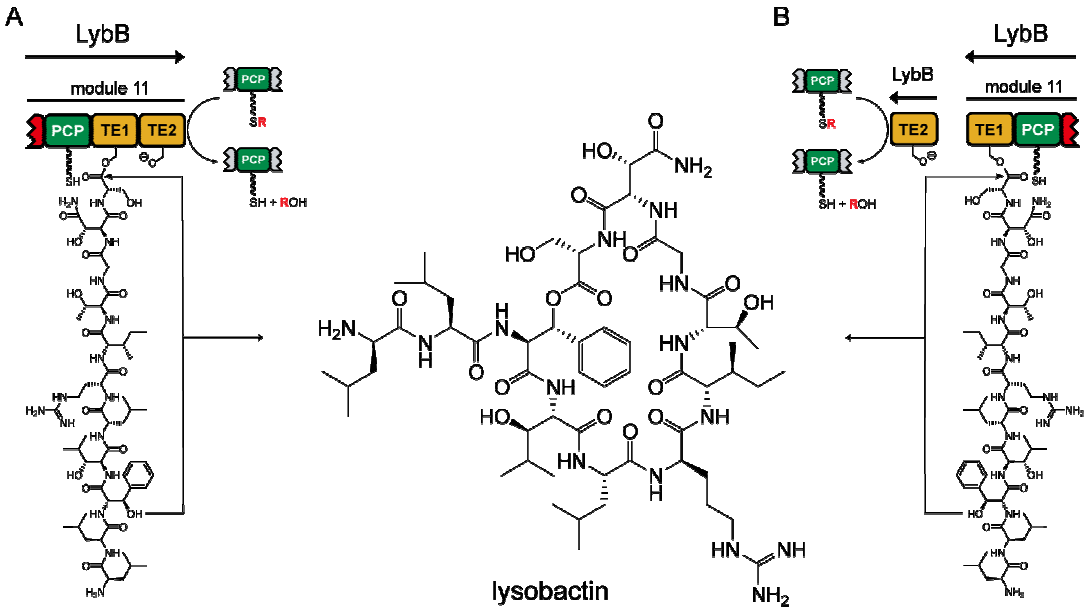
Identification and Characterization of the Lysobactin Biosynthetic Gene Cluster and Its Unusual Termination Module

---

Identifikation und Charakterisierung des Biosynthesegenclusters von Lysobactin und seines ungewöhnlichen Terminationsmoduls



Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften im Fachbereich Chemie der Philipps-Universität Marburg

Dipl.-Chem. Jie Hou

aus Beijing / China

Marburg an der Lahn, 2012

Vom Fachbereich Chemie

der Philipps-Universität Marburg (Hochschulkennziffer: 1180) als Dissertation

am _____ angenommen

Erstgutachter : Prof. Dr. M. A. Marahiel

Zweitgutachter : Prof. Dr. A. Geyer

Tag der Disputation: _____

Dedicated to my family

Summary

Nonribosomal peptides (NRPs) constitute a class of structurally and functionally diverse natural products, which are synthesized by nonribosomal peptide synthetases (NRPSs). NRPs exhibit a wide range of bioactivities, including antimicrobial, antifungal, antiviral, immunosuppressive and antitumor properties. Numerous of these compounds have been discovered via screening of microbial extracts. In recent years, increasing knowledge of the biosynthesis of natural products and development of new sequencing techniques lead to the identification of gene clusters, which are putatively involved in the biosynthesis of nonribosomal peptides.

Based on the sequencing result of the genome of *Lysobacter sp*. ATCC 53042 and the former work from Bernhard et al.[1] on the gene fragment involved in the biosynthesis of lysobactin, the entire biosynthetic gene cluster of lysobactin was identified and characterized. The cluster encodes two multimodular nonribosomal peptide synthetases (LybA and LybB). Due to the correlation of the number of modules found within the lysobactin gene cluster and the primary sequence of lysobactin, the biosynthesis of lysobactin follows the colinearity principle. Investigation of the adenylation domain substrate specificities confirmed the direct association between the synthetases and lysobactin biosynthesis.

Furthermore, an unusual tandem thioesterase domain architecture (PCP-TE_1-TE_2) of the LybB termination module was identified. Biochemical characterization of the individual thioesterases *in vitro* proved that the first thioesterase is responsible for the cyclization and the release of the final product, while the second thioesterase showed a type II TE activity, which is responsible for the regeneration of the mis-primed peptide carrier protein during the biosynthesis of lysobactin. Together with the observation of the proteolytic degradation during the heterologous production of LybB-PCP-TE_1-TE_2 giving rise of LybB-PCP-TE_1, we have proposed that the LybB is also cleaved to generate lone-standing LybB-TE_2 prior to lysobactin synthesis in the native strain. The resulting lone-standing TE_2 serves as external type II TE to regenerate mis-primed peptide carrier protein via hydrolytic cleavage of the PCP-bound noncognate substrates.

Additionally, the sequence of the genome of *Lysobacter sp.* ATCC 53042 was bioinformatically analyzed. The analysis result delivered further potential NRPS and PKS-NRPS hybrid gene clusters. Based on the proposed substrate specificities of the adenylation domains, the chemical structures of the products were proposed. However, further experiments are needed to confirm the production of these compounds.

Zusammenfassung

Nichtribosomale Peptide (NRP) konstituieren eine Klasse strukturell und funktionell diverser Naturstoffe, welche durch nichtribosomale Peptidsynthetasen (NRPS) synthetisiert werden. Nichtribosomale Peptide weisen eine Vielfalt biologischer Aktivitäten auf, sowie antimikrobielle, antifungielle, antivirale, immunsuppressive und antitumore Eigenschaften. Durch Screening mikrobieller Extrakte wurden zahlreiche solcher Verbindungen entdeckt. In den letzten Jahren ermöglichten die Kenntnisse über die Biosynthese von Naturstoffen und die Entwicklung der Sequenzierungstechnologien Identifikation von Genclustern, die vermutlich an der Biosynthese von nichtribosomalen Peptiden beteiligt sind.

Basierend auf der Sequenzierung des Genoms von *Lysobacter sp.* ATCC 53042 und der früheren Arbeit von Bernhard et al.[1] an dem Gen-Fragment, das an der Biosynthese von Lysobactin beteiligt ist, wurde der gesamte Gencluster der Lysobactinbiosynthese identifiziert und charakterisiert. Der Gencluster kodiert zwei multimodulare nichtribosomale Peptidsynthetasen (LybA und LybB). Wegen der Korrelation zwischen der Anzahl der Module innerhalb des Genclusters und der primären Struktur von Lysobactin wurde festgelegt, dass die Biosynthese von Lysobactin dem Kolinearität-Prinzip folgt. Die Untersuchung der Spezifitäten der Adenylierungsdomäne bestätigte den direkten Zusammenhang zwischen den Synthetasen und der Biosynthese von Lysobactin.

Darüber hinaus wurde eine ungewöhnliche Tandemthioesterase-Architektur im Terminationsmodul von LybB identifiziert. Die biochemische Charakterisierung der einzelnen Thioesterasen *in vitro* bewies, dass die erste Thioesterase für die Zyklisierung und Freisetzung des Endproduktes zuständig ist, während die zweite Thioesterase eine Typ II TE Aktivität zeigte, die für die Regeneration des Peptide-Carrier-Proteins verantwortlich ist. Zusammen mit der Beobachtung des proteolytischen Abbaus, der während der heterologen Produktion von LybB-PCP-TE_1-TE_2 die Produktion von LybB-PCP-TE_1 verursacht, haben wir vorgeschlagen, dass die LybB auch im nativen Stamm gespalten wird. Somit wird die dadurch erzeugte allein stehende TE_2 produziert, die als externer Typ II TE dient.

Zusätzlich wurde die Sequenz des Genoms von *Lysobacter sp.* ATCC 53042 bioinformatisch analysiert. Die Analyse lieferte weitere potenzielle NRPS und PKS-NRPS Hybridgencluster. Basierend auf den vorhergesagten Spezifitäten der Adenylierungsdomänen wurden Strukturvorschläge für die putativen Produkte gemacht. Es sind jedoch weitere Untersuchungen nötig, um die Produktion dieser putativen sekundären Metaboliten zu bestätigen und um die vorgeschlagenen Strukturen zu verifizieren.

The majority of the work presented herein has been published:

Hou, J., Robbel, L., Marahiel, M.A., (2011), Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture, *Chem Biol*, 2011. **18**(5), 655-64.

Additional publications:

Miethke, M., **Hou, J.**, Marahiel, M.A., (2011), The siderophore-interacting protein YqjH acts as a ferric reductase in different iron assimilation pathways of *Escherichia coli*, Biochemistry, 2011, **50**, 10951-64.

Abbreviation

3',5'-ADP	3',5'-adenosinediphosphate
A-domain	adenylation domain
AcOH	acetic acid
ACP	acyl carrier protein
AL	acyl-CoA ligase
AMT	amino transferase
AT	acyltransferase
AT_L	loading AT domain
ATP	adenosine 5'-triphosphate
BLAST	Basic Local Alignment Search Tool
bp	base pair
C-domain	condensation domain
CDA	calcium-dependent antibiotic
CDS	coding sequence
CHS	chalcone synthase
CMA	coronamic acid
CoA	coenzyme A
Cy-domain	cyclization domain
DCM	dichlormethane
DEBS	6-deoxyerythronolide synthase
DH	dydratase
DIPEA	diisopropylethylamine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
dsDNA	double-stranded DNA
EDTA	ethylenediaminetetraacetic acid
ER	enolreductase
ESI	electrospray ionization
EtBr	ethidium bromide
F-domain	formylation domain
FAS	fatty acid synthase
fH_4F	formyltetrahydrofolate
FMN	flavine-mononucleotide
Fmoc	9-fluorenylmethyloxycarbonyl
Fmoc-Cl	fluorenylmethyloxycarbonyl chloride
FT	Fourier transformation
E-domain	epimerization domain
Gtf	glycosyltransferase
HBTU	*O*-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate

Abbreviation

HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HOBt	1-Hydroxybenzotriazole hydrate
HPLC	high-performance liquid chromatography
HyPhe	L-*threo*-3-phenylserine
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl-β-D-thiogalactopyranoside
KR	ketoreductase
KS	keto-synthase
LC-MS	liquid chromatography-mass spectrometry
LDD	loading didomain
LTQ	linear trap quadrupole
MALDI	Matrix Assisted Laser Desorption Ionization
MCoA	malonyl-CoA
MeOH	methanol
MIC	minimum inhibitory concentration
mMCoA	methylmalonyl-CoA
MRSA	methicillin-resistant *Staphylococcus aureus*
MT	methyltransferase
Ni-NTA	Ni-nitriloacetic acid
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetases
OD	optical density
Ox-domain	oxidation domain
PCP	peptide carrier protein
PCR	polymerase chain reaction
Ppan	phosphopantetheine
PPtase:	phosphopantetheinyl-transferase
R-domain:	reduction domain
RT	room temperature
SAM	*S*-adenosylmethionine
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SFF	standard flowgram format
Sfp	4'-phosphopantetheine transferase involved in surfactin production
SNAC	*N*-acetylcysteamine
SPPS	solid phase peptide synthesis
T-domain	thiolation domain
TCA	trichloroacetic acid
TE	thioesterase
TEA	triethylamine

Abbreviation

TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
TIPS	triisopropylsilane
TOF	time of flight
Tris	tris(hydroxymethyl)aminomethane
VRE	vancomycin-resistant *enterococcus*

Table of amino acid abbreviations

Amino acid	1-letter-abbreviation	3-letter-abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glumatic acid	Glu	E
Glutamine	Gln	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
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table of contents

Table of Contents

Table of Contents

Table of Contents

Inhaltsverzeichnis

Table of Contents

1. Introduction

Since ancient times people have learned to prepare traditional medicine from flora and fauna to protect themselves against diseases. From these “crude drugs”, different medicines were discovered and isolated exemplified by artemisinin and quinine. These natural products with complex structures are considered to play a highly significant role in the drug discovery and development process.[2] Microorganisms such as bacterial or fungi are rich sources for discovering new drugs or lead compounds. Under selection pressure, these microorganisms evolved the ability to produce natural products which are optimized for chemical defense or communication and show a broad bioactivity spectrum. Natural products discovered from microorganisms show a broad structural diversity and comprise peptides, polyketides, steroids, glycosphingolipids, terpenes etc. In the following chapter, natural product assemblies such as nonribosomal peptide synthetase (NRPS) and polyketide synthetase (PKS) will be discussed in detail.

1.1 Non-ribosomal peptide synthetase

Discovery of penicillin from fungal host organism *Penicillium notatum* by Sir Alexander Fleming marks a new epoch in research into biologically active natural products. Numerous compounds with antimicrobial, antifungal, antiviral, immunosuppressive and antitumor activities were discovered via screening of microbial extracts.[3-5] Moreover, knowledge about gene clusters encoding enzymes for biosynthesis of natural products has greatly increased, which suggests greatly underestimating of natural product biosynthetic capacity by detecting products in fermentations.[6] Many of these compounds are produced by polyketide synthases or nonribosomal peptide synthetases. NRPSs are large, multi-modular enzyme complexes that catalyze the biosynthesis of NRPs. In contrast to ribosomally synthesized peptides, NRPs are synthesized independent of messenger RNA template and following the multienzyme thiotemplate

mechanism.[4] NRPs exhibit unique structural features such as heterocyclic element, non-proteingenic and D-amino acids, glycosylation and N-methylation (Figure 1.1). A common feature of many NRPs is their cyclic structure, which is achieved normally via oxidative cross-linking like vancomycin, or heterocyclization, exemplified by β-lactam antibiotics, or macrocyclization such as Daptomycin.[3] Their structures are strictly connected with their biological activities and specific interaction with the corresponding molecular targets in the cell.

