, which was subsequently stored at -80 °C until use. The ligation reaction (10 µl) was added to the remaining 25 µl of the MaxPlax™ Packaging Extract, which was kept on ice for the whole time. The solution was then mixed by gentle pipetting (with cut tips), avoiding the introduction of air bubbles, as

this might damage the phages. The mixture was held at 30 °C for 1.5 hours in an incubator (3.1.1). This incubation step was repeated after thawing and adding the second half of the MaxPlax™ Packaging Extract. The mixture was then diluted in Phage Dilution Buffer (PDB) (3.1.12) to a final volume of 1 ml. 25 µl of chloroform were added to prevent bacterial contaminations and the tube was inverted gently several times. The final solution was then stored at 4 °C until use. For long term storage at -80°C, sterile glycerol was added to a final concentration of 20%.

To determine an appropriate titer for easier clone picking, a 1:10, 1:50 and a $1:10^4$ dilution of the packaging reaction were made in PDB. 100 µl of each dilution were mixed with 100 µl of competent EPI300-T1R cells (3.2.10).The mixture was incubated for 30 min at 37 °C to allow the phage particles to infect the *E. coli* cells and to transfer the linear fosmid DNA to the cell interior by transduction. The infected EPI300-T1R cells were spread on LB agar plates containing the selective marker chloramphenicol in a final concentration of 12.5 µg/ml. The plates were incubated at 37 °C overnight. Inside the *E. coli* cells, the cohesive ends (cos sites) enable the circulization of the linear DNA to form a plasmid bearing a 36 kb fragment of the genomic DNA of *N. pusilla* B150. In contrast to cosmids, the pCC1FOS vector does not have the drawback of partial deletion of the insert DNA, because it contains a single copy origin of replication based on the F plasmid. A second origin of replication oriV requires induction of desired clones to high copy number to obtain high yield of DNA. This procedure is described in 3.2.15.

3.2.13.7 Plating the genomic library

For each dilution prepared in 3.4.14.6, the colonies per LB plate were counted to determine the appropriate amount of host cell suspension for plating. The 1:10 dilution yielded about 100 colony forming units (cfu) per plate, and therefore this dilution was chosen to plate the fosmid library. Altogether, 2,290 colonies were cultivated on LB plates containing chloramphenicol (12.5 µg/ml). 29 microtiter plates, each having 96 wells, were prepared to contain 100 µl of LB liquid medium and 12.5 µg/ml chloramphenicol in every well using the Transferpette®-8 multipipette (table 3.2). The clones were picked and suspended in the wells of the microtiter plates using tooth picks. Afterwards, the plates were incubated overnight at 37 °C. Copies of the produced cultures were made using the Boekel Replicator (table 3.2). For long term storage, an equal volume of 100 % glycerol was added to the bacterial culture

and mixed by pipetting applying the Transferpette®-8 (table 3.2) prior to freezing at -80° C. To facilitate pipetting of the highly viscous glycerol the eppendorf pipette tips were cut prior to sterilization.

3.2.14 Screening of the fosmid gene library

The fosmid gene library was screened by whole cell PCR for the presence of NRPS A domain positive clones using specific primer pairs designed from PCR fragments, obtained in performing a PCR with general NRPS A domain. In order to make the fosmid DNA accessible for the Taq-polymerase, a heating step has to be integrated in the standard PCR program described in 3.4.9.2. This denaturation step at the beginning of the PCR procedure was performed at 97 °C for 20 min to initiate cell lysis (whole cell PCR). To identify the gene encoding a phenylalanine activating NRPS A domain the primer pair A3rev.1/ LGDD.S was used in a PCR on genomic DNA (3.1.9). From the obtained sequence the primer pair NRPS-F1/NRPS-R1(table 3.12) was deduced and applied for the screening of the fosmid library. To facilitate fast screening of the library, all 96 clones of each microtiter plate were united in 0.5 ml Eppendorf tubes (p-pool). Standard PCR reactions (3.2.7) were set and performed with 1 µl of each pool at the appropriate annealing temperature including an initial heating step for cell breakage as described above. PCR probes were subsequently analyzed by agarose gel electrophoresis (3.2.5). Localization of single clones in positive pools was achieved by splitting the p-Pool into 8 sub-pools each containing 12 clones (the clones of one line of the microtiter plate). Subsequently, the right clone was localized in one of the 8 pools by PCR of the 12 single clones (one line).

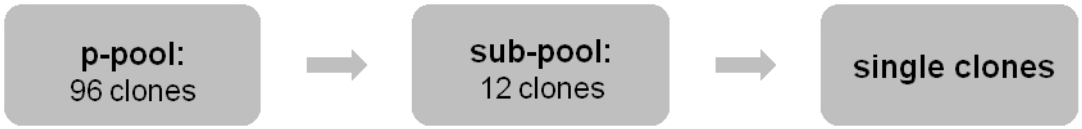


Figure 3.2 Screening schemata: 96 clones were combined in a plate pool (p-pool), a positive p-pool was then divided in 8 sub-pools à 12 clones. In case of a positive sub-pool the 12 single clones were analyzed respectively.

3.2.15 Induction and isolation of identified positive clones

For the isolation of great amounts of fosmid DNA from PCR positive clones, it was necessary to induce a promoter (*ori*V) by addition of the CopyControl™ induction solution, which was supplied with the CopyControl™ Fosmid Library Production Kit (3.1.4). This induction step initiates expression of the *trfA* gene product, which enables the amplification of fosmid clones at high copy number. A pre-culture was set by supplementing 5 ml of LB medium with 12.5 mg/ml chloramphenicol in a test tube. After inoculation with a single fosmid clone, the culture was shaken overnight at 180 rpm and 37 °C. 45 ml of fresh, sterile LB medium containing 12.5 mg/ml chloramphenicol were mixed with 5 ml of the pre-culture. 50 µl of the 1000 x CopyControl™ induction solution (Epicentre, Madison, U.S.A.) were added to the mixture and subsequently shaken overnight at 30 °C and 180 rpm. Fosmid isolation was accomplished by applying the QIAgen Plasmid Midi Kit following the manufacturer's instruction manual (3.1.4).

3.2.16 Genome sequencing using the 454 sequencing procedure

Within this study, results of a 454 sequencing performed for the *N. pusilla* B150 genome were analyzed (4.5). The sequencing was carried out by GATC Biotech AG (Konstanz, Germany) using 5 µg of DNA. In this method, the genomic DNA is fragmented, ligated to universal adapter primers and subsequently linked to beads at a dilution that promotes one DNA molecule per bead. These bead-DNA complexes are then encapsulated into emulsion based droplets. By emulsion PCR, the DNA fragments are enriched within these droplets. The beads are then allotted to individual picotiterplate wells, in which a bioluminescence method, referred to as pyrosequencing is performed. In this process, nucleotides are not added coincidently as done in dideoxy sequencing methods (Sanger method and variants thereof) but successively. If a matching nucleotide is added, a cascade of enzymatic reactions visualizes the formation of ATP from pyrophosphate via a light emitting conversion of luciferin into oxyluciferin (Metzker, 2010, Sanger *et al.*, 1977).

3.2.17 Isolation, cultivation and long term storage of myxobacteria

The myxobacterial strain 150 was isolated by B. Ohlendorf from a soil sample, collected from the intertidal region of Crete (Ohlendorf *et al.*, 2008). For bacterial growth 250 µl of a kryo-culture of *N. pusilla* B150 was placed on VY/2 agar plates. For cultivation in liquid medium, slices of agar, containing fruiting bodies of the

myxobacterium, were inoculated in MD-1 glucose medium (table 3.13) and grown at 30 °C for 7 days under shaking conditions of 140 rpm. For long term storage the liquid bacterial cultures were mixed with an equal volume of 1 % casitone. This mixture was stored at -80°C.

3.2.18 Cultivation and long term storage of recombinant *E. coli* strains

For plasmid preparations *E. coli* strain XL1-Blue was cultivated in LB medium containing ampicillin (100 µg/ml) at 37 °C and 180 rpm for 12 h. *E. coli* XL1 Blue MR was cultivated in LB medium at 37 °C and 180 rpm for about 4 h. For long term storage an equal volume of 100 % glycerol was added to the bacterial culture, and the mixture was stored at – 80 °C.

3.2.19 Determination of bacterial cell density

The bacterial growth in liquid culture was followed by measuring the optical density at 600 nm (OD_{600}) with a spectrophotometer (UV mini 1240, table 3.2). The optical density or turbidity of a cell suspension is directly related to the cell number. An OD_{600} of 1.0 at an optical path length of 1 cm corresponds to approximately 8 x 10^8 cells (Eckert and Kartenbeck, 1997). The sensitivity of this method is limited to about 10^7 cells per ml for most bacteria.

3.2.20 Determination of nucleic acid concentration and purity of DNA

To quantify obtained nucleic acids and determine the quality of DNA solutions, the OD was measured at the wavelengths 260 nm (A_{260}) and 280 nm (A_{280}) using a UV/VIS spectrophotometer (table 3.2). The OD value at 260 nm (A_{260}) refers to the amount of DNA or RNA in the solutions to be quantified. The Lambert-Beer law describes the linear relation between the concentration of an absorbing material and its light absorption at a given wavelength. This linear correlation is only valid for absorption values between 0.2 and 1.2. Therefore, highly concentrated DNA solutions had to be diluted to an appropriate measurable concentration. The true concentration of the original nucleotide solution was then extrapolated by including the dilution factor (DF) in calculations. The absorption coefficient is a material-dependent parameter that describes the specific absorption of the respective material. For solutions containing double stranded DNA, an average extinction coefficient of 50 (µg/ml)-1 cm-1 was applied (Sambrook & Russell, 2001).

The concentration of nucleic acids was calculated according to the following equation:

$$conc._{nucleic\ acid} = DF \times A_{260} \times 50\ \mu g/ml$$

Aromatic amino acids of proteins show significant absorption at 280 nm and therefore the ratio of A_{260}/A_{280} is a common value to specify the purity of a DNA solution. A ratio value of 2.0 indicates 100% nucleic acids and 0 % proteins. A_{260}/A_{280} values between 1.8 and 2 are favorable for DNA used in downstream applications.

3.2.21 Heterologous expression of proteins

For the in vitro investigation of protein functions the protein of interest was overexpressed in a heterologous host. 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 6xHis (consisting of 6 successive histidines). For the expression vector pet151/D-TOPO® the 6xHis-tag is expressed at the N-terminus of the protein, which facilitates the purification via affinity chromatography (3.6.3).

The DNA sequence of the protein of interest was amplified by PCR (3.2.7) using a *pfu*-Polymerase to generate a PCR product with blunt-ends. The pet151/D-TOPO® vector is a topoisomerase I activated expression vector that enables fast and directional cloning of blunt-end PCR fragments. The forward primer for the AMP-ACP PCR product was set at the start of the ORF, 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 One Shot® TOP10 Chemically Competent *E. coli* cells (3.1.5). To exclude the possibility of mutations in the reading frame, the chosen clones were submitted for sequencing. Plasmid DNA was isolated from confirmed clones and transformed into BL21 Star™(DE3) One Shot® Chemically Competent *E. coli* cells. For a pre-culture, 10 ml LB medium with 100 µl/ml ampicillin were inoculated with 500 µl of transformed cells and grown overnight at 37 °C. For the main culture, 100 ml LB medium containing 100 µl/ml ampicillin were inoculated with 2 ml of the overnight culture. After inoculation, the main culture was grown to an OD_{600} of ~ 0.5. Protein expression was induced by adding IPTG to a final concentration of 0.1–1.0 mM. Generally, cultivation was carried out for 4h at 18-37°C or overnight at 16°C.

3.2.22 Cell lysis by sonication

Cells from protein expression were harvested by centrifugation at 8,500 rpm for 5 min at 4°C using 50 ml centrifugation tubes (3.1.1). The cell pellet was resuspended in 2-5 ml lysis buffer 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 4-5 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 10 min at 8,500 rpm and 4°C. The supernatant containing the soluble proteins and the pellet were collected.

3.2.23 Purification of recombinant proteins by Ni-NTA-columns

Proteins harboring 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 with 4 ml lysis buffer. The lysate was applied on the column and allowed to pass the matrix. The column was washed twice with 4 ml washing buffer. Elution was carried out in five steps each with 500 µl of elution buffer with 100, 150, 200, 300 and 500 mM imidazole. All fractions were collected in 1.5 ml eppendorf tubes and stored on ice to prevent protein degradation.

3.2.24 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 polymerization of bis-acrylamide to polyacrylamide, which forms a molecular sieve. The concentration of polyacrylamide can be chosen in dependence of the analyzed protein size. For a better focusing 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 polymerization 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 1.5 mm. During polymerization, the separating gel was covered with isopropanol to ensure a smooth surface. Isopropanol was removed before the addition of the stacking gel.

SDS-stacking gel (5%)		**SDS-separating gel (12%)**	
Tris/HCL pH 6.8 (1M)	375 µl	Tris/HCL pH 6.8 (1M)	2500 µl
SDS (10%)	30 µl	SDS (10%)	100 µl
Bis-acrylamide (30%)	510 µl	Bis-acrylamide (30%)	4000 µl
Water	2040 µl	Water	3300 µl
APS (10%)	30 µl	APS (10%)	100 µl
TEMED	3 µl	TEMED	4 µl

For each gel run, the reservoirs of the electrophoresis assembly were filled with fresh 1x SDS gel electrophoresis buffer. Protein samples were mixed with 4x denaturing loading buffer and boiled at 90°C for 5 min, before loading them on the gel. Electrophoresis was performed in a XCell SureLock® Mini-Cell (3.1.1) at 100 V until the samples reached the separating gel, then voltage was increased to 130 V. As a reference, a molecular size marker (3.1.3) was loaded in one well. After electrophoresis the gels were analyzed by coomassie-staining.

3.2.24.1 Coomassie-staining

Staining with coomassie brilliant blue was used to analyze the protein bands on a polyacrylamide gel. Gels were immersed in the staining solution (3.1.13) and shortly heated in a microwave oven for quicker staining and subsequently incubated several minutes on a horizontal shaker. Destaining of the background colour was performed by shaking with destaining solution (3.1.13) for several hours or even over night. The gels were documented with the Intas iX Imager (3.1.1).

3.2.25 Concentration of purified proteins and buffer exchange

Imidazole, which was used for the elution of the protein from Ni-NTA columns, may interfere with following enzyme assays. Thus, Amicon Ultra centrifugal filters were used to remove imidazole from the purified protein from heterologous expression (3.2.21) and to concentrate the latter. The exchange of buffer was performed in several centrifugation steps with 50 mM Tris/HCl (pH8), following the manufacturer's instructions. With a start volume of 3.5 ml purified protein (3.2.23), the sample was concentrated with Amicon Ultra centrifugal to 500 µl of purified protein.

3.2.26 ATP- PP_i exchange assay

As the adenylation of amino acids requires ATP, the functionality of the A domain can be investigated with an ATP-PP_i exchange assay, described by Phelan and colleagues in 2009. It measures the consumption of γ-$^{18}O_4$-labelled ATP and the formation of $^{16}O_4$-ATP by an excess of unlabelled PP_i. The resulting mass shifts are detected by MALDI-TOF-MS. 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 with the help of Max Crüsemann in the laboratory of Jörn Piels' group (Kekulé-institute for Organic Chemistry and Biochemistry, University of Bonn, Germany) according to the described method. For this purpose 2 µl of the purified A domain (3.2.25) were incubated with 1 mM γ-$^{18}O_4$-ATP, 5 mM PP_i, 5 mM $MgCl_2$ and 1 mM substrate, in a reaction volume of 6 µl for 90 min at RT. Exchange rate by MALDI-TOF-MS was determined by comparison of the ratio of γ-$^{16}O_4$-ATP (m/z 506) to the sum of all ATP species, including unlabelled (m/z 506), partially labeled (m/z 508, 510, 512), fully labeled (m/z 514), and monosodium-coordinated ions (m/z 528, 530, 532, 534, 536). Determination of percent exchange was normalized with the following modifier:

$$\%_{exchange} = (100/0.833) \times {}^{16}O/({}^{18}O+{}^{16}O)$$

4 Results

4.1 Isolation and cultivation of the myxobacterial strain 150

The bacterial strain 150, collected from the intertidal region of Crete, was previously isolated on WCX / *E. coli* plates and identified as a bacteriolytic myxobacterium (Ohlendorf, 2008). This bacterium with the strain collection number 150 showed some features that indicated that the microorganism belongs to the genus of *Nannocystis*. Some of these prominent characteristics include the almost spherical form of the vegetative cells, the ability to corrode agar and to swarm in colonies. 16S rDNA analysis was performed to identify this myxobacterial strain on the species level (4.3).

The myxobacterium was cultivated in MD1 liquid medium supplemented with glucose (Behrens *et al*., 1976) (3.1.11). It grows slowly in form of orange spherical agglomerates and is best harvested for DNA isolation after approximately seven days of cultivation in liquid medium.

4.2 Isolation of genomic DNA from myxobacterial strain 150

Genomic DNA from the bacterium was isolated for the purpose of taxonomic identification, PCR screening studies, genomic library construction and whole genome sequencing. DNA isolation turned out to be extremely tedious, resulting mostly in sheared, small DNA fragments.

Several molecular biological kits were tested, as well as the standard isolation procedure with phenol-chloroform (3.2.3). The tested molecular biological kits from Promega, Machery-Nagel, Fermentas, Invisorb and Analytic Jena (3.1.4) resulted mostly in sheared DNA, too small for the construction of a genomic library or whole genome sequencing. Even modifying the duration and speed of the centrifugation step did not improve DNA quality, when using molecular biological kits.

The phenol-chloroform extraction technique resulted in less sheared DNA, especially after modifying this procedure. Decreased centrifugation speed and cut tips were the key to larger DNA fragments, crucial for the establishing of a fosmid library. A

modified protocol of the standard procedure was therefore used for the isolation of genomic DNA (3.2.3).

4.3 Strain identification of myxobacterial strain 150

For the phylogenetic identification of the investigated strain, a 16S rDNA analysis was carried out. For this the primer pair pA/pH (Edwards *et al*. 1989) was used in a PCR with genomic DNA of the myxobacterial strain (3.2.3). The so obtained 1.5 kb DNA fragmentwas then analyzed using nBLAST (3.1.14). The 16S rDNA sequence (8.4) of the myxobacterial strain was closely related to that of *Nannocystis pusilla* strain DSM 14622T (GenBank accession: FR749907.1) with an identity between the two sequences of 98%. This outcome supports the previous classification (4.1; Ohlendorf, 2008) of this strain to the genus of *Nannocystis*. Whereas B. Ohlendorf however suggested the species to be *N. exedens*, the current result points towards *N. pusilla* B150. Construction and screening of a fosmid library

4.4 Fosmid library production

To elucidate and investigate the complete sequence for the biosynthetic gene cluster of phenylnannolone A, the construction of a genomic library was necessary (3.2.13).In general the required number of recombinants in a library depends on the genome size of the particular organism and on the insert size of the DNA cloned into the fosmids (3.2.13.4).The genome size of myxobacteria varies between 8 and 13 Mb (*Myxococcus xanthus*, Goldmann *et al.*, 2006; *Sorangium cellulosum*, Schneiker *et al.*, 2007; *Plesiocystis pacifica*, Shimkets *et al.*, 2010, *Stigmatella aurantiaca*, Huntley *et al.*, 2011), and most of the already sequenced myxobacteria have genome sizes between 8 and 10 Mb. Thus, it was estimated that 2,300 clones were sufficient for a complete coverage of the genome (Sambrook & Russell, 2001; Clarke & Carbon, 1976). It was decided to use a fosmid vector, i.e. pCC1FOS, for the library production. Fosmid vectors contain in contrast to cosmid vectors two origins of replication, the F plasmid, a single-copy origin of replication, and oriV, the high-copy origin of replication. The advantages of these two features are on the one hand to maintain the clone with a single-copy of the recombinant, and on the other hand there is the possibility to induce positive clones to a high copy number to obtain high yields of fosmid DNA.

During the DNA isolation procedure the chromosomal DNA was sheared, so that the isolated DNA revealed a size of approximately 35-40 kb. This was verified by agarose gel analysis (3.2.5) using fosmid control DNA (3.1.3), supplied with the CopyControl™ Fosmid Library Production Kit (3.1.4), as a marker. DNA fragments smaller than 25 kb should be avoided, as the possibility of chimeric clones is then increased. This is undesired as it may impede the localization and mapping of genes (see manufacturer's manual). Another issue that makes small DNA inserts undesirable is that they remain unrecognized by phages, and thus would skirt the process of packaging DNA pieces into phage particles.

Size selecting the sheared DNA by excising it from a LMP agarose gel (figure 4.2) (3.2.13.1) and subsequent purification (digest of the agarose and precipitation of DNA) led to DNA fragments of approximately 36 kb. The following end repair of the size selected DNA fragments resulted in blunt ends, required for the ligation into the pCC1FOS vector.

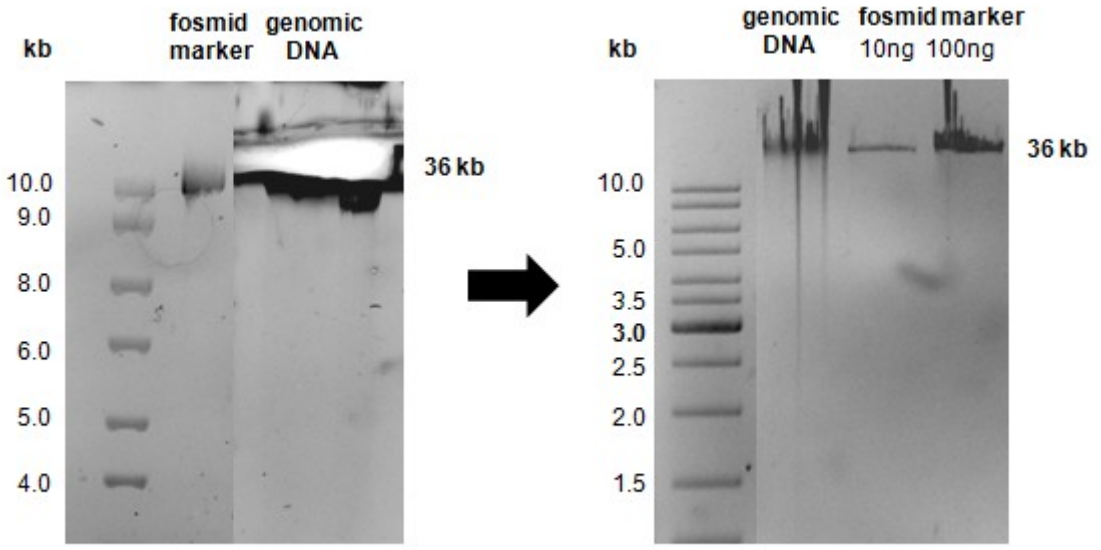


Figure 4.1 Gel extraction from LMP agarose: 120 µl of genomic DNA of *N. pusilla* B150 were size selected with 36 kb fosmid control DNA as marker (left picture). The right picture shows purified and end repaired genomic DNA fragments with an average size of 36 kb. The 36 kb fosmid control DNA marker was loaded on the gel in concentrations of 10 ng and 100 ng to estimate of the preparated DNA concentration.

After several attempts the blunt end ligation of the prepared genomic DNA fragments into the fosmid vector was successful. Several modifications in duration and temperature in the ligation step were tested as the described conditions of two hour

ligation at room temperature did not result in a successful ligation. Slow ligation over night at 4°C seemed to solve the problems with the ligation step (3.2.13.4).

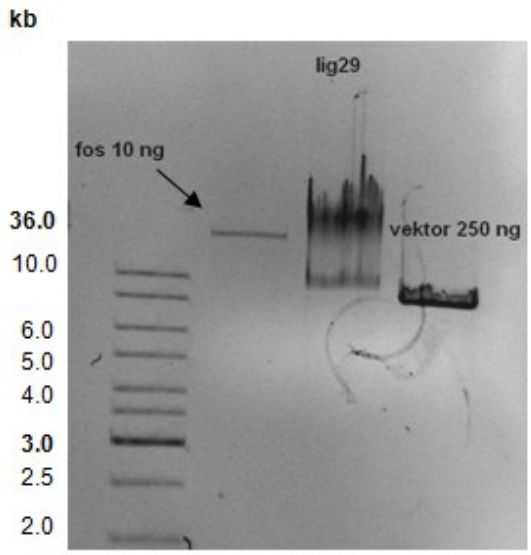


Figure 4.2 Agarose gel electrophoresis of the successful ligation (lig29) of *N. pusilla* B150 DNA into pCC1FOS vector. To detect a successful ligation the fosmid control DNA marker (36 kb) and the pCC1FOS vector were also applied to the gel besides the ligation reaction.

The ligation reaction was stopped by heat inactivation and then packaged into the phage heads using the MaxPlax™ Lambda Packaging Extracts in following the corresponding instructions (3.2.13.6). To determine the adequate titer of the phage particles, the transfection of competent EPI300™-T1R *E. coli* cells with neat and diluted phage particles was necessary. The infected *E. coli* cells were spread on LB plates and selected with chloramphenicol, as the fosmid vector carries a gene (*cat*) for resistance (3.1.7). In counting the number of clones on each plate, the titer of a 1:10 dilution, calculated in cfu/ml (Colony forming units), could be set as 10^5 cfu/ml.

A number of 100 clones per plate was desirable, to facilitate the isolation and storage preparations for the fosmid clones. This was achieved in using the 1:10 dilution of the phage particles, so that 2,300 fosmid clones with an average insert length of 36 kb were generated from the *N. pusilla* B150 genome.

4.5 Genome sequencing

A step towards the identification of the gene cluster for phenylnannolone A biosynthesis, was a whole genome sequencing, performed by GATC biotech (3.2.16). Sequence data were obtained with the Roche Genome Sequencer FLX Titanium (F. Hoffmann-La Roche AG, Basel, Switzerland). Raw sequences were assembled and resulted in 3804 contigs with an average contig size of 3 kb. The largest contig assembled had a size of 68 kb. The number of aligned bases was 166.220.937. These data would offer a 16.6 fold coverage of the *N. pusilla* genome with the imputed genome size of 10 Mb (4.4.1). The contigs were analyzed via BLAST search. In this analysis eleven contigs containing gene sections encoding NRPS adenylation (A) domains were identified. Subsequently their substrate specificity was predicted with the NRPSpredictor (3.1.14).

Table 4.1 Selected and relevant contigs from 454 sequencing of the *N. pusilla* genome. BLAST / CLUSEAN search analysis revealed the putative function of the gene products for these contigs. The substrate specificity for the NRPS adenylation (A) domains was predicted by the NRPSpredictor / CLUSEAN. KS=ketosynthase domain PCP= peptidyl carrier protein, C= condensation domain, TE=thioesterase domain.

Contig	Size (kb)	Putative function	Putative specificity of A domain
1153	6.4	PCP , C , A , PCP, C , A	Leu; Tyr (His)
3732	4.9	A , PCP , C , A	Ser; Tyr (Val)
770	5.0	C , A , PCP , C	Ser
1648	1.9	A , PCP , C	Ser (Cys)
544	4.7	A , PCP , TE	no prediction
845	7.1	A , PCP , TE	Ile
3177	1.1	A	Cys
2768	0.7	A	Pro (Cys)
3602	1.8	C , A	Tyr (Val)
1592	1.5	A , PCP	Phe
1241	3.0	A , PCP, KS	Pro

Only one of the examined contigs, contig 1592, comprised sequences for an A domain specific for the activation of phenylalanine and considered most interesting in

the light of the phenylnannolone A structure. Further analysis for PKS genes in the 454 assembly revealed contigs with DNA information for ketosynthase, acyltransferase, dehydratase, ketoreductase and thioesterase domains, which is listed in table 4.2. As the contigs had only a size of 1-2 kb, they encoded merely one or two PKS domains.

Table 4.2 Analysis of 454 assembly on PKS genes: PKS genes were grouped according to their corresponding function.

PKS domain	Ketosynthase	Acyltransferase	Dehydratase	Ketoreductase	Thioesterase
Number of contigs	**15**	**9**	**4**	**4**	**3**

The presence of three PKS-TE domains indicates that at least three gene clusters, encoding for either a PKS or a NRPS/PKS hybrid, are present on the *N. pusilla* genome.

Feeding experiment with ^{13}C labeled acetate and butyrate suggested large parts of the phenylnannolone A structure to be derived from a PKS (Ohlendorf *et al*., 2008). Thus, the last domain of the biosynthetic machinery is probably a PKS-TE domain. As the analysis of the 454 assembly revealed three contigs encoding PKS-TE domains, the domain order of these contigs was examined (table 4.3).

Table 4.3 PKS domain order of contigs encoding for PKS-TE domains

Contig	PKS domain order
1642	**AT – ACP - TE**
3326	**AT – ACP - TE**
1540	**KR – ACP - TE**

The domain order on contig 1540 with a ketoreductase domain (KR), an acyl carrier protein (ACP) and a thioesterase domain was coincident with our hypothesis of the phenylnannolone A biosynthesis. The domain order of the other contigs identified to encode for a PKS-TE did not fit the hypothesis. This is due to the fact that they lack a reductive domain, necessary for the formation of the last double bond in of phenylnannolone A biosynthesis.

4.6 Screening for NRPS/PKS encoding genes in the fosmid library

Previous feeding experiments (Ohlendorf et al., 2008)[8] with ^{13}C labeled precursors showed the incorporation of acetate and butyrate, but also of phenylalanine into phenylnannolone. Thus, a mixed NRPS/PKS system was suggested to be responsible for the formation of this natural product.

A degenerated primer pair LGDD.S/A3rev.1 (3.1.9) deduced from the alignment of several sequences encoding for NRPS-A domains was used for a PCR with genomic DNA of *N. pusilla* B150. Analyzing the so obtained sequences revealed a sequence for a phenylalanine activating A domain. From this sequence a specific primer pair NRPS-F1/ NRPS-R1 (3.1.9) was deduced, used for the detection of this genetic region in the fosmid library. The fosmid gene library was screened by whole cell PCR. To facilitate the screening procedure 98 fosmid clones were combined in a pool. For the entire fosmid library 24 pools were generated and screened in a first step. Subsequently, a positive pool was then analyzed in detail (3.2.14). Amongst the 2300 fosmid clones, two fosmid clones, 11A3 and 21H12, could be identified carrying the gene encoding for a phenylalanine activating A domain. Due to the NRPS/PKS nature of the suspected phenylnannolone A cluster, degenerated KS primers KS 1up/KS d1 (3.1.9) were used in a PCR with DNA from fosmid 11A3 and 21H12. For both fosmid clones distinct DNA fragments (700bp), identified to encode for a ketosynthase, could be amplified.

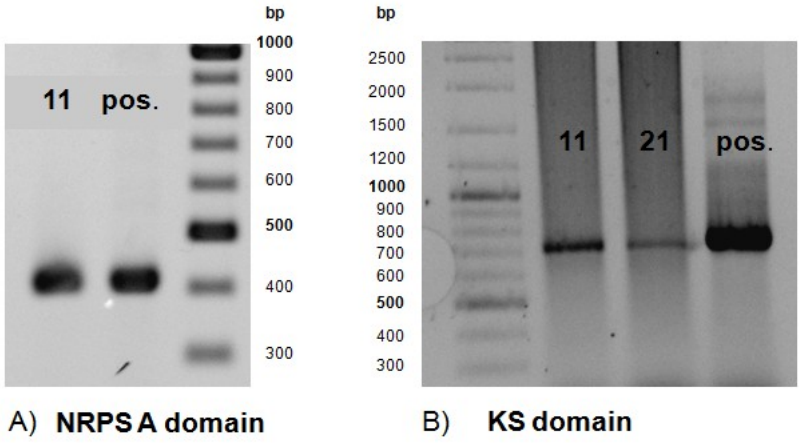
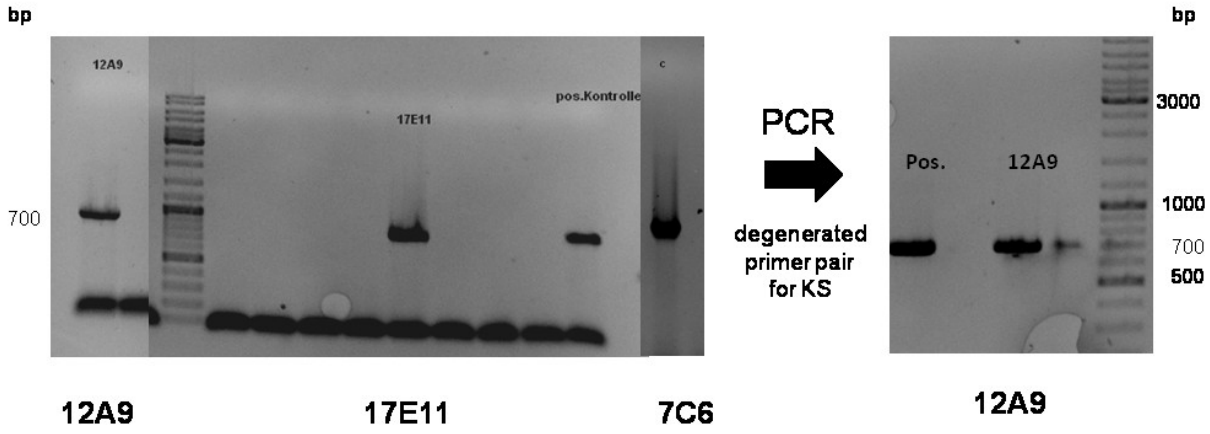


Figure 4.3 Fosmid library screening for NRPS A domains and PKS KS domains. A) PCR-screening for a NRPS A domain resulted in a PCR product with 400 bp for the positive control (pos.) and fosmid clone 11A3 (11). B) Fosmids 11A3 and 21H12 were screened with degenerated primers for a KS domain. Both fosmids harbor a KS, as amplificates of 700 bp were achieved.

The fosmid library was also screened for a thioesterase encoding gene. Such a gene had been identified on contig 1540 from the 454 genome sequencing (4.5). The fosmid library was thus screened with the specific primer pair 1540.F/1540.R (3.1.9) for the TE domain of contig 1540 and scored three fosmid clones: 12A9, 17E11 and 7C6. Those clones were then analyzed in a second PCR with the previously used KS primer (3.1.9) for existing ketosynthase domains, necessary for the PKS assembly line.



4.4 Screening the fosmid library for sequences encoding a TE domain: A specific primer pair was deduced from the sequence of a TE domain from contig 1540. PCR resulted in a product with 800 bp length, positive for three fosmid clones 12A9, 17E11 and 7C6. These fosmids were subsequently tested with degenerated primers for a KS domain, and only one fosmid clone, 12A9, was positive for a KS, as an amplificate of 700 bp was achieved.

For the fosmid clone 12A9 a specific DNA fragment could be amplified through PCR. Sequence analysis of construct pKS12A9, containing the 700 bp amplificate verified the presence of a ketosynthase gene section. Thus a further investigation on this fosmid clone followed.

4.7 Fosmid sequencing

4.7.1 End sequencing of fosmids

Since fosmids 12A9, 21H12 and 11A3 were found to contain NRPS and PKS encoding sequences, the end sequences of the myxobacterial DNA in these fosmids were analyzed. Therefore, an end sequencing reaction was performed at GATC Biotech AG. The investigated fosmid clones 12A9, 21H12 and 11A3 were sequenced with two primers, T7 a forward primer that binds to a T7 promoter site and Epi-RP a reverse sequencing primer (3.1.9). The obtained sequences were translated into amino acid sequences and analyzed via BLAST (3.1.14). The BLAST results are listed in the table below (table 4.4) showing the highest identity with homologous proteins from GenBank.

Table 4.4 BLAST search results for the end sequencing of fosmids 12A9, 21H12 and 11A3 with forward (T7) and reverse (Epi-RP) sequencing primers. The obtained sequences were searched for identities with contigs from the 454 assembly

Sequencing primer	Fosmid clone	Putative function	Identity of aligned amino acids	Genbank acession number	Contig
T7	12A9	SRPBCC superfamily protein	59/147 (40%)	YP_299911.1	2843
Epi-RP	12A9	SARP family transcriptional regulator	44/154 (29%)	YP_004813558.1	1429
T7	21H12	hypothetical protein	144/235 (61%)	YP_001612581.1	289
Epi-RP	21H12	succinate dehydrogenase, cytochrome b558 subunit	52/163 (32%)	EFW42511.1	none
T7	11A3	uncharacterized peroxidase-related enzyme	58/128 (45%)	ZP_06975623.1	1794
Epi-RP	11A3	hypothetical protein	54/94 (57%)	YP_001618118.1	289

For all of the three fosmid clones the end sequencing revealed that the borders of their DNA inserts are not part of genes involved in secondary metabolism, thus the conclusion was that the respective biosynthetic genes are completely located on the fosmids. The obtained sequences were searched for identities with contigs from the 454 assembly of the whole genome sequencing. Identical sequences for the border regions from fosmid 11A3 (Epi-RP) and 21H12 (T7) are present on contig 289, located in a distance of 800 bps from each other. As no contig could be identified for

the other end of fosmid 21H12, suggestions about the distance of those gene sections on the genomic level were not possible.

4.7.1.1 Comparison of fosmids 11A3 and 21H12

PCR studies showed for both of the fosmid clones, 11A3 and 21H12, the presence of genes for an NRPS adenylation and a PKS ketosynthase domain. Also, in both cases end sequencing of the fosmids evidenced that genes for the mixed NRPS/PKS system are not present at the borders of the fosmid insert and are completely located in midst of the DNA section.

To compare these fosmid clones with each other and to prove that they share nearly the same sequence information a restriction digest of the isolated fosmid DNA was undertaken (figure 4.5).