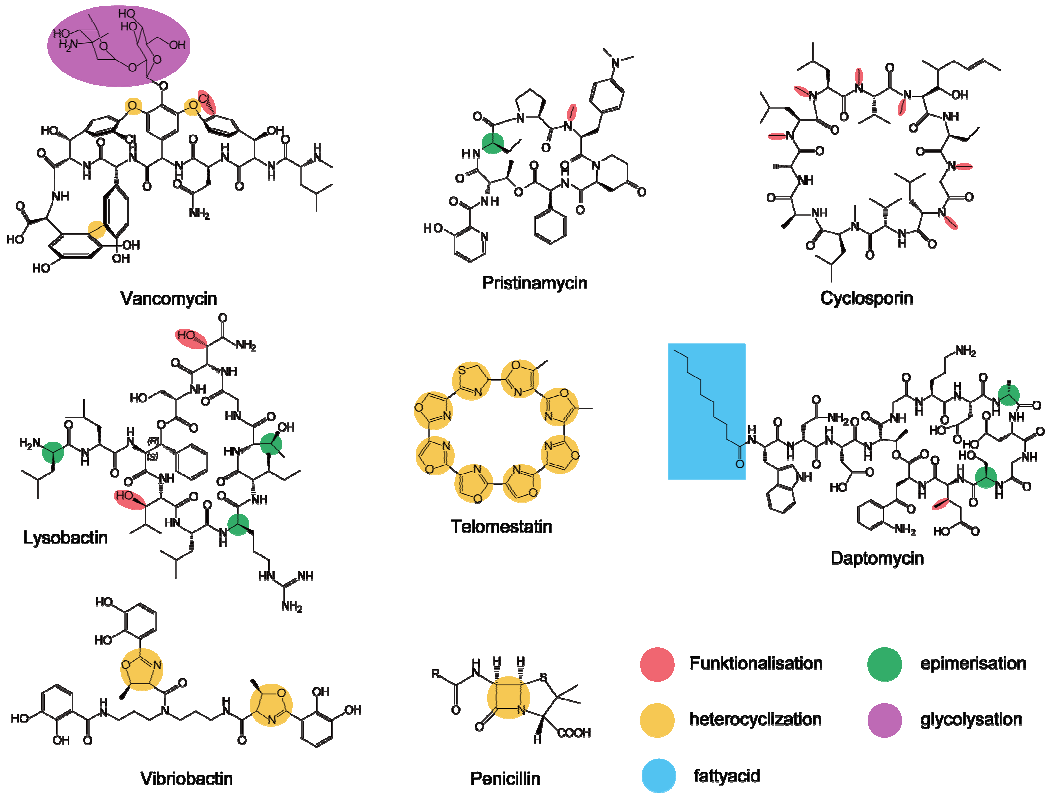


Figure 1.1: Structural diversity of the NRPs.

1.1.1 Essential domains

NRPSs are large, multi-modular enzyme complexes. Each module is responsible for the incorporation of one building block into the growing peptide chain and could be further subdivided into different domains responsible for substrate recognition and activation, binding, condensation, modification and product release.[4] These domains can be

identified by analyzing their highly conserved amino acid sequence motifs. Domains required for incorporating of building blocks into the peptide chain are considered essential and named as essential domains or core catalytic domains. These domains are adenylation (A)-domain, condensation (C)-domain and peptidyl-carrier-protein (PCP, also named as thiolation (T)-domain). This chapter gives detailed information of these essential domains.

1.1.1.1 Adenylation (A)-domain

Adenylation domain recognizes, activates and loads the corresponding building block (amino acid or organic acid) onto the sulfhydryl-group of the phosphopantetheine cofactor (ppan) of the downstream neighboring PCP. Substrate activation is accomplished in a 2-step-reaction. Firstly, A-domain recognizes specifically the amino acid and catalyzes the formation of an aminoacyl adenylate intermediate in the presence of Mg^{2+} and with consumption of ATP. Secondly, the carbonyl group of the oxo-ester of aminoacyl-O-AMP is converted to a thioester by the nucleophilic attack of the sulfhydryl-group of the ppan residue of the downstream neighboring PCP and thus covalently attached to the downstream PCP. Although the reaction shares the same mechanism with aminoacyl-tRNA synthesis catalyzed by aminoacyl-tRNA synthetase, the two enzyme families share neither sequence nor structural similarities.[7]

A-domains separated from the original NRPS assembly show a high degree of substrate specificity, thus, A-domain acts as the gatekeeper for incorporating monomeric building blocks. Due to the lack of proof-reading-mechanisms, A-domain shows in general a lower specificity comparing to aminoacyl-synthetases.[8] Some A-domains, such as TycA-A1[9] and LybA-A1[10], are reported to have a relaxed substrate tolerance. This substrate tolerance could lead to misincorporation of non-cognate building blocks and results in synthesis of differing NRPs by the same NRPS machinery.

A-domains usually consist of about 550 amino acids and can be subdivided into an N-terminal core domain consisting of about 450 amino acids and a C-terminal

subdomain of about 100 amino acids. Crystallization study of phenylalanine activating subunit of gramicidin synthetase I (PheA) in complex with AMP and phenylalanine (Figure 1.2) shed light on the enzyme-substrate-interaction and the residues responsible for substrate recognition.[11] Combined with biochemical and bioinformatic analysis of a large number of adenylation domains, the specificity-conferring code of A-domains was established.[12] This code contains ten amino acids responsible for substrate binding within the active-site and allows the prediction of A-domain specificities and primary sequences of NRPs synthesized by newly found NRPS-clusters.

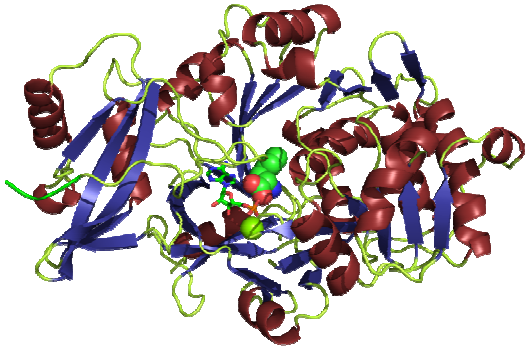


Figure 1.2: Structure of the phenylalanine activating subunit of gramicidin synthetase I (PheA) in complex with AMP and Phenylalanine (PDB 1AMU). AMP is shown in sticks form, phenylalanine is shown in sphere form, the green ball represents the Mg^{2+} ion in the complex.

1.1.1.2 Condensation (C)-domain

The condensation (C)-domain is a further essential domain in NRPS machinery comprising about 450 amino acids and catalyzes peptide bond formation. C-domains contain an acceptor site and a donor site, which accepts downstream aminoacyl-S-PCP and upstream peptidyl-S-PCP (aminoacyl-S-PCP for first C-domains in NRPS assembly, acyl-S-PCP for initial C-domains in NRPS producing lipopeptides) as substrates. The peptide bond formation is initiated by the nucleophilic attack of the α-amino group of the acceptor site substrate onto the carbonyl group of the thioester of the donor site substrate. Therefore, the upstream peptide chain is transferred onto the downstream PCP-domain and the elongated peptide chain serves as donor site substrate for the next

C-domain.

The structural and bioinformatic analysis of C-domains identified a highly conserved catalytic histidine residue in the core-motif MHHxxxDG(WV)S. Although exact mechanism remains to be elucidated, mutational studies suggested that the second histidine residue may catalyze the deprotonation of α-amino group of the acceptor site substrate and thus enhance the nucleophilicity of the amino group to ease the nucleophilic attack onto the thioester of the donor site substrate[13]. A Crystal structure study of the lone-standing C-domain VibH from the vibriobactin synthetase (Figure 1.3) revealed the V-shaped form of the C-domain, which enables the substrates tethered on the up- and downstream neighbouring PCP-domains to reach the highly catalytic histidine in the active site.[14] PCP-misloading experiments showed that the acceptor and donor site have different substrate specificities. The C-domain has restricted substrate specificity at its acceptor site. Neither enantiomeric substrate nor different aminoacyl

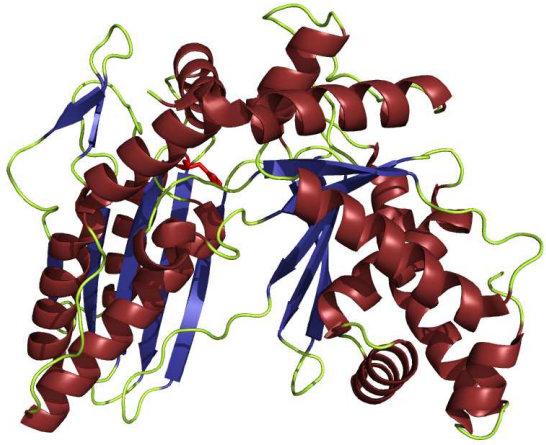


Figure 1.3: Structure of alone-standing C-domain VibH from vibriobactin synthetase shows the typical V-form of the condensation domains from NRPS (PDB 1L5A). The highly conserved histidine residue in active site in highlighted in red.

side chain length are accepted, while the donor site accepted different side chains.[13, 15] Various functional subtypes of C-domains were reported, including $^{L}C_{L}$-domains catalyzing peptide bond formation between an upstream peptide with a C-terminal L-amino acid and a downstream L-amino acid; $^{D}C_{L}$-domains catalyzing peptide bond

formation between upstream peptide with a C-terminal D-amino acid and a downstream L-amino acid; starter C-domains acylating the first amino acid activated by the downstream A-domain with a fatty acid[16]; heterocyclization Cy-domains catalyzing both peptide bond formation and subsequent heterocyclization of cysteine, serine or threonine residues[17]; and C/E dual domains catalyzing epimerization of the C_α-atom of the C-terminal amino acid of the donor site substrate and peptide bond formation between the donor site and acceptor site substrates[18].

1.1.1.3 Peptidyl-carrier-protein (PCP)

The Peptidyl-carrier-protein (PCP) or thiolation domain (T-domain) has a size of about 80 amino acids and is responsible for covalent tethering of monomeric building blocks or growing peptidyl intermediates and transportation of elongating peptide chain towards the C-terminus of the synthetase. Thus, PCP domains play a supreme role in NRPS system. PCPs are posttranslationally modified by phosphopantetheinyl-transferases (PPtases) by attaching the phosphopantetheine cofactor to the highly conserved serine residue of the core-motif GGxS. PPtases catalyze the nucleophilic attack of the hydroxyl-group of the highly conserved serine residue onto the β-phosphate group of coenzyme A, releasing 3',5'-adenosinediphosphate (3',5'-ADP). The substrate or intermediate is covalently attached on the sulfhydryl-group of the ppant cofactor of holo-PCP as thioesters. NMR studies of $TycC_3$-PCP (Figure 1.4) revealed three different conformations, namely apo (A), holo (H) and A/H states of the PCP. A and A/H coexist when the PCP is in apo-state; while H and A/H coexist as the PCP is in holo-state. The study also showed that the sulfhydryl group of the ppant cofactor could move approximately 16 Å, which enables the delivery of tethered acyl/peptidyl-substrates to the catalytic site of neighbouring NRPS domains, for peptide bond formation, modification or cyclization.[19]

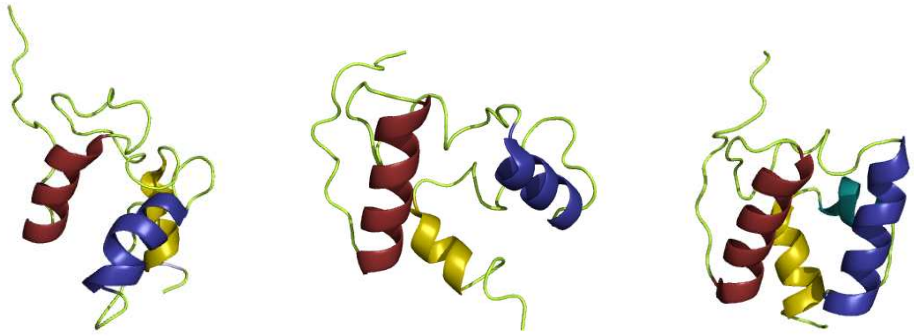


Figure 1.4: Structure of $TycC_3$-PCP domain in different states. (PDB-codes: A: 2GDY; H: 2GDX; A/H: 2GDW)

1.1.2 Non-proteinogenic building block synthesis

A significant hallmark of NRPs comparing to the ribosomally produced peptides is that NRPs contain a large number of non-proteinogenic building blocks, which are essential for their conformation and biological activities. Many of these unique structural features, such as D-amino acids, N-methylation, heterocyclic rings etc, are derived from proteinogenic amino acids. Different mechanisms were reported for introduction of these structural features. Considering the time of the modification event during the biosynthesis process of the NRPs, these modifications could be classified into three types: non-proteinogenic building block precursor synthesis, tailoring on PCP-bound substrates and post-assembly modification. In this chapter, these mechanisms are discussed in detail with concrete examples.

1.1.2.1 Non-proteinogenic building block precusor synthesis

In some NRPs, α-keto and α-hydroxy acids are incorporated beside α-amino acids exemplified by cyclodepsipeptides such as PF1022A produced by *Rosellinia* sp. PF1022[20] and enniatin found in *Fusarium*[21, 22], [23]. The α-hydroxy acids in these depsipeptides are produced by certain pathways prior to activation by the corresponding NRPS A-domain[20].

β-Hydroxy or β-amino acids are also frequently observed in NRPs. In the biosynthesis of calcium-dependent antibiotic (CDA), a non-heme Fe^{2+}/α-ketoglutarate-dependent oxygenase, AsnO, is responsible for direct hydroxylation of L-asparagine to L-β-OH-asparagine, which is subsequently activated by the corresponding A-domain and incorporated into CDA.[24], [25] In lipopeptide antibiotic friulimicin produced by *Actinoplanes friuliensis*, two 2,3-diaminobutyric acids are observed. Gene disruption and feeding compensation experiments showed that DabA similar to cysteine synthase and DabB similar to a fusion protein containing a putative ligase and an argininosuccinate lyase should be involved in biosynthesis of 2,3-diaminobutyric acid prior to activation by the NRPS A-domain.[26]

Methylation is also a common modification observed in NRPs. A SAM-dependent catechol 4'-O-methyltransferase (SafC) is involved in synthesis of a 4'-O-methyl-L-dopa precursor in the biosynthesis of the antitumor agent saframycin.[27] There is also a number of C-methylations observed in NRPs exemplified by a glutamate mutase delivering β-methylaspartate in friulimicin biosynthesis.[28]

Some NRPs contain *allo*-threonine or *allo*-isoleucine, which has an altered chirality at β-carbon, exemplified by enduracidin and Coronatine. Studies of the corresponding A-domain specificities suggested that these amino acids are epimerized prior to incorporation into the final product[29], [30]. It was also reported that certain NRPSs utilize external racemase to provide a D-monomer prior to activation by the corresponding A-domain in NRPS exemplified by cyclosporin biosynthesis.[31]

1.1.2.2 Tailoring enzyme acting on PCP-bound substrates

Additional to the precursor synthesis, modifications are also observed occuring on PCP-tethered intermediates catalyzed by catalytic domains embedded in NRPS assembly or additional enzymes acting *in trans*. The former is exemplified by well studied N-methylation domains and epimerization domains, the latter is observed in case of some hydroxylations, halogenations and α-β-desaturations.

Epimerization. D-amino acids are found in many NRPs and are not only important for the conformational stability and biological activity of NRPs, but also contribute to the protection of the peptide against proteolytic degradation. Addition to the C/E-dual domain and external Racemase described above, a more common strategy to generate D-amino acids is utilizing an in NRPS assembly embedded epimerization (E)-domain located downstream after the adjacent PCP precisely in the module responsible for incorporating D-isomer into the growing peptide chain. E-domains show similarity with C-domains with a size of 450 amino acids.[4] E-domains in initial modules catalyze the racemization and produce rapidly an equilibrium mixture of PCP-*S*-L,D-monomer. The specific incorporation of the D-isomer is established by the D-specific selectivity of the donor site of the downstream C-domain.[32] For E-domains embedded in elongation modules, it was shown that the epimerization does not occur at the L-monomer-*S*-PCP stage before condensation with the upstream peptidyl-*S*-PCP, but rather at the peptidyl-*S*-PCP stage before the condensation with downstream monomer-*S*-PCP. The adjacent downstream C-domains are expected to have a D-specific selectivity for donor site peptide chirality.[15]

Methylation. Methylated amino acids can be found in many NRPs, most of them are introduced by methyltransferases (MT), which catalyze the transfer of a methyl group from S-adenosylmethionine (SAM) on carbon, nitrogen or oxygen atoms on backbone of NRPs. Therefore, MTs are subclassified into C-MT, N-MT and O-MT depending on the site of methylation. Studies showed that N-MT is typically 450 amino acids long, while C- and O-MT normally have a size of 300 amino acids. Generally, MT show a bidomain structure with the N-terminal domain responsible for methyl-donor (SAM) binding and the C-terminal domain responsible for methyl-acceptor binding. Normally, MT-domains embed in the C-terminal region of an A-domain and contain three highly conserved motifs.[33, 34] An outstanding example of N-methylated NRP is cyclosporin, a cyclic peptide consisting of 11 amino acids, 7 of which are N-methylated. The corresponding

seven modules of cyclosporin synthetase show typical order of C-A(MT)-PCP. N-methylation occurs on the aminoacyl-*S*-PCP intermediate prior to condensation.[35] O-methylation is observed in the biosynthesis of kutznerides and perthamide.[36, 37] C-methylation was reported in the biosynthesis of yersiniabactin and melithiazol.[38, 39]

Formylation. Formylation of NRPs was reported in linear gramicidin produced by *Bacillus brevis* ATCC 8185[40] and anabaenopeptilides produced by *Anabaena* strain 90 [41]. Formylation of the N-terminal amino group is catalyzed by a formylation (F)-domain located at the N-terminus of the corresponding synthetase with a size of about 200 amino acids. F-domains catalyze the transfer of a formyl group from formyltetrahydrofolate (fH_4F) on the α-amino group of the amino acids using both cofactors N10- and N5-fH_4F. It was reported, that the F-domain of linear gramicidin synthetase LgrA catalyzes the α-N-formylation of PCP-bound L-Val or L-Ile. It was also observed that the formylation of the starter unit is necessary for initiation of gramicidin biosynthesis.[40]

Heterocyclization. In some NRPs, heterocycles such as thiazoline, oxazoline or methyloxazoline are observed. These structural features result from the heterocyclization of cysteine, serine or threonine side chains catalyzed by cyclization (Cy)-domains, which are structurally and mechanistically related to C-domains.[4] The Cy-domain conducts in the first step a nucleophilic attack of the amino group of the acceptor site aminoacyl-*S*-PCP onto the thioester of the donor site peptidyl- or aminoacyl-*S*-PCP, respectively, resulting in peptide bond formation. Subsequently, the Cy-domain catalyzes in the second step the nucleophilic attack of the sulfhydryl group of cysteine or the hydroxyl group of serine or threonine onto the carbonyl group of the peptide bond and delivers a thiohemiaminal or hemiaminal intermediate, which is then dehydrated to yield the thiazoline or oxazoline ring.[17] These heterocycles improve the structural diversity and backbone rigidity of NRPs and are important for chelating metals or interaction with proteins, DNA or RNA.[42]

Oxidation and Reduction. The oxidation state of heterocycles can be altered in NRPSs using oxidation or reduction domains. In bleomycin, epothilone or myxothiazol synthetases, oxidation (Ox)-domains comprising approximately 250 amino acids were observed.[43-45] Ox-domains were found located within A-domains or after PCP-domains as observed in myxothiazol synthetase.[45] Studies showed that the Ox-domain needs flavine-mononucleotide (FMN) as cofactor and molecular oxygen to reoxidize reduced FMN.[43] NADPH-dependent reduction (R)-domains were observed in yersiniabactin and pyochelin synthetase, which are responsible for reduction of heterocycles.[46]

Hydroxylation *in trans*. In addition to the hydroxylation occurring on free monomers described above, hydroxylation on PCP-tethered substrates was also observed. A family of heme protein hydroxylases exemplified by NovI and NikQ catalyzing hydroxylation of PCP-tethered Tyr and His residues were observed. KtzO and KtzP in the family of non-heme iron dioxygenases involved in Kutznerides biosynthesis catalyze hydroxylation of PCP-tethered L-Glu substrates and generate *threo*- and *erythro*-diastereomers of β-hydroxy-glutamate respectively.[47]

Halogenation *in trans*. Several NRPs contain halogenated residues, whereas chlorination is most prevalent. In biosynthesis of syringomycin (*Pseudomonas syringaea*), the nonheme Fe(II)/α-KG-dependent halogenase SyrB2 catalyzes the γ-chlorination of L-Thr-*S*-PCP intermediate.[48] The non-heme Fe(II)-dependent halogenase KtzD catalyzes the γ-chlorination of L-Ile-*S*-PCP intermediate and delivers PCP-bound γ-chloroisoleucyl intermediate.[49]

1.1.2.3 Post assembly tailoring

Post assembly tailoring refers to modifications of the NRPs after release from the assembly line such as glycosylation and oxidative cross-linking. Antibiotics in the vancomycin group undergo oxidative cross-linking, which are probably catalyzed by cytochrome-450-type heme proteins. In vancomycin and chloroeremomycin, three cross-links are formed, while a fourth cross-link is formed in the biosynthesis of

taicoplanin.[50] Through these covalent connections, the flexible peptides are converted into rigid aglycone scaffolds which enables the five-hydrogen-bond network to N-acyl-D-Ala-D-Ala termini of peptidoglycan strands.[51] Recent research showed that the cross-linking takes place on the peptide tethered on the last PCP-domain of the assembly.[52]

Glycosylation belongs to group transfer modifications. The addition of a monosaccharide and its iterative elongation to an oligosaccharide increases significantly the diversity of natural products and their water solubility.[52] Occasionally, common hexoses such as glucose and mannose are added to the peptide. More often deoxy- and deoxyaminosugars are attached to the peptides providing sites for hydrogen bond or further tailoring of the hydroxyl or amino groups.[53, 54] Two examples are the glycosyltransferases GtfD and GtfE, which construct the L-vancosaminyl-1,2-D-glucosyl disaccharide attached to the phenolic hydroxyl of residue 4 of the cross-linked heptapeptide aglycone core in the biosynthesis of vancomycin.[50]

1.1.3 Mechanisms of peptide release

The essential domains discussed above are repeating units contributing to linear peptide synthesis by adding monomer after monomer to the growing peptide chain tethered to the enzyme. The mature peptide has to be released to regenerate the synthetase. In most cases the peptide release is conducted by a thioesterase (TE)-domain located at C-terminus of NRPS.[55] TE-domains contain approximately 250 amino acids and show high degree of sequential and structural homology to α/β-hydrolases. Structural studies of the TE-domain in the surfactin synthetase, SrfC-TE, revealed the formation of the active site consisting of a catalytic triad Ser-His-Asp, which leads to deprotonation of the active site serine in the core motif GxSxG (Figure 1.5).[56] Nucleophilic attack of the resulting serine oxyanion onto the peptidyl thioester tethered to the upstream PCP generates a peptidyl-*O*-TE oxoester intermediate.[4] Following nucleophilic attack of water or of intramolecular nucleophile results in hydrolysis or in the generation of a

cyclic peptide.[57] In the case of formation of a cyclic peptide, it was suggested that an α-helical lid region shields the peptide binding pocket from solvent to prevent hydrolysis. An open conformation of lid region allows access to the active site, while in a closed conformation the substrate is cyclized.[58], [59] Another model (edge-on binding mechanism) suggested that the substrate itself shield the active serine from solvent molecule and thus prevent the hydrolysis.[59]

TE-domains are catalytically independent subunits. Cyclization of chemically synthesized peptidyl-*S*-*N*-acetylcysteamines[60], which mimic the peptidyl-S-Ppant-PCP substrate, and

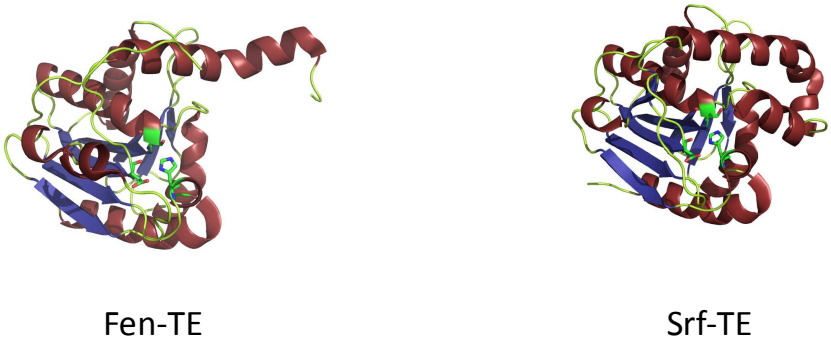


Figure 1.5: Structure of the Fen-TE and Srf-TE. (PDB-codes: Fen-TE: 2CB9; Srf-TE: 1JMK) The catalytic triads are presented in stick form.

peptidyl-S-thiophenols, which offer a reactive leaving group due to delocalization of the thiolate electrons throughout the aromatic ring[61], using excised TE-domains was employed to study their substrate specificities. Studies showed that the nucleophile and electrophile positions are most critical for substrate recognition.[62], [63] There are also TE-domains showing relaxed substrate specificities exemplified by the TE-domain from tyrocidin synthetase TycC TE, that showed tolerance to substitution of the nucleophile and peptide chain length.[64], [65] These relaxed substrate specificities of TE-domains offer a possible access to new structurally diverse NRPs.

In several NRPSs, unusual tandem-TE architecture was reported to be observed at the C-terminal end of the synthetases, including Arthrofactin-, Massetolide A- and

Lysobactin-synthetase[66, 67]. It was reported that inactivation of the first TE-domain could totally abolish the production of NRP, while inactivation of the second TE-domain decreased the production significantly.[66]

In addition to the TE-domain, some NRPSs employ R-domain or C-domain for product release. The NADPH-dependent R-domains at the C-terminus in LgrD and MxcG for the biosynthesis of linear gramicidin and myxochelin were reported to reduce the peptide chain tethered to the adjacent PCP to an alcohol and through that release the product from the synthetase.[68, 69] In cyclosporine synthetase it was reported that a C-terminal C-domain conducts a head-to-tail cyclization of the mature peptide tethered on the upstream PCP to release the final product[70].

1.1.4 Related enzymes (Sfp and TE II)

In addition to the described essential and modifying subunit of NRPS and the modifying enzymes acting *in trans*, there are still enzymes playing a central role in the biosynthesis of NRPs. In this chapter, these enzymes are discussed in detail.

1.1.4.1 Ppant transferase

All acyl-carrier or peptidyl-carrier proteins in PKS or NRPS systems contain an essential cofactor 4'-phosphopentetheine to bind the growing peptide chain or monomeric building block covalently.[71] Phosphopantetheinyl transferases catalyze posttranslationally the transfer of this cofactor from Coenzyme A onto the highly conserved serine residue in ACP or PCP and thus activate these enzymes. PPTases can be classified into two groups, namely acyl carrier protein synthase type PPTases and Sfp type PPTases, which activate ACPs from PKSs and NRPS/PKS multienzymes respectively[71]. Sfp type PPTase is named after the PPTase Sfp in surfactin biosynthesis in *Bacillus subtilis*, which exhibits extraordinarily broad substrate specificity [72]. The crystal structure of Sfp in complex with CoA and Mg^{2+} showed that the pantetheinyl part of CoA does not interact with the Sfp (Figure 1.6), which could explain the broad substrate

specificity of Sfp.[71] Sfp was shown to be able to activate a wide set of PCPs and accepts many different CoA-derivatives, ranging from acyl- or aminoacyl-CoA to bulky peptidyl- or fluoresceinyl-CoA. This property was widely used to generate artificial loading of PCPs. This method was applicated in many works to misprime PCPs intentionally for evaluation of C-domains, on-line tailoring enzymes etc.[73-76]

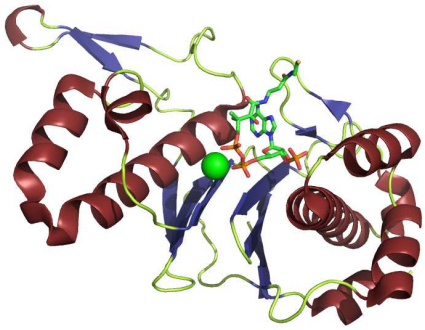


Figure 1.6: Crystal structure of Sfp in complex with Mg^{2+} and CoA. (PDB-code: 1QR0) CoA is shown in stick form and Mg^{2+} ion is represented with the green ball.

1.1.4.2 Type II thioesterase

In addition to the TE-domain at C-terminus of synthetases discussed above responsible for product release, a second type of TE-domain (TE II) was found in PKS or NRPS systems encoded normally by a distinct gene. Gene disruption experiments showed that these TE-domains are important for efficient production of the natural products.[76] Further studies showed that TE II could efficiently regenerate misacylated thiol groups of Ppant cofactor tethered on PCPs or ACPs.[76, 77] This editing function is vital, because ca. 80% of CoA is acetylated in bacteria, which could lead to possible misacylation of the ACPs and PCPs and block the synthetase. Figure 1.7 shows the structure of surfactin TE II exhibiting a typical α/β-hydrolase fold. The active site residues of the TE II (Ser_{86}, Asp_{189} and His_{216}) are found on the surface of the enzyme and are more accessible comparing to that of the TE-domain. This explains the relaxed substrate specificity of TE II.

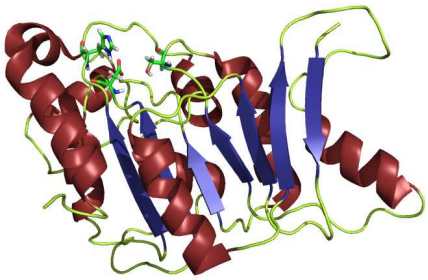


Figure 1.7: Crystal structure of the type II thioesterase surfactin TE II. (PDB-code: 2RON) The catalytic triad is shown in stick form.