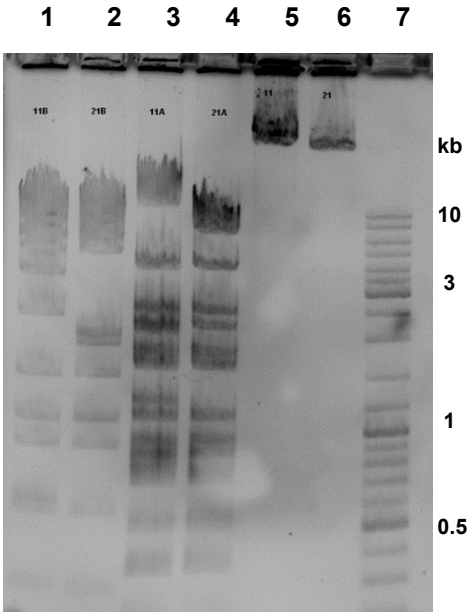


Figure 4.5 Restriction digest of fosmid 11A3 and 21H12

The agarose gel displays restriction digests of fosmid 11A3 with *Bam*HI (1) and *Apa*I (3) as well as the restrictions for the fosmid 21H12 with *Bam*HI (2) and *Apa*I (4) and the undigested fosmid DNA of both fosmids 11A3 (5) and 21H12 (6). The Gene Ruler^{TM} DNA Ladder Mix (7) is shown for size estimation.

The restriction digest showed many identical DNA fragments for the fosmid clones 11A3 and 21H12, thus only one fosmid clone, F21H12, was chosen to be further investigated.

4.7.2 Complete sequencing of fosmid 21H12 and 12A9

The contig sizes of the 454 genome sequences were too small to delineate the complete biosynthetic gene cluster. Thus, it was decided to sequence fosmid 21H12, shown to encode a NRPS A domain, specific for phe activation, as well as a KS domain, which suggested a mixed NRPS/PKS system. In addition to this another fosmid clone believed to harbor a PKS gene cluster was sequenced, i.e. fosmid 12A9, known to encode KS and TE domains. Complete sequencing of fosmid 21H12 and 12A9 was performed by IIT-Biotech GmbH (Bielefeld, Germany). For this purpose HT (high-throughput)-libraries with insert lengths of 500 bp were constructed and subsequently analyzed in a pyrosequencing reaction on a Roche GS FLX Titanium +. Sequence data previously identified from the 454 genome sequencing and PCR studies on the fosmid clones 21H12 and 12A9 (4.7.1.1) were supported and completed in the fosmid sequencing analysis. The sequences obtained in the pyrosequencing were assembled, and this resulted for fosmid12A9 in one contig of 45720 kb and for fosmid 21H12 in two contigs with 32221bp and 6755bp. To obtain a continuous sequence for the latter, PCR was performed to close the gap.

The sequence for fosmid 12A9 revealed 12 open reading frames (ORF) (4.8.1) of which two, *phn1* and *phn2*, are putatively involved in the biosynthesis of phenylnannolone A (4.8). The fosmid sequencing of clone 21H12 revealed *sb1*, a gene encoding for a NRPS/PKS mixed system of an unknown natural product (4.9).

4.8 Elucidation of the putative phenylnannolone A biosynthetic gene cluster

4.8.1 Sequencing results for fosmid 12A9

The sequencing result of fosmid 12A9 revealed 12 ORFs, of which two, *phn1* and *phn2*, are putatively involved in the biosynthesis of phenylnannolone A. These biosynthetic genes, i.e. *phn1* and *phn2* are surrounded by genes for hypothetical proteins of unknown function (*orf2* - *orf10*). A gene for a transcriptional regulator is displayed in *orf1*. The gene product for *phn1* showed in a BLAST analysis an identity of 70% to a ß-subunit of a propionyl-CoA carboxylase, known for acting as a carboxyl-transferase in precursor supply for fatty acid or polyketide biosynthesis. With an identity of 47% *phn2* could be identified as a gene encoding for a polyketide

synthase, consisting of a whole assembly line with five modules. The BLAST search results are shown in table 4.5

Table 4.5 BLAST search results for sequences of fosmid clone 12A9. The deduced amino acid sequences of ORFs found on fosmid 12A9 were aligned with homologous proteins by BLAST. The genes putatively responsible for the phenylnannolone A biosynthesis are highlighted and marked with bold letters.

Gene	Size (kb)	Highest homology (protein level)	Putative function	Identity of aligned amino acids[1]	GenBank accession number
orf1	1.7	SARP family transcriptional regulator [*Streptomyces violaceusniger* Tu 4113]	transcriptional regulator	44/154 (29%)	YP_004813558.1
orf2	1.7	hypothetical protein M23134_06735 [*Microscilla marina* ATCC 23134]	ANK superfamily SMI1-KNR4 family protein	48/117 (41%)	ZP_01694204.1
orf3	1.3	unnamed protein product [*Oikopleura dioica*]	ANK superfamily protein	35/106 (33%)	CBY24175.1|
orf4	0.5	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	SMI1-KNR4 family protein	86/124 (69%)	YP_001613106.1
orf5	0.6	hypothetical protein PPSIR1_23304 [*Plesiocystis pacifica* SIR-1]	unknown	43/103 (42%)	ZP_01910555.1
orf6	2.1	sigma-24 (FecI-like) protein [*Plesiocystis pacifica* SIR-1]	RNA-Polymerase	79/185 (43%)	ZP_01911328.1
orf7	0.3	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	unknown	42/99 (42%)	YP_001611977.1
orf8	0,8	hypothetical protein STIAU_4586 [*Stigmatella aurantiaca* DW4/3-1]	unknown	45/141 (32%)	ZP_01464984.1
orf9	0.8	hypothetical protein Lnog2_01990 [*Leptospira noguchii* str. 2006001870]	unknown	71/267 (27%)	ZP_09260961.1
phn1	**1.7**	**propionyl-CoA carboxylase, β-subunit [*Plesiocystis pacifica* SIR-1]**	**Carboxyl-Transferase**	**373/535 (70%)**	**ZP_01908434.1**
phn2	**22**	**beta-ketoacyl synthase [*Streptomyces violaceusniger* Tu 4113]**	**PKS**	**2174/4630 (47%)**	**YP_004817600.1**
orf10	1.2	unnamed protein product [*Ralstonia eutropha* JMP134]	SRPBCC superfamily protein	59/147 (40%)	YP_299911.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value)

4.8.2 Phn1, a Carboxyl Transferase

BLAST search analysis revealed for the gene product of *phn1* an identity of 70% to a putative ß-subunit of a propionyl-CoA carboxylase from *Plesiocystis pacifica* SIR-1. Indeed, all proteins in table 4.6, homologous to Phn1, are from myxobacterial origin and showed identities of 69-70%. Carboylases are known to act as a carboxyl-transferase in precursor supply for fatty acid or polyketide biosynthesis. The carboxyl transferases as acyl-CoA carboxylases are described to carboxylate either acetyl-CoA, propionyl-CoA or butyryl-CoA to malonate, methylmalonate or ethylmalonate, respectively. Acyl-CoA carboxylases accepting acetyl-CoA as a substrate are called acetyl-CoA carboxylase. Propionyl-CoA carboxylases accept several substrates, predominantly propionyl-CoA (Arabolaza *et al*., 2010).

Table 4.6 BLAST search results for *phn1*: The deduced amino acid sequences of *phn*1 were aligned with homologous proteins by BLAST.

Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
propionyl-CoA carboxylase, beta subunit [*Plesiocystis pacifica* SIR-1]	373/535 (70%)	435/535 (81%)	ZP_01908434.1
propionyl-CoA carboxylase subunit beta [*Myxococcus xanthus* DK 1622]	341/493 (69%)	407/493 (83%)	YP_629373.1
propionyl-CoA carboxylase beta chain [*Stigmatella aurantiaca* DW4/3-1]	343/493 (70%)	407/493 (83%)	ZP_01461629.1
propionyl-CoA carboxylase subunit beta [*Myxococcus fulvus* HW-1]	338/493 (69%)	407/493 (83%)	YP_004663613.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

There are three conserved residues in the beta subunit of propionyl-CoA carboxylases: glycine (419), aspartate (422) and serin (426). Important for the substrate specificity is the aspartate (D) residue. For *phn1* this residue comprises an alanine instead of the conserved aspartate, this mutation is discussed to direct substrate specificity towards butyryl-CoA (Arabolaza *et al*., 2010).

M. xanthus	G	A	Y	D	V	M	A	S	417
M. fulvus	G	A	Y	D	V	M	A	S	417
S. aurantiaca	G	A	Y	D	V	M	A	S	419
P. pacifica	G	A	Y	D	V	M	A	S	441
Phn1	G	A	Y	A	V	M	S	S	457
	*			*				*	

Figure 4.6. Multiple sequence alignment of Phn1 with amino acid sequences of propionyl-CoA carboxylases from *Plesiocystis pacifica* SIR-1, *Myxococcus xanthus* DK 1622, *Stigmatella aurantiaca* DW4/3-1 and *Myxococcus fulvus* HW-1. GenBank accession numbers are given in table 4.6. Conserved residues are highlighted and marked with an asterisk.

4.8.3 A Polyketide Synthase assembly line encoded by *phn2*

The BLAST analysis of the gene *phn2* revealed 47% identity to a beta-ketoacyl synthase from *Streptomyces violaceusniger* Tu 4113. The beta-ketoacyl synthase (KS) is a part of a PKS and responsible for the formation of polyketides.

Table 4.7 BLAST search results for the gene *phn*2. The deduced amino acid sequence for this gene was aligned with the respective homologous proteins by BLAST (3.1.14).

Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
beta-ketoacyl synthase [*Streptomyces violaceusniger* Tu 4113]	2174/4630 (47%)	2752/4630 (59%)	YP_004817600.1
beta-ketoacyl synthase [*Micromonospora sp*. L5]	2132/4607 (46%)	2694/4607 (58%)	YP_004085506.1
beta-ketoacyl synthase [*Micromonospora aurantiaca* ATCC 27029]	2127/4628 (46%)	2697/4628 (58%)	YP_003835580.1
short-chain dehydrogenase/reductase SDR [*Streptomyces bingchenggensis* BCW-1]	2157/4689 (46%)	2746/4689 (59%)	ADI05097.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets
[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

Phn2 has a gene size of 22 kb and its gene product consists of 5 modules: A loading module and four modules for the extension of the polyketide chain. The loading module comprises an AMP-ligase and an ACP domain, followed by the first extending module with a KS, AT, DH, KR and ACP domain. The same domain structure is given for the third module, whereas the fourth module lacks the reductive domains DH and KR. The last module shows again the same domain architecture as

module 1 and 2, with the difference of an additional downstream TE domain, responsible for the release of the molecule.

All domains encoded by *phn2* are examined in detail (see 4.8.3.1-4.8.3.7) and grouped according to their function, i.e. AMP-ligase, ACP, KS, AT, DH, KR and TE domains.

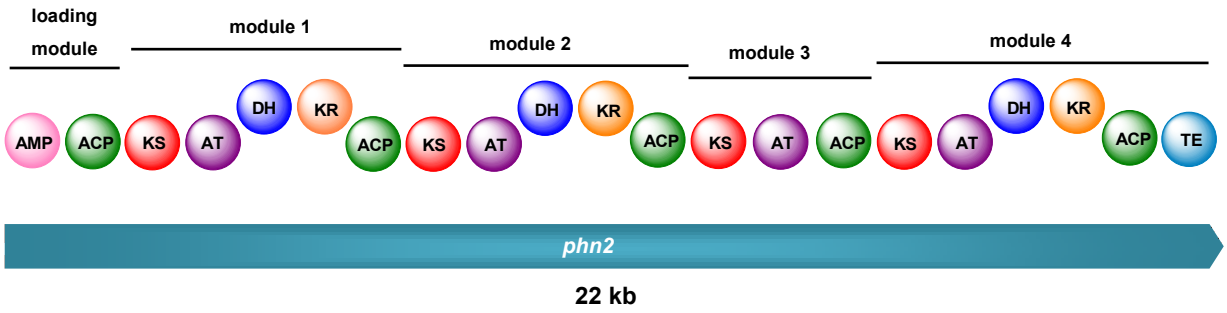


Figure 4.7 *Phn2*, a PKS gene encoding 5 PKS modules AMP=AMP-ligase, ACP=acyl carrier protein, KS=ketosynthase domain, AT=acyltransferase domain, DH=dehydratase domain, KR=ketoreductase domain, TE=thioesterase

4.8.3.1 AMP-ligase

The first domain encoded by the *phn 2* gene was analyzed by BLAST search as well as Pfam, and showed the highest homology to an AMP-dependent synthetase/ligase from *Nostoc punctiforme PCC 73102.* Detailed results are listed in table 4.8.

Table 4.8 BLAST search results for the first domain encoded by the *phn2* gene. The deduced amino acid sequence for this domain was aligned with the respective homologous proteins by BLAST (3.1.14**).**

Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
AMP-dependent synthetase and ligase [*Nostoc punctiforme* PCC 73102]	219/461 (48%)	285/461 (62%)	YP_001865690.1
AMP-dependent synthetase and ligase [*Cyanothece sp.* ATCC 51142]	211/461 (46%)	289/461 (63%)	YP_001805172.1
amino acid adenylation domain-containing protein [*Acetonema longum* DSM 6540]	221/462 (48%)	286/462 (62%)	ZP_08622841.1
Long-chain-fatty-acid--CoA ligase [*Cyanothece sp*. ATCC 51472]	210/459 (46%)	291/459 (63%)	ZP_08974929.1
AMP-dependent synthetase and ligase [*Cyanothece sp. ATCC* 51142]	208/449 (46%)	288/449 (64%)	YP_001801621.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets
[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

AMP-binding enzymes belong to the ANL superfamily (Pfam: CL0378) and catalyze the initial adenylation of a carboxylate to form an acyl-AMP intermediate. This reaction is followed by the formation of a thioester (Guilck, 2009). A multiple sequence alignment of homologous protein sequences reveals the presence of the typical three conserved motifs reported for the AMP-ligases.

The first highly conserved consensus region [T/S/G]-[S/G]-G-[S/T]-[T/S/E]-[G/S]-X-[P/M]-K-[G/L/F] (Chang et al., 1997) was identified in Phn2 as T-S-G-S-T-G-D-P-K-G (motif I).

Motifs II and III, which are [Y/L/W/F]-[G/S/M/W]-X-[T/A]-E and [Y/F/L]-[R/K/X]-[T/S/V]-G-D could also be identified as Y-G-M-A-E and L-R-T-G-D. Crystal structure analysis of the firefly luciferase, belonging to the same ANL superfamily, showed the

involvement of the three conserved motifs in substrate binding (Ingram-Smith *et al.*, 2006).

Motif I

C. sp. 51142	L	A	F	L	Q	Y	T	S	G	S	T	G	N	P	K	G	V	202
C. sp. 511472	I	A	F	L	Q	Y	T	S	G	S	T	G	N	P	K	G	V	203
N. punctiforme	L	A	F	L	Q	Y	T	S	G	S	T	G	T	P	K	G	V	182
A. longum	L	A	F	L	Q	Y	T	S	G	S	V	G	I	P	K	G	V	186
Phn2	L	A	T	L	I	Y	T	S	G	S	T	G	D	P	K	G	V	145

Motif II

C. sp. 51142	Y	P	C	Y	G	M	A	E	A	350
C. sp.511472	Y	P	C	Y	G	M	A	E	A	351
N. punctiforme	Y	P	C	Y	G	M	A	E	T	330
A. longum	Y	P	C	Y	G	M	A	E	A	334
Phn2	Y	P	C	Y	G	M	A	E	A	293

Motif III

C. 51142	F	L	R	T	G	D	L	G	463
C. 511472	F	L	R	T	G	D	L	G	466
N. punctiforme	F	F	R	T	G	D	L	G	441
A. longum	F	L	R	T	G	D	L	G	446
Phn2	Y	L	R	T	G	D	L	G	417

Figure 4.8 Multiple sequence alignment of amino acid sequences of AMP-synthetases and ligases from Phn1 (*N. pusilla*) , *Nostoc punctiforme, Cyanothece sp*. and *Acetonema longum*. Motif I-III show highly conserved regions for the AMP-synthetases/ligases. GenBank accession numbers are given in table 4.8

AMP-ligases and NRPS A domains are biochemically closely related in that both activate acids to their derivatives. For the NRPS A domains, there are apart from the three already mentioned motifs at least five additional ones (Konz and Marahiel, 1999). These could not be identified for the first domain encoded by *phn2* (see also 4.9.2.2).

The NRPSpredictor can be useful for the prediction of substrate specificity of A domains in NRPS, for the AMP-ligase of Phn2 no precise prediction could be made. Nevertheless, the analysis with the NRPSpredictor2 revealed the substrate of this enzyme to be probably similar to phenylalanine (40%, see Appendix).

4.8.3.2 Acyl Carrier Proteins (ACPs)

For the PKS gene product Phn2 five domains were identified as acyl carrier protein domains, namely ACP_{LM}, ACP_1, ACP_2, ACP_3 and ACP_4. ACP_{LM} is the acyl carrier protein of the loading module (LM), whereas ACP_1 to ACP_4 are the respective proteins of extension modules 1 to 4. ACP domains are responsible for the delivery of intermediates to the various catalytic domains. The conserved residue for the ACP is a serine, the putative 4'-phosphopantetheine binding site (Findlow *et al*., 2003). This could be identified for all of the ACP domains present in this PKS gene. BLAST results revealed identities to ACP domains from *Actinomycetes* (ACP_{LM}), *S. cellulosum* (ACP_1, ACP_2, ACP_3) and even to one of fungal origin (ACP_4). The respective results are listed in table 4.9 and table 4.10.

Table 4.9 BLAST search result for ACP_{LM} The deduced amino acid sequences of the ACP domain of the loading module was aligned with homologous proteins by BLAST.

Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
acyl transferase [*Catenulispora acidiphila DSM 44928*]	34/62 (55%)	74/124 (60%)	YP_003117673.1
lasalocid modular polyketide synthase [*Streptomyces sp*.]	35/69 (51%)	42/62 (68%)	ZP_07291842.1
polyketide synthase [*Streptomyces cattleya NRRL 8057*]	35/68 (51%)	48/68 (71%)	YP_004912779.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

ACP_3	G	M	N	S	L	M	A	V	Q	34
ACP_1	G	L	D	S	L	M	A	V	E	37
ACP_2	G	L	D	S	L	M	A	V	D	37
AmbB	G	L	D	S	L	T	A	V	E	46
JerB	G	L	D	S	L	T	A	V	E	46
ACP_{LM}	G	F	D	S	L	A	A	V	E	29
ACP_4	G	V	D	S	L	M	A	F	E	30
				*						

Figure 4.9 Multiple sequence alignment of amino acid sequences of ACP domains from *Streptomyces sp*. (JerB) and *Sorangium cellulosum* (AmbB). GenBank accession numbers are given in table 4.9 and table 4.10. Conserved residues are highlighted and marked with an asterisk. Suffixey indicate the module number. LM=loading module.

Table 4.10 Blast search results for the ACP domains of modules 1-4 encoded by the PKS gene *phn2*. The deduced amino acid sequences of the ACP domains were aligned with homologous proteins by BLAST

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
ACP_1	AmbB [*Sorangium cellulosum*]	47/74 (64%)	57/74 (77%)	ABK32256.1
	JerB [*Sorangium cellulosum*]	47/74 (64%)	57/74 (77%)	ABK32288.1
	JerD [*Sorangium cellulosum]*	46/70 (66%)	53/70 (76%)	ABK32290.1
ACP_2	JerB [*Sorangium cellulosum*]	46/74 (62%)	58/74 (78%)	ABK32288.1
	AmbB [*Sorangium cellulosum*]	46/74 (62%)	58/74 (78%)	ABK32256.1
	JerC [*Sorangium cellulosum*]	43/74 (58%)	56/74 (76%)	ABK32289.1
ACP_3	JerE [*Sorangium cellulosum*]	45/72 (63%)	59/72 (82%)	ABK32291.1
	TgaC [*Sorangium cellulosum*]	47/72 (65%)	59/72 (82%)	ADH04641.1
	polyketide synthase [*Sorangium cellulosum*]	45/70 (64%)	57/70 (81%)	CAL58682.1
ACP_4	Beta-ketoacyl synthase [*Cordyceps militaris* CM01]	22/48 (46%)	32/48 (67%)	EGX96860.1
	ATP-dependent serine activating enzyme [*Xanthomonas fuscans subsp. aurantifolii* str. ICPB 11122]	25/64 (39%)	41/64 (64%)	ZP_06703845.1
	hypothetical protein LEMA_P006610.1 [*Leptosphaeria maculans* JN3]	25/66 (38%)	40/66 (61%)	CBY01874.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

The analysis with the CLUSEAN Software (3.1.14) helped to answer the question whether the first ACP after the AMP-ligase, i.e. ACP_{LM}, is an acyl carrier protein, or a peptidyl carrier protein (PCP). With a score of 78.5 for this phosphopantetheine binding enzyme the domain was annotated as an acyl carrier protein. In contrast the score for a PCP domain was only 57.5.

4.8.3.3 Ketosynthase domains (KS)

The homodimeric KS domains in PKS assemblies are responsible for the formation of a carbon-carbon-bond. The mechanism for this reaction is a Claisen condensation that requires the triad C-H-H, catalyzing the steps of acyl transfer, decarboxylation and condensation (Zhang *et al.,* 2006).

For Phn2 four KS domains were identified with identities of over 60% when compared with corresponding enzymes of bacteria of the order *Actinomycetes*. The result of the BLAST analysis (table 4.11) showed for KS_1, KS_2 and KS_3 highest homologies to proteins of type I modular PKS found in bacteria belonging to the *Streptomycetes, Saccharomonospora and Saccharopolyspora*. For KS_4 high identities with homologous proteins of the modular PKS genes *tugA* and *tgaC* from *Chondromyces crocatus* and *Sorangium cellulosum*, all belonging to the order of *Myxobacteria*, were observed. Multiple sequence alignment of the deduced protein sequences of the KS domains proved the presence of the highly conserved catalytic triad C-H-H for all four KS domains (Zhang et al., 2006).

Figure 4.10 Multiple sequence alignment of amino sequences of KS domains from Phn2 and TugD (*Chrondromyces crocatus*). GenBank accession numbers are given in table 4.11. Conserved residues are highlighted and marked with an asterisk.

Table 4.11 BLAST analysis revealed homologous proteins for KS domains of Phn2 (KS_1-KS_4)

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
KS_1	type I PKS modular polyketide synthase [*Saccharopolyspora erythraea* NRRL 2338]	271/431 (63%)	327/431 (76%)	ZP_06568061.1
	lasalocid modular polyketide synthase [*Streptomyces lasaliensis*]	278/433 (64%)	334/433 (77%)	CAQ64692.1
	modular polyketide synthase [*Streptomyces bingchenggensis* BCW-1]	292/426 (69%	340/426 (80%)	ADI12313.1
KS_2	modular polyketide synthase [*Streptomyces bingchenggensis* BCW-1]	294/428 (69%)	346/428 (81%)	ADI12313.1
	modular polyketide synthase [*Streptomyces avermitilis* MA-4680]	287/429 (67%)	337/429 (79%)	NP_822727.1
	6-deoxyerythronolide-B synthase, partial [*Saccharomonospora paurometabolica* YIM 90007]	261/425 (61%)	310/425 (73%)	ZP_09035064.1
KS_3	modular polyketide synthase [*Streptomyces bingchenggensis* BCW-1]	291/426 (68%)	343/426 (81%)	ADI12313.1
	lasalocid modular polyketide synthase [*Streptomyces lasaliensis*]	276/426 (65%)	328/426 (77%)	CAQ64692.1
	6-deoxyerythronolide-B synthase, partial [*Saccharomonospora paurometabolica* YIM 90007]	262/425 (62%)	310/425 (73%)	ZP_09035064.1
KS_4	TugD [*Chondromyces crocatus*]	286/424 (67%)	328/424 (77%)	ADH04660.1
	TgaC [*Sorangium cellulosum*]	291/427 (68%)	327/427 (77%)	ADH04641.1
	polyketide synthase [*Sorangium cellulosum*]	285/424 (67%)	324/424 (76%)	CAL58684.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

A phylogenetic tree was constructed with Phylogeny.fr in a multiple sequence alignment of protein sequences of ketosynthase domains from *N. pusilla* (KS_1-KS_4) and a ketosynthase domain from TugD, the latter being part of a PKS from

Sorangium cellulosum. KS_2 and KS_3 show the highest similarity to each other. KS_1 is also very closely related to the latter, whereas KS_4 seems to be closer related to a KS domain from TugD.

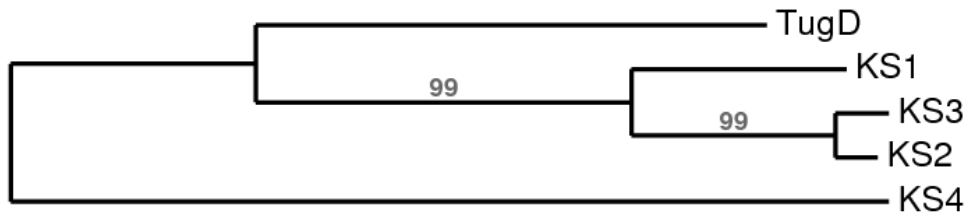


Figure 4.11 Phylogenetic tree of KS domains obtained from a multiple sequence alignment with phylogeny.fr. KS domains from *N. pusilla* (KS_1-KS_4) and from *S. cellulosum* (TugD) were aligned in a multiple sequence alignment. The distance of tree arms shows the respective phylogenetic relation.

4.8.3.4 Acyltransferase domains

Acyltransferase domains (AT) are responsible for the selection of extender units used in a PKS assembly line, and for the transfer of the latter onto the phosphopantetheine arm of the ACP domain. For Phn2 four AT domains could be identified with identities of over 50% when compared with corresponding enzymes of bacteria of the order *Myxobacteria,* except AT_4. The latter domain is more related to that of bacteria from the *Actinomycetes*. The result of the BLAST analysis is shown in table 4.12.

The catalytic serine, positioned in the highly conserved motif GHSxG and the conserved histidine from the conserved motif xAxH catalyze in a ping-pong bi-bi catalytic mechanism the transfer of acyl moieties (Ruch and Vagelos, 1973, Smith and Tsai, 2007). The presence of these highly conserved motifs could be proved in a multiple sequence alignment for all of the AT domains from Phn2. As all acyltransferase domains were categorized as active domains, the sequence elements important for the substrate activity were investigated.

Sequence analysis of numerous AT domains revealed conserved amino acids motifs that can be correlated to substrate specificity (Yadav *et al*., 2003, Smith and Tsai, 2007). In this aspect the best characterized extender units are the bicarboxylic malonyl-CoA and methylmalonyl-CoA. Three conserved motifs in AT domains are described to obtain informations about substrate specificity towards malonyl-CoA or methylmalonyl-CoA.

Table 4.12 BLAST analysis revealed homologous proteins for AT domains from Phn2 (AT_1-AT_4).

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
AT_1	TgaA [*Sorangium cellulosum*]	169/295 (57%)	201/295 (68%)	ADH04639.1
	polyketide synthase type I [*Stigmatella aurantiaca* DW4/3-1]	156/296 (53%)	196/296 (66%)	YP_003954820.1
	oxidoreductase, zinc-binding dehydrogenase family [*Stigmatella aurantiaca* DW4/3-1]	156/296 (53%)	196/296 (66%)	ZP_01467322.1
AT_2	type I modular polyketide synthase [*Streptomyces bingchenggensis* BCW-1]	169/300 (56%)	203/300 (68%)	ADI03772.1
	polyketide synthase [*Polyangium cellulosum*]	173/301 (57%)	208/301 (69%)	CAD43450.1
	amino acid adenylation protein [*Haliangium ochraceum* DSM 14365]	169/298 (57%)	208/298 (70%)	YP_003267364.1
AT_3	polyketide synthase [*Polyangium cellulosum*]	178/303 (59%)	212/303 (70%)	CAD43450.1
	polyketide synthase [*Polyangium cellulosum*]	178/303 (59%)	212/303 (70%)	CAD43451.1
	amino acid adenylation protein [*Haliangium ochraceum DSM 14365*]	173/299 (58%)	211/299 (71%)	YP_003267364.1
AT_4	modular polyketide synthase [*Streptomyces avermitilis* MA-4680]	169/302 (56%)	207/302 (69%)	NP_821594.1
	modular polyketide synthase [*Streptomyces avermitilis*]	169/302 (56%),	207/302 (69%)	BAB69303.1
	beta-ketoacyl synthase, partial [*Saccharomonospora azurea* SZMC 14600]	162/301 (54%),	203/301 (67%)	EHK81054.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

The conserved region for motif I in ATs is completely different for methylmalonyl-CoA as a substrate ([R/Q/S/E/D]-V-[D/E]-V-V-Q), when compared to the one for malonyl-CoA (Z-T-X-$-[A/T]-[Q/E]). The abbreviations Z and $ in motif I (malonyl-CoA) stand for a hydrophilic (Q) and an aromatic residue (Y, H), respectively (Yadav *et al*., 2003).

AT_2, AT_3 and AT_4 showed the conserved motif I for malonyl-CoA (QTLYTQ and QTGHTG), whereas AT_1 shared more amino acids with the methylmalonyl-CoA motif (RIEVVQ).

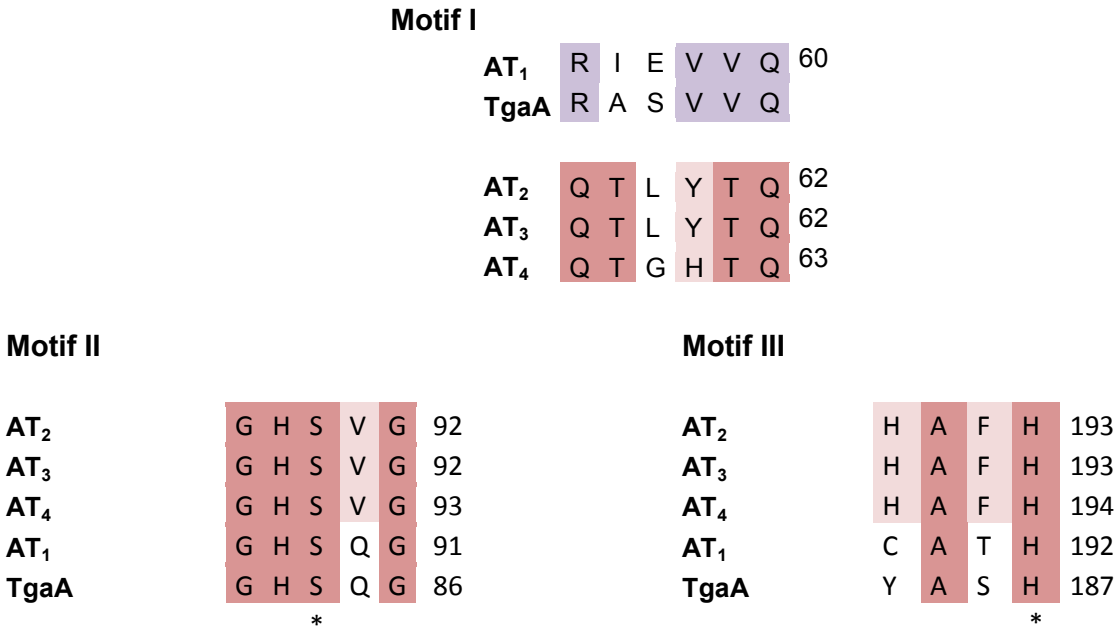


Figure 4.12 Multiple sequence alignment of amino sequences of AT domains from Phn2 and TgaA (*S. cellulosum*). Conserved residues from the highly conserved motifs are highlighted and the catalytic serine and histidine are marked with an asterisk. Amino acids identical in all compared sequences are shaded in red and violet (methylmalonyl-CoA motif); identities for up to three compared amino acids are highlighted in rose.

The second motif (motif II), described as GHSxG, is highly conserved. Sequence alignments and crystal structure indicate that hydrophobic residues as valine, leucine or isoleucine in position five(x) refers to a malonyl-specific AT, whereas less bulky residues as glutamine or methionine are observed for the other AT domains (Haydock *et al.*, 1995, Smith and Tsai, 2007). For three of the four AT domains, i.e. AT_2, AT_3 and AT_4, the presence of valine in the GHSVG motif suggests substrate specificity towards malonyl-CoA, whereas the first AT domain (AT_1) seems to possess a different substrate specificity.

The xAxH motif (motif III) with a conserved histidine is present in methylmalonyl-specific AT domains described as YASH motif and for the malonyl-specific AT as HAFH motif. As the previous sequence motifs of AT_2-AT_4 already showed a bias for malonyl-CoA as a possible substrate, the HAFH motif, present for AT_2, AT_3 and AT_4,

confirms this suggestion. The CATH motif for the AT_1 domain is an as yet undescribed motif for an AT.

A phylogenetic tree was built with phylogeny.fr in a multiple sequence alignment of protein sequences of the AT domains AT_1-AT_4 from *N. pusilla* and an AT domain from TgaA of *Sorangium cellulosum*, as well as putative AT domains from *Haliangium ochraceum* and *S. cellulosum*. AT_2 and AT_3 show the highest similarity to each other, AT_4 is also very closely related to them, whereas AT_1 seems to be closer related to the AT domain from TgaA. The AT domain from TgaA is reported to select and transfer a methylmalonyl-CoA to the adjacent ACP domain. AT domains from *H. ochraceum* and *S. cellulosum* are more related to AT_2, AT_3 and AT_4 as they putatively share the same substrate specificity for malonyl-CoA. The latter was already deduced from sequence motif HAFH.

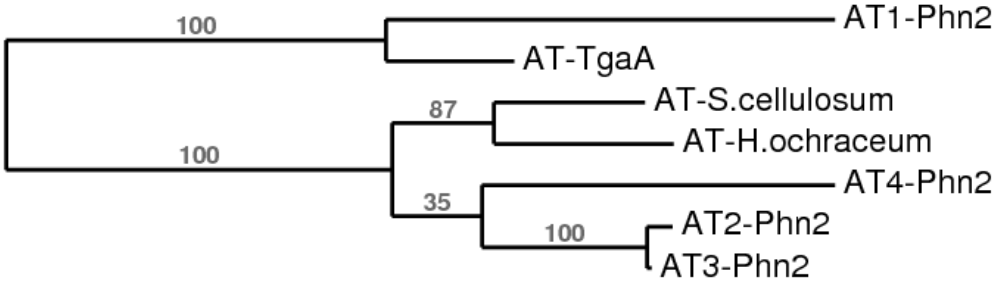


Figure 4.13 Phylogenetic tree of AT domains obtained from a multiple sequence alignment with phylogeny.fr. AT domains from *N. pusilla* (AT_1-AT_4), *H .ochraceum* and *S. cellulosum* were aligned in a multiple sequence alignment. The distance of tree arms shows the respective phylogenetic relation

AT_1 domain loading ethylmalonyl-CoA

BLAST search (table 4.12) showed that AT_1 is closely related to an AT domain from TgaA, a PKS gene from *S. cellulosum.* This AT domain is localized in the second module of the thuggacin A biosynthesis in *S. cellulosum* and is reported to recognize a methylmalonate as its substrate (Buntin *et al*., 2010). Multiple sequence alignment (figure 4.12) revealed for the AT_1 domain three conserved motifs: motif I (RIEVVQ), motif II (GHSQG) and motif III (CATH). Motif I and motif II matches more a methylmalonate-CoA-specific AT. Motif III does not fit the methylmalonate-CoA-specific motif YASH. Previous feeding experiments in *N. pusilla* B150 with labeled precursors showed, that the ethyl side chain of phenylnannolone A is derived from a butyrate unit. Therefore, ethylmalonyl-CoA was suggested to be the favored

substrate for the acyltransferase AT_1. Yadav *et al.* claim the third amino acid (residue 200) in motif III to be crucial for the substrate prediction, especially for the distinction between malonyl-CoA and methylmalonyl-CoA. In malonyl-CoA specific AT's, residue 200 comprises a phenylalanine, in contrast to this methylmalonyl-CoA specific AT's harbor a serine in this position. Serine is a less bulky amino acid, which enables the binding of methylmalonyl-CoA (Yadav *et al*., 2009). Ethylmalonyl-CoA specific ATs harbor at this position small amino acids, i.e. glycine, threonine, serine and alanine (figure 4.14). For AT_1, this amino acid is a threonine, just as in ethylmalonate specific ATs from oligomycin and ascomycin biosynthesis. A multiple sequence alignment could also show a close phylogenetic relationship between AT_1 and other ethylmalonate specific ATs, from biosynthetic gene clusters of phoslactomycin, monensin and ascomycin. These results support the theory that AT_1 accepts ethylmalonyl-CoA as its preferred substrate.

	11	58	59	60	61	62	63	90	91	92	93	94	117	198	199	**200**	201
OlmA3 AT_8 (Oligomycin)	Q	R	I	E	V	I	Q	G	S	S	Q	G	R	V	P	T	H
IdmL-AT_3 (Indanomycin)	Q	R	I	E	V	L	Q	G	H	S	Q	G	R	V	A	S	H
TiaA2-AT_4 (Tiacumicin)	Q	R	I	D	V	V	Q	G	S	S	Q	G	R	V	A	S	H
TylG-AT_6 (Tylosin)	Q	R	V	D	V	V	Q	G	H	S	Q	G	R	T	A	G	H
ConE-AT_{10} (Concanamycin)	Q	R	I	D	V	V	Q	G	H	S	Q	G	R	V	A	G	H
Cfa-AT_2 (Crononatine)	Q	R	I	D	V	V	Q	G	H	S	Q	G	R	V	A	G	H
MonA IV-AT_5 (Monensin)	Q	R	I	D	V	V	Q	G	H	S	Q	G	R	V	A	G	H
Plm-AT_4 (Phoslactomycin)	Q	E	V	D	V	L	Q	G	H	S	Q	G	R	I	A	A	H
Phn1-AT_1 (Phenylnannlone)	Q	R	I	E	V	V	Q	G	H	S	Q	G	R	C	A	T	H
FkbB-AT_4 (Ascomycin)	Q	R	V	D	V	V	H	G	H	S	Q	G	R	C	P	T	H
		motif I						**motif II**							**motif II**		

Figure 4.14 Multiple sequence alignment of ethylmalonyl-CoA specific ATs

Computational analysis with AntiSMASH, where AT substrate prediction is based on two methods: the twenty four amino acid signature sequence for active site described by Yadav *et al*. and the pHMM based method of Minowa *et al*., yielded contradictory predictions. The forecast with Yadav's method points in direction of methoxymalonyl-CoA, whereas Minowa's method predict either methylmalonyl-CoA (score: 175.2) or ethylmalonyl-CoA (score: 173.9) as possible substrate (figure 8.3).

4.8.3.5 Ketoreductase domains

Ketoreductase domains (KR) of modular polyketide synthases, belong to the short-chain dehydrogenase/reductase (SDR) superfamily of enzymes and catalyze the stereospecific reduction of the ß-keto function to a ß-hydroxyl group. This NADPH dependent reduction occurs three times in the biosynthesis for phenylnannolone A, supposedly by KR_1, KR_2 and KR_3. The KR domains of Phn2 were analyzed with BLAST (results see table 4.14) and in a multiple sequence alignment and were characterized by the consensus Rossmann fold motif GxGxxG, which is required for NADPH binding, and the catalytic triad K-S-Y (Smith and Tsai, 2007). All KR domains of the phenylnannolone A gene cluster posses this typical NADPH-binding motif (figure 4.15). The identification of several conserved motifs in KR domains, allowing the stereochemical prediction ('A-' or 'B-type') of the resulting hydroxyl bearing carbon atom (see table 4.13), helped to characterize KR_1 and KR_2 as B1-type ketoreductase domains. Such a B1-type KR domain followed by a DH domain results in a *trans* double bond by *syn* dehydration. KR_4 could not be definitely classified by the conserved motifs to one of the B-type ketoreductases as the conserved motif LDD was not present. The activity of this KR domain is, however given as it harbors the Rossmann fold motif and the catalytic triad K-S-Y (Figure 4.15) (Caffrey *et a*l., 2003, Reid *et al*., 2003, Tsai and Ames, 2009, Kwan and Schulz, 2011).

Table 4.13 KR signature sequences described by Kwan and Schulz (2011)

type	Conserved residues	Resulting intermediate
A1	W^{141}	(2*R*,3*S*) 4 hydroxy-2-methylacyl
A2	W^{141}x x x x H^{146}	(2*S*,3*S*) 4 hydroxy-2-methylacyl
B1	$L^{93}D^{94}D^{95}$;P^{144} x x x N^{148}	(2*R*,3*R*) 4 hydroxy-2-methylacyl
B2	$L^{93}D^{94}D^{95}$;P^{144} x x x N^{148} x x P^{151}	(2*S*,3*R*) 4 hydroxy-2-methylacyl

Table 4.14 Blast search results for the KR domains of the PKS gene Phn2

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
KR_1	polyketide synthase [*Streptomyces platensis*]	117/180 (65%)	137/180 (76%)	BAH02271.1
	FosE [*Streptomyces pulveraceus*]	115/181 (64%)	134/181 (74%)	AEC13071.1
	polyketide synthase pks17 [*Mycobacterium tuberculosis* CDC1551A]	108/178 (61%)	132/178 (74%)	EGB28539.1
KR_2	polyketide synthase [*Streptomyces platensis*]	116/180 (64%)	137/180 (76%)	BAH02271.1
	FosE [*Streptomyces pulveraceus*]	116/181 (64%)	135/181 (75%)	AEC13071.1
	polyketide synthase pks17 [*Mycobacterium tuberculosis* CDC1551A]	110/178 (62%)	133/178 (75%)	EGB28539.1
KR_4	StiD protein [*Stigmatella aurantiaca*]	98/194 (51%)	127/194 (65%)	CAD19088.1
	6-deoxyerythronolide-B synthase [*Paenibacillus curdlanolyticus* YK9]	91/194 (47%)	123/194 (63%)	ZP_07386547.1
	lichenysin synthetase A [*Acetonema longum* DSM 6540]	92/194 (47%)	123/194 (63%)	ZP_08622844.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

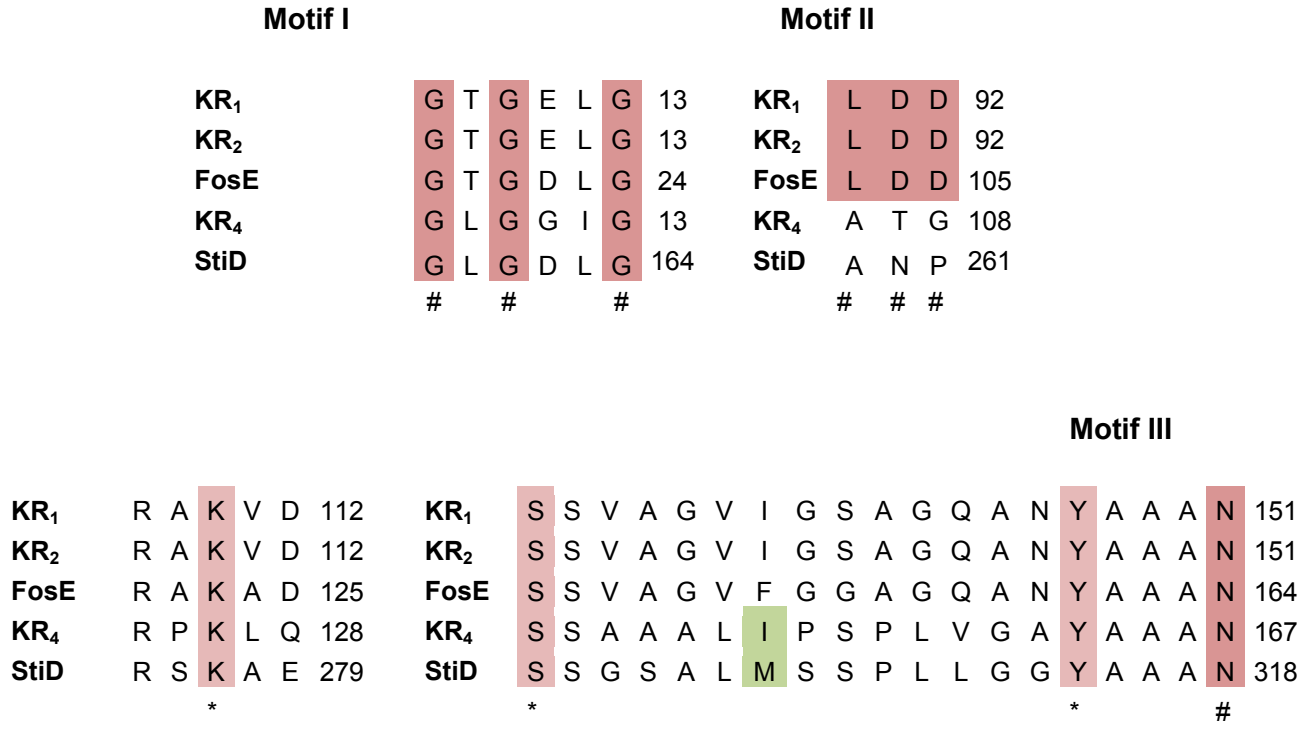


Figure 4.15 Multiple sequence alignment of KR domains from Phn2 and from *Streptomyces pulveraceus* (FosE) revealed the conserved sequence motifs I-III (highlighted in red and marked with a number sign) and the catalytic triad with the residues K-S-Y(highlighted in rose and marked with an asterisk).

A phylogenetic tree (figure 4.16) built with Phylogeny.fr in a multiple sequence alignment of protein sequences of ketoreductase domains from *N. pusilla* (KR_1, KR_2, KR_4) and KR domains from *Streptomyces pulveraceus* (FosE) and *Stigmatella aurantiaca* (StiD), showed high similartity between KR_1 and KR_2. KR_4 seems not closely related to KR_1 and KR_2, but similar to StiD, involved in stigmatellin biosynthesis in *Stigmatella aurantiaca* (Gaitatzis *et al*., 2002). The ketoreductase domain from StiD and KR_4 from Phn2 lack both the tryptophan (W), which is found in most of the KR domains from the A-type.

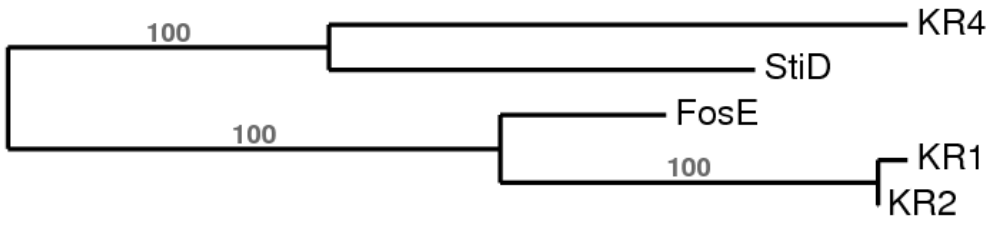


Figure 4.16 Phylogenetic tree of KR domains obtained from a multiple sequence alignment with phylogeny.fr

4.8.3.6 Dehydratase domains

The dehydratase domains are responsible for double bond formation in the growing polyketide chain. The dehydration of the ß-hydroxyl group to an α,ß unsaturated moiety is catalyzed through a catalytic dyad of histidine (H^{44}) and aspartate (D^{206}) conserved in the active site motif LxxHxxxGxxxxP and Dxxx[Q/H] (Tsai and Ames, 2009). BLAST search results and multiple sequence alignment (results see table 4.15 and figure 4.17) showed that the catalytic activity is given for all three DH domains of the Phn2 protein, as the conserved histidine and aspartate are present for DH_1, DH_2 and DH_4. The highly conserved active site motif Dxxx(Q/H) could be characterized for all DH domains. For motif LxxHxxxGxxxxP mutations of the conserved leucine (DH_2) and glycine (DH_1 and DH_4) were observed. Glycine (G^{48}) is responsible for the 'hot-dog-folding' that enables van der Waals interactions between histidine (H^{44}) and proline (P^{53}) of the conserved region. For the DH_1 domain this gylcine is replaced by a aspartate, for DH_2 it is substituted with a glutamate. The effect of these mutations for the folding of the respective proteins is not predictable yet.

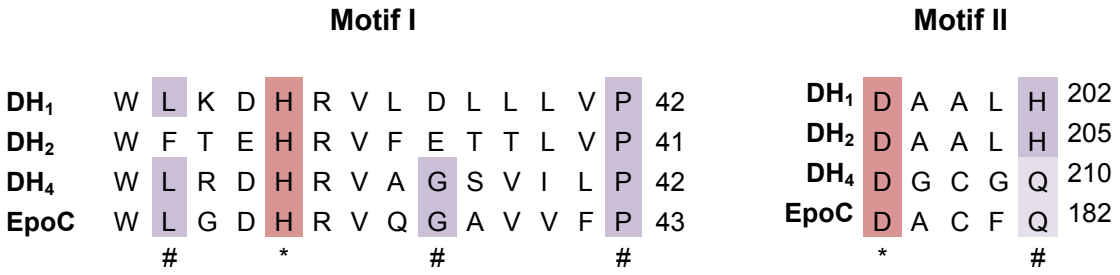


Figure 4.17 Multiple sequence alignment of DH domains from Phn2 and from EpoC (*S. cellulosum*) revealed the conserved sequence motifs I and II (highlighted in violet and marked with a number sign) and the catalytic dyad with the residues H and D (highlighted in red and marked with an asterisk).

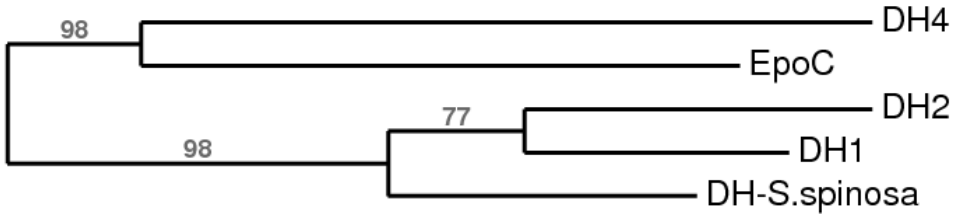


Figure 4.18 Phylogenetic tree of DH domains obtained from a multiple sequence alignment with Phylogeny.fr

A phylogenetic tree (figure 4.18) built with Phylogeny.fr in a multiple sequence alignment of protein sequences of dehydratase domains from *N. pusilla* (DH_1, DH_2, DH_4) and KR domains from *Sorangium cellulosum* (EpoC) and *Saccharopolyspora spinosa*, showed that DH_1 and DH_2 show more homology to each other than to DH_4, which is more related to a KR domain from EpoC.

Table 4.15 Blast search results for the DH domains of the polyketide synthase Phn2

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
DH_1	predicted protein [*Streptomyces sp.* AA4]	77/165 (47%)	104/165 (63%)	ZP_07281889.1
	beta-ketoacyl synthase, partial [*Saccharopolyspora spinosa* NRRL 18395]	82/165 (50%)	99/165 (60%)	ZP_08880287.1
	polyketide synthase extender module 2 [*Saccharopolyspora spinosa*]	82/165 (50%)	99/165 (60%)	AAG23265.1
DH_2	modular polyketide synthase [*Streptomyces clavuligerus* ATCC 27064]	73/163 (45%)	100/163 (61%)	ZP_06769490.1
	type I polyketide synthase [*Streptomyces clavuligerus* ATCC 27064]	73/163 (45%)	100/163 (61%)	ZP_08214249.1
	polyketide synthase extender module 2 [*Saccharopolyspora spinosa*]	77/163 (47%)	93/163 (57%)	AAG23265.1
DH_3	polyketide synthase [*Streptomyces longisporoflavus*]	54/181 (30%)	87/181 (48%)	ACR50774.1
	EpoC [*Sorangium cellulosum*]	57/171 (33%)	85/171 (50%)	AAF62882.1
	EpoC [synthetic construct]	57/171 (33%)	85/171 (50%)	ABB92692.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

4.8.3.7 Thioesterase domain

For the last domain of Phn2 the BLAST search (results see table 4.16) uncovered a thioesterase domain belonging to the prolyl oligopeptidase family, a family member of the AB hydrolase clan (pfam: CL0028). The active site triad S-D-H of TE domains is reported to catalyze the hydrolytic release of the polyketide (PK) chain that may result in an open-chained or in a cyclic PK. Serine, the first catalytic residue, is conserved in the motif G-X-S-X-G. The catalytic aspartate is located 28 amino acids upstream to this active site. The third catalytic residue is located 77 amino acids upstream to the aspartate. For the TE domain from Phn2 the catalytic triad and the conserved motif could be identified in a multiple sequence alignment (figure 4.19).

Table 4.16 BLAST search results for the TE domain in the gene product of *phn2*

Highest homology (protein level)	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
polyketide synthase type I [*Saccharopolyspora erythraea* NRRL 2338]	99/220 (45%)	132/220 (60%)	YP_001104811.1
prolyl oligopeptidase family protein [*Pseudoxanthomonas suwonensis* 11-1]	68/206 (33%)	103/206 (50%)	YP_004147749.1
prolyl oligopeptidase family protein [*Xanthomonas oryzae pv. oryzicola* BLS256]	57/171 (33%)	98/207 (47%)	AEQ94842.1
prolyl oligopeptidase family protein [*Stenotrophomonas sp.* SKA14]	75/235 (32%)	114/235 (49%)	ZP_05136892.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

P. suwonensis	G	A	S	Y	G	72	V	Y	D	L	P	100	G	H	G	F	177
X. oryzae	G	A	S	D	G	72	V	Y	D	L	D	100	G	H	G	F	177
TE (*N. pusilla*)	G	T	S	R	G	78	F	Y	D	P	L	106	G	H	D	I	196
S. eryhtraea	G	T	S	R	G	60	L	Y	D	P	G	88	G	H	D	N	146
	#		*		#				*					*					

Figure 4.19 Multiple sequence alignment of the TE domain of Phn2 (*N. pusilla*) and homologous proteins from *Saccharopolyspora erythraea*, *Pseudoxanthomonas suwonensis* and *Xanthomonas oryzae pv. oryzicola* revealed the catalytic triad with the residues S-D-H (highlighted in red and marked with an asterisk) and the conserved sequence motifs (highlighted in rose and marked with a number sign)

4.9 The biosynthetic NRPS/PKS-type gene *sb1*

4.9.1 Sequencing result for the fosmid 21H12

The sequencing result of fosmid 21H12 revealed 17 open reading frames (ORF). Amongst these ORFs is *sb1*, a gene encoding a mixed NRPS/PKS system, putatively involved in the biosynthesis of an unknown natural product. This biosynthetic gene (*sb1*) is surrounded with genes for hypothetical proteins of unknown function (*orf2, orf3, orf5, orf8, orf9, orf11, and orf16*), as well as transcriptional regulator and sensor proteins (*orf6, orf7, orf12, orf13, and orf14*), a cytochrome P450 enzyme (*orf15*) and a peroxidase-related enzyme (*orf4*). The gene for a succinate quinone oxidoreductase (*orf1*) on the boundary to the fosmid vector might be involved in primary metabolism. The BLAST search results for all of the ORFs are imaged in table 4.17. The gene product for *sb1* showed in the BLAST analysis an identity of 39% to a NRPS/PKS from *Myxococcus xanthus* DK 1622.

Table 4.17 BLAST search results for fosmid sequences of clone 21H12. The deduced amino acid sequences of open reading frames found on fosmid 21H12 were aligned with homologous proteins by BLAST. The results for the gene of the mixed NRPS/PKS system are in bold and highlighted in blue.

Gene	Size (kb)	Highest homology (protein level)	Putative function	Identity of aligned amino acids[1]	GenBank accession number
orf1	0.5	predicted protein [*Capsaspora owczarzaki* ATCC 30864]	Succinate: quinone oxidoreductase	52/163 (32%)	EFW42511.1
orf2	0.3	hypothetical protein PPSIR1_34148 [*Plesiocystis pacifica* SIR-1]	unknown	36/53 (68%)	ZP_01911596.1
or3f	0.5	unnamed protein product [*Mycobacterium ulcerans* Agy99]	unknown	39/102 (38%)	YP_907810.1
orf4	0.6	uncharacterized peroxidase-related enzyme [*Ktedonobacter racemifer* DSM 44963]	Peroxidase related enzyme	81/171 (47%)	ZP_06975623.1
orf5	0.8	putative toxin-antitoxin system, toxin component [*Streptomyces hygroscopicus* ATCC 53653]	unknown	141/284 (50%)	ZP_07294672.1
orf6	0.4	ArsR family toxin-antitoxin system, antitoxin component [*Streptomyces hygroscopicus* ATCC 53653]	Transcriptional regulator	79/104 (76%)	ZP_07294673.1

orf7	0.5	hypothetical protein SPV1_00677 [*Mariprofundus ferrooxydans* PV-1]	Sensor globin protein (gene regulation)	60/138 (43%)	ZP_01453014.1
orf8	0.4	roadblock/LC7 family protein [*Roseiflexus sp.* RS-1]	unknown	65/116 (56%)	YP_001275112.1
orf9	1.0	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	unknown	41/112 (37%)	YP_001618803.1
orf10	1.1	short-chain dehydrogenase/reductase SDR [*Variovorax paradoxus* S110]	DH/KR	28/84 (33%)	YP_002942254.1
orf11	1.9	serine/threonine protein kinase [*Haliangium ochraceum* DSM 14365]	unknown	97/371 (26%)	YP_003266021.1
sb1	**10.3**	**non-ribosomal peptide synthetase/polyketide synthase [*Myxococcus xanthus* DK 1622]**	**NRPS/PKS**	**1235/3160 (39%)**	**YP_631961.1**
orf12	0.7	hypothetical protein HMPREF0005_05065 [*Achromobacter xylosoxidans* C54]	glutathione S-transferase	139/208 (67%)	EFV87494.1
orf13	0.6	transcriptional regulator, TetR family protein [*Labrenzia alexandrii* DFL-11]	Transcriptional regulator	82/190 (43%)	ZP_05112352.1
orf14	5.4	ATP-binding region ATPase domain protein [*Microcoleus vaginatus* FGP-2]	Proteinkinase	677/1779 (38%)	ZP_08494285.1
orf15	1.3	cytochrome P450 [*Ktedonobacter racemifer* DSM 44963]	CYP450	159/409 (39%)	ZP_06973265.1
orf16	1.9	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	unknown	374/582 (64%)	YP_001612581.1

1: ratio of identical amino acids (first value) to all compared amino acids (second value)

4.9.2 Analysis of *sb1*, a NRPS/PKS mixed biosynthetic gene

The sequencing of the fosmid clone 21H12 revealed a gene encoding for a mixed NRPS/PKS multi-enzyme as the BLAST search results in table 4.17 and table 4.18 show. The NRPS/PKS gene *sb1* comprises 10 kb sequence information. The corresponding gene product consists of 3 modules: The first two modules belong to the NRPS part, whereas the third module forms the PKS part. Module1 and module 2 reveal the same domain architecture consisting of a condensation (C), an adenylation

(A) and a peptidyl carrier protein (PCP) domain. Module 3 shows the architecture of a minimal PKS module with a ketosynthase, an acyltransferase, an acyl carrier protein and a thioesterase domain. The diverse domains encoded by s*b1* are examined in detail (see 4.9.2.1-4.9.2.4) and grouped according to their function, i.e. C, A, PCP, KS, AT, and TE domains.

Table 4.18 BLAST search results for the gene *sb1*. The deduced amino acid sequence for this gene was aligned with the respective homologous proteins by BLAST (3.1.14).

Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
non-ribosomal peptide synthetase/ polyketide synthase *[Myxococcus xanthus* DK 1622]	1236/3165 (39%)	1695/3165 (54%)	YP_631961.1
nonribosomal peptide synthetase/ polyketide synthase hybrid [*Lysobacter lactamgenus*]	1252/3113 (40%)	1670/3113 (54%)	ABB80392.1
non-ribosomal peptide synthetase/+polyketide synthase [*Myxococcus fulvus* HW-1]	1054/2634 (40%)	1434/2634 (54%)	YP_004668379.1
hypothetical protein PAU_02675 [*Photorhabdus asymbiotica subsp.asymbiotica* ATCC 43949]	1048/3092 (34%)	1534/3092 (50%)	YP_003041510.1
amino acid adenylation domain protein [*Methylobacter tundripaludum* SV96]	861/1975 (44%)	1172/1975 (59%)	ZP_08782282.1

1: ratio of identical amino acids (first value) to all compared amino acids (second value) (percentage in brackets).
2: Positives are amino acid residues that are similar to each other concerning their chemical properties.

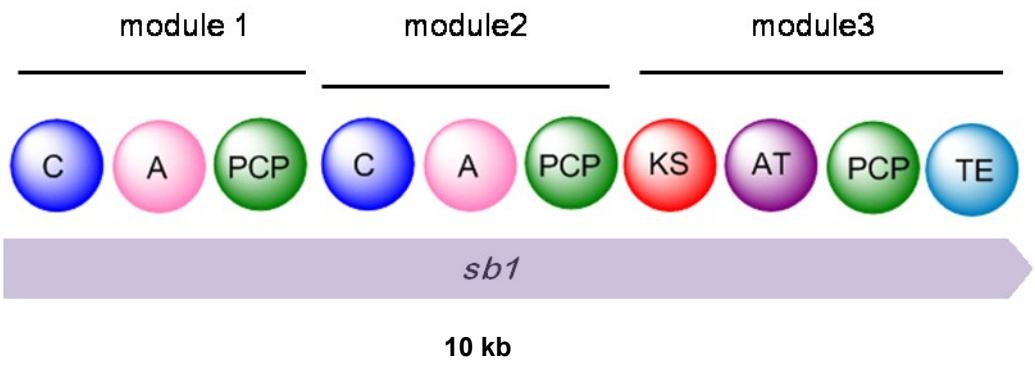


Figure 4.20 Modular NRPS/PKS hybrid encoded by *sb1* consisting of two NRPS modules and one PKS module: *C*= condensation domain, A= adenylation domain, PCP= peptidyl carrier protein, KS= ketosynthase domain, AT= acyltransferase domain, TE= thioesterase domain.

4.9.2.1 Condensation domains

The gene product of *sb1* features two C domains located in the first and second module of the NRPS/PKS system. The C domains were analyzed in a BLAST search (see table 4.19). The first C domain (C_1) is related to an amino acid adenylation domain protein from *Fischerella* sp. and to two unnamed protein products from *Sorangium cellulosum*. The second C domain (C_2) is related to the same unnamed protein product from *Sorangium cellulosum* as C_1, but with a higher identity (43%). It shows also similarities to peptide synthetases from different *Ralstonia solanacearum* strains.

Table 4.19 BLAST search results for C domains of *sb1*

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
C_1	amino acid adenylation domain protein [*Fischerella sp.* JSC-11]	86/302 (28%)	139/302 (46%)	ZP_08985494.1
	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	85/269 (32%)	122/269 (45%)	YP_001617472.1
	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	94/280 (34%)	131/280 (47%)	YP_001610982.1
C_2	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	137/319 (43%)	187/319 (59%)	YP_001610982.1
	putative non ribosomal peptide synthetase protein [*Ralstonia solanacearum* CMR15]	139/308 (45%)	185/308 (60%)	CBJ40245.1
	peptide synthetase [*Ralstonia solanacearum* GMI1000]	139/308 (45%)	185/308 (60%)	NP_522203.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

The C domains of Sb1 were aligned with related C domains from *Fischerella*, *Ralstonia* and *Sorangium* species revealing 7 signature sequences (Konz and Marahiel, 1999), whereby the histidine and the aspartate residues in the conserved motif H-H-X-X-X-D-G (motif III) are considered to be essential for the peptide bond formation (Bergendahl *et al.*, 2002, Sieber and Marahiel, 2005). Regarding the C domains C_1 and C_2, C_1 seems to be inactive due to the lack of the critical histidine

residue in motif C3, present for the second C domain (C_2). As for the other core motifs for C domain C_2, all motifs were present with few mutations, except motif C2, which could not be located. Phylogenetic analysis with NaPDos (3.1.14) classified the C domains as LCL-C domains, connecting two L-amino acids with each other.

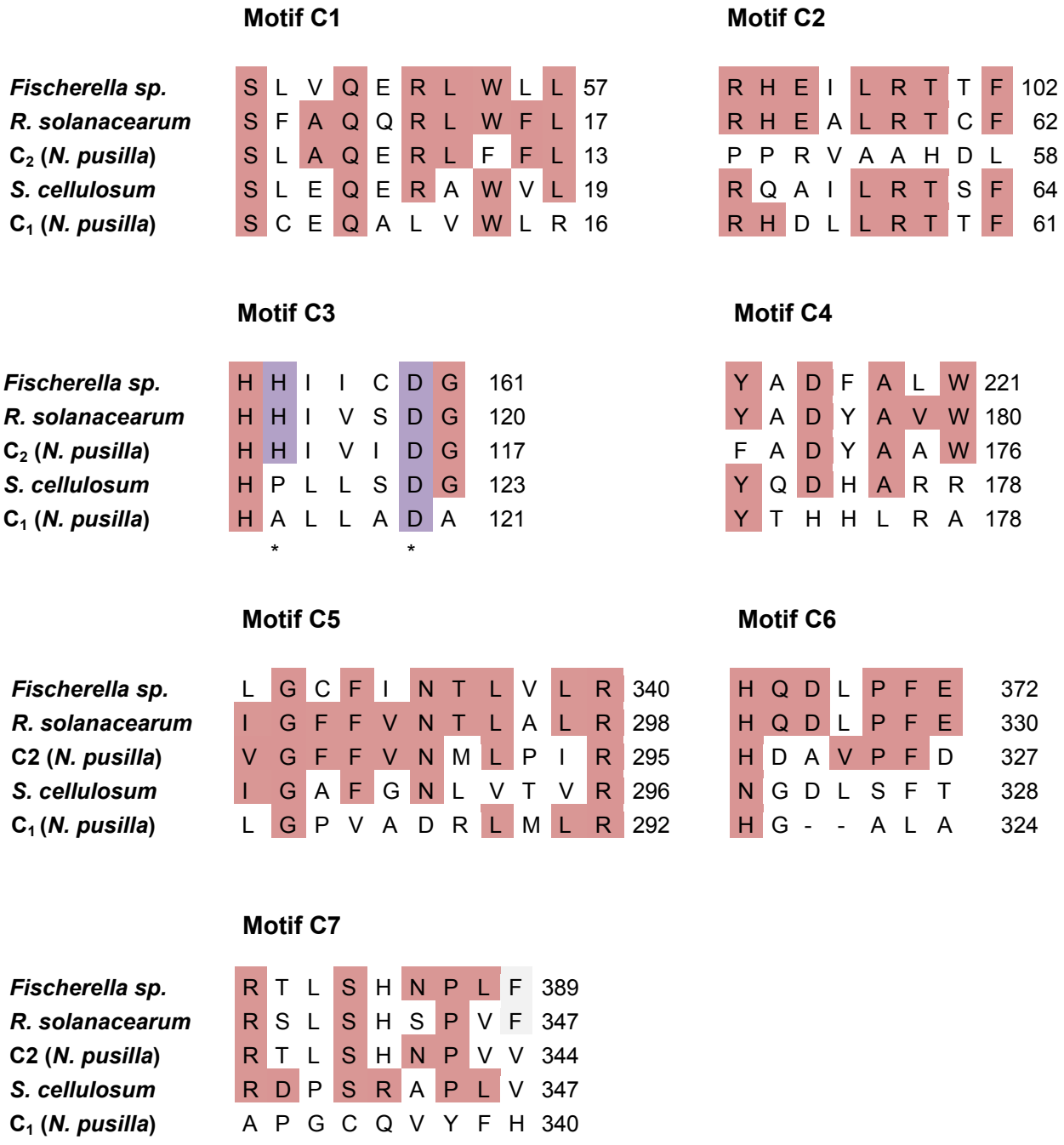


Figure 4.21 Multiple sequence alignment revealing the 7 conserved motifs for C domains.

4.9.2.2 Adenylation domains from Sb1

The NRPS adenylation (A) domain recognizes amino acids and activates them through adenylation. For the NRPS/PKS mixed gene product Sb1 two A domains were identified via BLAST analysis, predominantly related to proteins from cyanobacteria (for detailed results see table 4.20). The A domain A_1 shows the highest identity to an amino acid adenylation domain protein from *Pedosphaera parvula* , A domain A_2 shares 51% identity with an amino acid adenylation protein from *Herpetosiphon aurantiacus.*

Table 4.20 BLAST search results for A domain A1 and A2 from Sb1

Domain	Highest homology (protein level)	Identity of aligned amino acids	Positives of aligned amino acids	GenBank accession number
A_1	amino acid adenylation domain protein [*Pedosphaera parvula strain Ellin514*]	211/409 (52%)	272/409 (67%)	ZP_03628586.1
	peptide synthetase [*Nodularia spumigena* CCY9414]	194/407 (48%)	277/407 (68%)	ZP_01632065.1
	AMP-binding enzyme domain protein [Microcoleus chthonoplastes PCC 7420]	198/405 (49%)	274/405 (68%)	ZP_05029395.1
A_2	amino acid adenylation protein [Herpetosiphon aurantiacus DSM 785]	205/402 (51%)	269/402 (67%)	YP_001545183.1
	amino acid adenylation protein [Cyanothece sp. PCC 7424]	199/400 (50%)	264/400 (66%)	YP_002381151.1
	amino acid adenylation domain-containing protein [Cyanothece sp. PCC 7822]	198/396 (50%)	269/396 (68%)	YP_003899892.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

The highly conserved regions for the A domains are described as ten regions with the following motifs: A1: L[T/S]YxEL, A2: LKAGxAYL[V/L]P[L/I])D, A3: LAYxxYTSG[S/T]TGxPKG, A4: FDxS , A5: NxYGPTE, A6: GELxIxGxG[V/L]ARGYL, A7: Y[R/K]TGDL, A8: GRxDxQVKIRGxRIELGEIE, A9: LPxYM[I/V]P and A10: NGK[V/L]DR. The A domains of SB1 showed the presence of 8 conserved motifs (figure 4.2), located in a multiple sequence alignment with homologous sequences of A domains (table 4.20).

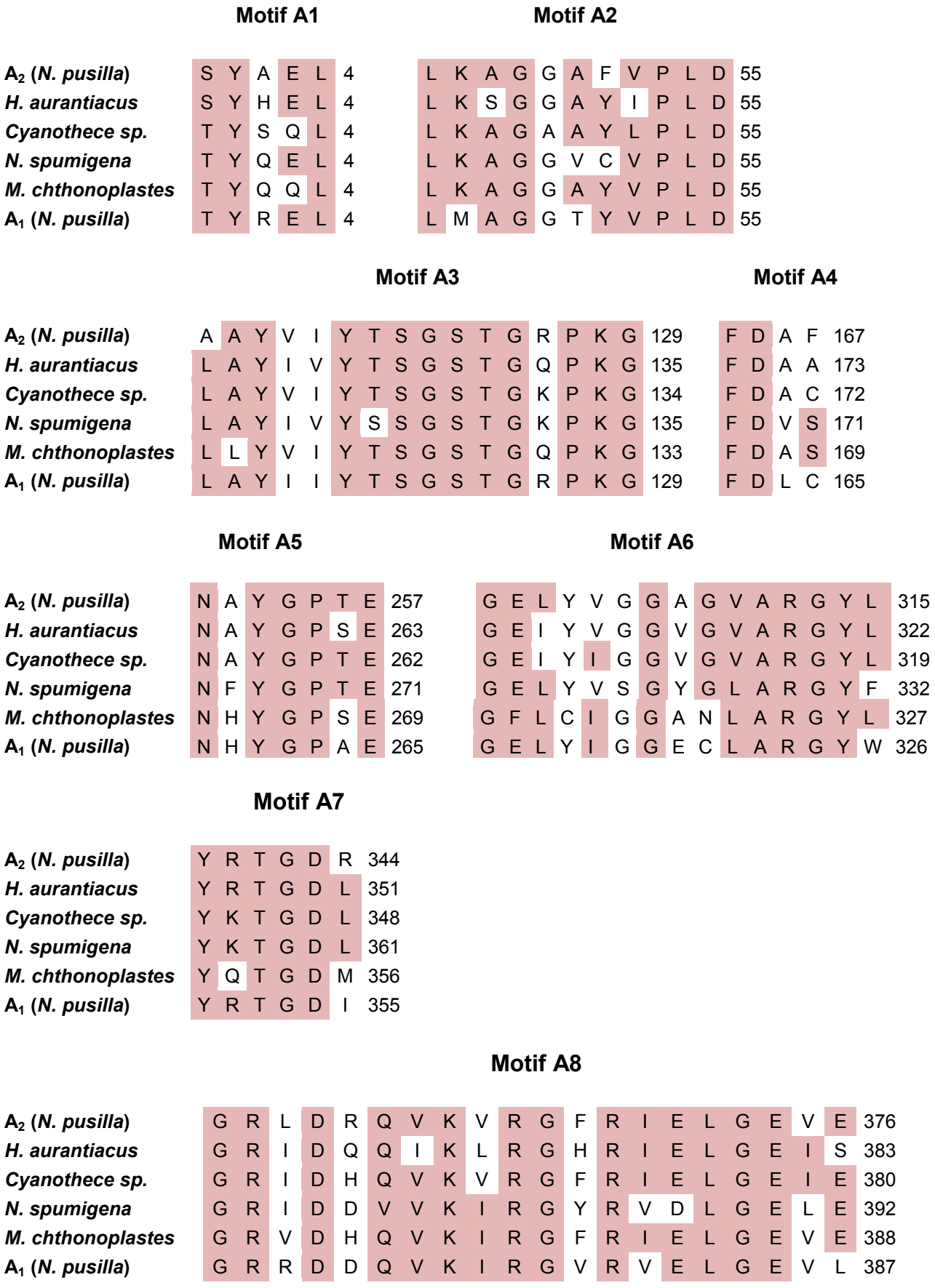


Figure 4.22 Multiple sequence alignment of A domains revealing 8 core motifs

Substrate specificity was predicted by AntiSMASH using the NRPS predictor2 (Rausch *et al*., 2005) and the Minowa HMM method (Minowa *et al*., 2007). For the first adenylation domain proline (Stachelhaus Code: DLYITSHVV-) could be predicted as possible substrate, and for the second one the consensus prediction pointed towards phenylalanine (Stachelhaus code: DAFVVAAVC-) (for more details see Appendix, figure 8.3).

4.9.2.3 PCP domains

The NRPS/PKS multienzyme Sb1 features three domains encoding for phosphopantetheine (PP) binding enzymes responsible for the delivery of the intermediates to other catalytic domains. BLAST search analysis showed similarities to PP-binding domains from proteobacteria and cyanobacteria (for results see table 4.21), all belonging to the ACP-like superfamily (clan: CL0314).