1.2 Lysobactin

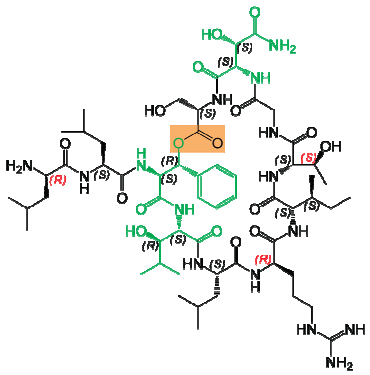
Antibiotic resistance caused by misuse or overprescription of antibiotics has become a serious threat to public health. The rise of multiresistant pathogens like methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *enterococci* (VRE) emphasizes the urgent requirement for the development of new antibiotics to ensure therapeutic efficiency against multiresistant pathogens in the future.[10, 78-81]

Lysobactin, also known as katanosin B, is one of the recently discovered potential antibiotics. This depsipeptide is produced by *Lysobacter sp*. ATCC 53042, which was first isolated at the Squibb Institute of Medical Research.[82, 83] This chapter gives detailed information of lysobactin, including its structure, biosynthesis and mode of action.

1.2.1 Structure and activity

Lysobactin is a branched cyclic depsipeptide consisting of 11 amino acid residues and features a 9-membered macrolactone ring with two N-terminal exocyclic residues.[84] The peptide core is composed of a set of non-proteinogenic amino acids including three β-hydroxylated amino acids ($hyPhe_3$, $hyLeu_4$ and $hyAsn_{10}$), two D-configured amino acids

(D-Leu$_1$ and D-Arg$_6$), and *allo*-Thr$_8$, that are shown in green and red in Figure 1.8, respectively. The macrolactone ring is constructed by forming an ester bond between the nucleophile β-hydroxyl group of hyPhe$_3$ and C-terminal carboxyl group of Ser$_{11}$, which is highlighted in orange.



D-Leu$_1$-Leu$_2$-hyPhe$_3$-hyLeu$_4$-Leu$_5$-D-Arg$_6$-Ile$_7$-aThr$_8$-Gly$_9$-hyAsn$_{10}$-Ser$_{11}$

Figure 1.8: Structure and primary sequence of lysobactin. The β-hydroxylated amino acids in the structure are highlighted in green, the chiral Cα-atom of the D-configured amino acids are labelled in red, the ester bond catalyzed by the C-terminal TE for the macrocyclization is highlighted with an orange rectangle.

Lysobactin shows a very strong activity against gram-positive bacteria as MRSA and VRE, with minimum inhibitory concentrations (MIC) of 0.39 and 0.78 µg/mL respectively, which are obviously lower (2 to 50-fold for different strains) than that of vancomycin.[85] It was shown that lysobactin inhibits the incorporation of *N*-acetylglucosamine into peptidoglycan of *Staphylococcus aureus* at concentration comparable to its MIC-value. *In vitro* studies showed that lysobactin inhibits the formation of lipid intermediates and nascent peptidoglycans with IC$_{50}$s of 2.2 and 0.8 µg/mL, respectively.[85] Vancomycin, a transglycosylation inhibitor widely used in the treatment of infections caused by Gram-positive bacteria, does inhibit the formation of nascent peptidoglycan but not the formation of lipid intermediates. Acetyl-Lys-D-Ala-D-Ala, an analog of the terminus of the lipid intermediates, was shown to suppress the inhibition of transglycosylation by vancomycin effectively, while no obvious suppress of that by lysobactin was observed. These observations suggest that lysobactin differs in the mode of action from

vancomycin and is considered as a high potential agent for the treatment of bacterial infections caused by resistant pathogens.[86]

1.2.2 Biosynthesis and organic synthesis of lysobactin and its derivatives

The macrocyclic structure and non-proteinogenic residues of lysobactin suggest that it could be a NRPS-product. In 1996 it was first confirmed that a NRPS is involved in the biosynthesis of lysobactin using hybridization of genomic libraries of *Lysobacter sp.* ATCC 53042 with oligonucleotides derived from core-motifs of ACV synthetases and the gramicidin S synthetase (*Bacillus brevis*).[1] Lysobactin non-producing mutants were generated via marker-exchange mutagenesis employing a 4.6 kbp NPRS-encoding DNA-fragment. This confirmed the identified DNA fragment to be part of the lysobactin synthetase. Further bioinformatic analysis of the 4.6 kbp DNA fragment revealed that it codes a truncated tetradomain NRPS with C-A-PCP-C organization.[1]

The main drawbacks of lysobactin are its higher toxicity compared to vancomycin[87] and lability in basic or neutral aqueous solution due to hydrolysis of ester bond in macrolactone structure.[88] Various synthetic pathways were elaborated trying to generate lysobactin derivatives with improved pharmaceutical properties. Some of them employed an altered macrocyclization strategy,[84-86] such as the macrolactam derivative of lysobactin, which has a much higher stability under neutral or basic pH-value.[88]

1.3 Polyketide synthase

Polyketide is one of the most remarkable classes of natural products showing a wide range of structural and functional diversity (Figure 1.9).[89] Polyketides obtain broad spectrum of medicinally important activities including antibacterial (tetracycline[90], rifamycin[91]), antifungal (amphotericin b[92], monensin[93]), antitumor (bleomycin[94, 95]), antiparasitic (avermectin[96]), immunosuppressant (rapamycin[97]) and cholesterol-lowering (lovastatin[98]). From a structural point of view, polyketides have structurally interesting carbon skeletons comprising polyphenols, macrolides, polyenes, enediynes,

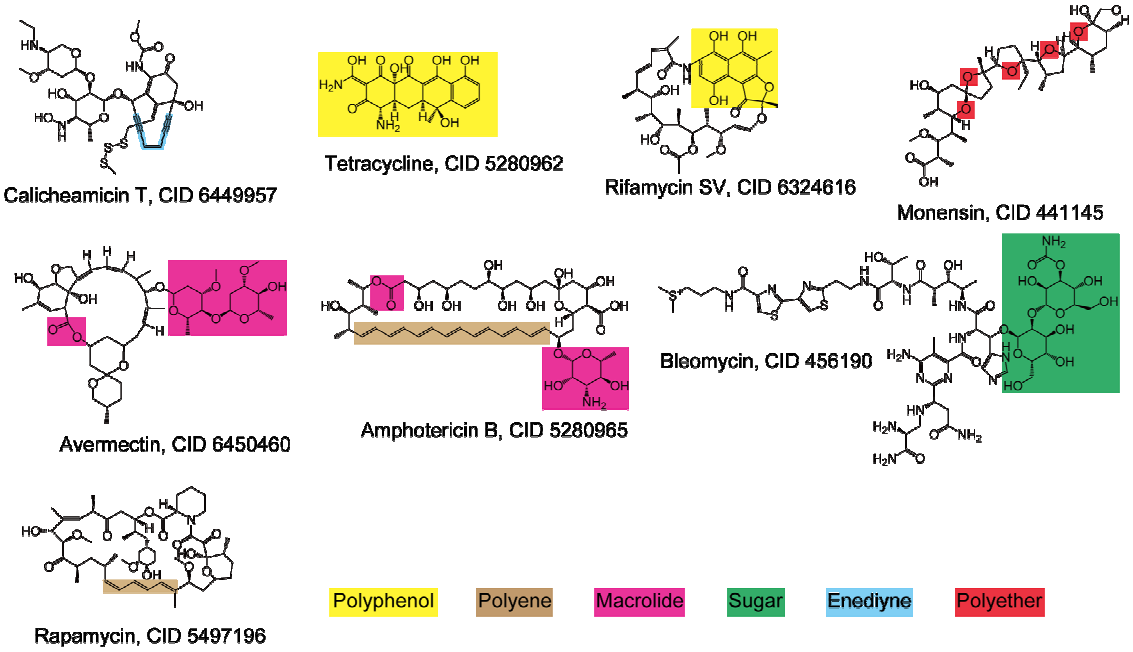


Figure 1.9: Structural diversity of polykeitde. Structural features found in polyketides are highlighted in different colours.

and polyethers.[99] The development of understanding the polyketide-biosynthesis began with Collie's pioneering work[100] in late 19th century and improved by Robinson with his "acetogenin" hypothesis.[89] The first widely accepted biosynthesis pathway was raised by Birch and Donovan in 1953.[101] Closely related to fatty acid biosynthesis (s. Figure 1.10[99]), polyketides are synthesized by repeated Claisen thioester condensations of an activated acyl starter unit with malonyl-CoA-derived extender units.[99] The biosynthese of fatty acids and polyketides show striking homology not only in the chemical mechanism of chain elongation, but also through utilizing simple precursors for the synthesis of complicated structure. However, the biosynthesis of polyketides differs in many ways from that of fatty acids. The major difference is that after the condensation of the extender unit, the fatty acid synthase catalyzes the full reduction of the β-keto group in the carbon chain, while the reduction in polyketide biosynthesis is optional and could be fully or partially omitted, resulting in a highly functionalized chain (s. Figure 1.10). Additionally , the broader range of utilized building blocks and the resulting highly

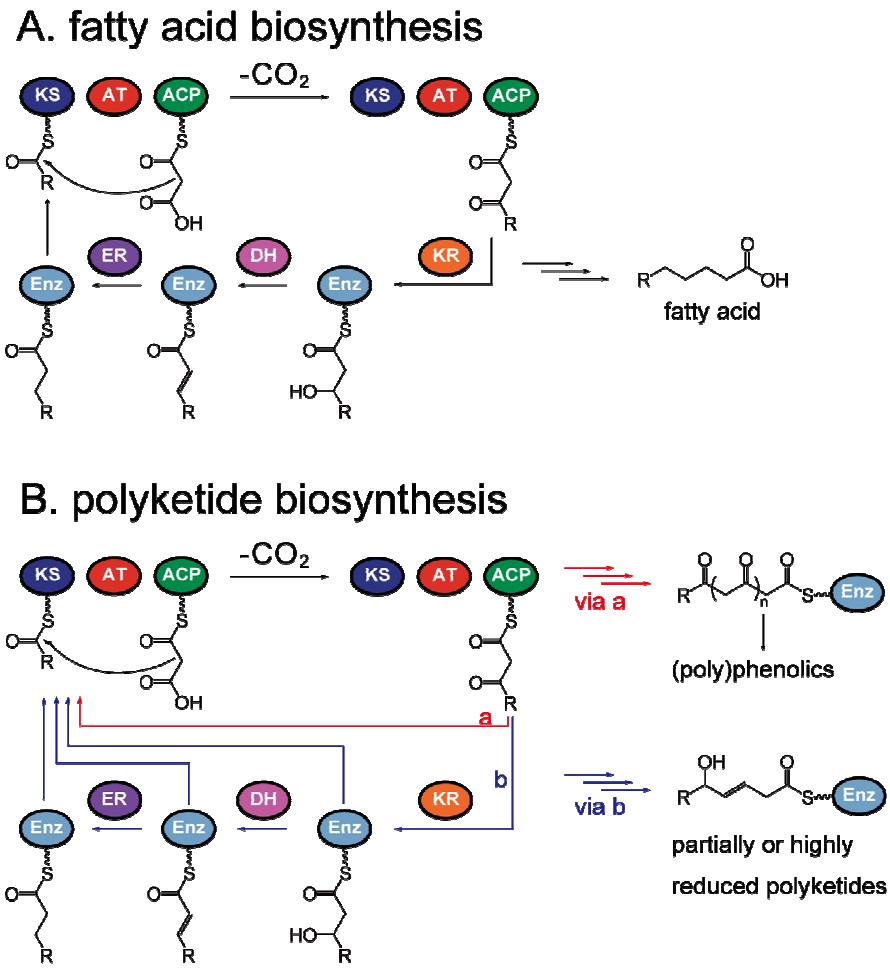


Figure 1.10: Mechanisms of fatty acid and polyketide biosynthesis. A: biosynthesis of fatty acids. B: biosynthesis of polyketides. KS: Ketosynthase, AT: Acyltransferase, ACP: Acyl carrier protein, KR: Ketoreductase, DH: Dehydratase, ER: Enoylreductase, Enz: Enzyme.

diverse structures of the mature product are also obvious differences between fatty acid and polyketide biosynthesis.[99]

The biosynthesis of polyketides can be divided into 3 steps: starting, elongation and termination. In the starting step, the AT-domain in the starter module catalyzes the loading of the starter group, usually acetyl-CoA or malonyl-CoA, onto the ACP-domain in the starter module. In the elongation step, the acetyl or malonyl group on the starter ACP-domain or the polyketide chain on the ACP-domain of the previous module is transferred onto the KS-domain of the current module, where the transfer is catalyzed by the KS-domain itself. Then the current AT-domain catalyzes the loading of the next

building block onto the current ACP-domain. After that, the ACP-bound building block reacts with the KS-bound polyketide chain in a Claisen condensation reaction releasing one molecule of CO_2. This makes the elongated polyketide chain move one step forward and set the KS-domain in the current module for the next round of elongation free. After the condensation, the polyketide chain undergoes optional modifications: reduction of the β-keto group to a β-hydroxy group catalyzed by the KR-domain, subsequent dehydration catalyzed by the DH-domain resulting an α,β-unsaturated intermediate, and reduction of the double bond to a single bond catalyzed by the ER-domain. After these optional modifications, the polyketide chain is ready for the next elongation cycle. In the termination step, the polyketide chain is released from the ACP-domain via macrocyclization or hydrolysis catalyzed by the TE-domain after reaching the appropriate chain length.[6] Based on the organization and mode of action of the enzymes, PKSs are classified into different types:[102] PKS-type I, II and III, which will be discussed in detail in the following section. (s. Table 1.1).