Table 4.21 BLAST search results for PCP domains from Sb1

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
PCP_1	amino acid adenylation [*Pseudomonas syringae pv. japonica str*. M301072PT]	37/63 (59%)	46/63 (73%)	EGH34330.1
	cyanopeptolin synthetase [*Planktothrix prolifica* NIVA-CYA 406]	37/63 (59%)	50/63 (79%)	CAQ77047.1
	cyanopeptolin synthetase [*Planktothrix prolifica* NIVA-CYA 401]	37/63 (59%)	50/63 (79%)	CAQ77045.1
PCP_2	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	43/63 (68%)	54/63 (86%)	YP_001617472.1
	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	44/63 (70%)	54/63 (86%)	YP_001610982.1
	peptide synthase [*Crocosphaera watsonii* WH 0003]	29/63 (46%)	44/63 (70%)	EHJ14209.1
PCP_3	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	33/54 (61%)	42/54 (78%)	YP_001610982.1
	Non-ribosomal peptide synthetase [*Pseudomonas putida* KT2440]	29/65 (45%)	43/65 (66%)	NP_746337.1
	hypothetical protein STIAU_0869 [*Stigmatella aurantiaca* DW4/3-1]	32/56 (57%)	38/56 (68%)	ZP_01462890.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

PP-binding domains are components of NRPS and PKS multienzymes and are named peptidyl carrier proteins (PCP) and acyl carrier proteins (ACP) domains respectively; this refers to the substrate they bind, i.e. either a peptide or an acyl

moiety. To distinguish between ACP and PCP domains CLUSEAN software was used. All three domains were annotated as PCP domains due to the calculated values obtained in an analysis with CLUSEAN. The results of this analysis are displayed in the table 4.22. Surprisingly, the last PP-binding enzyme was predicted to be a PCP and not an ACP domain.

Table 4.22 CLUSEAN prediction for the phosphopantetheine binding enzymesPCP_1 to PCP_3

domain	PCP-C (Score/E-value)	PCP-E (Score/ E-value)	PKSI-ACP (Score/ E-value)
PCP_1	99.9/1.8e-29	67.4/1.1e-19	31.3/ 8e-09
PCP_2	90.9/ 9e-27.	49.4/2.9e-14	46.2/ 2.6e-13
PCP_3	68.5/ 5.1e-20.	29.0/ 3.2e-10	40.3/1.6e-11

The most important conserved residue for the PCP domain is a serine, the putative 4'-phosphopanthetine binding site, that could be identified for all of the three putative PCP domains in a multiple sequence alignment (figure 4.23). For PCP domains this serine residue is conserved in the motif D-X-F-F-X-X-L-G-G-[H/D]-S-[L/I] (Konz and Marahiel,1999).

PCP_1 (*N. pusilla*)	D	D	F	F	A	L	G	G	H	S	L	26
P. putida	D	D	F	F	E	L	G	G	H	S	L	26
PCP_3 (*N. pusilla*)	D	N	F	F	D	L	G	G	T	S	L	26
S. cellulosum [1]		N	F	F	D	L	G	G	H	S	L	20
PCP_2 (*N. pusilla*)	A	P	F	F	E	L	G	G	H	S	L	26
S. cellulosum [2]	D	P	F	F	E	I	G	G	H	S	L	24
										*		

Figure 4.23 Multiple sequence alignment of PCP domains and homologous proteins from *S. cellulosum* and *P. putida* [S. cellulosum [1](GenBank: YP_001610982.1), S. cellulosum [2](GenBank: YP_001617472.1)]

4.9.2.4 Ketosynthase domain (KS)

KS domains in PKS assemblies belong to the superfamily of condensing enzymes (GenBank accession: cl09938).They form carbon-carbon-bonds, using a decarboxylative Claisen condensation. This requires an active centre, the triad C-H-H, catalyzing the steps of acyl transfer, decarboxylation and condensation (Zhang *et al.,* 2006). For the NRPS/PKS system Sb1 a single KS domains was identified. The protein sequence of this KS shares 66% identity with a KS domain from an unnamed protein product from *S. cellulosum* (table 4.23). Multiple sequence alignment of the deduced protein sequence of the KS domain proved the presence of the highly conserved catalytic triad C-H-H for KS_{Sb1} (Zhang *et al*., 2006).

Table 4.23 BLAST search result for KS domain from Sb1

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
KS_{Sb1}	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	280/427 (66%)	316/427 (74%)	YP_001610982.1
	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	268/427 (63%)	313/427 (73%)	YP_001617472.1
	polyketide synthase [*Oscillatoria sp.* PCC 6506]	221/427 (52%)	292/427 (68%)	ACJ46057.1
	6-deoxyerythronolide-B synthase [*Fischerella sp*. JSC-11]	229/428 (54%)	303/428 (71%)	ZP_08985496.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

***Soce*001610982.1**	T	A	C	S	T	170	E	G	H	G	T	G	301	G	H	L	G	A	348
***Soce*001617472.1**	T	A	C	S	T	172	E	A	H	G	A	G	303	G	H	L	R	A	350
KSSb1	S	A	C	S	T	169	E	A	H	G	T	A	300	G	H	L	S	A	347
Oscillatoria sp.	T	T	C	S	T	175	E	A	H	G	T	G	306	G	H	L	N	T	353

Figure 4.24 Multiple sequence alignment of KS domain from Sb1 with homologous protein sequences[S. cellulosum [1](GenBank: YP_001610982.1), S. cellulosum [2](GenBank: YP_001617472.1)]

4.9.2.5 Acyltransferase (AT)

In PKS assembly lines AT domains are responsible for the selection of extender units and the transfer of the latter onto the phosphopantetheine arm of the ACP domain. Sb1, a single AT domain, showed highest homology to an unnamed protein product from *Sorangium cellulosum*. The result of the BLAST analysis is shown in table 4.24.

Table 4.24 BLAST search results for AT domain from Sb1

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
AT_{Sb1}	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	163/302 (54%)	197/302 (65%)	YP_001610982.1
	polyketide synthase [*Myxococcus xanthus* DK 1622]	146/302 (48%)	186/302 (62%)	YP_631807.1
	hypothetical protein [*Planktothrix rubescens* NIVA-CYA 98]	126/303 (42%)	187/303 (62%)	CAQ48282.1
	ProE [*Planktothrix rubescens*]	126/303 (42%)	187/303 (62%)	ACG63859.1

[1]: ratio of identical amino acids (first value) to all compared amino acids (second value) percentage in brackets

[2]: Positives are amino acid residues that are similar to each other concerning their chemical properties.

The catalytic serine, positioned in the highly conserved motif GHSxG and the histidine from the conserved motif xAxH catalyze in a ping-pong bi-bi catalytic mechanism the transfer of acyl moieties (Ruch and Vagelos, 1973, Smith and Tsai, 2007). The presence of these highly conserved motifs could be proved in a multiple sequence alignment for the AT domain in Sb1.

Sequence analysis of numerous AT domains revealed conserved amino acids motifs that can be correlated to substrate specificity (Yadav *et al*., 2003, Smith and Tsai, 2007). Three conserved motifs in AT domains are described to obtain informations about substrate specificity towards malonyl-CoA or methylmalonyl-CoA.

The conserved region for motif I in AT_{Sb1} was found to be K-P-S-R-N-M. It was neither compatible with the motif I for methylmalonyl-CoA ([R/Q/S/E/D]-V-[D/E]-V-V-Q), nor with the one for malonyl-CoA (Z-T-X-$-[A/T]-[Q/E]), where the abbreviations Z and $ stand for a hydrophilic (Q) and an aromatic residue (Y, H), respectively (Yadav *et al*., 2003).

The highly conserved second motif GHSxG, is present in AT_{Sb1} as G-H-S-L-G. Sequence alignments and crystal structure indicate that hydrophobic residues as valine, leucine or isoleucine in position five (x) refers to a malonyl-specific AT, whereas less bulky residues as glutamine or methionine are observed for the other AT domains (Haydock *et al.*, 1995, Smith and Tsai, 2007). The leucine in motif II of AT_{Sb1} indicates malonyl-CoA to be a possible substrate for this AT.

The xAxH motif (motif III) with a conserved histidine is present in methylmalonate-specific AT domains described as YASH motif and for the malonate-specific AT as HAFH motif. TASH, the present sequence motifs for AT_{Sb} , shares the conserved serine from the YASH motif. Yadav *et al.* claim this third amino acid in motif III to be crucial for the substrate prediction, especially for the distinction between malonyl-CoA and methylmalonyl-CoA. For all 77 investigated methylmalonate specific AT residue 200 was identified as a serine. Only one from 98 malonate specific ATs was observed to carry a serine at this position (Yadav *et al*., 2003). The serine motif III of AT_{Sb1} suggests methylmalonyl-CoA to be a possible substrate for this AT.

M. xanthus	G	H	S	L	G	97		V	A	A	H	194
ProE	G	H	S	I	G	97		H	A	F	H	195
AT_{Sb1}	G	H	S	L	G	101		T	A	S	H	198
S. cellulosum	G	H	S	L	G	97		T	A	A	H	195

Figure 4.25 Multiple sequence alignment of AT domain from *N. pusilla* with AT domains from homologous proteins (for GenBank accession no. see table 4.24)

As the three motifs showed different results for the substrate specificity, a reliable prediction could not be made on base of these results. Therefore the amino acid sequence of AT_{Sb1} was analyzed with two online tools: antiSMASH and ASMPKS. AntiSMASH identified the AT to select malonyl-CoA for the acyl transfer to the ACP domain. The identities shown by AntiSMASH with sequences from other malonate

specific ATs were only about 54%, whereas a previous analysis with sequences from Phn2 showed more than 80% identities for the malonate specific ATs. ASMPKS predicted it as a methylmalonate specific AT, but a closer to the detail informations, concerning the homologues sequences of this domain, indicated AT_{Sb1} to share more identities with malonate specific ATs. The results of these queries lead to the assumption that AT_{Sb1} accepts malonate as its substrate.

4.9.2.6 Thioesterase domain

For the downstream domain encoded on *sb1* the BLAST search (results see table 4.25) uncovered a thioesterase activity. The TE domain belongs to the family of thioesterases, a member of the AB hydrolase clan (pfam: CL0028). TE domains are responsible for the hydrolytic release of the polyketide or non-ribosomal peptide chain resulting in an open-chained or in a cyclic PK or NRP.

Table 4.25 BLAST search results for TE domain from Sb1

Domain	Highest homology (protein level)	Identity of aligned amino acids[1]	Positives of aligned amino acids[2]	GenBank accession number
TE	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	106/225 (47%)	147/225 (65%)	YP_001610982.1
	unnamed protein product [*Sorangium cellulosum* 'So ce 56']	98/230 (43%)	132/230 (57%)	YP_001617472.1
	NRPS/PKS hybrid [*Nocardia brasiliensis* ATCC 700358]	71/220 (32%)	118/220 (54%)	EHY29995.1
	polyketide synthase type I [*Crocosphaera watsonii* WH 0003]	73/252 (29%)	111/252 (44%)	EHJ09264.1

1: ratio of identical amino acids (first value) to all compared amino acids (second value) (percentage in brackets).

2: Positives are amino acid residues that are similar to each other concerning their chemical properties.

The hydrolytic action is mediated by the active site triad S-D-H. The catalytic serine is located in the conserved motif GXSXG, 27 amino acids upstream of the serine the catalytic aspartate is found. The third catalytic residue, a histidine is located in the motif GXHF, reported to be conserved in NRPS-TE domains (Reimmann *et al.*, 2004).

Multiple sequence alignment of the TE domain from Sb1 with homologous amino acid sequences revealed the catalytic triad and the conserved motifs GHSAG and GDHF (figure 4.26).

S. cellulosum [1]	G	H	S	A	G	74	L	L	D	T	G	101	G	N	H	F	206
S. cellulosum [2]	G	H	S	A	G	77	L	L	D	T	G	105	G	N	H	F	208
N. brasiliensis	G	H	S	S	G	69	L	L	D	T	T	97	G	N	H	F	202
TE (*N. pusilla*)	G	H	S	A	G	75	I	L	D	A	P	102	G	D	H	F	207
C. watsonii	G	S	S	F	G	84	M	V	D	P	P	112	G	D	H	F	242
			*						*						*		

Figure 4.26 Multiple sequence alignment of TE domain from *N. pusilla* with TE domains from homologous proteins (for GenBank accession no. see table 4.22) S. cellulosum [1](GenBank: YP_001610982.1), S. cellulosum [2] (GenBank: YP_001617472.1)

4.10 Antibiotic selectivity test for *N. pusilla* B150

For further experiments such as the construction of knock-out mutants selection markers, e.g. antibiotic resistance genes are required. Before choosing a selection marker for cloning experiments in bacterial strains, the resistance of the bacterium towards several antibiotics should be evaluated.

To achieve this aim, *N. pusilla* B150was cultured with several antibiotics. Co-culturing with ampicillin, carbenicillin, gentamycine, kanamycin and nalidixic acid did not affect the growth of *N. pusilla* on solid agar.

The growth of *N. pusilla* B150 in liquid was tested for chloramphenicol, streptomycin, tetracycline and carbenicillin. After 4 days, good growth was observed for the carbenicillin culture, but no growth was visible for the other antibiotic-containing cultures. After 14 days the liquid culture complemented with streptomycin, however, showed a slight growth. Table 4.26 shows the results of the antibiotic selectivity test:

Table 4.26 Antibiotic selectivity test

Antibiotic	Solid agar	Growth in liquid
Ampicillin	++	n.t.
Apramycin	-	n.t.
Carbenicillin	++	+
Chloramphenicol	-	--
Gentamycine	+	n.t.
Kanamycin	+	n.t.
Nalidixic acid	+++	n.t.
Streptomycin	-	+
Tetracycline	-	--

n.t.: not tested, +++= fast & good growth, ++=mediate growth,+=moderate growth; -=no growth

4.11 Functional proof of the AMP-ligase from the PKS gene, *phn2*

For the further elucidation of the biosynthesis of phenylnannolone A, functionality and substrate selectivity of the AMP-ligase of Phn2 were investigated. Therefore, the expression of the recombinant protein and a subsequent ATP exchange assay (Phelan *et al*., 2009) were necessary.

4.11.1 Protein expression of the AMP-ligase from Phn2

For the evaluation of the functionality of the AMP domain (4.8.3.1) a recombinant protein comprising the AMP and the proximate ACP domain was expressed in BL21 Star™ *E. coli.* To achieve this, the pet151/D-TOPO® cloning vector kit was used. A PCR fragment (AMP-ACP) of 2.26 kb encoding for the two domains was cloned into the pet151/D-TOPO® cloning vector (figure 4.27) and transferred into BL21 Star™ *E. coli* cells (3.2.21).

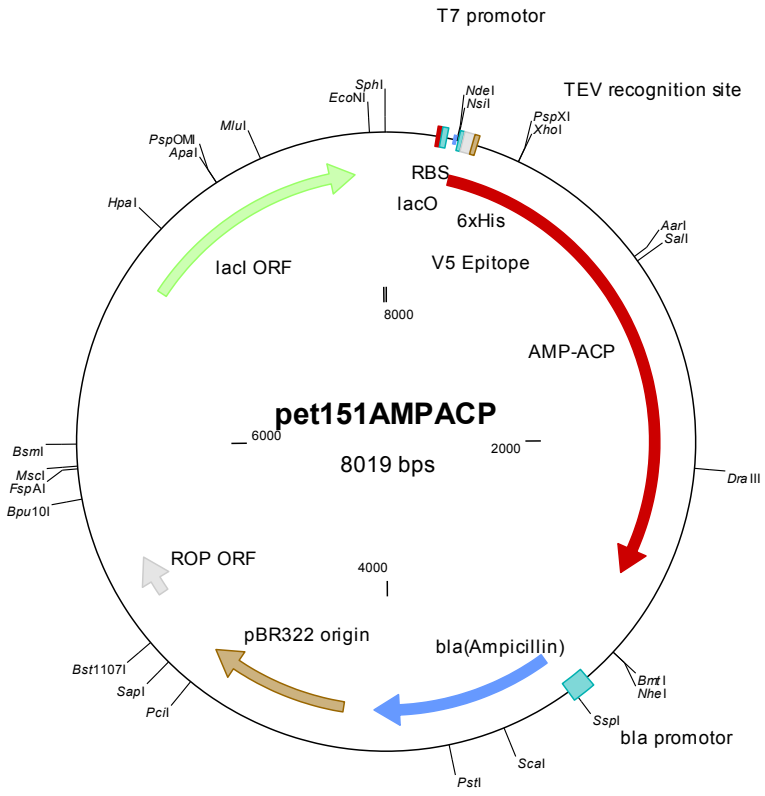


Figure 4.27 Construct for protein expression of the AMP-ligase for subsequent A domain-assay
A PCR fragment (AMP-ACP) with the sequences encoding for the AMP-ligase and the proximate ACP was cloned into pet151/D-TOPO® cloning vector resulting in pet151AMP-ACP, the construct for heterologous protein expression.

The expression of the recombinant protein with an add-on of 0.5mM IPTG took place at 37 °C and was stopped after 4 h. The obtained protein had a length of 785 amino acids, including a N-terminal 6xHis-tag, and was purified on Ni-NTA columns (3.2.23) concentrated and re-buffered (3.2.25). Determination of the 85 kDa AMP-ACP protein was possible by SDS-page gel electrophoresis (3.2.24).

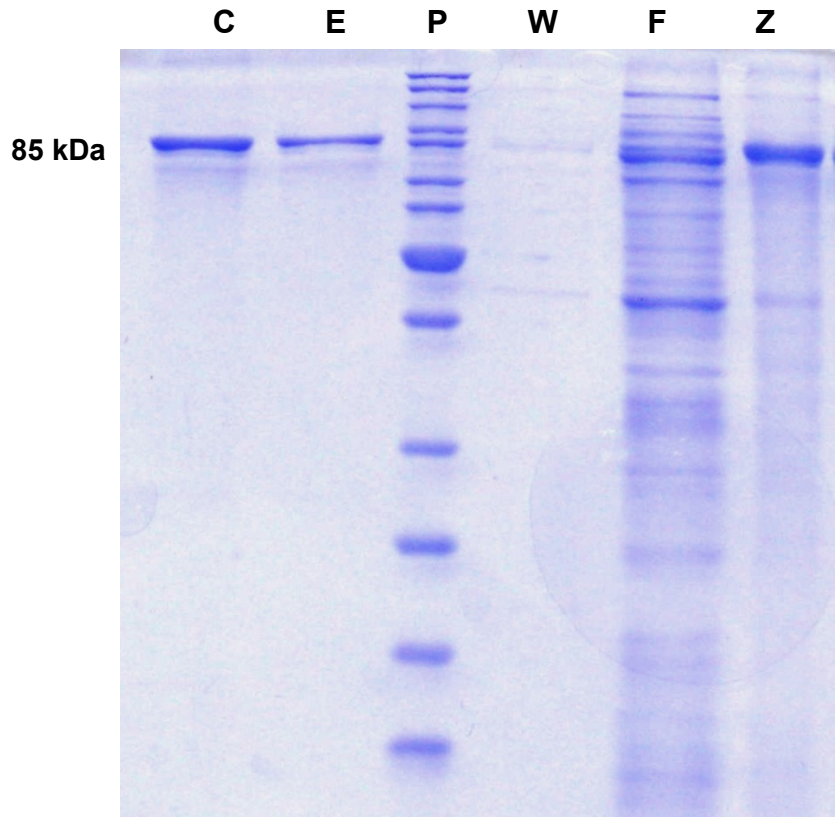


Figure 4.28 SDS-Page electrophoresis gel picture of AMP-ACP protein. Purified (E) and concentrated (C) protein, protein ruler (P) for size estimation, wash fraction (W) and flow through fraction (F), were loaded onto a SDS-Page gel electrophoresis. As a reference for the insoluble fractions, the lysate of the cell pellet (Z) was applied.

4.11.2 ATP exchange assay

With the γ- $^{18}O_4$-ATP pyrophosphate exchange assay the activity of the AMP domain towards various amino acids was determined (3.2.26). This method measures the isotopic back exchange of unlabeled pyrophosphate into γ–$^{18}O_4$-labeled ATP via MALDI-TOFMS (Phelan *et al.*, 2009). The AMP-ACP protein showed activity for two tested substrates, namely cinnamic acid and hydrocinnamic acid, proving the

functionality of the *phn2*-encoded AMP domain. The activity for cinnamic acid was significant, as an exchange of 47.4 % was measured. Hydrocinnamic acid showed only a third of this exchange activity (17.7%). For the other amino acids tested, no significant ATP/PPi exchange was detectable. The assay results are listed in table 4.27 and depicted in a diagram (figure 4.29).

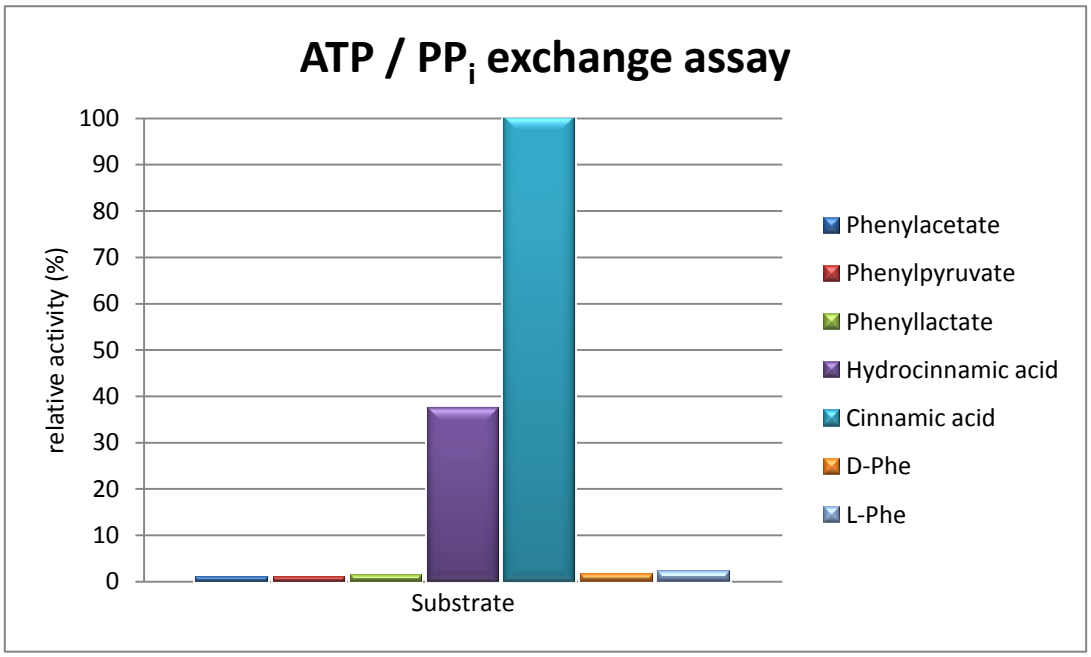


Figure 4.29 Relative activity of the AMP domain from Phn2, measured in an γ-$^{18}O_4$-ATP pyrophosphate exchange assay.

Table 4.27 Results for the ATP/PPi exchange assay: Monoisotopic peak areas of ATP species were determined (506-536) for each substrate. Percent exchange was determined by comparison of the ratio of γ–$^{16}O_4$-ATP (506) to the sum of all ATP species normalized with this modifier: % exchange = (100/0.833) x $^{16}O/(^{18}O + ^{16}O)$. For the relative activity the exchange of cinnamic acid was set as 100%.

Substrate	Peak areas for ATP isotopes										Exchange (%)	Relative activity
	506	508	510	512	514	528	530	532	534	536		
Phenylacetate	18	36	33	241	3368	11	46	24	28	463	0.5	1.1
Phenylpyruvate	21	67	86	345	3772	8	20	27	38	456	0.5	1.1
Phenyllactate	39	62	25	429	5454	13	37	32	29	322	0.7	1.5
Hydro-cinnamic acid	1030	137	53	394	4457	111	63	48	84	603	17.7	37.4
Cinnamic acid	2085	122	23	119	2334	228	49	54	75	189	47.4	100.0
D-Phe	30	72	37	285	3503	25	47	23	61	468	0.8	1.7
L-Phe	28	50	38	176	2595	16	29	6	16	240	1.1	2.2

5 Discussion

Myxobacteria produce a wide range of structurally diverse compounds with unique modes of actions, and thus are an important source of novel secondary metabolites (Gerth *et al*., 2003). Most of these secondary metabolites that are synthesized by myxobacteria employ PKS and NRPS assembly lines or hybrids of both (Weissman and Müller, 2009). The understanding of the pathways leading to these unique myxobacterial compounds is of utmost interest, as the biosynthesis in these microorganisms deviates from those in other organisms and gives insight into new biochemical reactions. The identification of the corresponding biosynthetic gene clusters are facilitated by genome sequencing or screening of genomic libraries.

The current study addressed the elucidation of the biosynthesis of phenylnannolone A from the myxobacterium *Nannocystis pusilla* B150, employing both strategies: genome sequencing and screening of a genomic library.

5.1 Construction of the genomic library

A genomic library is a set of clones that together represent the entire genome of an organism. Each clone of the library carries a DNA fragment from the desired organism. These libraries are used for the elucidation of biosynthetic gene clusters, the heterologous expression of the latter and for studies on the function of selected genes.

The construction of a gene library of *N. pusilla* B150 turned out to be a real challenge. One of the major reasons, which made the fosmid library construction so difficult, was due to the delicate nature of the myxobacterial DNA. This could be observed during the DNA isolation step at the very beginning of the procedure. The latter turned out to be extremely tedious, as it resulted mostly in sheared, small DNA fragments. To obtain DNA fragments of around 36 kb, several adjustments were made to the DNA isolation procedure. The use of cut tips and decreased centrifugation speed finally led to less sheared DNA. The delicate character of this myxobacterial DNA was not only observed during the DNA isolation process, but also in several cloning and sequencing experiments. Myxobacteria are microorganisms with high GC-content in their DNA, which leads to secondary intramolecular

structures and hair-pin loops that resist denaturation (Bravo *et al*., 2010). However, this could not be the sole reason for the delicate nature of *N. pusilla*'s DNA, as handling other DNA from myxobacterial strains with high GC-contents was not as difficult as experienced in the current case.

In addition to the DNA isolation, another factor, i.e. the ligation reaction, made the library construction difficult. In general, the blunt-end ligation of large fragments (~36 kb) is quite sensitive and challenging, and ligation of these fragments to the fosmid vector DNA determines the success of gene library construction. Several factors may complicate or abolish this ligation reaction. Therefore, it is important to remove contaminants that inhibit the activity of the T4-DNA ligase. Salt residues resulting from DNA recovery from LMP-agarose or former enzyme reactions may interfere with the ligation reaction, as they change buffer conditions and hence abolish enzyme activity of the T4-DNA ligase. An incomplete end-repair reaction of the selected DNA fragments impedes ligation, as only blunt-end DNA fragments are ligated to the blunt-ends of the fosmid vector. In this work the end-repair enzyme kit provided with the CopyControl™ Fosmid Library Production kit (3.1.2) was, after several unsuccessful trials, replaced by the Quick blunting kit from NEB (3.1.2). The benefit that this kit provided was the immediate use of the reaction mixtures after enzyme reaction and subsequent heat-inactivation for ligation. As each step of purification led to a loss of DNA quantity, the immediate use of processed DNA avoided further waste of DNA. The molar ratio of vector to template is another parameter for a successful ligation that must be strictly followed.

Even if genome sequencing is getting easier and cheaper, the availability of genomic libraries is still important. They are necessary for physically obtaining biosynthetic gene clusters, which allows further investigations of the different biochemical steps involved in the biosynthesis of natural products.

5.2 Sequencing of DNA from *N. pusilla* B150

During this project different sequencing strategies were employed to elucidate phenylnannolone's biosynthetic gene cluster, such as fosmid sequencing and genome sequencing with Roche 454 or sequencing of PCR fragments with the aim to close gaps between contigs with the Sanger method.

As already mentioned, handling the DNA of *N. pusilla* B150 was not an easy task. This was also observed during sequencing. A possible reason for this might be the high GC-content (70%) of this strain, which leads to intramolecular secondary structures and hair-pin loops hindering the polymerase, and hence stopping the chain elongation and sequencing reaction.

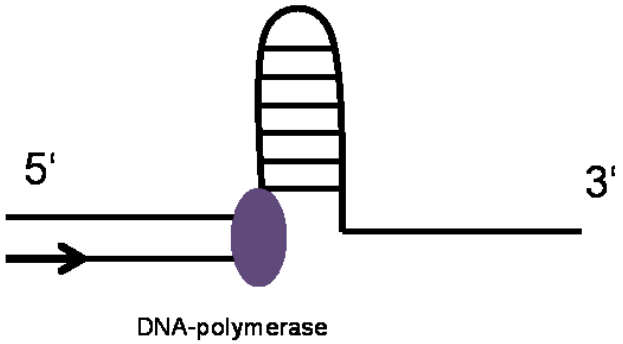


Figure 5.1Seconary intramolecular structure (hair-pin loop) stops DNA-polymerase and hence sequencing

High GC-content in DNA demands special sequencing conditions and chemical additives. Dimethyl sulfoxide (DMSO), glycerol, polyethylene glycol, formamide, betaine, 7-deaza-dGTP and dITP improve the amplification of GC-rich DNA sequences. A combination of betaine, DMSO and 7-deaza-dGTP showed best results in the prevention of intramolecular stable stem loops in GC-rich templates. Each additive acts differently: DMSO is believed to disrupt base pairing, betaine acts as an isostabilizing agent, whereas 7-deaza-dGTP is partially substituting dGTP and hence reduces the number of hydrogen bonds with complementary dCTP (Musso *et al*., 2006).

In our laboratory PCR amplifications of GC-rich DNA were supplemented with DMSO, which helped to overcome the problem of intramolecular secondary structures and resulted in distinct bands of PCR fragments. High-throughput sequencing of cloned PCR fragments or partial sequencing (BIGrun) on fosmid DNA

(performed by GATC Biotech AG, Konstanz) mostly failed. Better results were obtained with customized cycling conditions and “special chemistry” used for sequencing (Eurofins MWG Operon, Ebersberg). The additives for the “special chemistry” service provided by Eurofins MWG Operon surely include some of the above mentioned compounds, but the exact composition remains unknown as every sequencing company keeps it as a well guarded secret. For next generation sequencing of the *N. pusilla* B150 genome and fosmid DNA on the Genome Sequencer (GS) FLX titanium (Roche/454), special kits for GC-rich DNA sequences also of unknown composition were applied for genome (Roche) and fosmid (IIT Biotech) sequencing.

The de-novo-sequencing of *N. pusilla’*s genome yielded a 16.6 fold coverage. The very limited average size (3 kb) of contiguous DNA sequences (contigs) in the draft genome complicated genome mining for the phenylnannolone biosynthetic genes. Biosynthetic gene clusters of secondary metabolites usually exhibit sizes between 30 and 80 kb. The high incidence of repetitive regions in PKS genes impedes the assembly of sequences encoding PKS domains. Hence contigs harboring genes that code for PKS domains were very short, in average 1-2 kb. Assembling these contigs to a coherent gene cluster was nearly impossible, as some sequences were falsely assembled and hence a lot of the sequence information necessary to build the PKS gene cluster was lost. Albeit genome sequencing could not be used for the identification of the phenylnannolone biosynthetic gene cluster, the sequence information of two contigs was used for a screening of the fosmid library. The screening resulted in two fosmid clones, one harboring a NRPS/PKS gene (21H12) and the other encoding the putative biosynthetic genes for the formation of phenylnannolone A (12A9).

The obtained 16.6 fold coverage of the *N. pusilla* B150 genome reflects a quite low coverage of a *de-novo* sequencing. Higher coverage might help to achieve better results for the data assembly and hence longer contiguous DNA. Genome coverages of 20-25x are generally preferred for the use of next generation sequencing devices. Indeed for our 454 sequence data from fosmid sequencing a better assembly was achieved. This resulted in one contig for fosmid 12A9 and two contigs for fosmid 21H12. The higher coverage of the fosmid DNA sequences (i.e. 29.3x and 19.4x) is the major reason for the improved results.

Comparing the different sequencing approaches during the current project, a high coverage and special additives (of unknown composition) are the main factors for success. Sequencing GC-rich DNA is still challenging and for some strains genome sequencing can hardly be accomplished, not even with the new techniques developed for sequencing (personal communication, Dr. Christian Rückert, CeBiTec, Uni Bielefeld).

5.3 The biosynthetic gene cluster for the phenylnannolones

The putative phenylnannolone gene cluster with an approximate size of 24 kb was identified on fosmid clone 12A9. It comprises a putative carboxyl-transferase gene (*phn1*) and a modular type I PKS gene (*phn2*), putatively responsible for the formation of phenylnannolone A.

The carboxyl-transferase encoded by *phn1* is probably responsible for the precursor supply, further discussed in section 5.3.1. The adjacent PKS enzyme comprises 5 modules, i.e. a loading module and 4 extension modules. From the domain order of the PKS encoded by *phn2* the following biosynthetic pathway is proposed for the formation of phenylnannolone A (figure 5.2). The loading module comprises an AMP-dependent ligase and synthetase (AMP) that adenylates the putative starter cinnamic acid (5.3.2) and hence enables the loading to the adjacent acyl carrier protein (ACP_{LM}) domain. In the following step the ketosynthase (KS_1) of the first extension module prolongs the starter unit with a butyrate moiety (5.3.3) in a decarboxylative Claisen condensation, which is followed by reduction and dehydration, caused by the corresponding ketoreductase (KR_1) and dehydratase (DH_1) domains, forming a *trans*-configurated C-C double bond. In the second elongation step (module 2) the growing chain is extended by an acetate unit, which is reduced and dehydrated in the same manner as in the step before. The third elongation module is once again responsible for the extension of the polyketide with an acetate, with the difference that the ß-keto group of the thioester remains and is not further reduced. This is important for the lactone ring formation, discussed in section 5.3.5. The extender unit of the last step is again malonyl-CoA, prolonging the polyketide chain with a final acetate unit. The corresponding KR reduces the ß-keto group to an L-hydroxyl-group, that is then dehydrated by DH_4, forming a *cis*-double bond. The so generated stereochemistry is

necessary forl lactone ring formation and thereby for the release of the polyketide from the assembly line through the thioesterase domain (5.3.5).

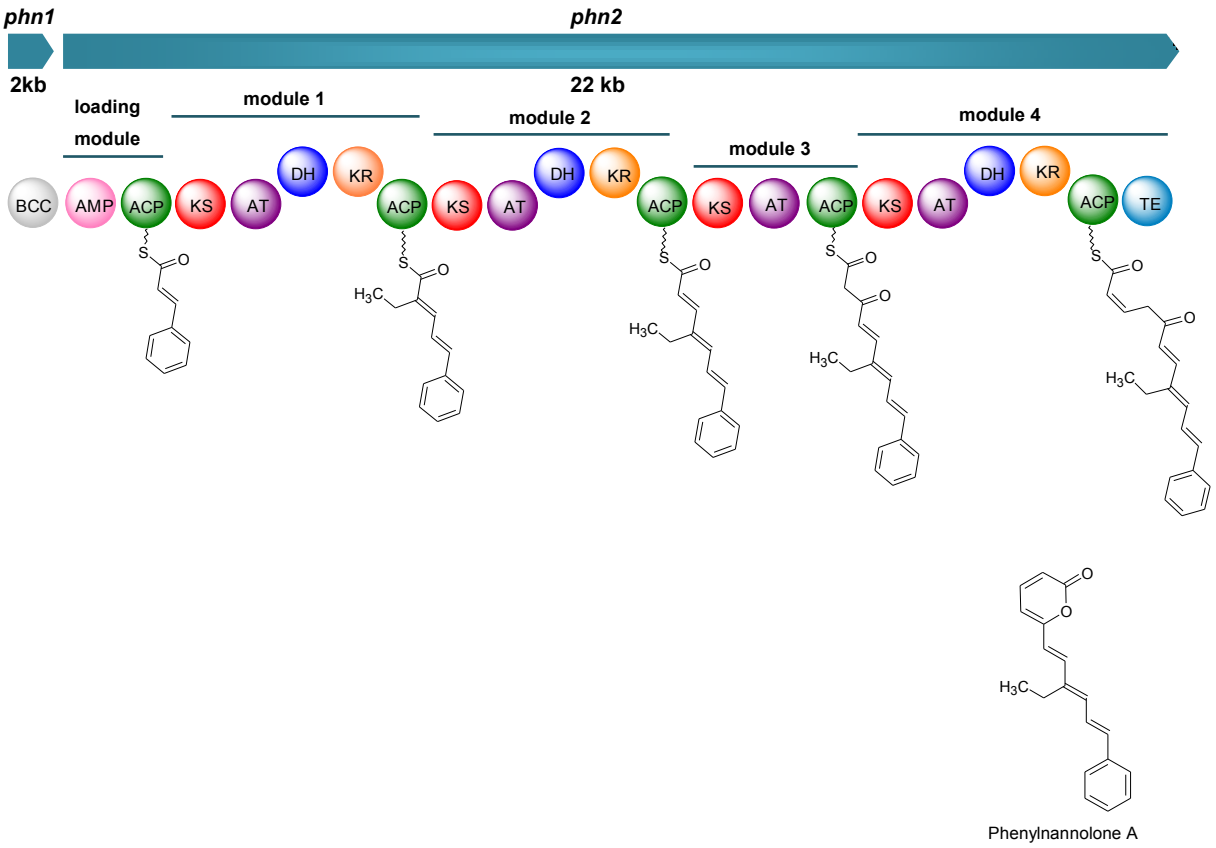


Figure 5.2 Hypothetical pathway for the biosynthesis of phenylnannolone A: BCC=butyryl-CoA carboxylase AMP=AMP-dependent ligase, ACP=acyl carrier protein, KS=ketosynthase domain, AT=acyltransferase domain, DH=dehydratase domain, KR=ketoreductase domain, TE=thioesterase domain

5.3.1 Precursor supply - ethylmalonyl-CoA formation

In contrast to polyketide starter units, which comprise a wide variety of structures, extender units are much less chemically diverse (Wilson and Moore, 2012). Most common PKS extender units are malonyl-CoA and methylmalonyl-CoA, derived from fatty acid metabolism and other primary metabolism pathways (Quade *et al*., 2011). PKS extender units such as methoxymalonyl-CoA, hydroxymalonyl-CoA and ethylmalonyl-CoA are less common in polyketide biosynthesis. The supply with specific precursors, e.g. ethylmalonyl-CoA, is a limiting factor in polyketide biosynthesis, as their availability is not comparable to that of the common malonyl-CoA or methylmalonyl-CoA.

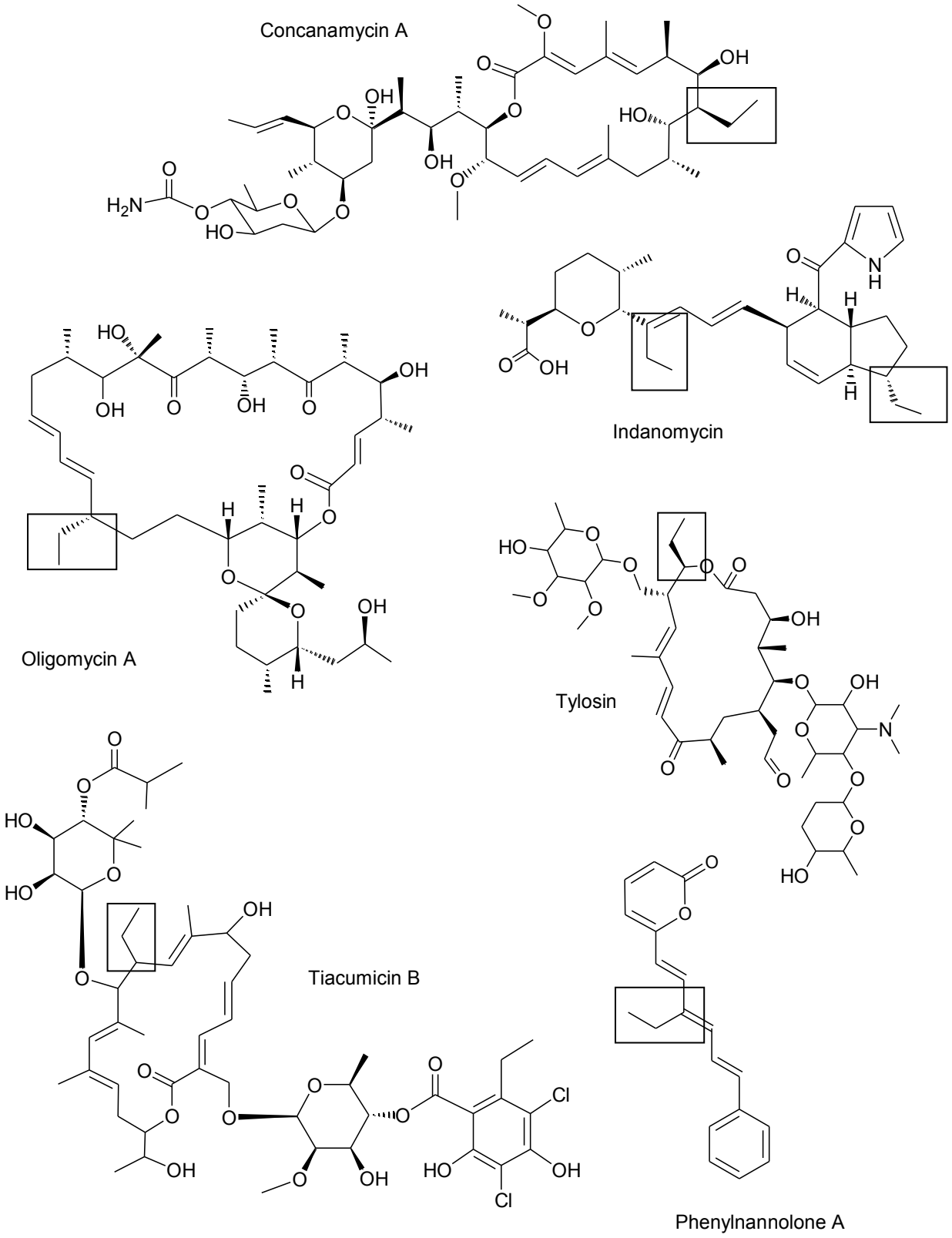


Figure 5.3 Polyketides with an ethyl-side chain built by modular PKS (ethy-side chains are framed)**.**

For the biosynthesis of ethyl-substituted natural products the supply of the PKS extender unit ethylmalonyl-CoA is crucial. The fact that ethylmalonyl-CoA can be derived from crotonyl-CoA through the action of the enzyme crotonyl-CoA carboxylase/reductase (CCR), was described for *Rhodobacter sphaeroides* by Erb *et al*. in 2007. These authors also defined later the complete CCR containing ethylmalonyl-CoA pathway, which is only represented in 5% (57 species) of the fully sequenced bacteria (1215 species) (Erb *et al*., 2009). As only few bacteria contain an operating ethylmalonyl-CoA pathway, most biosynthetic gene clusters of ethylsubstituted polyketides, harbor a gene encoding a CCR homolog. The associated CCR supplies the PKS with ethylmalonyl-CoA or even longer chained acyl-CoA, such as hexylmalonyl-CoA, for the biosynthesis of the respective natural product. It is thought that these enzymes are capable to form dicarboxylic acids with any side chain from a respective unsaturated fatty acid precursor (enoyl-CoA). Analysis of the crystal structure of CinF, a CCR homolog responsible for the supply with hexylmalonyl-CoA in cinnabaramide biosynthesis, revealed two positions in the binding pocket that determine substrate specificity. Thus mutations in these active site residues may lead to a variety of PKS extender units (Quade *et al*., 2012).

Besides the mentioned CCR pathway, there are also other pathways proposed for the formation of ethylmalonyl-CoA. They require the carboxylation of butyryl-CoA to ethylmalonyl-CoA, in which butyryl-CoA may be derived either from L-valine that is catabolized to isobutyryl-CoA and then converted to butyryl-CoA or from the ß-oxidation of even-chain fatty acids (Vrijbloed *et al.*, 1999, Chan *et al.*, 2009). For both pathways acyl-CoA carboxylases are possible enzymes for the carboxylation reaction.

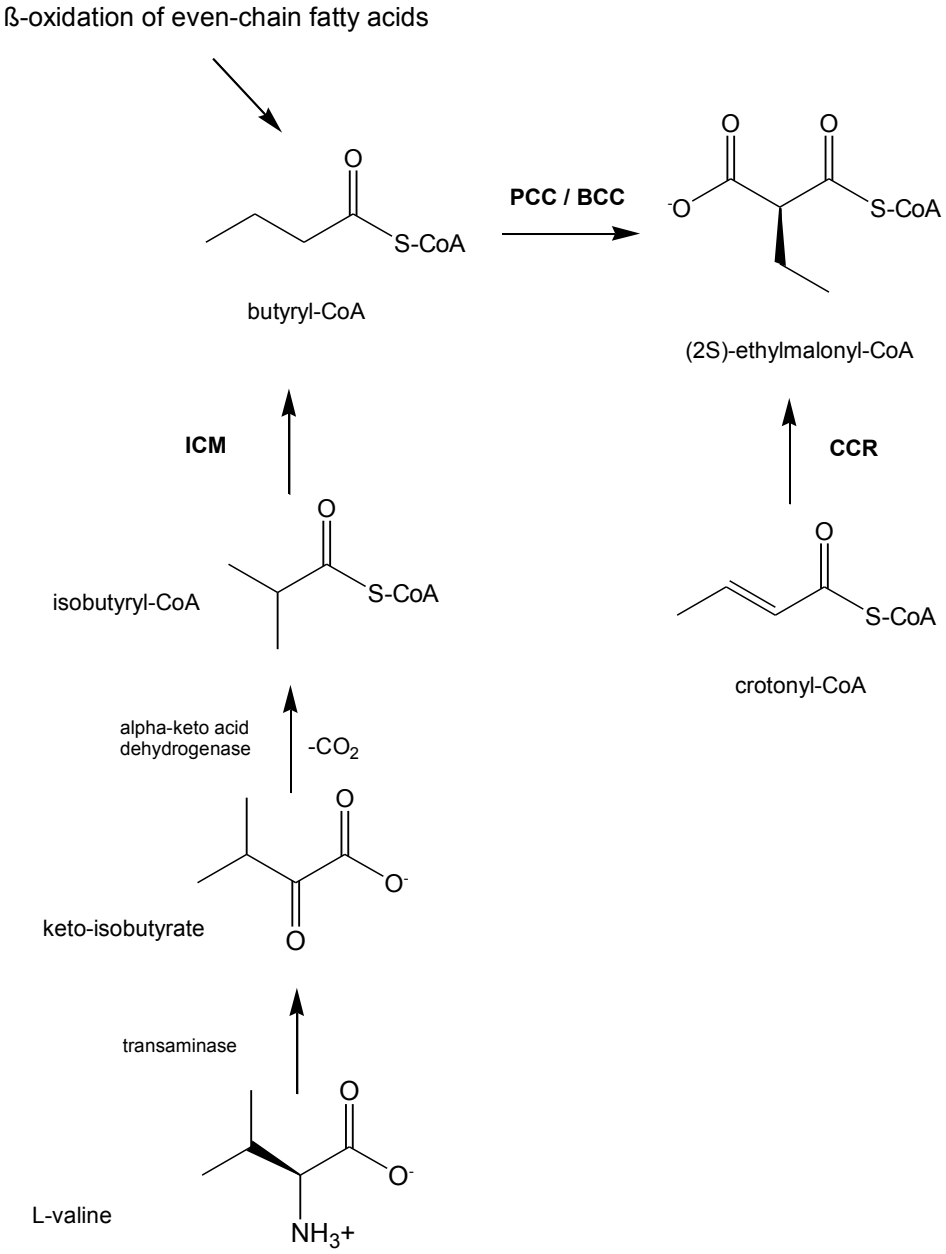


Figure 5.4 Biosynthetic pathways leading to the formation of ethylmalonyl-CoA (adapted from Chan et al., 2009); CCR=crotonyl-CoA reductase/dehydrogenase, PCC=propionyl-CoA carboxylase, BCC=butyryl-CoA carboxylase, ICM=isobutyryl-CoA mutase

Carboxyl transferases such as the acyl-CoA carboxylases (ACCs) are known to carboxylate acetyl-, propionyl- or butyryl-CoA to malonyl-, methylmalonyl or ethylmalonyl-CoA. Acetyl-CoA carboxylases (AcCCs) are carboxyl transferases that accept besides acetyl-CoA also propionyl-CoA and butyryl-CoA as substrates. The active site residue in ACC is an isoleucine in position 422. Propionyl-CoA carboxylases (PCC) are known to accept propionyl-CoA, and with a lower affinity also butyryl-CoA, but not acetyl-CoA (Diacovich *et al.*, 2004). Butyryl-CoA as a substrate is favored in PCC, when a mutation transforms the active site residue aspartate

(D422) into alanine or cysteine. This leads to changes in the binding pocket and directs the substrate specificity towards butyryl-CoA (Arabolaza *et al*., 2010).

As mentioned before, most biosynthetic gene clusters of ethyl-substituted polyketides are associated with CCRs, like concanamycin, oligomycin, tylosin and several others (figure 5.3, table 5.1).

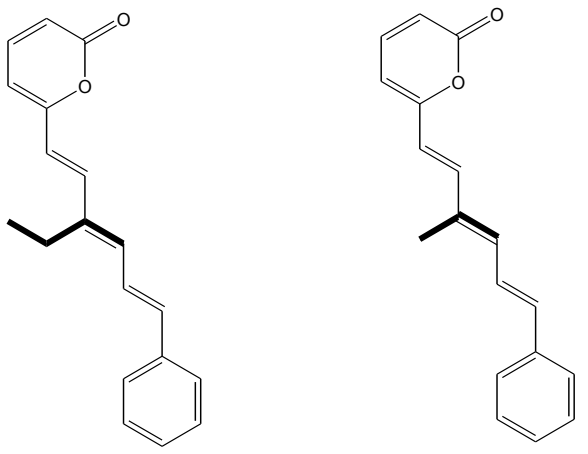
Table 5.1 Biosynthetic gene clusters of ethylsubstituted natural compounds that harbor crotonyl-CoA reductases (CCRs) and acyl-CoA carboxylases (ACCs) for supply with ethylmalonyl-CoA

Natural compound	CCR	ACC
Elaiophylin	+	
Tylosin	+	
Spiramycin	+	
Concanamycin	+	
Indanomycin	+	+
Tautomycetin	+	
Midecamycin	+	
Rosaramicin	+	
Phoslactomycin A	+	
Oligomycin	+	
Tiacumicin B	+	+
Lasalocid A	+	
Kirromycin	+	
Ascomycin (FK520)	+	
Sangifehrin	+	
Divergolides	+	
Monensin A	+	

The recently elucidated biosynthetic gene clusters of indanomycin and tiacumicin were found to harbor both types of enzymes, i.e. the CCR and the ACC. In the case of indanomycin Kelly and colleagues discussed an acyl-CoA carboxylase (IdmB) to be responsible for the supply of both, ethylmalonyl-CoA and methylmalonyl-CoA. This carboxyl-transferase harbors an isoleucine at position 422, instead of the conserved aspartate, which changes the specificity of the binding pocket of IdmB. Thus, they postulated IdmB to accept substrates such as acetyl-CoA, propionyl-CoA and butyryl-CoA with an almost equal affinity, just like acetyl-CoA carboxylases (Li *et*

al., 2009). For the precursor supply in tiacumicin B biosynthesis three genes were identified to encode for enzymes of the ethylmalonyl-CoA pathway: a hydroxybutyryl-CoA dehydrogenase (TiaJ), a crotonyl-CoA hydratase (TiaN) and a crotonyl-CoA carboxylase (TiaK). Beside these three enzymes that form ethylmalonyl-CoA, another enzyme, a PCC encoded by *tiaL*, is believed to supply the PKS with methylmalonyl-CoA (Xiao *et al*., 2010). Multiple sequence alignment (figure 4.6) with sequences of different PCCs revealed the active site residue for TiaL to be an isoleucine. Thus the affinity of this PCC is suggested to be comparable to the one of IdmB. This conclusion is supported by crystal structure and mutational analysis of the ß-subunit of a PCC from *S. coelicolor*, which revealed similar kinetic parameters for ACC and PCC_{I422D} mutants (Diacovich *et al.*, 2004, Arabolaza *et al*., 2010).

As for the biosynthesis of phenylnannolone A the supply with ethylmalonyl-CoA putatively results from Phn1, a protein with an identity of 70% to a putative ß-subunit of a propionyl-CoA carboxylase from *Plesiocystis pacifica* SIR-1. Acyl-CoA carboxylases in actinomycetes consist of two polypeptides: an α and β-subunit. The α subunit harbors the biotin-carboxylase (BC) and biotin-carboxylase-carrier protein (BCCP) domain, whereas the β-subunit acts as a carboxyl-transferase (Diacovich *et al.*, 2004). The ACC from *S. antibioticus* (IdmB) also displays only the ß-subunit of an acyl-CoA carboxylase. Li *et al.* mentioned that the formation of an active complex may be possible with partnering BC and BCCP, encoded elsewhere within the genome of the respective strain (Li *et al*., 2009).



Phenylnannolone A 2.8mg/L

Phenylnannolone B 0.4 mg/L

Figure 5.5 Structures of Phenylnannolone A and B. The concentration of the produced amount of natural products in the culture broth is given.

The active site residue of Phn1 comprises an alanine instead of the conserved aspartate. Mutational experiments for a PCC in *S. coelicolor* showed that mutations of this type led to a clear shift in substrate preference from propionyl-CoA to butyryl-CoA. The kinetic parameters for the mutant PCC_{A422D} showed a 2.3 fold higher substrate specificity towards butyryl-CoA as compared to propionyl-CoA (Diacovich *et al.*, 2004, Arabolaza et al., 2010). Thus, Phn1 is more likely a buytryl-CoA carboxylase (BCC) than a PCC, and hence provides preferentially ethylmalonyl-CoA for the PKS assembly line. This is also corroborated by the ratio of in vivo produced amounts of phenylnannolone A (2.8 mg/ l) and phenylnannolone B (0.4 mg/ l), showing a 7x higher prevalence for ethylmalonyl-CoA than methylmalonyl-CoA (Ohlendorf *et al.*, *2008*). Phn1 is to our knowledge the first BCC found in a biosynthetic gene cluster for an ethylsubstituted natural compound.

5.3.2 AMP-ligase activating cinnamic acid (starter unit)

The PKS assembly line encoded on *phn2* starts with a loading module, a didomain that comprises an AMP-ligase and an acyl carrier protein. AMP-ligases belong to the same class of enzymes as adenylation (A) domains, called the AMP binding enzymes. Both act in the same way, i.e. they adenylate their substrates, which enables the loading onto PKS or NRPS assembly lines.

Previous feeding experiments (Ohlendorf, 2008) with labeled building blocks showed that eight carbon atoms from phenylalanine, i.e. a C_6C_2 (phenylethyl) unit, was incorporated into the structure of phenylnannolone A, whereas the carbonyl-C atom (C1) of phenylalanine was not retrieved in the labeled molecule.

To investigate the substrate selectivity of the AMP-ligase heterlogous expression of the recombinant protein and subsequent performance of an ATP-PP_i exchange assay (Phelan *et al*., 2009) were necessary. For the latter several phenylpropanoids (C_6C_3) were tested as well as phenylacetate (C_6C_2) and some amino acids. The selectivity for cinnamic acid was striking, as an exchange of 47% was measured. Hydrocinnamic acid showed only a third of this exchange activity (17.7%). All the other measured substrates did not reach more than 1.1% exchange activity, hence the selectivity for the AMP-ligase in Phn2 is significantly pointing towards cinnamic acid. This outcome supports the hypothesis for phenylnannolone A biosynthesis

(figure 5.2) that can be postulated from the domain order of the PKS assembly line and the molecule's structure.

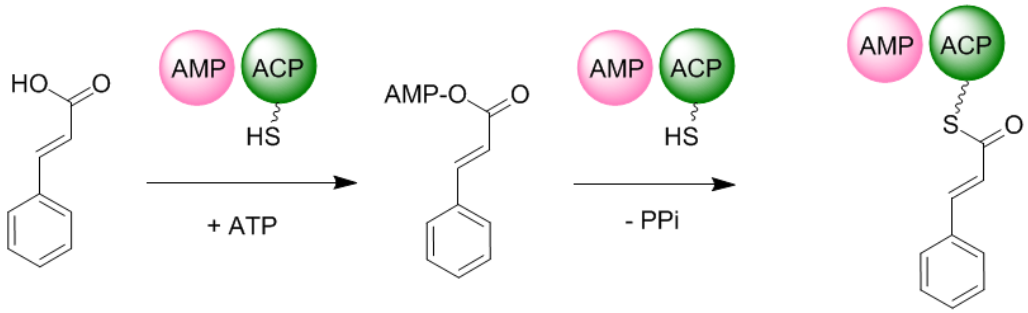


Figure 5.6 Activation of cinnamic acid by the AMP-ligase and loading onto the ACP domain

As far as we can conclude from the results of the ATP-PP_i exchange assay on the *apo*-enzyme, cinnamic acid is adenylated by the AMP-ligase and subsequently loaded to the the neighboring ACP domain (figure 5.6). Imitating the *in vivo* conditions, where the activated *holo*-enzyme loads its substrate and is subsequently analyzed with mass spectrometry, would give more clarity regarding the starter unit. For this purpose, the purified *apo*-AMP-ACP didomain has to be incubated with an ACP transferase or with a phosphopantetheinyl transferase to form the holo-form. When a mixture of substrates is then offered to this holo-enzyme, the favored substrate should then be bound to the pantetheine arm. A subsequent trypsin digest should then enable the characterization of the ACP active site via MALDI-TOF, as it was described for the didomain from the NRPS/PKS system forming microcystin (Hicks *et al*., 2006).

Regarding the feeding experiments reported by Ohlendorf *et al.* however, there are still some open questions to be answered. Similar results for feeding experiments were observed in the cyanobacterium *Microcystis aeruginosa*, where phenylalanine was incorporated in the same manner into microcystin. The incorporation of the C_6C_2 unit gave rise to the hypothesis that phenylacetate might be the correct starter unit. This hypothesis could be dismissed as feeding experiments with labeled phenylacetate failed for microcystin. This was also the case for phenylnannolone A (Ohlendorf, 2008, Hicks *et al*., 2006). Same applies to the starter units in nodularin and cryptophycin biosynthesis, both isolated from cyanobacteria (Rinehart *et al*., 1994; Moffitt and Neilan, 2004; Magarvey *et al*., 2006) (figure 5.7).

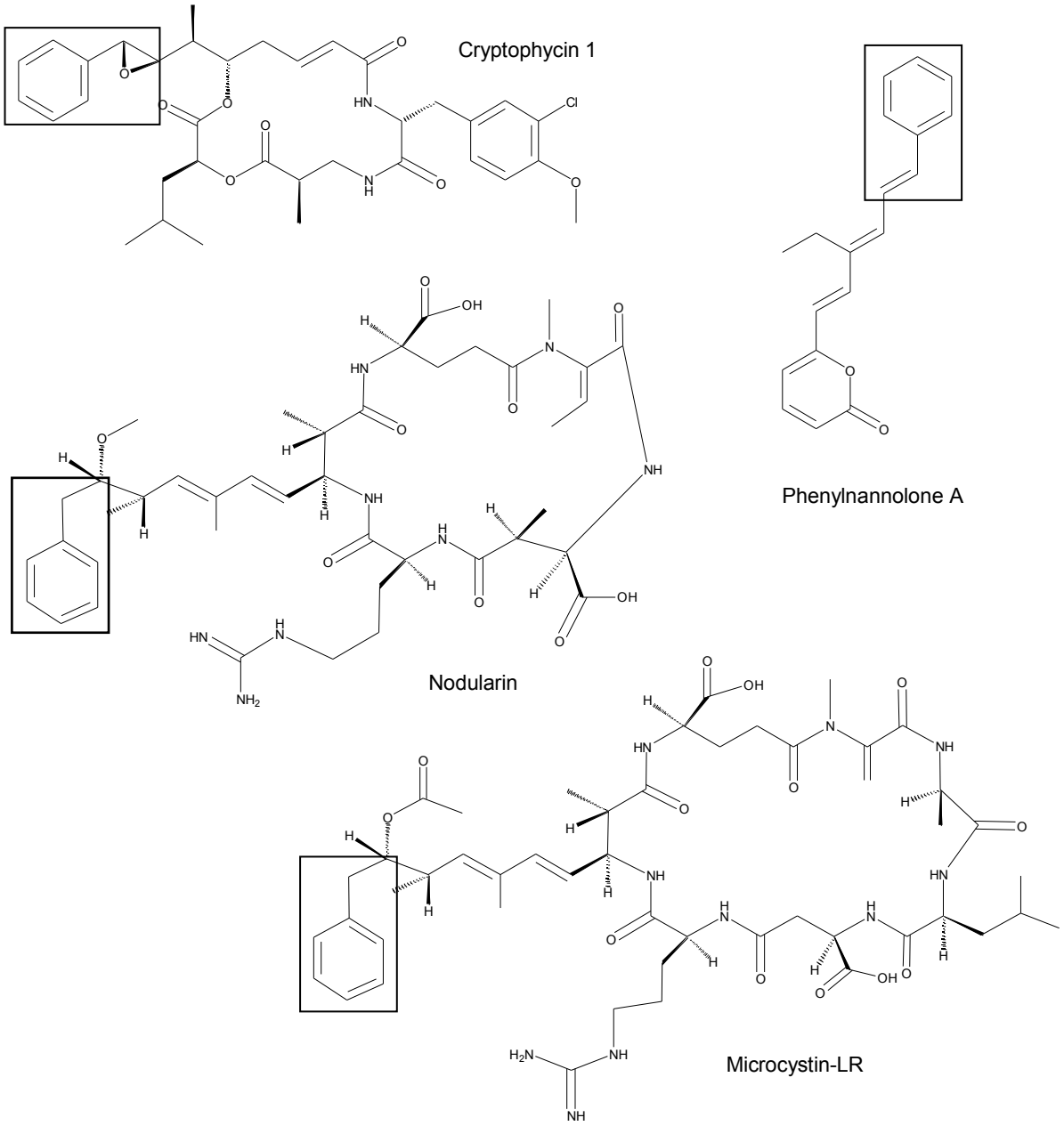


Figure 5.7 Natural compounds believed to initate biosynthesis with a phenylalanine-derived C_6C_2 starter unit (C_6C_2 unit is framed)

Biochemical characterization of an A-PCP didomain from McyG, a NRPS/PKS system involved in microcystin biosynthesis, revealed that several phenylpropanoids (C_6C_3) are preferentially activated and loaded onto the PCP domain, and not the estimated phenylacetate (C_6C_2). Interestingly the result of the ATP-PP_i exchange assay for the *apo* MycG A-PCP didomain did not coincide with the loading experiments of the *in vitro* and *in vivo* generated *holo*- A-PCP. To analyze substrate loading onto the PCP of the didomain Hicks and co-workers generated the *holo* A-PCP didomain *in vitro* by adding the PPTase Sfp from *Bacillus subtilis* or co-

expressed the didomain with the Svp PPTAse protein from *Streptomyces verticillus* to obtain the *in vivo* generated *holo* form. The ATP-PP_i exchange assay with the *apo*-didomain revealed *trans*-cinnamic acid as a putative substrate, whereas the *in vitro* generated *holo* A-PCP seemed to load substrates such as phenyllactate, hydrocinnamate and D-phenylalanine likewise. However, the results of the *in vivo* loading experiments pointed towards 3-phenyllactate, which is considered to be the true starter unit. The latter result was achieved by changing the Terrific Broth (TB) fermentation medium with a glucose-defined medium, as Hicks and co-workers suspected the hydrocinnamic acid to be derived from the TB medium (Hicks et al., 2006). Even though these results are to some extent inconsistent, it seems clear that initially a C_6C_3 unit is employed as a starter.

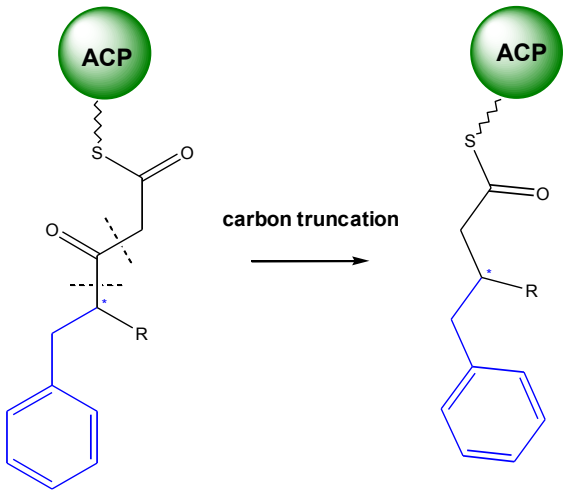


Figure 5.8 Carbon truncation of the C_6-C_3 starter unit, proposed for microcystin and cryptophycin biosynthesis: The mechanism for this reaction remains unclear according to Hicks *et al*., 2006 (asterisk marks C-2 of phenylalanine)

A remaining mystery is the mechanistic basis for an unprecedented one-carbon truncation within the biosynthesis of cryptophycin, microcystin and nodularin. If the phenylpropanoid is chain extended with malonyl-CoA by the PKS machinery, then the phenylalanine-derived carboxyl carbon would need to be excised, resulting in the cleavage of two C–C bonds and the formation of a new C–C bond (Hicks *et al*., 2006; Magarvey *et al*., 2006) (figure 5.8). In contrast to these cyanobacterial secondary metabolites, the phenylalanine derived-carboxyl carbon should not only be cleaved, but after the extension with an acetate only a single carbon atom (C-2) should remain, to explain the results of former feeding studies. This requires then a cleavage

of two carbon atoms, the carbonyl-atom from the C_6C_3 starter unit and the carbonyl-thioester (C-1).

From the domain order of the PKS encoded by *phn2* and the molecule's structure the starter unit should be a cinnamic acid that is then elongated in four steps. The biosynthetic origin of cinnamic acid in bacteria as myxobacteria and cyanobacteria is reported to be a phenylalanine ammonia lyase (PAL). In a single reaction the PAL reduces phenylalanine to *trans*-cinnamic acid, while every carbon atom is preserved and no carbon cleavage takes place. However, this does not explain the unusual results of feeding experiments, where a single carbon atom derived from acetate is incorporated. As the feeding experiments were only carried out once, feeding with labeled precursor should be repeated, considering also cinnamic acid as a putative precursor.

The Stachelhaus code of the A domains, involved in nodularin (NdaC), microcystin (McyG) and cryptophycin (CrpA) biosynthesis, all with nearly identical signatures, was compared with the AMP-ligase of Phn2, putatively involved in phenylnannolone's biosynthesis. The comparision showed no similarities between the AMP-ligase from Phn2 and the A domains of NdaC, CrpA or McyG. This supports the hypothesis that the starter units for nodularin, microcystin and cryptophycin biosynthesis are the same, but different from the one of phenylnannolone (table 5.2).

Table 5.2 Stachelhaus code of A domains activating phenylalanine-derived starter units

Protein	Stachelhaus code	Natural compound	Expected starter unit *
NdaC	**L W V A A S G K**	Nodularin	Phenylacetate
McyG	**L W V A A S G**	Microcystin	Phenylacetate
CrpA	**L W V A A S G**	Cryptophycin	Phenylacetate
Phn2	**D L V G F G C G M**	Phenylnannolone	Cinnamic acid

*: from domain order and molecule structure

This outcome corroborates that the AMP-ligase in Phn2 activates a different starter unit, than those activated by the A domains of the cyanobacterial compounds and might not share the mysterious mechanism for carbon truncation, postulated for cyanobacteria.

In type III PKS of plants and bacteria cinnamoyl-CoA is a quite common starter unit for stilbenes, curcumin and flavonoid biosynthesis (Dewick, p.109, Hertweck, 2009).

For the modular type I PKS however, it was not yet described to be involved in the PKS assembly line.

5.3.3 AT_1 domain loading ethylmalonyl-CoA

Acyltransferases (AT) in type I modular PKS extender units recognize, and load specific substrates onto the corresponding acyl carrier proteins (ACP). Each AT shows a high selecitivity towards its desired substrate, which is based on the protein's sequence and structure. In general, extender units as malonyl-CoA and methylmalonyl-CoA are used for chain elongation. The motifs in ATs specific for malonyl-CoA and methylmalonyl-CoA are quite well known, as they were intensely investigated and a large number of sequence data for those ATs is available (Yadav *et al*., 2003). Albeit distinguishing between a malonyl (m) - and methylmalonyl-CoA (mm) specific AT on the molecular level is relatively reliable, there are always exceptions to the rule. However, there are less differences between methylmalonyl-CoA and ethylmalonyl-CoA (em) specific ATs; hence distinction between these two types of ATs is rather difficult.

Three motifs are described to determine substrate specificity in ATs: The 'RDVVQ' motif, the GHSXG motif and the YASH motif. The first and the second motif are nearly identical and thus not useful for the discrimination between mmATs and emAT. For the third motif (YASH) there are some slight variations between mmATs and emATs. The serine in the YASH motif is highly conserved in mmATs, in some emATs it is also present but often replaced by some other small amino acid, i.e. glycine, threonine and alanine (see figure 4.14). In the case of AT_1, this amino acid is a threonine, just as in emATs from oligomycin and ascomycin biosynthesis. The tyrosine (Y) in this motif is also indicative for a mmAT, which is required for the formation of a hydrophobic pocket that is important for the binding of methylmalonyl-CoA. In some cases this residue is altered and replaced by a valine or tryptophan (Yadav *et al*., 2003, Smith and Tsai, 2007). Multiple sequence alignments revealed for emATs at this position amino acids such as valine, cysteine, threonine or isoleucine (see figure 4.14).

Multiple sequence alignment with malonyl-, methylmalonyl- and ethylmalonyl-CoA specific ATs showed that mATs and mmATs are clearly divided into different clusters in phylogenetic trees. But ATs specific for ethylmalonyl-CoA and methoxymalonyl-

CoA are in the same cluster of mmATs, therefore a clear distinguishing between these three types of ATs is not possible (Minowa *et al*., 2007). Phylogenetic analysis of three types of ATs (mAT, mmAt and emAT) with those from phenylnannolone's gene cluster showed that AT_2, AT_3 and AT_4 cluster with those specific for malonyl-CoA, whereas AT_1 clusters with ATs specific for methylmalonyl-CoA as well with those specific for ethylmalonyl-CoA.

Computational analysis of the AT domains from Phn2 with AntiSMASH, using two different methods for substrate prediction, yielded contradictory results. The forecast with Yadav's method points in the direction of methoxymalonyl-CoA, whereas Minowa's method predicted either methylmalonyl-CoA (score: 175.2) or ethylmalonyl-CoA (score: 173.9) as possible substrate (figure 8.3). Yadav and co-workers used 13 diagnostic amino acids and 11 conserved residues, extracted from multiple sequence alignments, for computational prediction of substrate specificity (Yadav *et al*., 2003). Minowa and collaborators used homology search with Hidden Markov Model (HMM) profiles, which is a more sensitive method then BLAST used by Yadav and co-workers (Thesis Zucko, 2010).

Summarizing the above, the methods available for substrate prediction could not clearly distinguish, whether the AT_1 domain in Phn2 has prevalence for methylmalonyl-CoA or ethylmalonyl-CoA. But the analysis of the YASH motif of AT_1 described above and in section 4.8.3.4, suggests it to be responsible for the loading of ethylmalonyl-CoA to the corresponding ACP. This is supported by the feeding experiments with 1-^{13}C-butyrate, which was incorporated in phenylnannolone A forming the ethylside chain (Ohlendorf *et al*., 2008).

Furthermore, substrate prevalence for ethylmalonyl-CoA is corroborated by the production rate of the different phenylnannolone derivatives. A 7x higher production rate was observed for the ethyl-branched phenylnannolone A than for the methyl-branched phenylnannolone B (Ohlendorf *et al*., 2008).This might not only result from the higher specificity of the AT towards ethylmalonyl-CoA, but also, as discussed in section 5.3.1, result from a higher supply of the precursor ethylmalonyl-CoA.

5.3.4 Stereochemistry of double bonds

Ketoreductase (KR) domains catalyze the stereospecific reduction of the ß-keto function to a ß-hydroxyl group. Subsequent to this NADPH dependent reduction, water elimination is initated by a dehydratase domain (DH) that forms the C-C double bond. The classification of KRs is depending on the resulting stereochemistry. An A-type KR generates L-hydroxyl-groups, whereas B-type KRs introduce a D-hydroxyl group (Keatinge-Clay, 2007, Caffrey, 2003). Some KR domains in extension modules with propionate extender units are able to epimerize the α-carbon during polyketide synthesis. These domains that introduce a stereochemical change of the methylgroup (α-substituent) are denoted with a '2' (A2 and B2). Whereas KRs, which perform a reduction on an unepimerized D-α-substituent, are denoted with a '1' (A1 and B1).

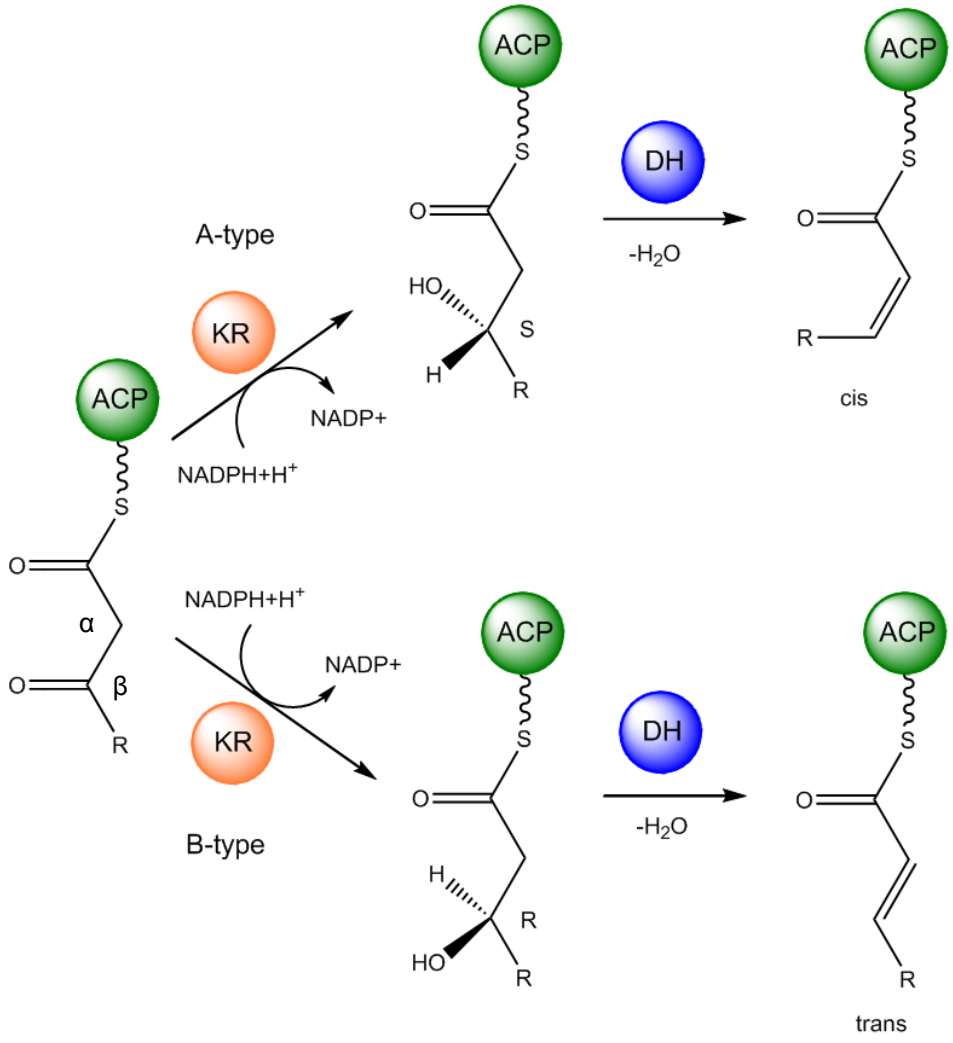


Figure 5.9 Stereochemistry of ketoreduction and double-bond formation

A third group of KRs are classified as ‘C-type’ and represent the reductase-incompetent domains present in some modular PKS. KRs denoted as ‘C1-type’ show no reduction and no epimerization and are only mentioned for the sake of completeness. The rare ‘C2-type’ KRs however, are able to epimerize D-α-substituents and act as racemases (Zheng *et al*., 2011). In general most of the KRs in PKS gene clusters are of ‘B1-type’. Therefore, most of the double bonds in polyketides are *trans* configurated. *Cis*-configured double bonds are rather rare in polyketides. Such Z double bonds are present in pikromycin, chivosazol, disorazole and etnangien (figure 5.10) (Kittendorf and Sherman, 2009; Perlova et al., 2006; Elnakady et al., 2004; Carvalho et al., 2005; Menche et al., 2008).