Table 1.1: Classification of PKS-types.[99]

PKS type	Iterative or non-iterative	Organisms
Type I	Non-iterative	Bacteria
	Iterative	Mainly fungi, some bacteria
Type II	Iterative	Bacteria
Type III	Iterative	Mainly plants, some fungi and bacteria
PKS/NRPS hybrid	Non-iterative	Bacteria
	Iterative	Fungi

Type I PKS. Type I PKSs are large enzyme complexes consisting of linearly arranged catalytic domains, which are covalently fused together. As shown in Table 1.1, type I PKSs can be subdivided into two classes, namely iterative and non-iterative type I PKSs. Non-iterative type I PKS are mainly found in bacteria and iterative type I PKS are mainly

found in fungi.[99] Non-iterative PKSs can be exemplified by 6-deoxyerythronolide synthase (DEBS) shown in Figure 1.11, which is mainly found in prokaryotes.[103] The

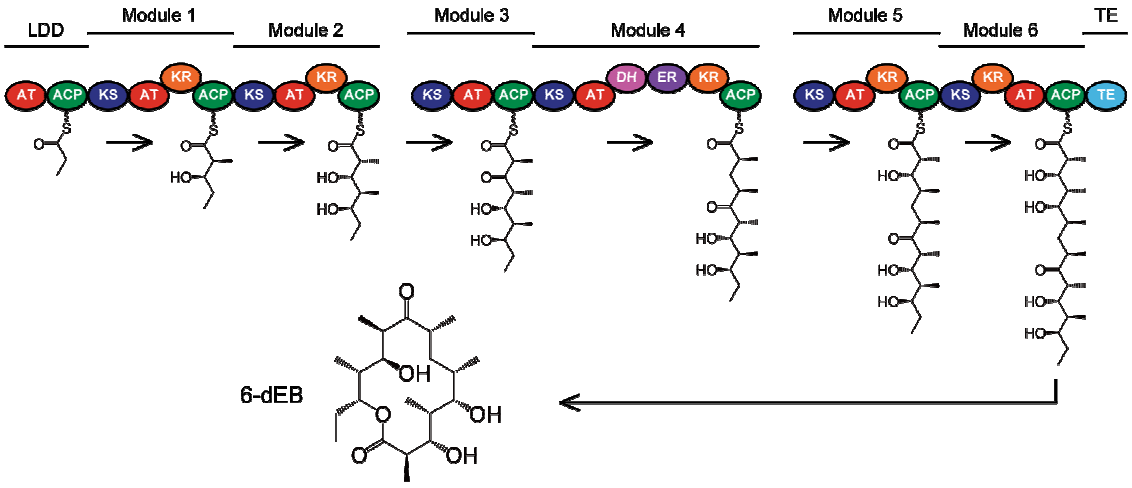


Figure 1.11: The 6-deoxyerythronolide-B-synthase (DEBS) responsible for erythromycin biosynthesis exemplifies non-iterative type I PKSs.

loading didomain (LDD) contains one AT-domain and one ACP domain and is responsible for initiating the biosynthesis by using propioyl-CoA for the loading reaction. Other modules are comprised of essential KS, AT and ACP domains and the optional β-keto processing enzymes DH, ER and KR domains located between the AT and ACP domains, that determine the reduction degree of the β-keto groups. The TE-domain at the C-terminal end of the last synthase is responsible for macrocyclization and product release. Generally, each module is responsible for the incorporation of only one building

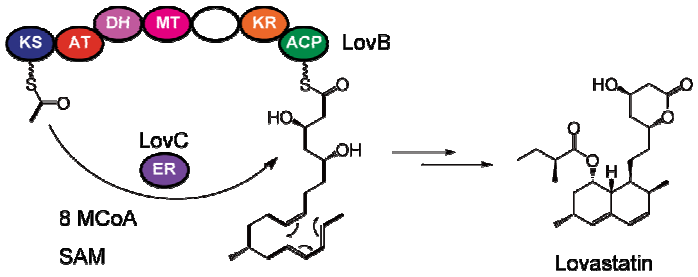


Figure 1.12: The lovastatin synthase LovB and LovC responsible for lovastatin biosynthesis in *Aspergillus terreus*. MCoA: malonyl-CoA, SAM: *S*-adenosylmethionone.

block and thus the number of modules found in the synthases correlates with the number of elongation cycles. This one-to-one correlation between the number of PKS modules and structure of the natural product is known as the colinearity principal,[104] which allows the prediction of metabolite structure from the enzyme organization and *vice versa*.[99]

Iterative type I PKSs are mainly found in fungi exemplified by the anticholesteremic agent lovastatin isolated from a strain of *Aspergillus terreus*.[98, 105] As shown in Figure 1.12,[99] lovastatin synthase has a covalent architecture, which is characteristic for type I

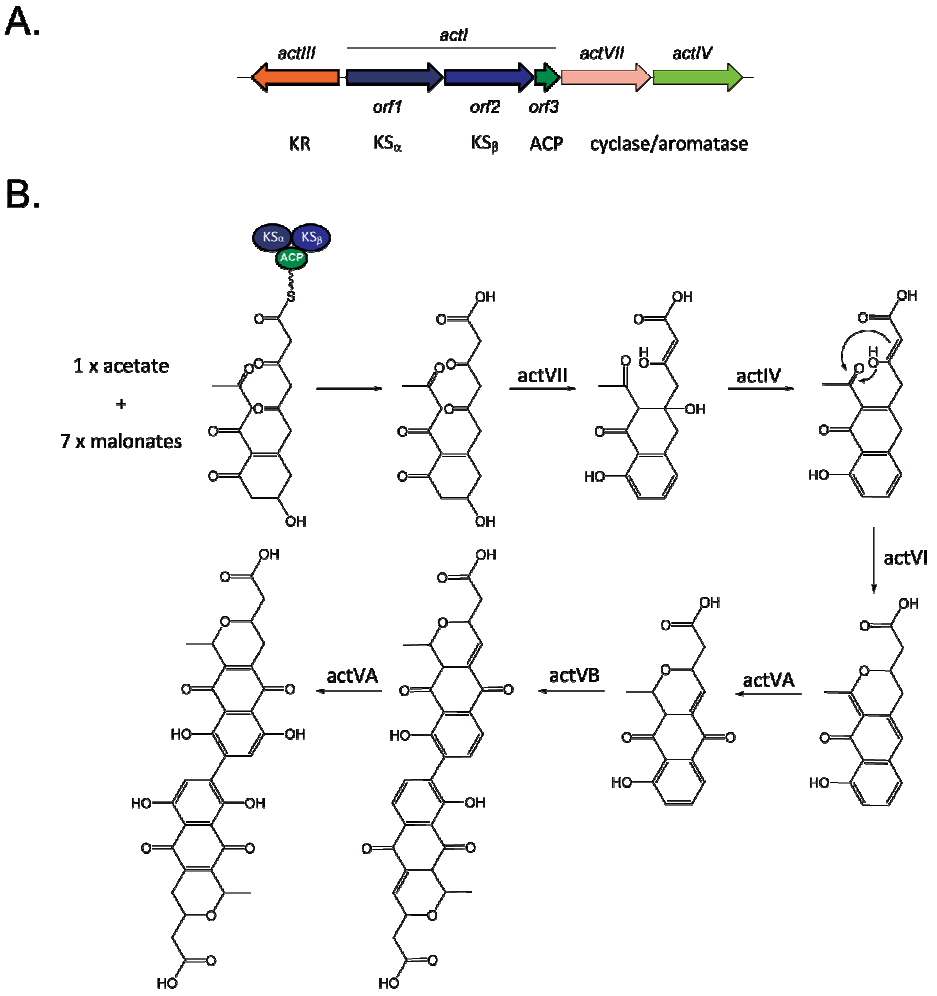


Figure 1.13: The gene cluster and predicted pathway of actinorhodin biosynthesis. A. Organisation of the act genes in Streptomyces coelicolor. B. The proposed biosynthesis pathway of actinorhodin.[89]

PKS. However, the catalytic domains are used repeatedly to incorporate the building blocks into the final product. Thus, the iterative PKSs do not obey the colinearity principal. The molecular basis of this complex programming is largely unknown.

Type II PKS. In contrast to type I PKSs, type II PKSs refer to dissociated enzyme complexes of monofunctional, discrete enzymes.[99] These monofunctional enzymes are encoded by discrete genes and organized into a multifunctional complex in their active state.[89] Type II PKSs are observed exclusively in bacteria and mainly in the actinomycetes. One example is actinorhordin, the biosynthetic gene cluster of which was identified in 1984.[106] As shown in Figure 1.13, three distinct genes *orf* 1-3 located in the region *actI* encode KS_α, KS_β and ACP. Genes (or region) *actVII, actIV* and *actVB* are located on the downstream side of *actI*. Genes (or region) *actIII*, *actII*, *actVA* and *actVI* are located on the upstream side of *actI*. *ActII* is the so-called central regulatory region, the gene products of which were characterized to be responsible for antibiotic export and regulatory mechanism for the biosynthetic genes. The KS_α, KS_β and ACP encoded by *orf* 1-3 in region *actI* assemble the minimal PKS in the actinorhodin biosynthesis. Together with the discrete KR encoded by *actIII*, these enzymes are predicted to synthesize the polyketide backbone from 1 unit of acetyl-CoA and 7 units of extender malonyl-CoA. Gene products of *actIV*, $actV_A$, $actV_B$, *actVI* and *actVII* are predicted tailoring enzymes involved in actinorhodin biosynthesis.[107]

Type III PKS. The first discovered type III PKS was chalcone synthase (CHS), which is responsible for the biosynthesis of chalcone, a central core of many important biological compounds, from *p*-coumaroyl-CoA and three malonyl-CoA.[99] In contrast to type I and II PKSs, type III PKSs maintain a much less complicated architecture exemplified by the homodimeric structure of CHS containing two identical KS monomeric domains. It was shown that these homodimeric enzymes could catalyze acyltransferase, decarboxylation, condensation, cyclization and aromatization reactions in the two independent active

sites.[108] The relatively simple structure makes them amenable to *in vitro* examination and structural analysis.[109] Type III PKSs have long been discovered in plants, but in the last decade many type III PKS were also found in bacteria[110, 111] and fungi.[112] An interesting feature is that the type III PKSs from plants show high identity (60-95%) to each other, while the bacterial type III PKSs show only 25-50% identity to each other or to those from plants.[109]

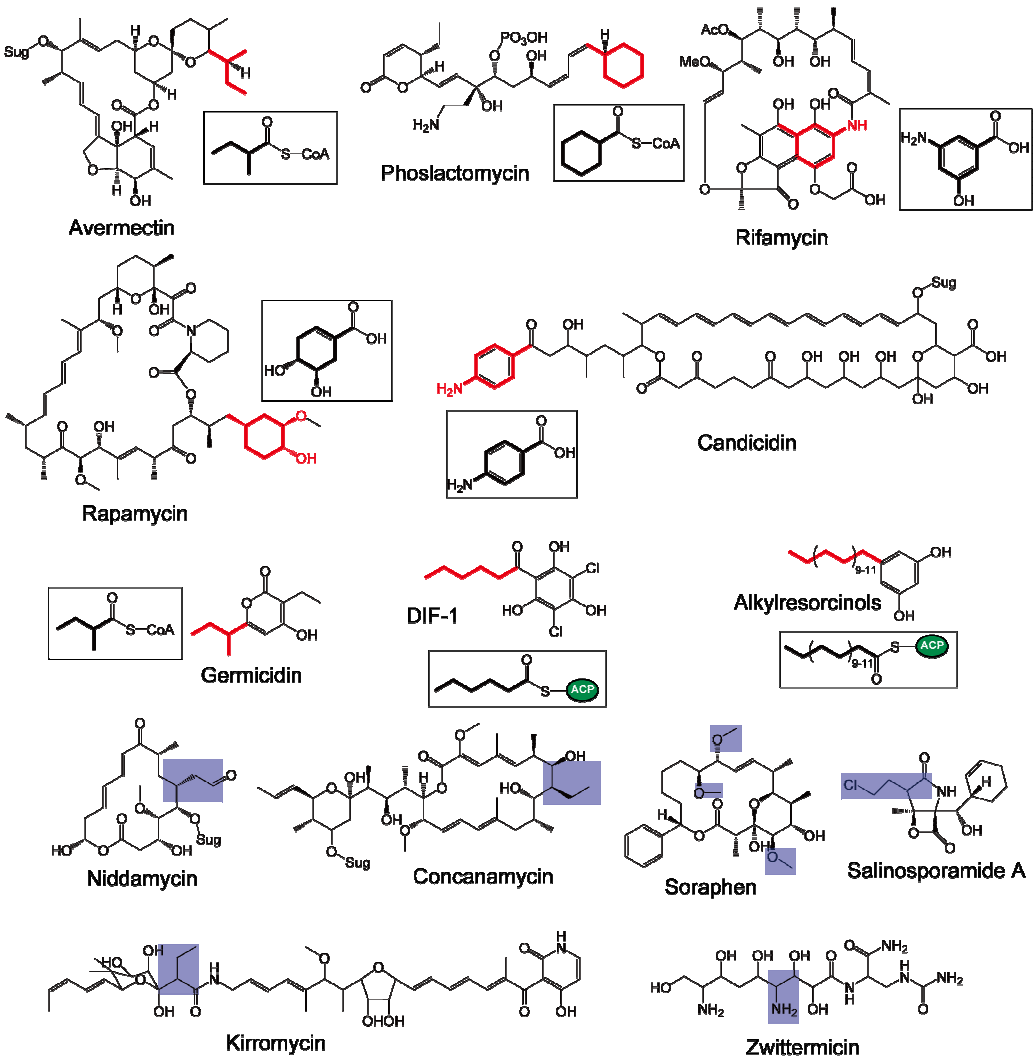


Figure 1.14: Type III PKSs using non-acetate starter units. M-CoA: malonyl-CoA, eM-CoA: ethyl-malonyl-CoA. Red: unusual building blocks in Polyketides with their non-acetate precursors. Blue: unusual malonyl-derived building blocks in polyketides.

Alternative building blocks

PKSs utilize routinely acetate/propionate as starter units and malonate/methylmalonate as extender units. However, unusual starter and extender units are observed in many cases. In some modular PKS systems, loading of unusual starter molecules is accomplished via a loading AT domain (AT_L). For example, isovaleryl-CoA is utilized as the starter in the biosynthesis of avermectin in *Streptomyces avermitilits*[113] and cyclohexanoyl-CoA is used as the starter in the biosynthesis of phoslactomycin.[114] If the starter is used as a free acid rather than a CoA thioester, a strategy similar to that of adenylation domains in NRPS, is utilized. In the biosynthesis of rapamycin, the starter unit, dihydroxycyclohexene carboxylic acid, is activated via an A-domain and loaded onto the adjacent downstream ACP domain.[115] The same strategy is observed in the biosynthesis of rifamycin and candicidin, which use 3-amino-5-hydroxybenzoic acid and *p*-aminobenzoate as starter units, respectively.[116, 117]

Type III PKSs also utilize a broad range of non-acetate starter units such as cinnamoyl derivatives (*p*-coumaroyl, caffeoyl and feruloyl in chalcone biosynthesis), benzoyl (biphenyl synthase) and fatty acids (unsaturated e.g. anacardic acid,[118] branched e.g. germicidin[119]).[99, 109] Transfer of building blocks from FAS to type III PKS was also reported exemplified by the biosynthesis of differentiation-inducing factor, DIF-1, via a type I FAS/type III PKS hybrid system in *Dictyostelium discoideum*,[120] and the biosynthesis of alkylresorcinols via “crosstalk” between an unusual type I FAS and a type III PKS,[121] as shown in Figure 1.14.[99]

In bacterial modular PKSs, MCoA or mMCoA are usually utilized as extender units. Other extender units are only rarely observed in bacterial PKS biosynthesis. 2-ethylmalonyl-CoA is used as an extender unit in the biosynthesis of the niddamycin,[122] concanamycin and kirromycin.[123] In the biosynthesis of FK506, 2-propenylmalonate is observed to be utilized as an extender unit. Other than alkylated malonylate extender units described above, heterosubstituted malonyl derivatives are observed in

polyketides such as hydroxyl and methoxy substitutions in soraphen,[124] FK520[125] and concanamycin[126] and aminomalonate in zwittermycin[127]. A halogenated extender unit was also observed exemplified by the incorporation of chloroethylmalonate into salinosporamide.[128]

1.4 PKS/NRPS hybrid

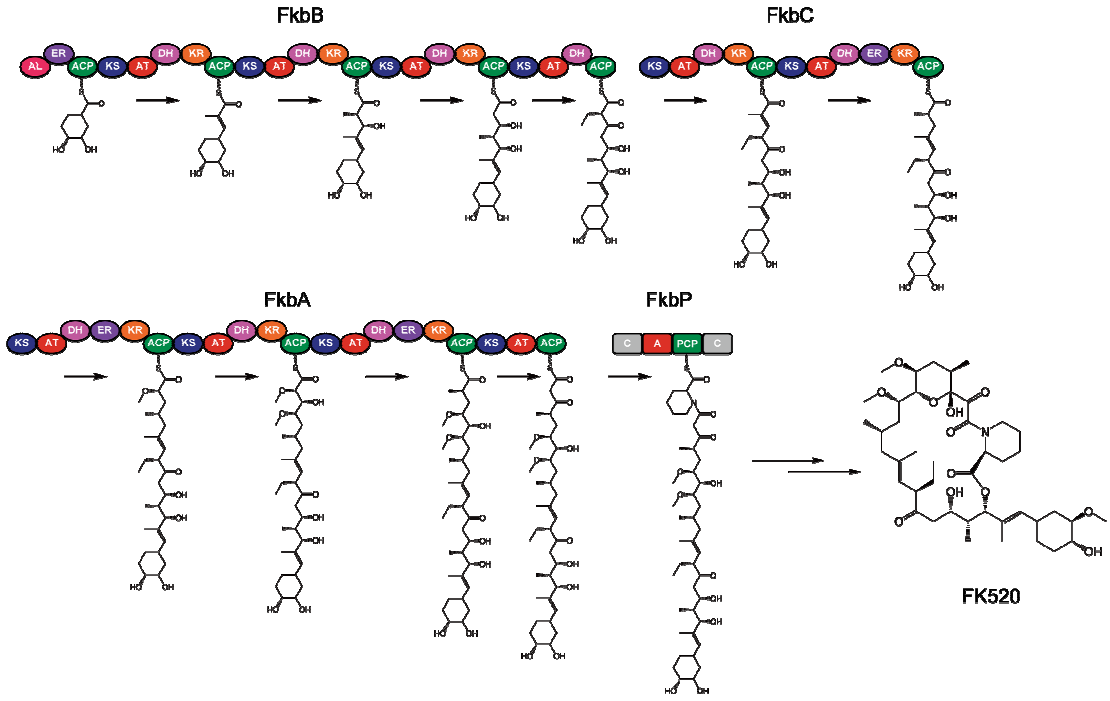


Figure 1.15: Structure and biosynthesis of the macrolide FK520.

By combination of the NRPS and PKS assembly lines, nature has developed a further strategy to synthesize complicated molecules. Some examples of PKS/NRPS hybrid assembly and the corresponding products are discussed in this chapter. The hybrid assemblies consist of different portion of NRPS and PKS machinery. Some assemblies consist of mostly PKS machinery, which is exemplified by FK520. FK520 is a macrolide produced by *Streptomyces hygroscopicus* var. ascomyceticus (ATCC 14981) that has immunosuppressive, neurotrophic and antifungal activities.[129] As shown in Figure 1.15[129], the FK520 assembly consists of 4 proteins: FkbB, FkbC, FkbA and FkbP. The first

PKS FkbB starts with a loading module which specifies a shikimate-derived starter units[130] followed by four extender modules. The second and the third PKS FkbC and FkbA consist of two and four extender units, respectively. The fourth protein in the assembly is FkbP, a NRPS, which activates pipecolic acid and catalyses the condensation of the intermediate synthesized by PKSs with the nitrogen of the pipecolic acid. After release from FkbP, the PK-NRP hybrid chain undergoes further hydroxylation catalyzed by FkbD, oxidation catalyzed by FkbO and methylation catalyzed by FkbM.[129]

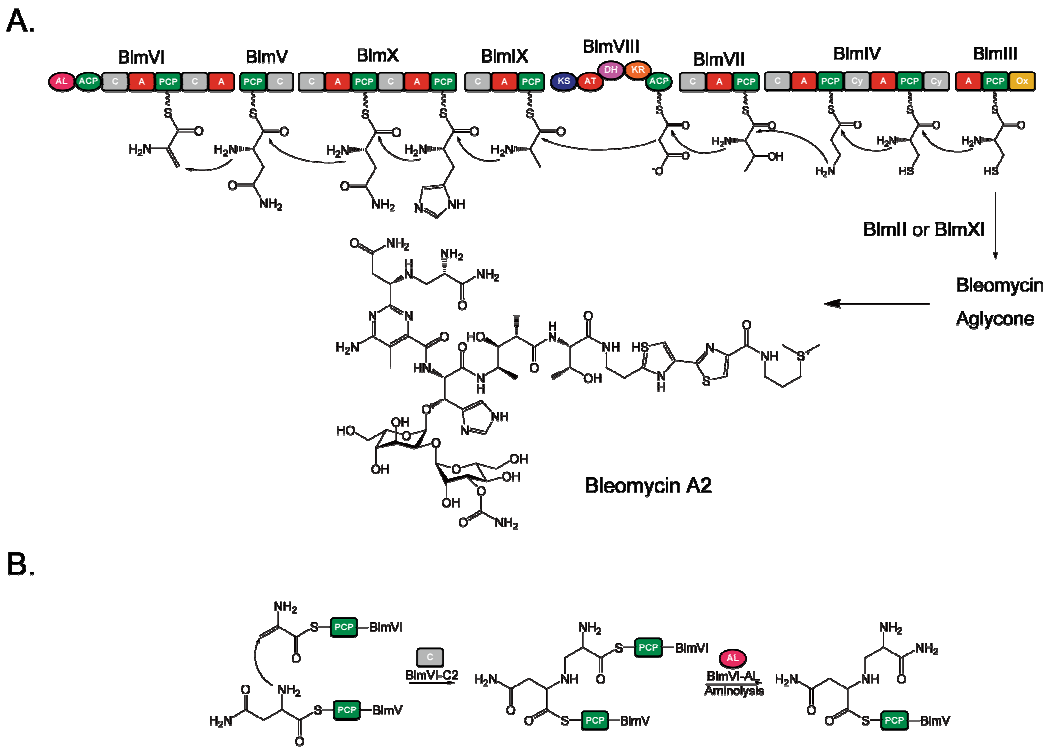


Figure 1.16: Structure and assembly of antitumor agent Bleomycin A2. A.: Structure and biosynthesis of Bleomycin A2. B.: The proposed mechanism of formation of the unusual β-aminoalaninamide moiety.

In contrast to FK520, the assembly of bleomycin consists of 10 NRPS modules distributed over 7 proteins (BlmIII, IV, V, VI, VII, IX and X) and 1 PKS module (BlmVIII) (s. Figure 1.16).[95, 131] Bleomycin is a natural hybrid peptide-polyketide metabolite produced by *Streptomyces verticillus* ATCC15003 and shows antitumor activity.[95] An unique feature observed in the bleomycin assembly is that BlmVI is a NPRS equipped with N-terminal

acyl CoA ligase domain (AL) followed by ACP-like domain. It was proposed that this AL in the starter module catalyzes the aminolysis of the acyl-*S*-PCP-BlmVI intermediate and delivers an unusual β-aminoalaninamide moiety (shown in Figure 1.16).[95]

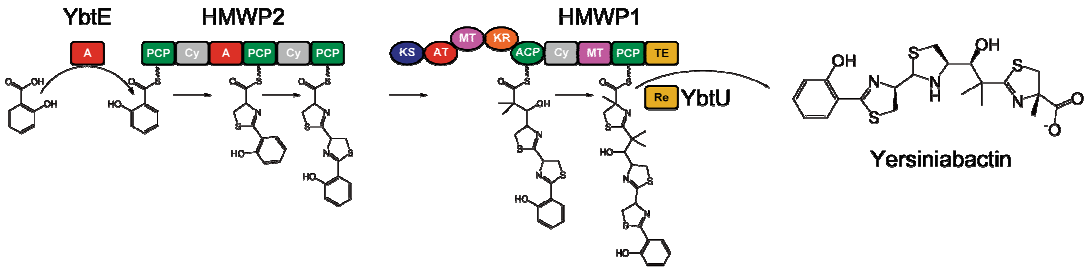


Figure 1.17: Structure and assembly of the siderophore yersiniabactin.

Yersiniabactin is a siderophore produced by different *Yersinia* species. Yersiniabactin synthetase comprises four proteins, YbtE, HMWP1, HMWP2 and YbtU, which can be subdivided into seventeen domains (s. Figure 1.17).[132] One molecule of salicylate, three cysteines and one malonyl moiety are activated and incorporated to build the final product.[132] YbtE is suggested to activate salicylate and load it onto the first PCP-domain of the NRPS HMWP2. HMWP2 is responsible for elongating the growing chain with two cysteines and cyclizes the two cysteines to two thiazoline rings. HMWP1 is a PKS/NRPS hybrid protein and contains 5 PKS-domains followed by 4 NRPS-domains. The PKS-part of HMWP1 is responsible for incorporating a further C_2-unit to the elongating chain using malonyl-CoA and and catalyzing the two methylation on the C_α-position. The NRPS-part of HMWP1 is responsible for incorporating a Cys buildingblock to the growing chain followed by methylation on the C_α-position.[132-134] The Cys residue is cyclized to thiazoline ring catalyzed by the upstream Cy. The middle thiazoling ring is then reduced by the NADPH-dependent reductase YbtU *in trans* to a saturated thiazolidine ring. In the last step, the mature product is released from the assembly line via hydrolysis catalyzed by the C-terminal TE-domain.

2. Objectives of this study

Different NRPSs were reported to employ tandem TE-domain for release of the final product. *In vivo* study of the tandem TE-domain in arthrofactin biosynthesis machinery showed that inactivation of the first TE-domain totally abolished the production of arthrofactin, while inactivation of the second TE-domain reduced the production remarkably. It was suggested that the second TE-domain could be added during the evolution in order to improve the macrocyclization efficiency. However direct proof of this inference was still absent. In this study, genomic DNA of *Lysobacter sp.* ATCC 53042 was sequenced and the entire biosynthetic gene cluster of lysobactin was identified and characterized, which also employs tandem TE-domain architecture for the cyclization and release of the final product. *In vitro* characterization of the individual thioesterases revealed the role of both TE-domains in lysobactin biosynthesis.

Bioinformatically analyze of genomic DNA of *Lysobacter sp.* ATCC 53042 delivered several other NRPS or PKS/NRPS gene clusters. These genes were bioinformatically analyzed and their potential natural products were predicted.

3. Materials

3.1 Chemicals, enzymes and consumables

Chemicals that are not listed below in the table were purchased from other manufacturers as standard compound in p.a. quality.

Table 3.1: Microorganisms, chemicals, enzymes and general materials

Chemical	Product Manufacture
[1-^{14}C]-acetyl-CoA	Amersham
2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate	Sigma-Aldrich
2-chlorotritylchlorideresin	Novabiochem
2,5-dihydroxybenzoic acid	Agilent Technologies
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	Roth
^{32}P-$Na_4P_2O_7$	Perkin-Elmer
Agar Roth 2266.2	Roth
Adenosine triphosphate	Applichem
Ampicillin	AppliChem
BBL™ Trypticase™ soy broth BD211768	BD
Chloroform	Roth
Coomassie Brilliant Blue R250	GE Healthcare
Dimethyl sulfoxide	Merck
Dimethylformamide	Acros
Ethidium bromide	Roth
Ethylenediaminetetraacetic acid	Serva Feinbiochemica
Endonucleases	NEB
Hi-Trap desalting column	GE-Healthcare
Hydroxybenzotriazole	Iris biotech GmbH
Isopropyl β-D-1-thiogalactopyranoside	GE-Healthcare
Isopropanol	Roth
Kanamycin	AppliChem
L-threo-3-phenylserine	Bachem
Liquid scintillation fluid	Roth
Methanol (HPLC-Grade)	Fisher Scientific
Ni-NTA	Qiagen
N-Fmoc-protected amino acids	Novabiochem
Oligonucleotides	Sigma-Aldrich

Perchloric acid	Riedel-Häen
Phusion polymerase	Finnzymes
Piperidine	Merck
Protein size markers	Fermentas
QIAquick gel extraction kit	Qiagen
T4 DNA ligase	NEB
Triisopropylsilane	Acros
amicon® Ultra centrifugal filter	Millipore

3.2 Equipments

Table 3.2: Equipment used in this thesis

Equipment	Manufacturer and type
Autoclave	Fedegari Autoclavi SPA
Cell density meter	Amersham Biosciences Ultrospec 10
Centrifuges	Thermo Scientific Sorvall RC5B Plus; Heraeus Biofuge pico; Heraeus Megafuge 1.0R
Clean Bench	Antair BSK
Fluidizer	Avestin EmulsiFlex-C5
FPLC	Pharmacia FPLC system
French-press	SLM Aminco French pressure cell press
HRMS system	Thermo Fischer Scientific LTQ-FT/Agilent 1100 HPLC system
Incubator	Infors HT Multiron
LC-MS system	Agilent series 1100 HPLC-system
Liquid scintillation counter	PerkinElmer
Lyophilization	Christ Alpha 2-4 LSC
MALDI-TOF	Bruker BiFlex III
NMR	Bruker AV 600
Peptide synthesizer	Advanced ChemTech Apex 396 synthesizer

pH meter	KOBE Seveneasy Mettler Toledo
Photometer	PEQLab Nanodrop ND-1000; Pharmacia Ultrospec 3000
Pipettes	Eppendorf Reference
Speedvac	Eppendorf concentrator 5301
Thermal cycler	Eppendorf Mastercycler Personal
Thermomixer	Eppendorf Thermomixer comfort
Vortexer	Neolab vortex mixer
Water deionizer	Seral seralpur pro 90 CN

3.3 Plasmid vectors

3.3.1 pET-28a(+)

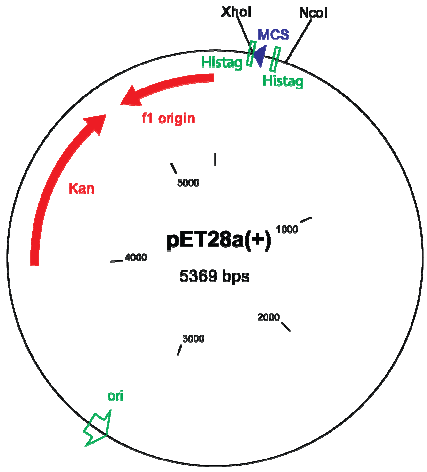


Figure 3.1 Physical map of pET28a(+)

The pET-28a(+) vector (*Novagen*) is used as a general expression vector for heterologous production of recombinant proteins. Transcription of the gene of interest is carried out by T7 RNA-polymersase and induced by IPTG. The pET-28a(+) vector provides the possibility to introduce N- or C-terminal His_6-tag upon in-frame ligation of the cloned genes, which enables the purification of recombinant protein using Ni-NTA affinity chromatography.