The KR domains of the phenylnannolone A gene cluster were characterized by several conserved motifs that allowed stereochemical prediction (‘A-’ or ‘B-type’) of the resulting hydroxyl bearing carbon atom. KR_1 and KR_2 were identified as B1-type ketoreductase domains. Such a B1-type KR domain results an D-hydroxyl-substituted thioester; a neighboring DH domain may then form in a *syn* dehydration a *trans* double bond. However, KR_4 lacks the LDD motif, present in KR_1 and KR_2 that is found in ‘B-type’ KRs. Although tryptophan is present in nearly all A-type KRs, there are a few examples of KR domains that generate L-hydroxyl-groups despite the missing tryptophan-motif. Such domains are found in stigmatellin, chivosazol and disorazol biosynthesis (Gaitatzis et al., 2002; Perlova et al., 2006; Carvalho et al., 2005). Hence, even if the ‘A-type’ characteristic tryptophan (W) is not present in KR_4 and is replaced by an isoleucine, it is still likely that this domain reduces the ß-hydroxyl group in an ‘A-type’-style and results a *cis* double bond after dehydration. KR_4 from Phn2 showed in a BLAST analysis (table 4.14) the highest identity to a KR from StiD, involved in stigmatellin biosynthesis in *Stigmatella aurantiaca*. This KR from StiD converts the ß-keto group to an L-hydroxyl-group, which is subsequently methylated via a methyltransferase (Gaitatzis *et al*., 1999). Multiple sequence alignment (figure 4.15) could show that the tryptophan-motif is replaced by a methionine in StiD’s KR.

In chivosazol biosynthesis KR4 and KR14 are classified as A-type KRs which catalyte the formation of *cis*-double bonds. Both KRs lack the tryptophan-motif, which is replaced by leucine and valine, respectively.

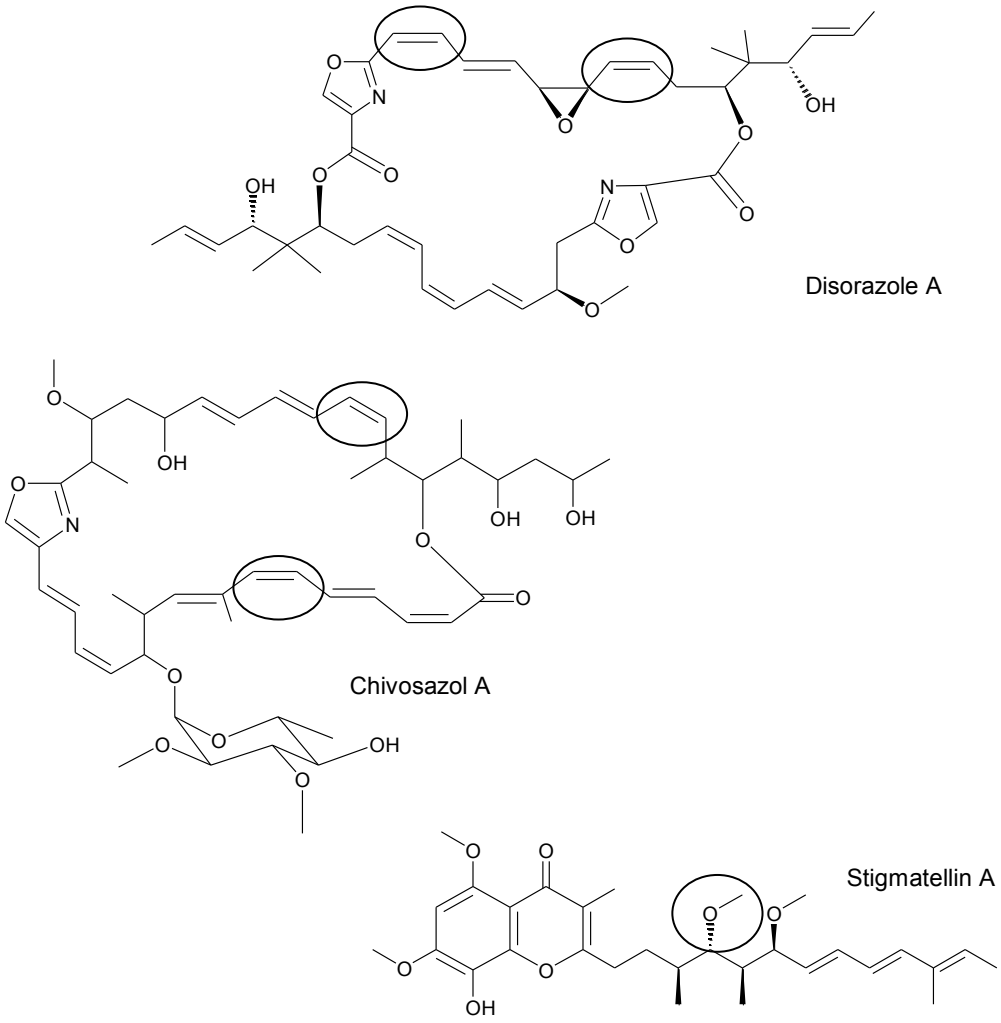


Figure 5.10 Natural compounds with *cis*-double bonds or L-methoxy-group derived from A-type KRs: Disorazole A and chivosazol A from *Sorangium cellulosum* and stigmatellin A from *Stigamtella aurantiaca*; cis double bonds and the L-methoxy group are circled.

Perlova and co-workers characterized the KRs in 2001 with a simpler method, where the presence or absence of the last aspartate in the LDD-motif was used for distinction between A-type and B-type KRs. A third KR, KR11, was due to the presence of the key aspartatic acid in position 151 predicted to be of B-type, but is based on the molecule's structure supposed to produce an L-hydroxyl- moiety. Albeit structural and functional analysis for ketoreductases was until then not available, they mentioned that the diagnostic aspartate residue might not always define the direction of reduction (Perlova *et al*., 2006). A multiple sequence alignment of KRs from gene

clusters of chivosazol, disorazole and phenylnannolone A (data not shown) revealed for this domain (KR11) the characteristic A-type tryptophan-motif.

A third example for A-type KRs lacking the tryptophan-motif is found in the disorazole gene cluster, where the tryptophan is replaced in KR4, KR5 and KR7 each by valine, methionine and leucine. If we appreciate all these examples of A-type KRs without tryptophan-motifs, it is very likely that KR_4 in Phn2 generates an L-hydroxyl-group, which is further reduced by DH_4 to a *cis*-double bond. To support this assertion the functional proof via heterologous expression of this domain and subsequent enzyme assay is necessary. Summarizing the above, it can be said that, most of the KRs can be annotated with the described classification system (table 4.13) but some exceptions are known.

5.3.5 Lactone ring formation through TE

Thioesterase domains in PKS release the polyketide chain from the PKS multienzyme either by hydrolysis (intermolecular release) or cyclization (intramolecular release). For lactonization a hydroxyl group of the polyketide chain acts as an internal nucleophile that attacks the carbonyl carbon of acyl-*O*-TE (Du and Lou, 2009). Structure elucidation of phenylnannolone A revealed a terminal pyrone ring. For the formation of this pyrone ring keto-enol tautomerism of the unreduced carbonyl group (module 3) is expected, resulting in an enol functionality. The so formed hydroxyl group acts therefore as a nucleophile, which forms together with the carbonyl carbon of the thioester the lactone ring (figure 5.11).

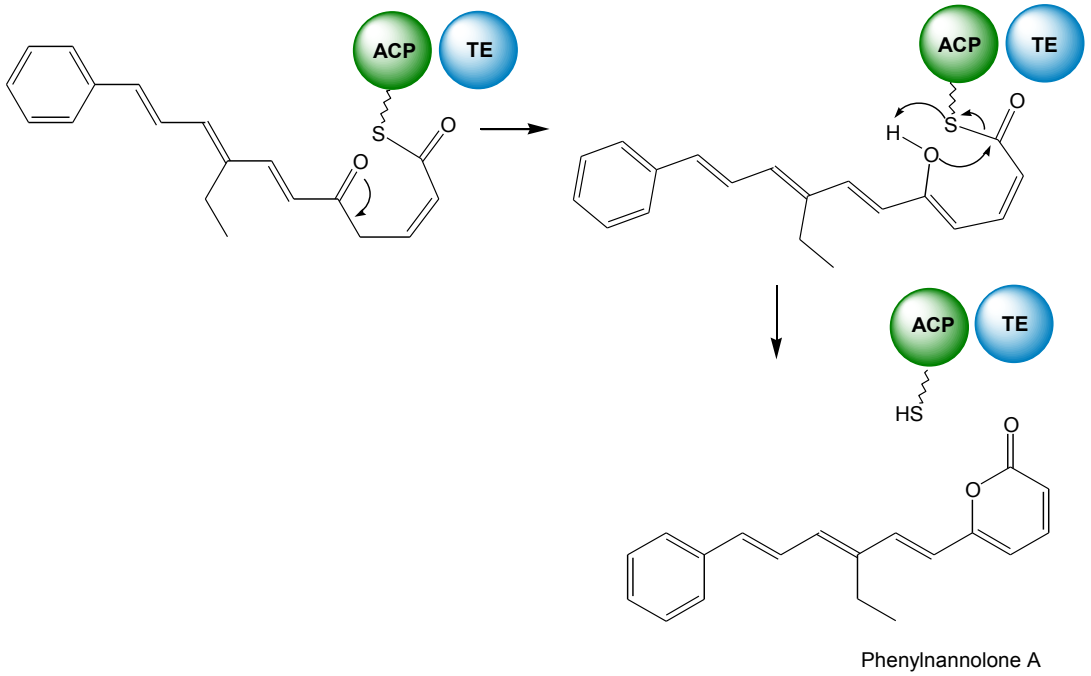


Figure 5.11 Proposed pyrone ring formation for phenylnannolone A by TE domain

A TE-induced pyrone ring formation is somehow uncommon for modular type I PKS. In literature triketide lactonization is studied for DEBS 1-TE, a truncated form of the erythromycin polyketide synthase (Bycroft *et al*., 2001, Moffet *et al*., 2006). Although TE-induced pyrone lactonization is not reported for modular type I PKS, there exists one example for pyrane formation in type I PKS, the jerangolid biosynthesis (figure 5.12), following the same mechanism as the one proposed for phenylnannolone A (Julien *et al*., 2006).

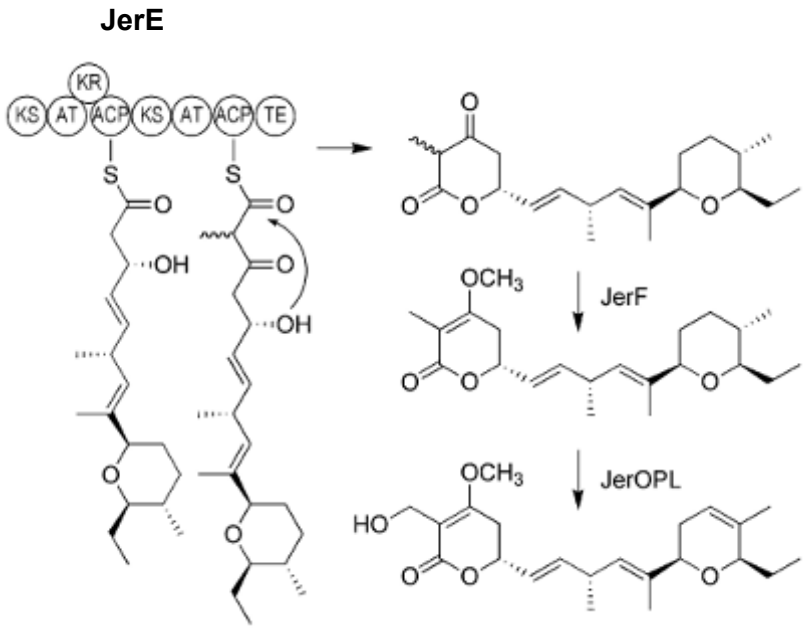


Figure 5.12 Hypothetical scheme for completion of jerangolid biosynthesis (figure origin: Julien *et al*., 2006)

From the results obtained, we conclude that the biosynthesis of phenylnannolone A in *N. pusilla* B150 involves a PKS and not a NRPS/PKS system. The bioinformatical analysis of the biosynthetic genes *phn1* and *phn2* showed that the domain order of the identified gene cluster is consistent with the putative biosynthesis that we postulated for phenylnannolone A. Biochemical studies on the loading module of the putative phenylnannolone gene cluster were performed, and indicate cinnamic acid to be the starter unit for phenylnannolone A biosynthesis. Further investigations on the formation of the precursor ethylmalonyl-CoA including heterologous expression of the putative butyryl-CoA carboxylase (BCC), would corroborate our findings and would represent the first BCC in a microbial biosynthetic gene cluster.

5.4 A mixed NRPS/PKS biosynthetic gene cluster

Looking for the putative biosynthetic gene cluster of phenylnannolone A, a small NRPS/PKS gene (*sb1*) was identified on fosmid 21H12. The gene *sb1* covers 10 kb sequence information. The corresponding gene product consists of 3 modules: The first two modules belong to the NRPS part, whereas the third module forms the PKS part. From the domain structure and rule of colinearity a putative biosynthetic pathway and a corresponding metabolite are proposed. The first module is comprised of a condensation (C), adenylation (A) and a peptidyl carrier protein (PCP). This is somehow unusual as loading modules in NRPS usually consist only of an A domain with an adjacent PCP. Sequence alignment revealed this C_1 domain to be inactive, as it lacks the conserved histidine. Computational analysis on substrate specificity of the A domain from the loading module revealed proline as a possible starter unit. The adenylated proline is thus expected to be initially loaded to the neighborring PCP domain. In the second module the A domain activates possibly a phenylalanine unit that will be attached to the pantetheine arm of the adjacent PCP. The C_2 domain then would form a peptide bond between the proline and phenylalanine. The peptide chain is elongated with an acetate or a propionate unit by the KS from the PKS module in the last step. The TE domain in the PKS module terminates chain elongation and releases the chain by cyclization from the multienzyme.

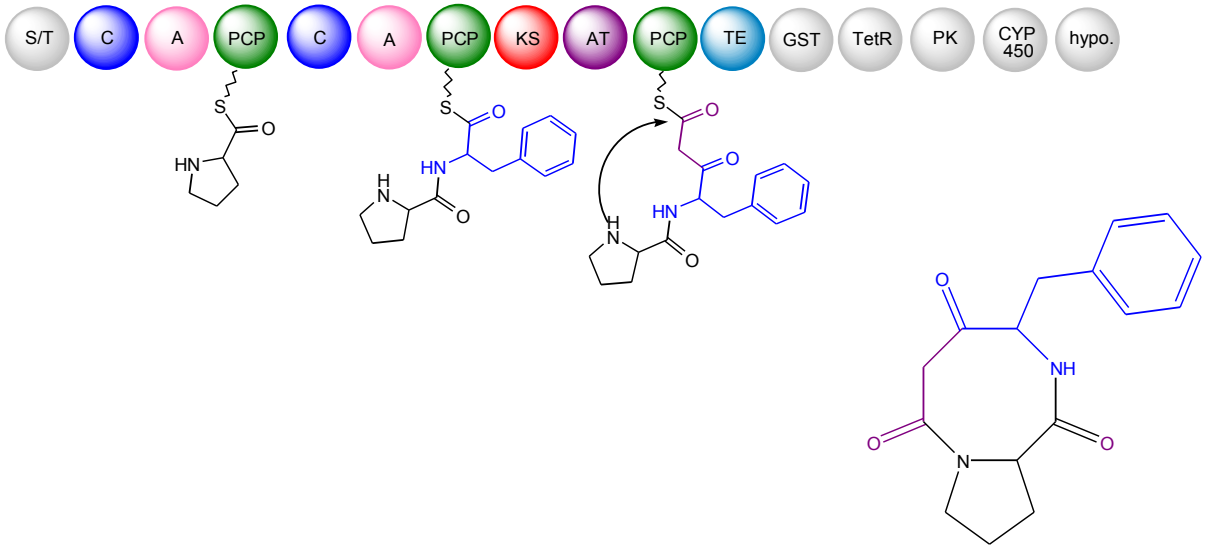


Figure 5.13 NRPS/PKS assembly line encoded by *sb1* with proposal for a corresponding structure
S/T=serine/threonine kinase, C=condensation domain, A= adenylation domain, PCP=peptidyl carrier protein domain, KS= ketosynthase, AT=acyltransferase, TE=thioesterase, GST=glutathione S-transferase, TetR=transcriptional regulator TetR, PK=protein kinase, CYP=cytochrome P450, hypo.=hypothetical protein

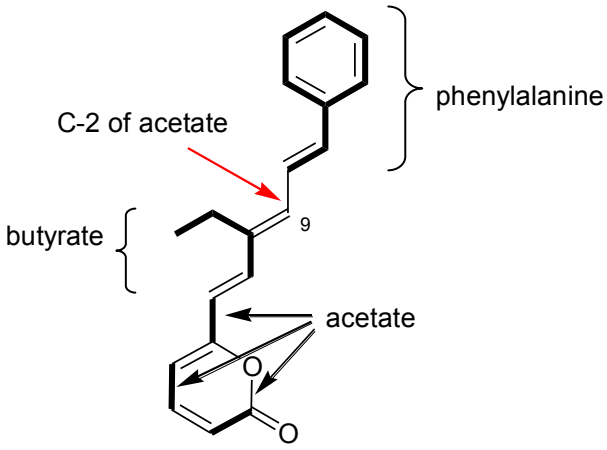
Post-modifications by the neighboring CYP 450 are possible, and may lead to another structure than the one proposed in figure 5.13. The other surrounding enzymes, such as the the transcriptional regulator TetR and the serine/threonine kinase (S/T) are usually present in biosynthetic gene clusters and are responsible for the transcriptional control of the biosynthetic pathway and signal transduction. They are usually located on the borders of biosynthetic gene clusters.

This small NRPS/PKS gene demonstrates the other way for the identification of a biosynthetic gene cluster, where genes are identified by bioinformatical analysis of genome sequences. From the bioinformatical analysis structure predictions for modular assembly lines are possible, following the rule of colinearity, but not always match the later identified structures. This approach, where the genomes of microorganisms are screened for biosynthetic genes with the aim to identify novel secondary metabolites is called “genomic mining”.

6 Summary

Myxobacteria are gliding microorganisms known for their unique biosynthetic capabilities. *Nannocystis pusilla* B150 is a myxobacterial strain that produces a new group of natural products, named phenylnannolones. Phenylnannolone A shows inhibitory activity towards the ABCB1 gene product p-glycoprotein and reverses daunorubicin resistance in cultured cancer cells. Previous feeding experiments with ^{13}C-labeled precursors showed the incorporation of acetate and butyrate, but also of phenylalanine into phenylnannolone A (Ohlendorf *et al.,* 2008). Thus, a mixed non-ribosomal peptide synthetase/polyketide synthase (NRPS/PKS) system or a polyketide synthase (PKS) was suggested to be responsible for the biosynthesis of this natural product.

The aim of this research project was to elucidate the genetic information encoding the biosynthetic enzymes responsible for phenylnannolone A formation. 454 sequencing of the *N. pusilla* B150 genome proved to be extremely difficult due to the GC-rich myxobacterial DNA. 3804 contigs were obtained, but were too small to delineate the complete biosynthetic gene cluster for phenylnannolone A. Therefore, a genomic library of *N. pusilla* B150 was constructed and screened with suitable primer pairs, deduced from sequences of the 454 assembly.



Phenylnannolone A

Figure 6.1 Molecular structure of phenylnannolone A with biosynthetic building blocks (in bold). Feeding experiments in *N. pusilla* B150 with ^{13}C-labeled building blocks showed incorporation of a phenylalanine derived starter unit, one butyrate and three acetate building blocks. Unprecedented and unusual is the incorporation of a C-2 of acetate in position 9

Fosmid sequencing revealed a PKS gene cluster (24 kb), whose domain order is consistent with the putative biosynthesis that we postulated for phenylnannolone A. It comprises a putative carboxyl-transferase gene (*phn1*) and a modular type I PKS gene (*phn2*), putatively responsible for the formation of phenylnannolone A. *Phn2* has a gene size of 22 kb and its gene product consists of 5 modules: A loading module and four modules for the extension of the polyketide chain (figure 6.2).

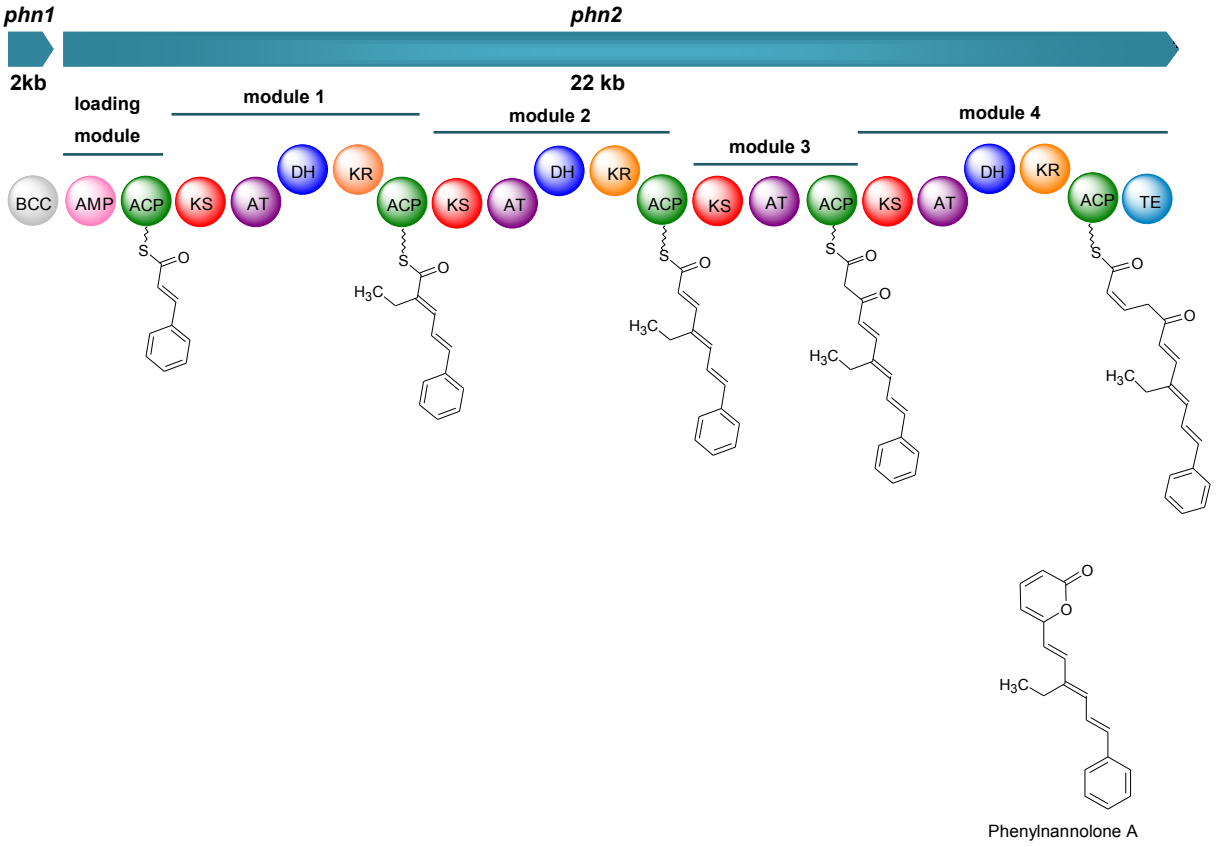


Figure 6.2 Hypothetical pathway for the biosynthesis of phenylnannolone A: BCC=butyryl-CoA carboxylase AMP=AMP-dependent ligase, ACP=acyl carrier protein, KS=ketosynthase domain, AT=acyltransferase domain, DH=dehydratase domain, KR=ketoreductase domain, TE=thioesterase domain

From the results obtained, we concluded that the biosynthesis of phenylnannolone A in *N. pusilla* B150 involves a PKS and not a NRPS/PKS system. However, the putative loading module comprises an AMP-synthetase/ligase that activates the aromatic starter unit. To investigate the functionality and substrate selectivity of the AMP-ligase, a recombinant protein was expressed. In a subsequent γ-$^{18}O_4$-ATP pyrophosphate exchange assay (Phelan *et al.*, 2009) with putative substrates the specific activation of cinnamic acid by this AMP-ligase was shown.

During the bioinformatic analysis of the genomic library for the phenylnannolone gene cluster, a small NRPS/PKS gene (10 kb) was identified on another fosmid clone. Sequence analysis of the gene and its corresponding protein product led to the prediction of an unknown peptide with three putative building blocks: proline, phenylalanine and acetate.

The current project led to the identification of two biosynthetic gene clusters from the myxobacterium *N. pusilla B150*. Biochemical studies on the loading module of the putative phenylnannolone gene cluster were performed, and indicate cinnamic acid to be the starter unit for phenylnannolone A biosynthesis.

7 References

Alber BE. (2011). Biotechnological potential of the ethylmalonyl-CoA pathway. *Appl. Microbiol. Biotechnol.* **89**: 17–25.

Ansari MZ, Yadav G, Gokhale RS, and Mohanty D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32**: W405–413.

Arabolaza A, Shillito ME, Lin T-W, Diacovich L, Melgar M, Pham H, *et al.* (2010). Crystal Structures and Mutational Analyses of Acyl-CoA Carboxylase β Subunit of *Streptomyces coelicolor*. *Biochemistry* **49**: 7367–7376.

Baker DD, and Alvi KA. (2004). Small-molecule natural products: new structures, new activities. *Current Opinion in Biotechnology* **15**: 576–583.

Bergendahl V, Linne U, and Marahiel MA. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry* **269**: 620–629.

Beyer S, Kunze B, Silakowski B, and Müller R. (1999). Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim. Biophys. Acta* **1445**: 185–195.

Bisang C, Long PF, Corte´s J, Westcott J, Crosby J, Matharu A-L, *et al.* (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**: 502–505.

Bollag DM, McQueney PA, Zhu J, Hensens O, Koupal L, Liesch J, *et al.* (1995). Epothilones, a New Class of Microtubule-Stabilizing Agents with a Taxol-Like Mechanism of Action. *Cancer Res* **55**: 2325–2333.

Bravo LTC, Tuohy MJ, and Shrestha NK. (2010). Successful pyrosequencing of GC-rich DNA sequences by partial substitution of deoxyguanosine with deoxyinosine. *Diagn. Mol. Pathol.* **19**: 123–125.

Broadhurst RW, Nietlispach D, Wheatcroft MP, Leadlay PF, and Weissman KJ. (2003). The Structure of Docking Domains in Modular Polyketide Synthases. *Chemistry & Biology* **10**: 723–731.

Buntin K, Irschik H, Weissman KJ, Luxenburger E, Blöcker H, and Müller R. (2010). Biosynthesis of Thuggacins in Myxobacteria: Comparative Cluster Analysis Reveals Basis for Natural Product Structural Diversity. *Chemistry & Biology* **17**: 342–356.

Buntin K. (2010). Studies on the biosynthesis of complex natural products from myxobacteria. Available from: http://scidok.sulb.uni-saarland.de/volltexte/2010/3152/

Butler MS. (2004). The Role of Natural Product Chemistry in Drug Discovery. *J. Nat. Prod.* **67**: 2141–2153.

Bycroft M, Weissman KJ, Staunton J, and Leadlay PF. (2000). Efficient purification and kinetic characterization of a bimodular derivative of the erythromycin polyketide synthase. *European Journal of Biochemistry* **267**: 520–526.

Byers DM, and Gong H. (2007). Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem. Cell Biol.* **85**: 649–662.

Caffrey P. (2003). Conserved Amino Acid Residues Correlating With Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **4**: 654–657.

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

Challis GL, Ravel J, and Townsend CA. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**: 211–224.

Chan YA, Podevels AM, Kevany BM, and Thomas MG. (2009). Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* **26**: 90–114.

Chang KH, Xiang H, and Dunaway-Mariano D. (1997). Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. *Biochemistry* **36**: 15650–15659.

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

Clarke L, and Carbon J. (1992). A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome. 1976. *Biotechnology* **24**: 179–187.

Conlin A, Fornier M, Hudis C, Kar S, and Kirkpatrick P. (2007). Ixabepilone. *Nature Reviews Drug Discovery* **6**: 953–954.

Dagert M, and Ehrlich SD. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. *Gene* **6**: 23–28.

Dereeper A, Audic S, Claverie J-M, and Blanc G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **10**: 8.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, *et al.* (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**: W465–469.

Diacovich L, Mitchell DL, Pham H, Gago G, Melgar MM, Khosla C, *et al.* (2004). Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. *Biochemistry* **43**: 14027–14036.

Du L, and Lou L. (2010). PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**: 255–278.

Du L, Sánchez C, Chen M, Edwards DJ, and Shen B. (2000). 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.

Du L, and Shen B. (2001). Biosynthesis of hybrid peptide-polyketide natural products. *Curr Opin Drug Discov Devel* **4**: 215–228.

Dworkin M. (2007). Lingering Puzzles about Myxobacteria. *Microbe* **2**: 18–24.

Eckert WA, and Kartenbeck J. (1997). Proteine: Standardmethoden Der Molekular- Und Zellbiologie. Springer DE.

Edwards U, Rogall T, Blöcker H, Emde M, and Böttger EC. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* **17**: 7843–7853.

Elnakady YA, Sasse F, Lünsdorf H, and Reichenbach H. (2004). Disorazol A1, a highly effective antimitotic agent acting on tubulin polymerization and inducing apoptosis in mammalian cells. *Biochemical Pharmacology* **67**: 927–935.

Erb TJ, Rétey J, Fuchs G, and Alber BE. (2008). Ethylmalonyl-CoA Mutase from Rhodobacter Sphaeroides Defines a New Subclade of Coenzyme B12-Dependent Acyl-CoA Mutases. *J. Biol. Chem.* **283**: 32283–32293.

Erol Ö, Schäberle TF, Schmitz A, Rachid S, Gurgui C, El Omari M, *et al.* (2010). Biosynthesis of the Myxobacterial Antibiotic Corallopyronin A. *ChemBioChem* **11**: 1253–1265.

Falkow S, Rosenberg E, Schleifer K-H, and Stackebrandt E. (2006). The Prokaryotes: Vol. 7: Proteobacteria: Delta and Epsilon Subclasses. Deeply Rooting Bacteria. Springer.

Findlow SC, Winsor C, Simpson TJ, Crosby J, and Crump MP. (2003). Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* **42**: 8423–8433.

Fischbach MA, and Walsh CT. (2006). Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *ChemInform* **37**: no–no.

Fox E, and Bates SE. (2007). Tariquidar (XR9576): a P-glycoprotein drug efflux pump inhibitor. *Expert Rev Anticancer Ther* **7**: 447–459.

Gaitatzis N, Silakowski B, Kunze B, Nordsiek G, Blöcker H, Höfle G, *et al.* (2002). The Biosynthesis of the Aromatic Myxobacterial Electron Transport Inhibitor Stigmatellin Is Directed by a Novel Type of Modular Polyketide Synthase. *J. Biol. Chem.* **277**: 13082–13090.

Garcia RO, Krug D, and Müller R. (2009). Chapter 3. Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods. *Meth. Enzymol.* **458**: 59–91.

Gehring AM, Mori I, and Walsh CT. (1998). Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**: 2648–2659.

Gerth K, Irschik H, Reichenbach H, and Trowitzsch W. (1982). The myxovirescins, a family of antibiotics from *Myxococcus virescens* (Myxobacterales). *J. Antibiot.* **35**: 1454–1459.

Gerth K, Pradella S, Perlova O, Beyer S, and Müller R. (2003). Myxobacteria: proficient producers of novel natural products with various biological activities--past and future biotechnological aspects with the focus on the genus *Sorangium*. *J. Biotechnol.* **106**: 233–253.

Gokhale RS, and Khosla C. (2000). Role of linkers in communication between protein modules. *Current Opinion in Chemical Biology* **4**: 22–27.

Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen J, *et al.* (2006). Evolution of sensory complexity recorded in a myxobacterial genome. *Proceedings of the National Academy of Sciences* **103**: 15200–15205.

Goodin S, Kane MP, and Rubin EH. (2004). Epothilones: Mechanism of Action and Biologic Activity. *JCO* **22**: 2015–2025.

Gulick AM. (2009). Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chem. Biol.* **4**: 811–827.

Hagelueken G, Albrecht SC, Steinmetz H, Jansen R, Heinz DW, Kalesse M, *et al.* (2009). The Absolute Configuration of Rhizopodin and Its Inhibition of Actin Polymerization by Dimerization. *Angewandte Chemie International Edition* **48**: 595–598.

Hahn M, and Stachelhaus T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 275–280.

Haydock SF, Aparicio JF, Molnár I, Schwecke T, Khaw LE, König A, *et al.* (1995). Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374**: 246–248.

He J, and Hertweck C. (2003). Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* **10**: 1225–1232.

Hertweck C. (2009). The Biosynthetic Logic of Polyketide Diversity. *Angewandte Chemie International Edition* **48**: 4688–4716.

Hicks LM, Moffitt MC, Beer LL, Moore BS, and Kelleher NL. (2006). Structural Characterization of in Vitro and in Vivo Intermediates on the Loading Module of Microcystin Synthetase. *ACS Chem. Biol.* **1**: 93–102.

Hopwood DA. (1997). Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **97**: 2465–2498.

Hung T, Mak K, and Fong K. (1990). A specificity enhancer for polymerase chain reaction. *Nucleic Acids Res* **18**: 4953.

Huntley S, Hamann N, Wegener-Feldbrügge S, Treuner-Lange A, Kube M, Reinhardt R, *et al.* (2011). Comparative genomic analysis of fruiting body formation in Myxococcales. *Mol. Biol. Evol.* **28**: 1083–1097.

Hutchinson CR. (1999). Microbial polyketide synthases: more and more prolific. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 3336–3338.

Iizuka T, Jojima Y, Fudou R, and Yamanaka S. (1998). Isolation of myxobacteria from the marine environment. *FEMS Microbiology Letters* **169**: 317–322.

Ingram-Smith C, Woods BI, and Smith KS. (2006). Characterization of the acyl substrate binding pocket of acetyl-CoA synthetase. *Biochemistry* **45**: 11482–11490.

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

Jurica Žučko. (2010). In silico analysis of polyketide synthases. Available from: http://bioinformatics.pbf.hr/dimplom_work/Jurica_Zucko_PhD_final.pdf

Kalaitzis JA, Cheng Q, Thomas PM, Kelleher NL, and Moore BS. (2009). In Vitro Biosynthesis of Unnatural Enterocin and Wailupemycin Polyketides. *J Nat Prod* **72**: 469–472.

Keatinge-Clay AT. (2007). A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chemistry & Biology* **14**: 898–908.

Khalil MW, Sasse F, Lünsdorf H, Elnakady YA, and Reichenbach H. (2006). Mechanism of action of tubulysin, an antimitotic peptide from myxobacteria. *Chembiochem* **7**: 678–683.

Kittendorf JD, and Sherman DH. (2009). The Methymycin/Pikromycin Biosynthetic Pathway: A Model for Metabolic Diversity in Natural Product Biosynthesis. *Bioorg Med Chem* **17**: 2137–2146.

Konz D, and Marahiel MA. (1999). How do peptide synthetases generate structural diversity? *Chem. Biol.* **6**: R39–48.

Kwan DH, and Schulz F. (2011). The stereochemistry of complex polyketide biosynthesis by modular polyketide synthases. *Molecules* **16**: 6092–6115.

Kwan DH, Sun Y, Schulz F, Hong H, Popovic B, Sim-Stark JCC, *et al.* (2008). Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chemistry & Biology* **15**: 1231–1240.

Li C, Roege KE, and Kelly WL. (2009). Analysis of the Indanomycin Biosynthetic Gene Cluster from *Streptomyces antibioticus* NRRL 8167. *ChemBioChem* **10**: 1064–1072.

Linne U, and Marahiel MA. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39**: 10439–10447.

Liu H, and Reynolds KA. (2001). Precursor supply for polyketide biosynthesis: the role of crotonyl-CoA reductase. *Metab. Eng.* **3**: 40–48.

Magarvey NA, Beck ZQ, Golakoti T, Ding Y, Huber U, Hemscheidt TK, *et al.* (2006). Biosynthetic Characterization and Chemoenzymatic Assembly of the Cryptophycins. Potent Anticancer Agents from *Nostoc* Cyanobionts. *ACS Chem. Biol.* **1**: 766–779.

Marshall CG, Hillson NJ, and Walsh CT. (2002). Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. *Biochemistry* **41**: 244–250.

Menche D, Arikan F, Perlova O, Horstmann N, Ahlbrecht W, Wenzel SC, *et al.* (2008). Stereochemical Determination and Complex Biosynthetic Assembly of Etnangien, a Highly Potent RNA Polymerase Inhibitor from the Myxobacterium *Sorangium cellulosum*. *J. Am. Chem. Soc.* **130**: 14234–14243.

Metzker ML. (2009). Sequencing technologies-the next generation. *Nature Reviews Genetics* **11**: 31–46.

Minowa Y, Araki M, and Kanehisa M. (2007). Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *Journal of Molecular Biology* **368**: 1500–1517.

Moffet DA, Khosla C, and Cane DE. (2006). Modular polyketide synthases: Investigating intermodular communication using 6 deoxyerythronolide B synthase module 2. *Bioorg. Med. Chem. Lett.* **16**: 213–216.

Moffitt MC, and Neilan BA. (2004). Characterization of the Nodularin Synthetase Gene Cluster and Proposed Theory of the Evolution of Cyanobacterial Hepatotoxins. *Appl Environ Microbiol* **70**: 6353–6362.

Moore BS, and Hertweck C. (2002). Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat Prod Rep* **19**: 70–99.

Mootz HD, Schwarzer D, and Marahiel MA. (2002). Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. *ChemBioChem* **3**: 490–504.

Müller R. (2004). Don't classify polyketide synthases. *Chem. Biol.* **11**: 4–6.

Musso M, Bocciardi R, Parodi S, Ravazzolo R, and Ceccherini I. (2006). Betaine, Dimethyl Sulfoxide, and 7-Deaza-dGTP, a Powerful Mixture for Amplification of GC-Rich DNA Sequences. *J Mol Diagn* **8**: 544–550.

Neumann B, Pospiech A, and Schairer HU. (1992). Rapid isolation of genomic DNA from gram-negative bacteria. *Trends Genet.* **8**: 332–333.

Newman DJ, Cragg GM, and Snader KM. (2003). Natural products as sources of new drugs over the period 1981-2002. *J. Nat. Prod.* **66**: 1022–1037.

O'Connor SE, Walsh CT, and Liu F. (2003). Biosynthesis of Epothilone Intermediates with Alternate Starter Units: Engineering Polyketide–Nonribosomal Interfaces. *Angewandte Chemie International Edition* **42**: 3917–3921.

Ohlendorf B, Leyers S, Krick A, Kehraus S, Wiese M, and König GM. (2008). Phenylnannolones A–C: Biosynthesis of New Secondary Metabolites from the Myxobacterium *Nannocystis exedens*. *ChemBioChem* **9**: 2997–3003.

Ohlendorf B. (2008). Chemistry, biosynthesis and bioactivity of secondary metabolites from *Nannocystis* and *Myxococcus species*.

Palaniappan N, Kim BS, Sekiyama Y, Osada H, and Reynolds KA. (2003). Enhancement and selective production of phoslactomycin B, a protein phosphatase IIa inhibitor, through identification and engineering of the corresponding biosynthetic gene cluster. *J. Biol. Chem.* **278**: 35552–35557.

Perlova O, Gerth K, Kaiser O, Hans A, and Müller R. (2006). Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *Journal of Biotechnology* **121**: 174–191.

Phelan VV, Du Y, McLean JA, and Bachmann BO. (2009). Adenylation Enzyme Characterization Using γ -18O4-ATP Pyrophosphate Exchange. *Chemistry & Biology* **16**: 473–478.

Piel J, Hertweck C, Shipley PR, Hunt DM, Newman MS, and Moore BS. (2000). Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate *'Streptomyces maritimus'*: evidence for the derailment of an aromatic polyketide synthase. *Chem. Biol.* **7**: 943–955.

Piel J, Wen G, Platzer M, and Hui D. (2004). Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *Chembiochem* **5**: 93–98.

Piel J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus beetles*. *PNAS* **99**: 14002–14007.

Pistorius D, and Müller R. (2012). Discovery of the Rhizopodin Biosynthetic Gene Cluster in *Stigmatella aurantiaca* Sg a15 by Genome Mining. *ChemBioChem* **13**: 416–426.

Quade N, Huo L, Rachid S, Heinz DW, and Müller R. (2012). Unusual carbon fixation gives rise to diverse polyketide extender units. *Nature Chemical Biology* **8**: 117–124.

Rachid S, Huo L, Herrmann J, Stadler M, Köpcke B, Bitzer J, *et al.* (2011). Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *Chembiochem* **12**: 922–931.

Rangaswamy V, Mitchell R, Ullrich M, and Bender C. (1998). Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. *J. Bacteriol.* **180**: 3330–3338.

Rausch C, Hoof I, Weber T, Wohlleben W, and Huson DH. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**: 78.

Reboll MR, Ritter B, Sasse F, Niggemann J, Frank R, and Nourbakhsh M. (2012). The Myxobacterial Compounds Spirangien A and Spirangien M522 Are Potent Inhibitors of IL-8 Expression. *ChemBioChem* **13**: 409–415.

Reichenbach H, and Höfle G. (2008). Discovery and development of the epothilones : a novel class of antineoplastic drugs. *Drugs R D* **9**: 1–10.

Reichenbach H. (1999). The ecology of the myxobacteria. *Environmental Microbiology* **1**: 15–21.

Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, *et al.* (2003). A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases. *Biochemistry* **42**: 72–79.

Reimmann C, Patel HM, Walsh CT, and Haas D. (2004). PchC Thioesterase Optimizes Nonribosomal Biosynthesis of the Peptide Siderophore Pyochelin in *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**: 6367–6373.

Richter CD, Nietlispach D, Broadhurst RW, and Weissman KJ. (2008). Multienzyme docking in hybrid megasynthetases. *Nature Chemical Biology* **4**: 75–81.

Rinehart K, Namikoshi M, and Choi B. (1994). Structure and biosynthesis of toxins from blue-green algae (cyanobacteria). *Journal of Applied Phycology* **6**: 159–176.

Ruch FE, and Vagelos PR. (1973). The isolation and general properties of *Escherichia coli* malonyl coenzyme A-acyl carrier protein transacylase. *J. Biol. Chem.* **248**: 8086–8094.

Sambrook J, and Russell DW. (2000). Molecular Cloning: A Laboratory Manual, 3 Vol. 0003 ed. Cold Spring Harbor Laboratory.

Sanger F, Nicklen S, and Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**: 5463–5467.

Schäberle TF, Goralski E, Neu E, Erol Ö, Hölzl G, Dörmann P, *et al.* (2010). Marine Myxobacteria as a Source of Antibiotics—Comparison of Physiology, Polyketide-Type Genes and Antibiotic Production of Three New Isolates of *Enhygromyxa salina*. *Marine Drugs* **8**: 2466–2479.

Schiefer A, Schmitz A, Schäberle TF, Specht S, Lämmer C, Johnston KL, *et al.* (2012). Corallopyronin A specifically targets and depletes essential obligate *Wolbachia endobacteria* from filarial nematodes in vivo. *J. Infect. Dis.* **206**: 249–257.

Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, *et al.* (2007). Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **25**: 1281–1289.

Shen B, Du L, Sanchez C, Edwards DJ, Chen M, and Murrell JM. (2001). The biosynthetic gene cluster for the anticancer drug bleomycin from *Streptomyces verticillus* ATCC15003 as a model for hybrid peptide-polyketide natural product biosynthesis. *J. Ind. Microbiol. Biotechnol.* **27**: 378–385.

Shimkets, L., Ferriera, S., Johnson, J., Kravitz, S., Beeson, K., Sutton, G., *et al.* (2010). *Plesiocystis pacifica* SIR-1, whole genome shotgun sequencing project - Nucleotide - NCBI. *direct submission*-Available from: http://www.ncbi.nlm.nih.gov/nuccore/149925296

Sieber SA, and Marahiel MA. (2005). Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chem. Rev.* **105**: 715–738.

Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, *et al.* (1999). New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274**: 37391–37399.

Smith S, and Tsai S-C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports* **24**: 1041.

Stachelhaus T, Mootz HD, and Marahiel MA. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology* **6**: 493–505.

Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, and Hranueli D. (2008). ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.* **36**: 6882–6892.

Staunton J, and Weissman KJ. (2001). Polyketide biosynthesis: a millennium review. *Natural Product Reports* **18**: 380–416.

Staunton J, and Wilkinson B. (1997). Biosynthesis of Erythromycin and Rapamycin. *Chem. Rev.* **97**: 2611–2630.

Steinmetz H, Irschik H, Kunze B, Reichenbach H, Höfle G, and Jansen R. (2007). Thuggacins, macrolide antibiotics active against Mycobacterium tuberculosis: isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chemistry* **13**: 5822–5832.

Strieker M, Tanović A, and Marahiel MA. (2010). Nonribosomal peptide synthetases: structures and dynamics. *Current Opinion in Structural Biology* **20**: 234–240.

Suo Z, Tseng CC, and Walsh CT. (2001). 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. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 99–104.

Tang L, Ward S, Chung L, Carney JR, Li Y, Reid R, *et al.* (2003). Elucidating the Mechanism of cis Double Bond Formation in Epothilone Biosynthesis. *J. Am. Chem. Soc.* **126**: 46–47.

Tsai S-C (Sheryl), and Ames BD. (2009). STRUCTURAL ENZYMOLOGY OF POLYKETIDE SYNTHASES. *Methods Enzymol* **459**: 17–47.

Vrijbloed JW, Zerbe-Burkhardt K, Ratnatilleke A, Grubelnik-Leiser A, and Robinson JA. (1999). Insertional Inactivation of Methylmalonyl Coenzyme A (CoA) Mutase and Isobutyryl-CoA Mutase Genes in Streptomyces cinnamonensis: Influence on Polyketide Antibiotic Biosynthesis. *J Bacteriol* **181**: 5600–5605.

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, *et al.* (2009). CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *J. Biotechnol.* **140**: 13–17.

Weissman KJ, and Müller R. (2009). A brief tour of myxobacterial secondary metabolism. *Bioorganic & Medicinal Chemistry* **17**: 2121–2136.

Weissman KJ, and Müller R. (2010). Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat Prod Rep* **27**: 1276–1295.

Weissman KJ. (2009). Chapter 1 Introduction to Polyketide Biosynthesis. In: Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates. Vol. Volume 459. Academic Press. p 3–16. Available from: http://www.sciencedirect.com/science/article/pii/S0076687909046011

Wenzel SC, and Müller R. (2009). The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Natural Product Reports* **26**: 1385.

Wilkinson CJ, Frost EJ, Staunton J, and Leadlay PF. (2001). Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. *Chemistry & Biology* **8**: 1197–1208.

Wilson MC, and Moore BS. (2012). Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat Prod Rep* **29**: 72–86.

Xiao Y, Li S, Niu S, Ma L, Zhang G, Zhang H, *et al.* (2011). Characterization of tiacumicin B biosynthetic gene cluster affording diversified tiacumicin analogues and revealing a tailoring dihalogenase. *J. Am. Chem. Soc.* **133**: 1092–1105.

Yadav G, Gokhale RS, and Mohanty D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**: 335–363.

Zhang Y-M, Hurlbert J, White SW, and Rock CO. (2006). Roles of the Active Site Water, Histidine 303, and Phenylalanine 396 in the Catalytic Mechanism of the Elongation Condensing Enzyme of *Streptococcus pneumoniae*. *J. Biol. Chem.* **281**: 17390–17399.

Zheng J, and Keatinge-Clay AT. (2011). Structural and Functional Analysis of C2-Type Ketoreductases from Modular Polyketide Synthases. *Journal of Molecular Biology* **410**: 105–117.

Zheng J, Taylor CA, Piasecki SK, and Keatinge-Clay AT. (2010). Structural and Functional Analysis of A-Type Ketoreductases from the Amphotericin Modular Polyketide Synthase. *Structure* **18**: 913–922.

8 Appendix

8.1 Amino acid sequences for the phenylnannolone A gene cluster

Sequences of genes *phn1* and *phn2* were obtained from 454 fosmid sequencing of the fosmid clone 12A9 and translated into protein sequences.

8.1.1 Amino acid sequence of Phn1, the putative butyryl-CoA Carboxylase

GARMSSELRNAAIKEISESQPVRELKRVPEHPHRVHDRRERLAHACAGPAQVAAG
DRGVAAAEARARDAEAKLHAGGRRTARERVEELLDPDSFVEMDEFVTHQCPDFDM
PDRKAYGDGVLTGHGTVDGRPVFVYAQQPQVLGGSLGLAHAMKICKVMDMAMTV
GAPVVGLNDSGGARIQEGVASLAGYAEIFKRNVLASGVVPQLSLILGPCAGGAVYSP
ALTDLVLMVEDHSYMFITGPEVIRAVTHEEVTKEQLGGAATHNERSGVAHFSCASET
VAMVMARELLSFMPSNNMEDAPRRATLDDPRRPTPELAEFVPVDTTKPYDMRTIVK
AIADDGYFFEVQESYARNIVIGFIRVDGRSVGVVANQPLVLAGCLNIDASVKAARFVR
MCDAFNIPLLTLVDVPGLLPGVEQEYAGIIRHGAKLLYAYVEATVPKITLVTRKAYGG
AYAVMSSKHIRGDVNLAYPTAEIAVMGAEGAVNIIYRREIAEAADPAARRAELLAEYR
DLFATPYKAAELGFIDKIIRPEETdRREIARSLALMGNKRQENPRRKHGNIPL

8.1.2 Amino acid sequence of the Phn2, the putative PKS assembly line

ASIRPRVPWLAPPDVNSETRSMQSQSHGRFERASRNLLDLLDEQATAFAGKIAYAF
LERGEHVTAQVSFAALRERAYVIAAELAKHCRPGDRAVLAFPPGIEFVEAFWGCLC
AGVIAVPSYLPEPGNQQRHLHRLYGTVDDARPAVILTSAGAHDRLQRLLLAAPNPVA
APCLATDAIREDPSAERPSLPASVSSETLATLIYTSGSTGDPKGVCVTHGNALHNIEL
VRDKFDNDEDATYVSWLPLFHDLGLIVVMLGAFQCGATCYLMDAVDFIRDPICWLR
AISRYRGRNAGGPNFAFELCARKVTPEQLGELDLRCWDVAFNGAEPVRADTMERF
AQVFAPCGFRSTAFYPCYGMAEATGMISGGASRAAPVVELFDEQSLGRGVVKVAD
KGGASADDDTVASRSHGRSLVGNGTAGGGQEIAIVDPDTYERCPPLTVGEVWLRG
PSIGPGYWQRPEATRETYGETIAHEDDGTYLRTGDLGFVRDGEVFLTGRLKDLIIIRG
ANHYPQDIELSVERSHAALRGGGCVVFPVETDTEERVGIIAETLSDRDPGEQSAEIL
RSIRAAVTREHEVGVAVIALVPPRTVLKTTSGKPRRRANRSTLFAGQFPVWAEWRD
AQRPAGDTAAPPDAPRPSPIPKHKLARRRDELFQLVHGVVAARLNAPEILPDRPLQE
LGFDSLAAVEIQAQLAKQLGVELPATLLYDHPTLRALTSFLLQTVLLAPHESAEEASE

DGRGLSTLADPDEPIAVIAMACRLPGGIKEPDAYWRVLAQALDVIAPFPADRWDAEA
LYDPDPDHKGKTYCTQGGFLADLEQFDAAFFGISPREAQAMDPQQRLVLETAWEA
LERAGIVPETLYGSQTGVYIGAMDSDYHLWGEDARALGLLDGYVSTGNTGSVLSGR
VSYVLGLQGPALTVDTACSSSLVALHLAAQGLRNGECDLALCGGVQVMNTPAAFVE
LSRLRALAPDGRSKSFSAQADGAGWSEGCGMVLLKRLSDAQRQGDRVLAVIRGTA
VNQDGRSQGLTAPNGPSQERVIRRALERSQLAPRDIDVVEAHGTGTALGDPIEAGA
LSAVFGSDRPAEQPLYLGSVKSNLGHTQAAAGVTAIIKMVLALQHEVLPRTLHAEEP
SPHIRWEGSPLALLQAAHSWPRGERPRRAGVSSFGISGTNAHVILEEAPAEGPHIEP
DVALAASPDAVPLVLSARSPEALRAQASQLGAHLSAHPEQSLVDVAHSLVTTRSLF
GHRAVVLATSREAATEKLAALALGATDSRTIAGVAAQTRGKLAFVFPGQGGQWPH
MARALLRGSQLFREQIEQCNEALSPYCDWSLIDMLDGTLGAEAWERIEVVQPALFA
MWVSLAALWRDMGVEPDGVIGHSQGEIAAACVAGALPLAEAARISALRSRTFKKLE
GRGAMAAVEMGEEELEARLRALGTGVEVAVNNGTNSCVVSGTPAAIEAFVAQLTR
DGVFARRLQTTCATHWAQAEEIRDELLTSLGAVRRADGSIPIYSTVLAAPIGGGELD
AQYWYANTRERVRMRQTVETMLADGFRYFVEVSSHPVLTMAVERSLAGAKLAGAA
VGSLRRDEGQPEKLLEGLAELHVAGLEVDWTRLLPPARVVELPTYPFERHRFWKKE
WREREPDVGTAGMTQVDHPLLVGATGLADAEGVLFTTELSQHTVPWLKDHRVLDL
LLVPGTAILEMMMVAGQLLGTPEVRDLTLMSGLVLPERGSLHVQFRASAGEPDGGR
RIALYSRPAEAPLELSWTCHATGHLGPASAARGASLDAWPAPGAEAVATDNLYDQL
AQQGLVYGPTFQGLKQLAVRDQAWFAEIKPRLTTHEASRYRIHPGLLDAALHGLSL
ALPSAGADALWLPFGFEDVAVWNTGAAELRVRVRLHQQPGDGDEVAQAELDLFD
GQGEAVGRVGTLRLRRASRAQVRASRTGEPPLHRLAWTARALEVTQNIATSEVVVL
GAGRLAARLGVRAVSEVAELLAGPAPRIVVIDASEAIERTGSNEHALKGMSGATYHS
PDELRAQLGAALEPLQQLLAQPSLEAAQLVWVTEQAVSTGPADPVASLAQAGLWG
MIRSARNETSRPLRLMDVEAGSSSELLWNALSQPEAPELAVRGSAVLRPRLEPCAA
DEATSPALRKDGTVLLTGGTGELGQTVARHLVREHGIRHLVLTSRRGRDAAGVEAL
LAELADLGAQVEVVRCDVSQREAVADLLATLDPPLSAVFHCAAVLDDGLVPALTPP
RMDVVLRAKVDSALHLDALTAELPLDAFVMFSSVAGVIGSAGQANYAAANTVLDAL
ASQRRARGQAGQSLAWGLWHPQGLGMTARLSPAQLQRLKRQGLNALRVATGMR
LLDAALRRPDALLVPIERATGELDESAPFLIRDGQPRKARAADAKVATGLWAALRGL
DAPARQQKLLEWVQTETAGVLGLQGRQAVAADRPLKELGLDSLMAVELRNRLAAR
AELSLPSTLAFDYPTARQIAEQIAARRAPETAAAPVATSRPATDVDAPIAIVAMSCRL
PGGVDTLEAFWELLEHGRDAIEPCPRGRWDLEALYDPDPEQRGTTYCTQGGFLAD
VAGFDASFFGISPREAQAMDPQQRLVLETAWEVLESAGIVPAQLQRSQTGVYIGAM

ASDYHVFGRRPDDLEALDGYRVTGNAGSVISGRIAYVLGLQGPALTVDTACSSSLV
ALHLAAQGLRNGECDLALCGGVQVMNAPTTFIEFSRLRGMAPDGRCKSFSADADG
AGWSEGCGMLLLKRLPDAERDGDRILAVLRGSAVNQDGRSQGLTAPNGPSQERVI
RRALELSHLEPQAIDAVEAHGTGTVLGDPIEAGALSAVFGARDEGPLYLGSVKSNLG
HTQAAAGVTGVIKMVLALQHELLPRTLHAQQATPHIAWEQSGLQLLQAPLPWPRGS
RVRRAGISSFGISGTNAHLILEEAPAPVEPARATPTDDKTAWPLLLSARSATTLAAQA
ARLQTHLRQHPEQSLVDVAFSLATARTQFEHRAVVLAASSEAAIEQLARLQQGRFD
RTILGPPRSDALEPSGKLALLFSGQGAQRPAMGRALRRYPVFARALEAVFAELADR
LPSLPAVMLADPDTPEAALLEQTLYTQTGLFALQVALYRLWESMGLVPDVLLGHSV
GEIAVAHVAGILSLADACTLVAERARLMQALPSGGAMVSVVASERAVLDVLADCPEP
VDIAGVNGPLSTVISGSEAAVLHAVHALELHGYKTTRLRVSHAFHSARMEPMLADFA
RVARGLTYHSPQRAIVSTRTGALVDPDELCDPEYWVRQIREPVRFADGVRALADLG
VSTYLELGPHGVLCGLASACLDAPRPSVFVPSLHADEPADHGFLAAAAALHTHGLD
LRWTALLPVDARRVELPTYPFERQSYWLDEQRTAELSAASLGMTAIEHPLLFGAIRP
ADTESVLFTSRLSQRAIPWFTEHRVFETTLVPGTALLELAMAAGQALDMPALNDLAL
VSALAWPETDAIHVQVQVSAEGPDGDRQVDLYSRPEDAAMASAWTRHATGRLGP
SSTSRPPQLGPWPPPDSDPVDLAGHYERTARLGLDYGPSFRGLRQLFRAEREPTL
WYAEASLLPEQVAEASRYRIHPALLDAALHGLSLALPAAGGDALWLPFGFEDVAVW
HTGAAELRVRVRLHQQPGDGDEVAQAELDLYDGQGEAIGRIGGLRLRRASRAQVR
ASRTGEPPLHRLAWSPSRALDITPSPSEVVVLGAGRLAARLGLRGVSEVAELLAGP
APRTVVIDATELARRADELRAQSSHTIEWTGSNGHGSMGMSDATGQDPDELRAQL
SAALMPLQQLLAQPSLEAAQLVWVTEQAVSTGPADPVVALAQAGLWGLIRSARNET
SRPLRLVDVEADSSSDLLWNALSQAEAPELAVRGAAVLRPRLEPCAADTETTPALR
TDGTVLLTGGTGELGQAVARHLVREHGIRHLVLTSRRGRDAAGVEALLAELANLGA
QVEVVRCDVSQREAVADLLAAIDPPLSAVFHCAAVLDDGLVPTLTPARMDVVLRAK
VDSALHLDALTAELPLDAFVMFSSVAGVIGSAGQANYAAANTVLDALASQRRARGQ
AGQSLAWGLWHPQGLGMTARLSSAQLQRLKRQGLSALRVPTGMRLLDAALRRPD
ALLVPIERATGELDESAPFLIRDGQPRKARPVDANVSTGLWAALRGLDASARQQKLL
EWVQTETAGVLGLQGRQAVAADRPLKELGLDSLMAVDLRNRLATRAELSLASTLAF
DYPTARQIAEHIAARRAPEAVAAAAPVTTSRPTDVDAPIAIVAMACRLPGGVDTLESF
WELLEHGRDVIEPCARWDLEALYDPDPEHRGKTYCTQGGFLADVAGFDASFFGISP
REAQAMDPQQRLVLETAWEVLERAGIVPAQLQRSQTGVYIGAMGSDYHVFGRGPD
DLEALDGYCVTGNAGSVISGRIAYVLGLQGPALTVDTACSSSLVALHLAAQGLRNGE
CDLALCGGVQVMNGPTTFIEFSRLRGVAPDGRCKSFSADADGAGWSEGCGMLLLK

RLPDAERDGDRILAVLRGSAVNQDGRSQGLTAPNGPSQERVIRRALEQSRLGPEAI
DAVEAHGTGTVLGDPIEAGALSAVFGARDDRPLYLGSVKSNLGHTQAAAGVTGVIK
MVLAMQHELLPRTLHAQQASPHIAWASSGLQLLQAPLPWPRGSRVRRAGISSFGIS
GTNAHLILEEAPAPVEPARTTITNAHLILEEAPAPVEPTQWTANDKTVWPLLISARSA
TTLVAQATRLQAHLRQHPEQSLRDVAFSLATARTQFEHRAVVLADSSQAAIERLAKL
QQDPRSDVLEPSGKLALLFSGQGAQRPAMGRALRRYPVFARALEAVFAELADRLP
SLPAVMLADADAPEAALLEQTLYTQTGLFALQVALFRLWDSMGLVPDVLLGHSVGEI
AVAHVAGILSLADACTLVAERARLMQALPSGGAMVSVVASERAVLDVLASCPEPVDI
AGVNGPLSTVISGSEAAVLHAVHALELQGYKTTRLRVSHAFHSARMEPMLADFARV
ARGLTYHSPQRAIVSTRTGALVDPDELCDPDYWVRQVREPVRFADGVRALADLGV
STFLELGPHGVLCGLASACLDAPRPSIFVPSLRADEPADHSFLTAAAALHTHGLDLR
WTGLLPVDARRVELPTYPFERQSYWLEPPSSATVSRAVDSQRATTTSRGPDSRGT
SMFLKHLDGFAREDRLRHLLQLVLGESAATLGHPDASQVDPDRGFADLGMNSLMA
VQLRERLQARTGLHIPATVAFDHPSPARVASFLLERSEATLHERVEAAPDVHPVNSS
ADDEPIAIVGVGLRLPGGCVDLTSLWRLLEHGVDAVTTIPADRWALDDFYDPVAGEP
GKSYVREAGFLSGIDEFDPGFFHISPREAKSIDPQHRMLLQVAWQTLEMAGVVPGS
LRDSQTGVFVACGASEYGLHMGAGEDPYEFTRALQSFSAGRLAYTLGLQGPALSV
ETACSSSLVALHLACQSLRSGECSLALAAGAQLMISPEPFLMLSRIRALAPDGRCKT
FSEQADGYGRGEGVVVLALERLSAARANGRRILALVRGSAINHDGASSGLTVPNGL
SQQKVLRAALRAAGVDAATVDYVECHGTGTHLGDPIEVQALGDVYGAARPLEQRV
KLGTIKPNIGHLEFAAGLAGVAKVIAAFQQQALPATIHTLPRNGHIEWDRLAVSVVDS
LTPWPSRPQEQPRRAGVSSFGLSGMNAHVILEEAPPAISPPAAATASEPAAVPLLLS
ARSEAALAMQASRLAEHLGQQPGQRLLDVAWSLATTRTHFEHRAVVLADSHESAR
ARLHELTQGARISSRVSRTAASGSVAFLFTGQGSQRCGMGQELARRFPVFRAALD
EVLAHLAPQLPTLRDIMFAPADSAEGRSLHQTGHTQPALFALQIALFRLWEDLGVRP
GWLLGHSVGEIAAAHAAGVLSLADACKLVAARGRLMQGLTARGVMYSVLAAEADV
REQLGERAGLVDIAGVNGPRSTVIAGDEHAVDELVRALAERGIKSSRLQVSHAFHSP
LMEPMLAELREVAQGLDYRAPALRLVSTRTGRAAAPEDFVTAEYWVRQAREAVQF
AAAMRTLADNNVDTYIELGPDGVLSRMGVKCLTEADQRRSRFLPSLRAGEAEAETL
LTAAGKLHTEGHEVNWASLFAGARSVELPTYAFERQRYWLDPRPSATAAAPRTLH
PFLSSRIDAADGNVTLFSGHLDLESTPWLRDHRVAGSVILPTTAYMDLMLSVGAKSL
GPVPLELEGLIVERPLLLAAEQSAEIQVDVQSTPGPDGGLTASVHSRRPDARAWQF
HARARVHHRPWQQPEPARKQRLDAIRQRAERVMEGSEFYAAWRARGNDWGPAF
QTIERLWIDGRTCVAQLRVRPTQSPRADGADFFAQPALLDGCGQALGVFALDRPAS

SFVGHRARVVRFKGALRGETFFCLIELDELDAAAPTLRGDIAIHDEHGELLGEALGAE
IELLEARRSDAAEPRRWLYEQQWVLAERAAERRQQVGHWLVLGSPVIAPLRDAMR
ERGVTTTFADSIRDPRALEALAHADLLGVSYLAPALEYALSPPEQAADAIDRVSRALE
ESLEPLRHLTASRPGFPVQLSWLTQGCWLVRPDDVPGAEGQAALWGLGKSCTAE
HPEYGGRLIDLPAGALLGEPSPTVIDHLVSSLLDEVAEKQLAIREDGIFLNRLTRHTPA
IGGEALRFRPDATYLVTGGLGGIGLALAEWLAERGACRLILVGRTPLPARRRWSALD
PRSKEGIAVDVIRRLERRGVSVQVAALDVANEAEVQRFLADHRDEAWPAIRGVVHA
AGVATGGPLRDMTATAFAQQLRPKLQAAWILQRALAEESLDFFVLCSSAAALIPSPL
VGAYAAANACLDALVERRRRAGQSGLSIAFGPWAEVGMAAEYQQQQARASTSPST
MTPMKTADALHLIGQLWQGTCGRPAILPIDWSLWQTRYRRLLEQPLFARFAAPAAA
RADDADLDPAADLASLPVAVQGQGLLAGLSRLIQQVLGHHGELDVEKSFSELGVDS
LMAFEAKEKIARVWSVDLPVLELLHAANIAELAQLVERRLAGSAPVDVRWPRLPTPC
EFPSLDGLTIHGHLSLPTGPGPHPAIVVHTADTGGALDDQGRYVHLFEHEPLVRAGF
AVLTVDQRGAPGHGDAYAGAADLGGADVEDLLAAARYLADRPDIDDARIGLVGTSR
GAYAGLLALQRAPDRFRAAVLRMGFYDPLEYGRNERQLRPETAPLLKLLPSWDAW
FERMGAPERNPLNRLDQVTASLFVIHGEDDRIADVEHSRRLVSAMTAAGRPTELRTI
PGVGHDIEEVHPVWREIWDEITSFLHNHLSPASETEAAE

Trimmed sequences of the identified domains on Phn2 gene

AMP-ligase

FAALRERAYVIAAELAKHCRPGDRAVLAFPPGIEFVEAFWGCLCAGVIAVPSYLPEP
GNQQRHLHRLYGTVDDARPAVILTSAGAHDRLQRLLLAAPNPVAAPCLATDAIREDP
SAERPSLPASVSSETLATLIYTSGSTGDPKGVCVTHGNALHNIELVRDKFDNDEDAT
YVSWLPLFHDLGLIVVMLGAFQCGATCYLMDAVDFIRDPICWLRAISRYRGRNAGG
PNFAFELCARKVTPEQLGELDLRCWDVAFNGAEPVRADTMERFAQVFAPCGFRST
AFYPCYGMAEATGMISGGASRAAPVVELFDEQSLGRGVVKVADKGGASADDDTVA
SRSHGRSLVGNGTAGGGQEIAIVDPDTYERCPPLTVGEVWLRGPSIGPGYWQRPE
ATRETYGETIAHEDDGTYLRTGDLGFVRDGEVFLTGRLKDLIIIRGANHYPQDIELSV
ERSHAALRGGG

ACP_{LM}

QLVHGVVAARLNAPEILPDRPLQELGFDSLAAVEIQAQLAKQLGVELPATLLYDHPTL
RALTSFLL

KS_1

EPIAVIAMACRLPGGIKEPDAYWRVLAQALDVIAPFPADRWDAEALYDPDPDHKGKT
YCTQGGFLADLEQFDAAFFGISPREAQAMDPQQRLVLETAWEALERAGIVPETLYG
SQTGVYIGAMDSDYHLWGEDARALGLLDGYVSTGNTGSVLSGRVSYVLGLQGPAL
TVDTACSSSLVALHLAAQGLRNGECDLALCGGVQVMNTPAAFVELSRLRALAPDGR
SKSFSAQADGAGWSEGCGMVLLKRLSDAQRQGDRVLAVIRGTAVNQDGRSQGLT
APNGPSQERVIRRALERSQLAPRDIDVVEAHGTGTALGDPIEAGALSAVFGSDRPAE
QPLYLGSVKSNLGHTQAAAGVTAIIKMVLALQHEVLPRTLHAEEPSPHIRWEGSPLA
LLQAAHSWPRGERPRRAGVSSFGISGTNAHVILEE

AT_1

FPGQGGQWPHMARALLRGSQLFREQIEQCNEALSPYCDWSLIDMLDGTLGAEAW
ERIEVVQPALFAMWVSLAALWRDMGVEPDGVIGHSQGEIAAACVAGALPLAEAARI
SALRSRTFKKLEGRGAMAAVEMGEEELEARLRALGTGVEVAVNNGTNSCVVSGTP
AAIEAFVAQLTRDGVFARRLQTTCATHWAQAEEIRDELLTSLGAVRRADGSIPIYSTV
LAAPIGGGELDAQYWYANTRERVRMRQTVETMLADGFRYFVEVSSHPVLTMAVER
SLAGAKLAGAAVGSLRRD

DH_1

AGMTQVDHPLLVGATGLADAEGVLFTTELSQHTVPWLKDHRVLDLLLVPGTAILEM
MMVAGQLLGTPEVRDLTLMSGLVLPERGSLHVQFRASAGEPDGGRRIALYSRPAE
APLELSWTCHATGHLGPASAARGASLDAWPAPGAEAVATDNLYDQLAQQGLVYGP
TFQG

KR_1

ATSPALRKDGTVLLTGGTGELGQTVARHLVREHGIRHLVLTSRRGRDAAGVEALLA
ELADLGAQVEVVRCDVSQREAVADLLATLDPPLSAVFHCAAVLDDGLVPALTPPRM
DVVLRAKVDSALHLDALTAELPLDAFVMFSSVAGVIGSAGQANYAAANTVLDALASQ
RRARGQAGQSLAWGLWHPQ

ACP_1

LRGLDAPARQQKLLEWVQTETAGVLGLQGRQAVAADRPLKELGLDSLMAVELRNR
LAARAELSLPSTLAFDYPTARQIAEQIAAR

KS_2

APIAIVAMSCRLPGGVDTLEAFWELLEHGRDAIEPCPRGRWDLEALYDPDPEQRGT
TYCTQGGFLADVAGFDASFFGISPREAQAMDPQQRLVLETAWEVLESAGIVPAQLQ
RSQTGVYIGAMASDYHVFGRRPDDLEALDGYRVTGNAGSVISGRIAYVLGLQGPAL
TVDTACSSSLVALHLAAQGLRNGECDLALCGGVQVMNAPTTFIEFSRLRGMAPDGR
CKSFSADADGAGWSEGCGMLLLKRLPDAERDGDRILAVLRGSAVNQDGRSQGLTA
PNGPSQERVIRRALELSHLEPQAIDAVEAHGTGTVLGDPIEAGALSAVFGARDEGPL
YLGSVKSNLGHTQAAAGVTGVIKMVLALQHELLPRTLHAQQATPHIAWEQSGLQLL
QAPLPWPRGSRVRRAGISSFGISGTNAHLILE

AT_2

FSGQGAQRPAMGRALRRYPVFARALEAVFAELADRLPSLPAVMLADPDTPEAALLE
QTLYTQTGLFALQVALYRLWESMGLVPDVLLGHSVGEIAVAHVAGILSLADACTLVA
ERARLMQALPSGGAMVSVVASERAVLDVLADCPEPVDIAGVNGPLSTVISGSEAAV
LHAVHALELHGYKTTRLRVSHAFHSARMEPMLADFARVARGLTYHSPQRAIVSTRT
GALVDPDELCDPEYWVRQIREPVRFADGVRALADLGVSTYLELGPHGVLCGLASAC
LDAPRPSVFVPSLHAD

DH_2

LGMTAIEHPLLFGAIRPADTESVLFTSRLSQRAIPWFTEHRVFETTLVPGTALLELAM
AAGQALDMPALNDLALVSALAWPETDAIHVQVQVSAEGPDGDRQVDLYSRPEDAA
MASAWTRHATGRLGPSSTSRPPQLGPWPPPDSDPVDLAGHYERTARLGLDYGPS
FRG

KR_2

TETTPALRTDGTVLLTGGTGELGQAVARHLVREHGIRHLVLTSRRGRDAAGVEALL
AELANLGAQVEVVRCDVSQREAVADLLAAIDPPLSAVFHCAAVLDDGLVPTLTPAR
MDVVLRAKVDSALHLDALTAELPLDAFVMFSSVAGVIGSAGQANYAAANTVLDALAS
QRRARGQAGQSLAWGLWHPQ