3.3.2 pCR®-XL-TOPO®

Topo® XL (*Invitrogen*) is a cloning vector supplied in linearized form bearing single 3'-T overhang and covalently bound topoisomerase. It offers a quick, ligase-free cloning of PCR-product bearing 3'-A overhang with high efficiency.

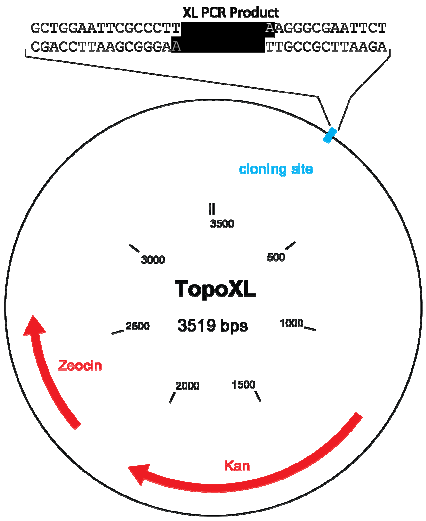


Figure 3.2 Physical map of Topo® XL

3.4 Oligonucleotides

Primers used for vector construction were purchased from Sigma-Aldrich and are list below in Table 3.3.

Table 3.3 Primers used in this study

Gene of interest	Primer name	Sequence	Restriction site
LybA-A$_1$	LybA-A$_1$ FP	AAAAAACCATGGtgagaacggaccaccaccgt	NcoI
	LybA-A$_1$ RP	AAAAAACTCGAGgtccggcgccggcaag	XhoI
LybB-A$_6$	LybB-A$_6$ FP	AAAAAACCATGGtcgattacttcaaggcgctgctg	NcoI
	LybB-A$_6$ RP	AAAAAACTCGAGgaaattgcgctgcgcgtagg	XhoI
LybB-A$_9$	LybB-A$_9$ FP	AAAAAACCATGGcgcaacgccacgagcaattg	NcoI
	LybB-A$_9$ RP	AAAAAAAAGCTTctcgtagccgcgctgcgcata	HindIII
LybB-PCP-TE$_1$	LybB-PCP FP	AAAAAACATATGTATGCGCAGCGTCCGTTC	NdeI
	LybB-TE$_1$ RP	AAAAAAGGATCCgttcaacagcccggcgatgg	BamHI
LybB-PCP-TE$_1$*-TE*$_2$	LybB-PCP FP	AAAAAACATATGTATGCGCAGCGTCCGTTC	NdeI
	LybB-TE$_2$ RP	AAAAAAGGATCCtcatggcgactcggcttgggt	BamHI

LybB-TE$_1$	LybB-TE$_1$ FP	AAAAAACATATGcatctggtgccgatccgctt	NdeI
	LybB-TE$_1$ RP	AAAAAAGGATCCgttcaacagcccggcgatgg	BamHI
LybB-TE$_2$	LybB-TE$_2$ FP	AAAAAACATATGaccgccgcgatcaccgtgca	NdeI
	LybB-TE$_2$ RP	AAAAAAGGATCCtcatggcgactcggcttgggt	BamHI
LybB-C-A-PCP$_8$	LybB-E FP	AAAAAACCATGGctgagccgcaagtcgcgtcc	NcoI
	LybB-PCP$_8$ RP	AAAAAAAAGCTTcaggaaccacaggcgcagctg	HindIII
pstD-PCP-TE	Friu-PCP-TE FP	AAAAAACATATGCAGTCGGCGGAGGGCCGG	NdeI
	Friu-PCP-TE RP	AAAAAAGGATCCTCAAACGCGGCCGCTGCGCAG	BamHI
pstD-TE	Friu-TE FP	AAAAAACATATGGTCCTGCCCCTGCGGACC	NdeI
	Friu-TE RP	AAAAAAGGATCCTCAAACGCGGCCGCTGCGCAG	BamHI

3.5 Microorganisms

3.5.1 One Shot® TOP10 Electrocomp™ *E coli*

One Shot® TOP10 Electrocomp™ *E coli* is a generally used host for cloning and plasmid propagation with transformation efficiency of 1X10^9 cfu/µg supercoiled DNA, allowing stable replication of high-copy number plasmids. The genotype of TOP10 cells is F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ-.

3.5.2 *E. coli* BL21 (DE3)

E. coli BL21 (DE3) (*Novagen*) is one of the most widely used host for expression of plasmide DNA. This strain has the genotype F$^-$ *omp*T *hsd*S$_B$ (r$_B^-$, m$_B^-$) *gal* [*dcm*] [*lon*] λ(DE3) and is deficient both in *lon* protease and *omp*T outermembrane protease, which reduces the proteolytic degradation and thus increases the expression level of the target protein. Furthermore this strain contains IPTG inducible T7 RNA polymerase gene, which is essential for the IPTG induction of genes under T7-promotor control.

3.5.3 *E. coli* Rosetta 1 DE 3

E. coli Rosetta 1 DE 3 strain (*Novagen*) is also used for expression of plasmid DNA. Like BL21, this Rosetta strain is also deficient both in *lon* and *ompT* proteases. Moreover, this strain supplies tRNAs for the rare codons AUA, AGG, AGA, CUA, CCC and GGA on a compatible chloramphenicol-resistant plasmid, which makes this strain capable of expression of DNA containing these rare codons, which would be otherwise limited by codon usage of *E. coli*. The genotype of this strain is F^- *ompT* *hsd*S_B($r_B^-m_B^-$) *gal dcm*(DE3) pRARE.

3.5.4 *Lysobacter sp.* ATCC 53042

Lysobacter sp. (ATCC 53042), a Gram-negative bacterium, is the lysobactin producing strain. It was cultivated for the subsequent preparation of genomic DNA.

3.6 Culture media

The media listed below were used for cultivation of microorganisms. Culture plate medium was prepared by adding 1.2% (by LB Medium plate) or 1.5% (by ATCC broth #18) to the medium followed by heating at 121°C and 1.5 bar for 30 min. After cooling down to 55°C, antibiotics in the following standard concentrations were added.

Kanamycin:	30 µg/mL
Chloramphenicol:	34 µg/mL

3.6.1 LB-medium

LB-medium was used for cultivation of *E. coli* strains.

Yeast-extract:	5 g/L
Bactotrypton:	10 g/L
NaCl:	5 g/L

3.6.2 ATCC #18 soy broth

ATCC #18 broth was used for cultivation of Lysobacter sp. ATCC 53042 strain.

Trypticase Soy Broth (BD 211825): 30 g/L

4. Methods

4.1 Molecular biology techniques

4.1.1 Cultivation of *lysobacter sp.* ATCC 53042, fermentation and isolation of lysobactin

Lysobacter sp. ATCC 53042 was purchased from LGC standards as freeze-dried culture sample. It was revived by resuspending little amount freeze-dried powder in 50 µL ATCC #18 soy broth, which was subsequently transferred to an ATCC #18 soy broth agar slant followed by incubation at 30°C for 2-3 days. The colonies were transferred to ATCC #18 soy broth and grown at 30°C for 48 hours. ATCC #18 soy broth was inoculated with the starter culture with a ratio of 1:100 and incubated at 30°C for 40 hours. After centrifugation of the culture at 7,000 rpm for 30 min, the supernatant was separated and extracted 3 times with butanol after adjusting the pH-value to 7.0. The solvent was removed using rotary evaporator after combination of the extracts to yield a yellow solid. Lysobactin was purified and separated via HPLC (*Agilent* 1100 system, Macharey-Nagel, VP 250/21 Nucleodur C18 HTec) with the following gradient: solvent A: water (0.1% TFA), solvent B: acetonitrile (0.1% TFA), 0 min, 10% B; 30 min, 60% B; 33 min, 95% B; 35 min, 95% B; 38 min, 10% B; 45 min, 10% B.

4.1.2 Genomic DNA preparation

2 mL cell culture of *lysobacter sp.* ATCC 53042 was centrifugated at 13,000 rpm for 3 min. The cell pellet was washed once with water and resuspended in 500 µL lysis buffer. Glass beads were added to the resuspension to reach a final volume of 1.25 mL. The mixture was vortexed for 2 min and the liquid was separated. After adding 275 µL 7 M ammonium acetate solution (pH 7.0), the mixture was incubated at 65°C for 5 min and subsequently 5 min on ice. After adding 500 µL chloroform, the mixture was centrifuged at 13,000 rpm for 5 min. The aqueous phase was transferred to a new reaction tube.

Genomic DNA was precipitated by adding 1 mL isopropanol. After incubating at RT for 5 min, the mixture was centrifuged at 13,000 rpm for 5 min. The pellet was washed with 200 µL 70% ethanol and dried at RT. The genomic DNA was dissolved in 50 µL H_2O and stored at 4°C.

Lysis buffer:

Tris:	100 mM
EDTA:	50 mM
SDS:	1%
pH 8.0	

4.1.3 Plasmid preparation

2 mL liquid cell culture of *E. coli* was centrifuged at 13,000 rpm for 3 min. After resuspension of the cell pellet in 300 µL resuspension buffer, 300 µL lysis buffer was added. The mixture was gently inverted 6 to 8 times and incubated for 5 min at RT. The cell debris was precipitated by adding 300 µL neutralization buffer. The mixture was gently inverted 6 to 8 times and incubated for 10 min on ice subsequently. After centrifugation at 13,000 rpm for 10 min, the supernatant was transferred to a new reaction tube and mixed with 600 µL isopropanol. After centrifugation at 4°C, 13,000 rpm for 30 min, the resulted plasmid DNA pellet was washed with 200 µL 70% ethanol and dried at 37°C for 30 min. The plasmid DNA was solubilized in 50 µL H_2O and stored at -20°C.

4.1.4 Construction of expression plasmids

Amplification of the interested genes from genomic DNA was performed by polymerase chain reaction (PCR) using Phusion™ High-Fidelity DNA Polymerase (*Finnzymes*) according to the protocol for GC-rich templates from manufacturer. The PCR products were purified via agarose gel electrophoresis and recovered using QIAquick gel

extraction kit (*Qiagen*) following the manufacturer's protocol. Digestion of the PCR products and corresponding plasmids was performed using restriction endonucleases (*NEB*). Subcloning was performed using T4 DNA ligase (*NEB*) following the manufacturer's protocol. Ligations were transformed into One Shot® TOP10 Electrocomp™ *E. coli* cells via electroporation. Transformants were plated on LB-agar slants containing the corresponding antibiotic(s). The plasmids were isolated as described in section 4.1.3 and verified by restriction mapping and dideoxy sequencing (*GATC-biotech*). Detailed information of the expression plasmids used in this study is listed in the following table.

Table 4.1 Plasmids used in this study

Plasmid	Gene	Primers	Restriction sites	Protein
pET28a(+)(*lybA-A_1*)	*lybA-A_1*	LybA-A_1 FP LybA-A_1 RP	NcoI/XhoI	lybA-A_1 (C-Histag)
pET28a(+)(*lybB-A_6*)	*lybB-A_6*	LybB-A_6 FP LybB-A_6 RP	NcoI/XhoI	lybB-A_6 (C-Histag)
pET28a(+)(*lybB-A_9*)	*lybB-A_9*	LybB-A_9 FP LybB-A_9 RP	NcoI/HindIII	lybB-A_9 (C-Histag)
pET28a(+)(*lybB-TE_1*)	*lybB-TE_1*	LybB-TE_1 FP LybB-TE_1 RP	NdeI/BamHI	lybB-TE_1 (N-Histag)
pET28a(+)(*lybB-TE_2*)	*lybB-TE_2*	LybB-TE_2 FP LybB-TE_2 RP	NdeI/BamHI	lybB-TE_2 (N-Histag)
pET28a(+)(*lybB-PCP-TE_1*)	*lybB-PCP-TE_1*	LybB-PCP FP LybB-TE_1 RP	NdeI/BamHI	lybB-PCP-TE_1 (N-Histag)
pET28a(+)(*lybB-PCP-TE_1-TE_2*)	*lybB-PCP-TE_1-TE_2*	LybB-PCP FP LybB-TE_2 RP	NdeI/BamHI	lybB-PCP-TE_1-TE_2 (N-Histag)
pET28a(+)(*lybB-E-A-PCP_8*)	*lybB-E-A-PCP_8*	LybB-E FP LybB-PCP_8 RP	NcoI/HindIII	lybB-E-A-PCP_8 (C-Histag)
pET28a(+)(*pstD-PCP-TE*)	*pstD-PCP-TE*	Friu-PCP-TE FP Friu-PCP-TE RP	NdeI/BamHI	Friu-PCP-TE (N-Histag)
pET28a(+)(*pstD-TE*)	*pstD-TE*	Friu-TE FP Friu-TE RP	NdeI/BamHI	Friu-TE (N-Histag)

4.1.5 Genome Pysosequencing

Genomic dsDNA was column purified and nebulized to fragments in a size of approximately 700 bp. The A- and B-adapters for sequencing with the Roche technology were ligated to the ends of the nebulized DNA fragments. The samples were run on a 2% agarose gel with TAE buffer and the band in a size range of 700 to 900 bp was excised and column purified. After concentration measurement the resulting library was immobilized onto DNA capture beads and the amplicon-beads obtained were amplified through emPCR according to the manufacturer's recommendations. Following amplification, the emulsion was chemically broken and the beads carrying the amplified DNA library were recovered and washed by filtration. The sample was sequenced on a GS FLX Pico-Titer plate device. The GS FLX produced sequence data as Standard Flowgram Format (SFF) file containing flowgrams for each read with basecalls and per-base quality scores.