ACP_2

LRGLDASARQQKLLEWVQTETAGVLGLQGRQAVAADRPLKELGLDSLMAVDLRNR
LATRAELSLASTLAFDYPTARQIAEHIAAR

KS_3

APIAIVAMACRLPGGVDTLESFWELLEHGRDVIEPCARWDLEALYDPDPEHRGKTY
CTQGGFLADVAGFDASFFGISPREAQAMDPQQRLVLETAWEVLERAGIVPAQLQRS
QTGVYIGAMGSDYHVFGRGPDDLEALDGYCVTGNAGSVISGRIAYVLGLQGPALTV
DTACSSSLVALHLAAQGLRNGECDLALCGGVQVMNGPTTFIEFSRLRGVAPDGRC
KSFSADADGAGWSEGCGMLLLKRLPDAERDGDRILAVLRGSAVNQDGRSQGLTAP
NGPSQERVIRRALEQSRLGPEAIDAVEAHGTGTVLGDPIEAGALSAVFGARDDRPLY
LGSVKSNLGHTQAAAGVTGVIKMVLAMQHELLPRTLHAQQASPHIAWASSGLQLLQ
APLPWPRGSRVRRAGISSFGISGTNAHLILE

AT_3

FSGQGAQRPAMGRALRRYPVFARALEAVFAELADRLPSLPAVMLADADAPEAALLE
QTLYTQTGLFALQVALFRLWDSMGLVPDVLLGHSVGEIAVAHVAGILSLADACTLVA
ERARLMQALPSGGAMVSVVASERAVLDVLASCPEPVDIAGVNGPLSTVISGSEAAV
LHAVHALELQGYKTTRLRVSHAFHSARMEPMLADFARVARGLTYHSPQRAIVSTRT
GALVDPDELCDPDYWVRQVREPVRFADGVRALADLGVSTFLELGPHGVLCGLASA
CLDAPRPSIFVPSLRAD

ACP_3

DGFAREDRLRHLLQLVLGESAATLGHPDASQVDPDRGFADLGMNSLMAVQLRERL
QARTGLHIPATVAFDHPSPARVASFLLER

KS_4

IAIVGVGLRLPGGCVDLTSLWRLLEHGVDAVTTIPADRWALDDFYDPVAGEPGKSYV
REAGFLSGIDEFDPGFFHISPREAKSIDPQHRMLLQVAWQTLEMAGVVPGSLRDSQ
TGVFVACGASEYGLHMGAGEDPYEFTRALQSFSAGRLAYTLGLQGPALSVETACS
SSLVALHLACQSLRSGECSLALAAGAQLMISPEPFLMLSRIRALAPDGRCKTFSEQA
DGYGRGEGVVVLALERLSAARANGRRILALVRGSAINHDGASSGLTVPNGLSQQKV
LRAALRAAGVDAATVDYVECHGTGTHLGDPIEVQALGDVYGAARPLEQRVKLGTIK
PNIGHLEFAAGLAGVAKVIAAFQQQALPATIHTLPRNGHIEWDRLAVSVVDSLTPWP
SRPQEQPRRAGVSSFGLSGMNAHVILEEAP

AT_4

FTGQGSQRCGMGQELARRFPVFRAALDEVLAHLAPQLPTLRDIMFAPADSAEGRSL
HQTGHTQPALFALQIALFRLWEDLGVRPGWLLGHSVGEIAAAHAAGVLSLADACKL
VAARGRLMQGLTARGVMYSVLAAEADVREQLGERAGLVDIAGVNGPRSTVIAGDE
HAVDELVRALAERGIKSSRLQVSHAFHSPLMEPMLAELREVAQGLDYRAPALRLVS
TRTGRAAAPEDFVTAEYWVRQAREAVQFAAAMRTLADNNVDTYIELGPDGVLSRM
GVKCLTEADQRRSRFLPSLRAG

DH_4

AAAPRTLHPFLSSRIDAADGNVTLFSGHLDLESTPWLRDHRVAGSVILPTTAYMDLM
LSVGAKSLGPVPLELEGLIVERPLLLAAEQSAEIQVDVQSTPGPDGGLTASVHSRRP
DARAWQFHARARVHHRPWQQPEPARKQRLDAIRQRAERVMEGSEFYAAWRARG
NDWGPAFQT

KR_4

PDATYLVTGGLGGIGLALAEWLAERGACRLILVGRTPLPARRRWSALDPRSKEGIAV
DVIRRLERRGVSVQVAALDVANEAEVQRFLADHRDEAWPAIRGVVHAAGVATGGP
LRDMTATAFAQQLRPKLQAAWILQRALAEESLDFFVLCSSAAALIPSPLVGAYAAAN
ACLDALVERRRRAGQSGLSIAFGPWA

ACP_4

LAGLSRLIQQVLGHHGELDVEKSFSELGVDSLMAFEAKEKIARVWSVDLPVLELLHA
ANIAELAQLV

TE

EFPSLDGLTIHGHLSLPTGPGPHPAIVVHTADTGGALDDQGRYVHLFEHEPLVRAGF
AVLTVDQRGAPGHGDAYAGAADLGGADVEDLLAAARYLADRPDIDDARIGLVGTSR
GAYAGLLALQRAPDRFRAAVLRMGFYDPLEYGRNERQLRPETAPLLKLLPSWDAW
FERMGAPERNPLNRLDQVTASLFVIHGEDDRIADVEHSRRLVSAMTAAGRPTELRTI
PGVGHDIEEVHPVWREIWDEITSFLHNHLSPASETEAAE

AMPPCP(protein)

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTASIRPRVPWLAPPDVNSETRSMQ
SQSHGRFERASRNLLDLLDEQATAFAGKIAYAFLERGEHVTAQVSFAALRERAYVIA
AELAKHCRPGDRAVLAFPPGIEFVEAFWGCLCAGVIAVPSYLPEPGNQQRHLHRLY
GTVDDARPAVILTSAGAHDRLQRLLLAAPNPVAAPCLATDAIREDPSAERPSLPASV
SSETLATLIYTSGSTGDPKGVCVTHGNALHNIELVRDKFDNDEDATYVSWLPLFHDL
GLIVVMLGAFQCGATCYLMDAVDFIRDPICWLRAISRYRGRNAGGPNFAFELCARK
VTPEQLGELDLRCWDVAFNGAEPVRADTMERFAQVFAPCGFRSTAFYPCYGMAEA
TGMISGGASRAAPVVELFDEQSLGRGVVKVADKGGASADDDTVASRSHGRSLVGN
GTAGGGQEIAIVDPDTYERCPPLTVGEVWLRGPSIGPGYWQRPEATRETYGETIAH
EDDGTYLRTGDLGFVRDGEVFLTGRLKDLIIIRGANHYPQDIELSVERSHAALRGGG
CVVFPVETDTEERVGIIAETLSDRDPGEQSAEILRSIRAAVTREHEVGVAVIALVPPRT
VLKTTSGKPRRRANRSTLFAGQFPVWAEWRDAQRPAGDTAAPPDAPRPSPIPKHK
LARRRDELFQLVHGVVAARLNAPEILPDRPLQELGFDSLAAVEIQAQLAKQLGVELP
ATLLYDHPTLRALTSFLLQTVLLAPHESAEEASEDGRGLSTLADPDEPIAVI

8.2 Amino acid sequences for the mixed NRPS/PKS protein Sb1

Sequences of the gene *sb1* were obtained from 454 fosmid sequencing of the fosmid clone 21H12 and translated into protein sequence.

NYLWLHDYGHFTLGATYPSPGGSCRWSALQRVPFAHKGGNVHLRRSDVYTGVDC
MDVDPPPRGDAGAGYTIAASCEQALVWLRLGLFGPSPVNHRCAGAWIDGPLDAAAI
EQATLRLGERHDLLRTTFCEQTDGNGEPRLVQVIAPTASPRLTRVDGGELGEEQVL
RRAGEAARQHLDPREEPPWRIYLIRQRPGRHLAVLVVHALLADADSDAASLLKEWQ
MLVEGTGRMTGQVTGPYTHHLRAQAESLESPQTRGSIAWWQAQHASAPPLLELPV
DRPRSPVPARAGARVSRRIDAGLLAALTSASTQLDADAVTLALTGLFAVLYRYTGQP
ELVVEARVPGGANGLGPVADRLMLRVAVAPELGFAALLERTRRAAQTAQEHGALA
SLVAAMTTRAPGCQVYFHGACAEPSAVALDLGFVAHELELAVDLRQGRLVAAFDRD
LFEHDTVQRLLGHWEVALAAALAEPQRAIAGLPLLVPAEQQTIVHAWNATGAAYGG
PDCLQGLGLFAAQAARTPEAPAVVDASGCLTYRELAARATRLAWRLRALGVGADV
PVGLFLDRSLDLAVGVLAILMAGGTYVPLDPAYPAERLQWIVGDSGALAMVTCSRLA
AALPPSQAQSVLVDVEVEAPAVTLPRTLPGDLAYIIYTSGSTGRPKGIGMPHAALVNL
IEWHAAELLGGARTLQFAALGFDLCFYELFTAWRTGGVVHMIDEAVRHDVARLGATI
AGEAIEKVILPVVVLQQLAEEYAERPEVLRSLKEVTTSGEQMHLTAPVIALFERLPGC

TLHNHYGPAESHVVTAYALPEARAQWVPHPPIGRPIANTRVLVLDPHMNLCPVGVP
GELYIGGECLARGYWNRPDLTAERFVPDPFAVAPGARLYRTGDITRQRADGTVEYL
GRRDDQVKIRGVRVELGEVLSALSRHPRVQEAVVIAREDAPGERRLVAYFVATGGE
VGDLASELRLFLRRSLTEAMIPTAFVALPAMPLSPNGKIDKRALPRPEDDTATSTSVP
PRSELEASIAGIWQDVLKVERVGAHDDFFALGGHSLTATQVLARVRRTFGVEVALA
QLYAGPTVAGLAGLVRAGRAPVQARAPIEPAPRSGPLPLSLAQERLFFLHRLAPDST
AYNCPHGFRVRGPLDAAALRAGRRRAGGPPRVAAHDLRRGRRGGAAGDRGARG
GAAAALRPAQPAGGAAGLELERRLATEVRRPFDLSRGPLVRGCLVQLADDEHFLLL
DLHHIVIDGWSMRLLLDELAALYDAERRGTPAALAPLGLQFADYAAWERARLQDGL
QAAIAWWKQTLADAPRLLPLPTDRPRPNTQSFRGGAITLKPDADTSAGLRRLGAPA
GLSASLVLLAGFAALLHRYTGEPTIVLGMPTAARDQVELEGLVGFFVNMLPIRVDLA
GDPSFVQLLAQVRRVTLDAFDHDAVPFDRIVQELRPERTLSHNPVVQVGFAPQPPG
ARDLRLAGLQVESVEADVPRAIFDLTLYVREDGGELTALLEYSTDLFERATIERMGG
HLLALLREVSARPQRRVSQVPLLGDEEREDVVVRWNTTGPSASPPEMVQALVSRA
AAVDPERPAVVIGEARLSYAELERRARSLARLLRARGVGPESLVAISVEKSLDLIVAV
VAVLKAGGAFVPLDPAYPAERLAFMLADSGARLVLTHARLAERFAGRELIRLDEPQE
EVEVELPPIADDPTRAAYVIYTSGSTGRPKGVVVEHRSAVNLARAQRVMFGLTTDS
VVLQFSALSFDAFVFELLLAWGVGATLCLVPPALPIPGREFSEMLRAQRVSICVLPPS
LLAAMSDDPLPDLVHVVAAGEACSAELVARWGRGRRFVNAYGPTETTVCATWAPC
EPGEGPPPIGKPLVDVQVYVLDPAGNPAPIGVFGELYVGGAGVARGYLARPELTAE
RFVPDPFSGAAGARLYRTGDRVRWRADGQLEFAGRLDRQVKVRGFRIELGEVEAV
VRAHPEVQDAVVVAREDLVAYVVPREPGTSLKTAVREFAQARLPGHMVPAAVVTV
EAFPTLPNGKVDVRALAERSPRSEARATEQPRSQIERTLAEIWREVLAIEAVGLDAP
FFELGGHSLLLAKVRTAIEARLGRRLDMVELFQHPTIRSLAARLSGGEQEVAPARTAI
RAEAGGDAIAIVGLAGRFPGAADVEALWSNLLAGVEGIHFAGADELAAAGVDPSLR
GSSGFVPAFGALADAFCFDAAFFGYSPQEARLMDPQQRVFLEAAFAALEDAGCDP
SRGQLRVGVFGGCDAPRHWLELGGASGIDEFQRGVANIPDNLTSRVAYKLGLRGP
AVTVLSACSTSLVAVHLASQSLRAGECDVALAGGAAVAPATVLGHVHEDGSILSAD
GHCRPFDADAGGTVAASGVGVVVLKRLADALADGDAIRAVIRGSAIGNDGADKVGY
TAPGLQGQVDVLRRAYAAADVAPASVALVEAHGTATRLGDPIEVAALTQVFGDATG
RTGWCALGSIKSNLGHLSAAAGVAGLIKATLAVERGQIPPTLHFRAPNPELRLEDSP
FFVNAGPVAWPDGGPRRAGVSAFGVGGTNAHVVLEEAPAPAPSEPGRAWQMLPL
SARTSAALATRAELLAGYLRERPTLPLADVAFTLQQRTALSPARCVVVCCDGEGAQ
AALTARRGPLVVSGQVRSRAPRLVFVFPGGGSQHVDMGRELYAQAPPFRAALDAC

AELFARELGEDIRGLMFAPAGEAEAAAQRLQKPSRNMAAIFAVEYALAHLLIAWGLR
PAAVLGHSLGGHSLGEYAAAAIAGVMSLADAVALVATRGALCDAAPPAAMLSVPLA
PAAATKLLRPGLSLAAVNGPEACVVAGAAADIEAFAATLAAAGVEARRLPIATASHS
ELMQPLVGRLVERARTVELPAPQIAMVSGVTGTWIGDEVRDPAYWGRHLRDTVRF
ADGVATLHADLDHVCIEVGPGAHLASLLRNHPDAGPERLVVSAMANPRAGRGELEA
LMLALAQLWCAGVTIDWQAVSAGERRRKVPLPTYPFARTRHALELRPAAPVLSLVT
PPPPVDAADEAMPRESPRDAVAQAVAQIWQENLGVAAVHDGDNFFDLGGTSLIAV
KIRARVRERLGVTTPVHALLEHPRFDAFVDSLRAHLGGLTEGRRSSLLVALRPGRP
GTRPIFLVQPIGGTVYTYLPLARHLDAGGAAIFGVRASGTDPGEPVIDDVPTMAERY
VEDILREQPVGPYTVGGHSAGGITAYEVARQLEARGHAVRVLILDAPSMPAVYDDVI
ETIDDFMRSHEAFATSESASYQSFVAALETDPALRQIVLATCLAEQRYRPRPIGAEV
VYFAASEQRDARDTHAGMYWLDLAEGPFSLYRTPGDHFTMMDEPRVETLARLIDQ
HLARDLYRRAS

8.3 Sequences obtained from end-sequencing of fosmid clones

end12A9

KLACLQVDSRGSHVERVRAVLDEPPVWPLFWPSLVDVERVGPRRNGVDGGRFLL
RWRAPGGYSLAHEVVITAVGERVIESLSHGQLEGAGRWELEPIEGGTRVRYHWHV
DTNRLWMAALRPVLAGWFAKNHDALMAEGGAGLAAYLGARLLAVRHHAGERARA
MWTRLPAPVADTGRSPGELGAAVLSAYAGPRPAGPWTTLQRWRDLVFLHWRVEP
GVLGPLLGGLAPEHIDGSPWVTLVCMRTPRIVTAHGVPLSVGEFSQVNLRTYVRHG
GRPMVSFLHVCCGSRLVAAMVRRLGRMPYRAARVATKRELLTHSYTCRGPGTLDV
DFRPRGPRPRSVGVEAELLERYAFVEARGRRVSVGELAHAPWEVCDADVAVRTNT
LLGALGIAGIDPLRPELVSFAPAASVVAW

end12A9rp

RQGFPSHDVVKRRPVNCNTTHYRANSSSVPGDPTLVELLTPLFPDGLREPPALDAH
VDQSRRSSDLSHPLAQVGREPEAAAMREVAVRLNLERRRPAGLINSLGDRSESLAA
AGRTHAAMHTWAVVDRLMHAHGEEHWGPLQRRAELATDLGRWDEAEADMQALG
KEHDTIVLDLCRAAIAFARGEDAVHFIKEASVYLTLGWSASIAARWNLLYGRIKLADG
EFAAALANLSNAVQLARRMGSDSARAQAWRAVCLANLDRPAEARQALADARLVAT
RHDNRRVAGILAAAHLALGERDLAEPLALAGYREAWADGPQYSWFDDLRVCEGVL
AELGLPPPELPTFDPTRAPPLPLADEIEAFIAELAARRPASGSRLLHLDPMLCEPPAQ

PDQGPPPAPGPPAFAVHERVKIATDAPWQADVNGLEGVVFWREPMQDHTAVAET
QRWKYVVYLPAQQVYRTLIERHLTAQGGSVEPSFFFAERAEISYDVVVTDASTIVEG
CLRLPGETWRTFYAEKTDDLDEARLETFDWENGIRGLTLEMPEDEPMNKAALEQAF
AALLGVEQIVEVPGPDSMMMRP

11.3T7

GcCGCCaTGGCCGCGGGatTCGAGCGCGAGCTGTTCGCGGCCTGGGTGTCGaan
nnnAangAGTGCGAGTTCTGAACGCGGGCCCACGGCGCGGTCGCGTCGAAGGT
GATGGGCGAGACGGCGGTCGCGGCGGCGCTGGCCGACCCCGACGGCGCCGA
CCTGCCCGAGAAGGTCAAGGCGACGCTGCGCCTGCTGACGAAGGTGACGCGC
GCCCGCGAGGCGCTCGGGCCCGAGGACTTCGAGCCGGTGCTCGCGGCCGGG
GTGTCGCGCCAGGGGGTGATCGACGCGCTGATGGTGGGGTTCGCGTTCAACG
TGATCACCCGCCTGGCCGACGCGTTCGAGTTCGAGGTCCCACCGCAGGCGTC
GTTCGACGCGTCGGCGAAGACGCTGCTGTCGCGCGGCTACAAGTGAAGGGAG
TGGTGACATGTCCGAAGGGACATGTCACCGCGCGCGCCCGCGGACGGACGCG
CCCTGCGGGGAGCCGTGACATGGTCGGAGGGCCATGTCCGTGCGAAGAGGTG
CGGGCGGTGCGAGGCCGTGGATCGTCTCGCCGACGTCGGCGAGAGGAGCCC
GTTCGGTCCTGTCgCCGTGGGTCGTCgACAGGACCgAanGGACnTGGGGCGATC

11A3RP

GCTATGAccnTGAtTACGCCaAGCTATTTAGGTGAGACTATAGAATACTCAAGCTT
GCATGCCTGCAGGTCGACTCTAGAGGATCCCACCGAACTTGCCGATCTTCTGAT
AGGCGTTGCGCAGGTGCGGGACCTTCATGGTCTGCACGAATTCGGTGCCGACA
TACGAACCGTCGGAGCCGAAGAAGCCCGGGCGATCGACGCCGTACTGGGCGT
TTCCGGTCGGGTCGAGCACATGACAGAAATTGCAGGTCATGCCGAGCGTGGCC
GGCTTGTTCATGTAGAGGTCGCGGCCGGCCACCTGATCGGGGGTGAGCGAGTT
GTCgAGGTTGCGGATCGGGTTCGGCGGgTAgGTCAgCTGCcncGCGAAGn

21H12T7

GCTGACTGGTCGGGGAGCACGTTCGGCTGGGCGTTGGGCGGGAAGTTCGGGC
TGGTCTTGTCGCCGCGCCAGTGCATCGCCCCGTGGTTGGCCATACCGCGCAGG
CTCTGGGTCGTCATCGGCCCCTTGATCGGGTGGAAGTCGATCGGGAAGTCGGG
CAGGAAGATGGTCTTGAACGGCCCGGGCGCGATCGTCACGTCGCCGTCGGGG
TTGCCGAGGTCCCACGCGAGGTGGTCCATGTCGCCGTAGACGTGGCAGGTGG

CGCAGGCCGAGTCGCCGCGGCCCGAGGTGAACCGCGCGTCGTAGAGGAACCG
GCGGCCCTGCGTCACGCTGGCCGGCTCGGGGTTGTGCATCGACAGGTGCTGG
ATCTCGTCGCCGTCGTCGGTGTCGATGATCGAGACCGCGTTGTCGAAGCGGGT
CATCACATACAGGCGGTCGCGCGACTCGTCGAGCAGCACCCCGGTCGGGCCG
CCGCCGCTGACCTGGAGCTGCTCGCACGGGTCGGGCCAGATCGCGTCGTTCT
CGAGCGCGGCGGTGTCGTAGACGCCGATCTTGTCGGAGCCGAGCGCGGCGAC
GTAGAGCGTGTTGCCGTTGCCCGTGATCGCCATGCCCTGCGGGAGGGCCATG
CTGAGCTCGTTGTCCTCGTTGGGCAGCGGCTGGAACGGCGCCTCGTAGTCGAG
GTGCGGGTTCATGTGATGGGTGGTGACTGGTTGGATCTGCGTGACGAGGAATA
ATTGCTGGTCCCCCTCATT

21H12RP

GCTATGaccnTGATTACGCCaaGCTATTTAGGTGAGACTATAGAATACTCAaGCTT
GCATGCCTGCAGGTCGACTCTAGAGGATCCCACGGAAGAAATAGGCGAAGAAC
CACACCATGGTGTAGGCGACGCCGTCGAAGCCGGGGAACACGTCGTAGGCCA
GCGAGGCGCCGCGGGTGGCGAGGACGTGGCCGGCGGTGAAGACGAGCAGGA
CGACGGCGGTGACCCGCTGCCAGCGCAGGCGCGGCGCGGCGGGCCCGGGCG
CGCCGCGACGGCGCCGCCGGACCATGCGCCACACGCTGACGACGATGTGCAC
GAGCAGCGGGGCGATCACGCAGGCGAGCTCGACCAGCGGGGCCTGATAGGC
GGCGCGCAGCTGGACCTGCATGTCGTCGTACACCGCAGGGCCGGCGGCGGCG
AACATCTGATTGACGAGGTGGGCCACGAGGAACACGGCGAACAGGGCGCCGC
TGGCGGCCTGGATGCGGATCAGCGCGCGTTCGCGCGCGGCGTCGCGGGAGG
GGTCGGTGGGGCTCATGGCGACCAGACGAGGCGGTCCTCGGCGCCGTGACAC
GGAAAAATCGCGCATCgATCGGcCGATGCAGGCGGGCCCGATCGCGgcCGTCgt
GCCggcgATcgCCGCGGcTGccgcagGt

8.4 16S rDNA sequence

tggcCGcgGnnttagaGTTTGATCCTGGCtcagAGCGAACGTTTGcngcggGCctaACACA
TGCAAGTcGaAcGGGttaGCAATAACCagtGGCGCAcGGgTGCgtaACacgtaGgTaATC
AACCCctCGGtTcGGGATAACGTTCTGAAagGaGCGctAATACCGGACGTGTCttCG
GgagCttcggctcCtgtcgAGAAAgaCCCtCAaGGGTTGCCnaGGGacgaGCCTGCGGCC
CATcAGctaGTTGGCGAGGTAATaGctCacCAAGGCGaagacGggtAGCTGgtCTGaga
GGaTGATCAgtCACacTGGaACTGAGACACGGTCCAGACTCCTACGGGagGcaGC
AGTGGGGAATATTGCGCAATGGGCGAAAGCCTGACGCAGCCACGCCGCGTGA
GCGATgaAGGCCTTCGGGTCgtaAAGCTCTGTGGGGAGAGGCGAAGaagcCTGTG
AAGagagGCCTTgaCGgttCTCCTtacAagacCGGCTAACTCCGTGCcacAGCCGcGna
atcgGAGgggCGAAcGttcTCGGAatattgGGCgtaAAGCGCACGTAGgcgCGGCGtagCG
GGATgTGAAAGCccngGGctCAACcctGGAAgtGCATCCCgnnacTGtgTCGCTTGantnt
CGGAGGgGGACAGAGAATTCCCGGTGTAGAGGTGAAattCGTAGATATCGGGAG
GAatnccAGTGGCGAAGGCGCTGTCCTGGACGAAGATTGACGCTGAGGTGCGAG
AGCGTGGGGAGCAAACAGGATTAGATACcCTGGTAGTCCACGCTGTAaACGATG
AGTGCTGGACGGTGGAGGATTTGACCCCTTCGCTGTCGAAGCTAACGCGTTAA
GCACTCCGCCTGGGGAGTACGGTCGCAAGACTAAAACTCAAAGGAATTGACGG
GGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGCAGAAC
CTTACCTGGGTTAAATCCACTGGAACCTGGCTGAAAGGCTGGGGTGCCCTTCG
GGGAGCCGGTGAGAAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATG
TTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCGCCAGTTGCCACCATTGA
GTTGGGAACTCTGGCGAGACTGCCGGTCTAAACCGGAGGAAGGTGGGGACGA
CGTCAAGTCCTCATGGCCCTTATGCCCAGGGCTACACATGTGCTACAATGGCTG
GTACAAAGAGCCGCAAGCCCGCGAGGGTGAGCAAATCTCAAAAAACCAGTCTC
AGTTCGGATTGCAGTCTGCAACTCGACTGCATGAAGCTGGAATCGCTAGTAATC
GGAGATCAGCACGCTCCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCG
TCACACCATGGGAGTCGGCTGCTCCAGAAGTAGGAACCTCAACCGCAAGGAAA
GGCCCTACCAAGGAGCGGTCGGTGACTGGGGTGAAGTCGTAaCAAGGTAGCCG
TagGGGAACCTGGGGCTGGannACCTCCTTAATCACTAGTGCGGCCGCCTGCAG
GtCGACCAtAtggaGAGctcCCaAACgcgtga

8.5 Mass spectrometry data of the A domain assay

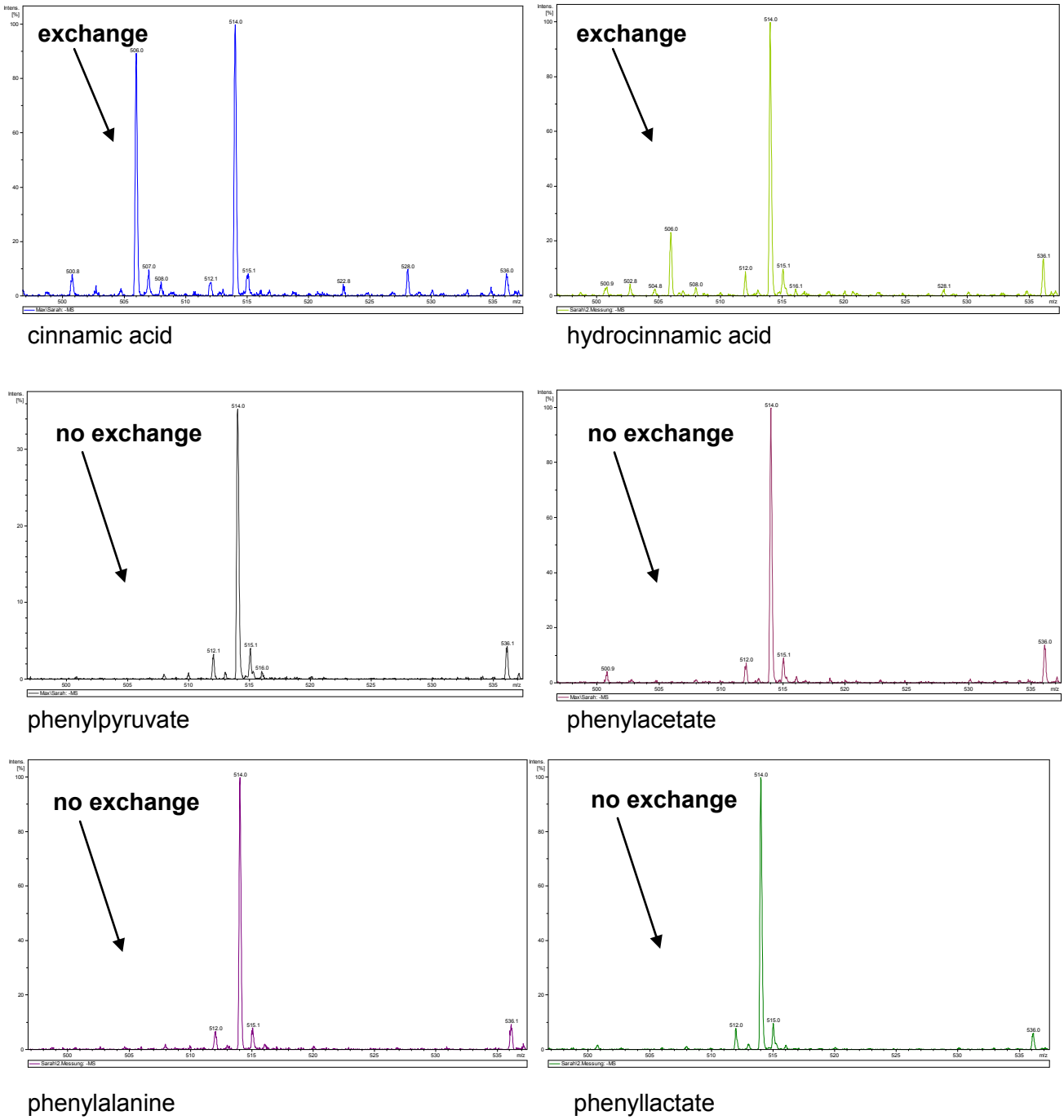
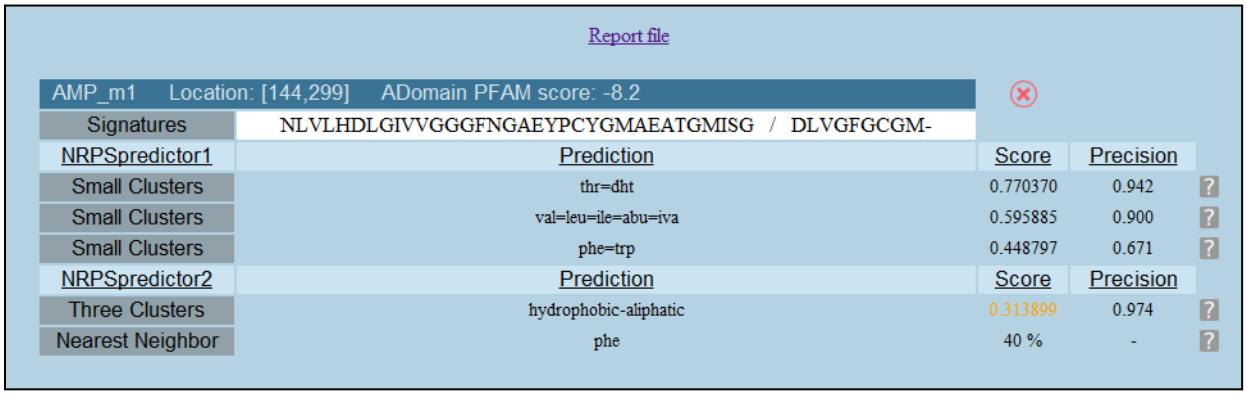


Figure 8.1 C) shows MS-spectra from the repeated ATP/PPi -exchange assay of AMPACP with cinnamic acid, phenylacetate and other phenylpropanoids (hydrocinnamic acid, phenylpyruvate, phenyllactate, and phenylalanine). Exchange activity was only observed for cinnamic acid and in a lower content also for hydrocinnamic acid. On the base of these data the relative exchange ATP/PPi was calculated as shown in section 4.11.2.

8.6 Substrate predictions of some domains by bioinformatical online tools

8.6.1 Prediction for the substrate specificity of the AMP-ligase from Phn2

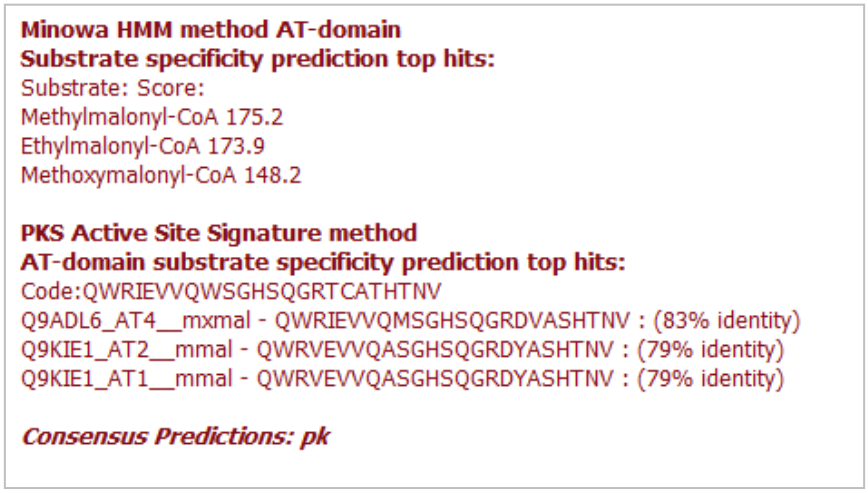


Report file

AMP_m1 Location: [144,299] ADomain PFAM score: -8.2			
Signatures	NLVLHDLGIVVGGGFNGAEYPCYGMAEATGMISG / DLVGFGCGM-		
NRPSpredictor1	Prediction	Score	Precision
Small Clusters	thr=dht	0.770370	0.942
Small Clusters	val=leu=ile=abu=iva	0.595885	0.900
Small Clusters	phe=trp	0.448797	0.671
NRPSpredictor2	Prediction	Score	Precision
Three Clusters	hydrophobic-aliphatic	0.313899	0.974
Nearest Neighbor	phe	40 %	-

Figure 8.2 Analysis of the AMP-Ligase of the loading module from Phn2 with the NRPS Predictor 2 online tool (http://nrps.informatik.uni-tuebingen.de).

8.6.2 Substrate prediction for the AT1 domain from Phn2



Minowa HMM method AT-domain
Substrate specificity prediction top hits:
Substrate: Score:
Methylmalonyl-CoA 175.2
Ethylmalonyl-CoA 173.9
Methoxymalonyl-CoA 148.2

PKS Active Site Signature method
AT-domain substrate specificity prediction top hits:
Code:QWRIEVVQWSGHSQGRTCATHTNV
Q9ADL6_AT4__mxmal - QWRIEVVQMSGHSQGRDVASHTNV : (83% identity)
Q9KIE1_AT2__mmal - QWRVEVVQASGHSQGRDYASHTNV : (79% identity)
Q9KIE1_AT1__mmal - QWRVEVVQASGHSQGRDYASHTNV : (79% identity)

Consensus Predictions: pk

Figure 8.3 Subtrate prediction for AT_1 by AntiSMASH, using two different methods for substrate prediction

8.6.3 Substrate prediction for the A domains of Sb1 the NRPS/PKS hybrid

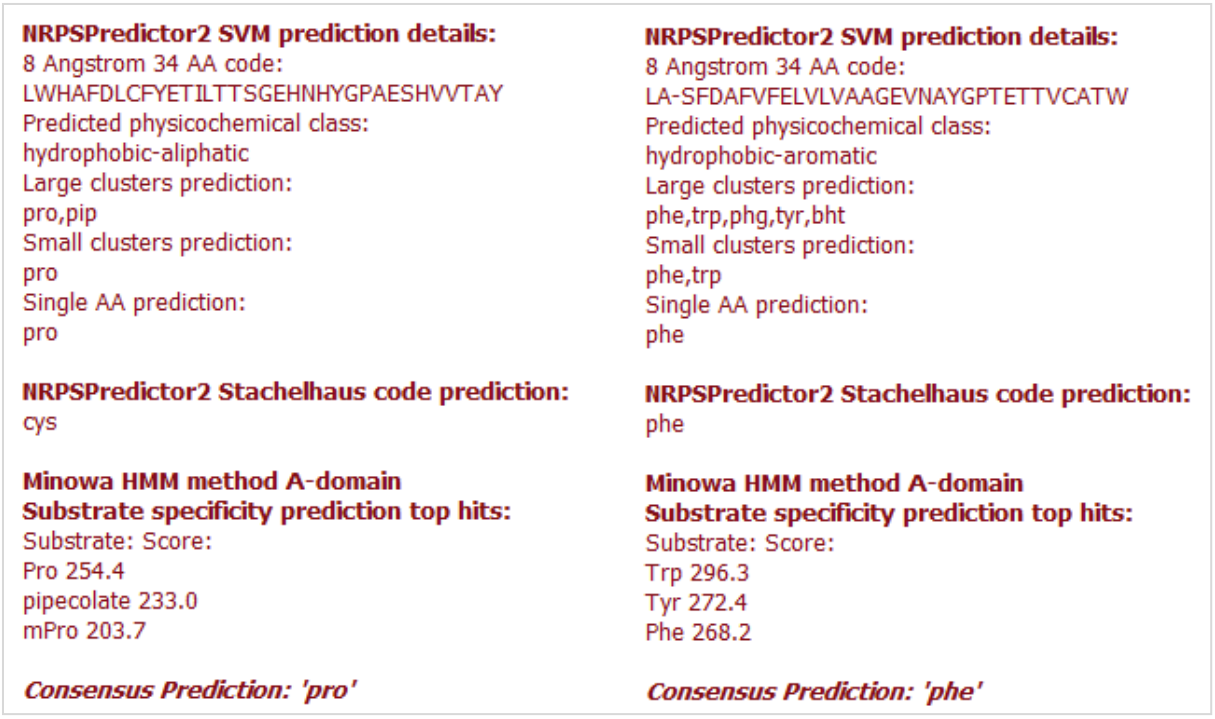


Figure 8.4 Subtrate prediction for the A domains of the NRPS/PKS mixed Sb1 by AntiSMASH, using two different methods for substrate prediction