4.2 Biochemical techniques

4.2.1 Protein expression

In this study, the proteins were expressed with the pET28a(+) system. 500 mL prewarmed (37°C) LB-medium containing corresponding antibiotics (30 µg/mL kanamycin, and additional 34 µg/mL chloramphenicol if Rosetta 1 DE3 cell was used) was inoculated with 5 mL overnight culture. The culture was incubated at 37°C and 220 rpm till the optical density (OD) reachs 0.6 (λ = 600 nm). After inducing with 0.1 mM IPTG, the culture was further cultivated for additional 16 h (or overnight) at 18°C. The cells were harvested by centrifugation (6,000 rpm, 4°C, 20 min), resuspended in 10 mL buffer (50 mM HEPES, 300 mM NaCl, pH 8.0) and either directly processed or stored at -20°C.

4.2.2 Protein purification

For purification of His_6-tagged recombinant protein, the cells were disrupted by the use of an EmulsiFlex®-C5 High Pressure Homogenizer (*Avestin*). Cell debris and insoluble components were precipitated through centrifugation (13,000 rpm, 4°C, 30 min). The supernatant was separated and the protein was purified via Ni^{2+}-NTA affinity chromatography using a FPLC system (*Amersham Pharmacia Biotech*). After equilibration of the column with the buffer HEPES A, the supernatant was applied onto the column with a flow-rate of 1 mL/min. The column was then washed with mixed buffer (90% HEPES A and 10% HEPES B) till the absorption of flowthrough at 280 nm reached baseline level. The His_6-tagged protein was then eluted by applying the following gradient of buffer HEPES A and B (0 min, 10% HEPS B; 30 min, 50% HEPES B; 40 min, 100% HEPES B) with a flow rate of 1 mL/min. Fractions containing the recombinant protein were monitored by SDS-PAGE, pooled together and concentrated using centrifugal filter (*Millipore*). After the concentrated protein was dialyzed against dialysis buffer using Hi-Trap™ desalting columns (*Amersham Biosciences*), the fractions were pooled and concentrated again with centrifugal filter. The recombinant protein can be either directly applied or stored at -80°C.

4.2.3 Protein concentration determination

The concentrations of the recombinant proteins were determined spectrophotometrically via NanoDrop using molar extinction coefficients which were calculated with the program “Protean” (*DNAStar*). The molar extinction coefficients of the recombinant proteins are listed below in Table 4.2.

Table 4.2 Calculated molar extinction coefficient of recombinant proteins used in this study

Protein	Molar extinction coefficient (λ=280 nm)
lybA-A_1 (C-Histag)	1.02 mg/mL
lybB-A_6 (C-Histag)	1.00 mg/mL
lybB-A_9 (C-Histag)	0.98 mg/mL
lybB-TE_1 (N-Histag)	0.91 mg/mL
lybB-TE_2 (N-Histag)	1.51 mg/mL
lybB-PCP-TE_1 (N-Histag)	1.03 mg/mL
lybB-PCP-TE_1-TE_2 (N-Histag)	1.14 mg/mL
lybB-E-A-PCP_8 (C-Histag)	1.13 mg/mL
Friu-PCP-TE (N-Histag)	1.57 mg/mL
Friu-TE (N-Histag)	1.27 mg/mL

4.3 Chemical synthesis

4.3.1 Synthesis of N_α-Fmoc-protected amino acids

L-*threo*-3-phenylserine hydrate (HyPhe) was purchased from Bachem. The primary amino group was protected with 9-fluorenylmethyloxycarbonyl (Fmoc) for subsequent solid phase peptide synthesis. 1 mmol L-*threo*-3-phenylserine (0.181g) was added to a vigorously stirred Na_2CO_3 solution (0.263 g in 5.7 mL H_2O) at 0°C. 2.76 mL 1,4-dioxane was added to the reaction mixture, forming an opaque mobile mixture. 1.05 mmol Fmoc-Cl (0.27 g) was dissolved in 3 mL 1,4-dioxane and dropped into the stirred solution over 40 min. Then the mixture was allowed to warm to ambient temperature followed by adding 26 mL H_2O. The reaction mixture was washed twice with 20 mL $CHCl_3$. The aqueous phase was acidified with HCl to pH = 2, providing a thick opaque mixture. After extraction with $CHCl_3$, the organic phase was combined, dried over Na_2SO_4 and evaporated in vacuo to yield a pale yellow solid residue (yield = 63%).

4.3.2 Solid phase peptide synthesis (SPPS)

The oligo peptide substrates used in this study were synthesized following a standard Fmoc-based solid phase peptide synthesis (SPPS) protocol. The synthesis was initialized by immobilizing the C-terminal building block of the peptide onto an insoluble polymer (2-chlorotritylchloride resin). Subsequently, the building blocks with protected α-amino group and side-chain were assembled to the immobilized initial building block. During the synthesis, the excess reagent and by-product could be easily separated via filtration. The peptide synthesis was carried out on an automated peptide synthesizer (APEX 396, *Advanced ChemTech*).

4.3.2.1 Intitiation

The initial step of the SPPS is to immobilize the C-terminal amino acid on resin. This step is essential for the final yield and purity of the peptide because unoccupied binding sites on the surface of resin at this step can be acylated in following steps and thus lead to truncated peptide products. Therefore excess amount of the initial building block is applied in order to saturate the binding sites on resin. After swelling of the 2-chlorotritylchloride resin in DCM, 2 eq. of C-terminal amino acid with protected α-amino group and side chain and 8 eq. DIPEA were added. The carboxyl functional group of the amino acid was deprotonated by the non-nucleophilic base and subsequently attacked the 2-chlorotritylcation, through which the first amino acid is attached to the resin (Figure 4.1). The reaction mixture was shaked at room temperature (2 h, 450 rpm). The solvents were removed by filtration and the resin was washed 3 times with DCM followed by incubation in MeOH to seal the remained active sites on the resin.

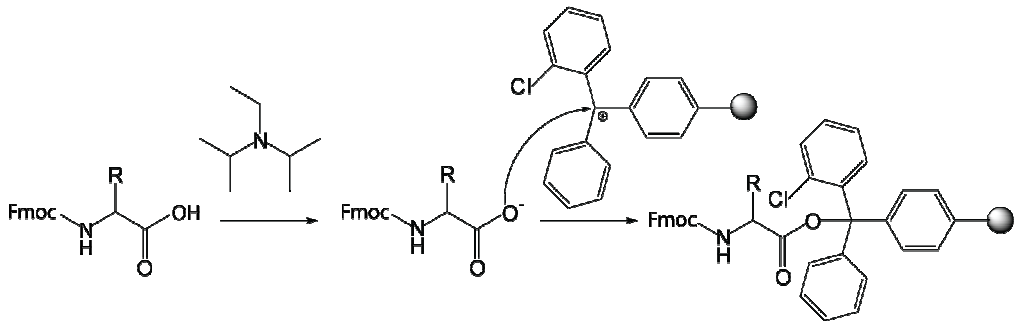


Figure 4.1 Initiation of SPPS. The C-terminal amino acid is loaded onto the 2-chlorotritylchloride resin.

4.3.2.2 Elogation

Before coupling with the next amino acid, the N-terminal Fmoc group was removed via incubation in 15%-20% piperidine in DMF for 20 min. The resulting aromatic cyclopentadiene-type intermediate rapidly eliminates to form the dibenzofulvene and carbon dioxide (Figure 4.2). The solid phase was washed with DMF to remove the excess of reagent and the by-products.

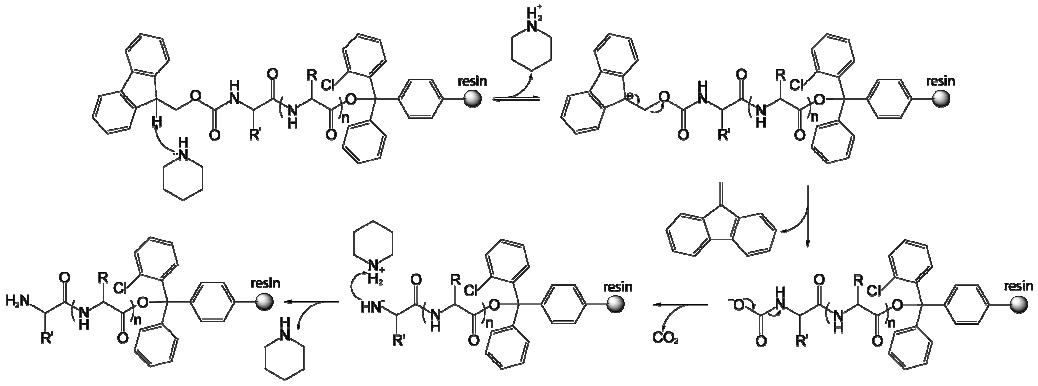


Figure 4.2 Deprotection of the N_{α}-Fmoc group.

In the next step, the N_{α}-Fmoc protected amino acid, HBTU, HOBt and DIPEA were added to the resin. The carboxyl group of the amino acid is deprotonated by the 10-fold excess of DIPEA and then attacked electrophilic carbenium ion of HBTU to generate a highly reactive tetramethylurea intermediate, which is converted into reactive benzotriazole ester in the presence of HOBt. The N-terminal amino group of the amino acid or peptide

tethered on resin attacks this benzotriazole ester to generate the elongated peptide chain. 3-fold excess of the N_α-Fmoc protected amino acid was applied to ensure a quantitative reaction. Excess of reagent and by-products were removed by washing the resin with DMF. The N-terminal Fmoc-group of the elongated peptide bound on resin can be deprotected again and couple with the next building block. The elongation reaction is repeated till the desired peptide is generated. As the N-terminal building block, N_α-Boc protected amino acids are normally used instead of N_α-Fmoc protected amino acids. Boc-group is acid-labile and can be deprotected together with the deprotection groups of the side chain in the subsequent reaction.

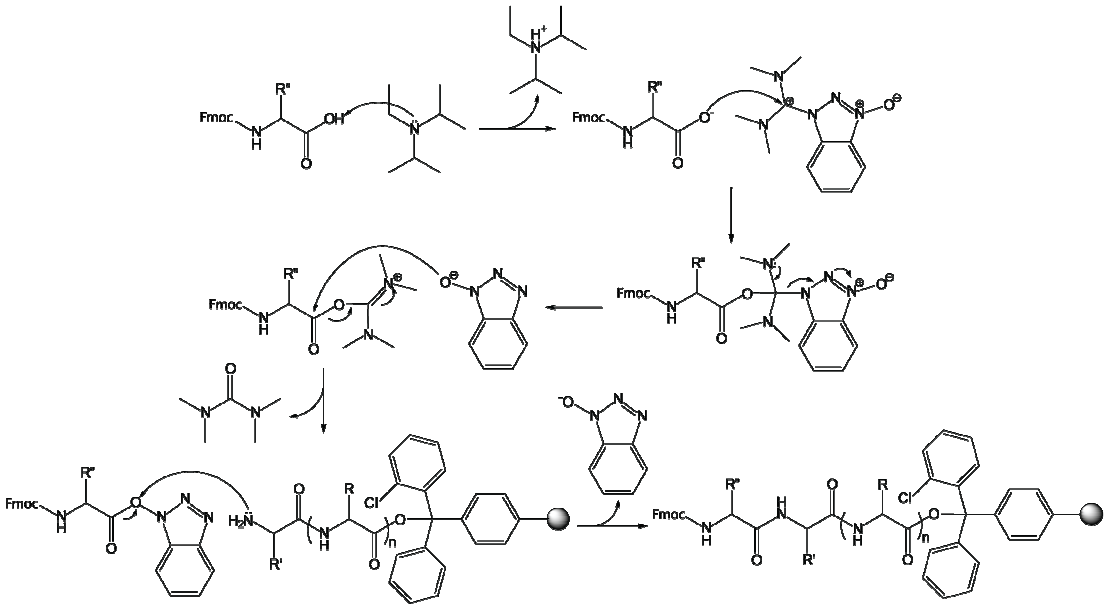


Figure 4.3 Peptide bond formation.

4.3.2.3 Termination

The mature peptide was cleaved off from the resin by incubation in the cleavage mixture DCM/TFE/AcOH (7:2:1) for 2 hours at 450 rpm. Under the applied mild reaction condition, the protection groups of the side chain are stable. The released peptide was separated from the resin via filtration and precipitated in *n*-Hexane. After removal of the solvents via rotary evaporator, the resulted white residue was stored at -20°C.

4.3.3 Synthesis of peptidyl thioester

1 eq. of side chain protected peptide, 2 eq. of HBTU and HOBt were dissolved in DCM. After addition of 10 eq. of N-acetylcysteamine or thiophenol, the reaction mixture was stirred at room temperature for 30 min. Catalytic amount of K_2CO_3 was added to the reaction mixture followed by further stirring for 2.5 hours at room temperature. After removal of the solvent via rotary evaporator, the white solid was dissolved in TFA/H_2O/TIPS (95:2.5:2.5) and stirred at room temperature for 2 hours to remove the side chain protection groups. The reaction mixture was then added to 30 mL ice-cold diethyl ether forming white precipitation. The precipitate was separated via centrifugation (4,000 rpm, 20 min), dissolved in DMSO and subsequently purified via semipreparative HPLC (*Agilent* 1100 system) with a reversed-phase column *Macherey and Nagel*, VP 250/21 Nucleodur C18 HTec) applying the following elution gradient: solvent A: H_2O (0.1% TFA), solvent B: MeCN (0.1% TFA), 0 min, 30% B; 30 min, 60% B; 33 min, 95% B; 40 min, 95% B; 42 min, 30% B; 50 min, 30% B.

4.4 Analytical methods

4.4.1 MALDI-TOF-MS

A Matrix Assisted Laser Desorption Ionization (MALDI) combined with a time-of-flight mass spectrometer (TOF) was applied to determine the molecular mass of peptides and proteins. 2,5-dihydroxybenzoic acid was employed as matrix. Sample preparing was achieved by pipetting 0.3 µL of peptide solution and 0.3 µL of matrix solution on certain position of a metallic probe target, the solution was well mixed and dried. The cocrystallized samples were then analyzed using a “Bruker FLEX III” (*Bruker Daltronic*).

4.4.2 HPLC-MS

High-performance liquid chromatography was used as a standard method to characterize substrate by retention time and mass. Reversed-phase column was mostly

applied, which bases on interaction between analyte and non-polar stationary phase. The non-polar stationary phase of reversed-phase column consists of carbon or alkyl chains (C_4, C_8 or C_{18}) immobilized on silica gel. Elution of compounds was performed with H_2O-MeOH or H_2O-MeCN containing 0.1% TFA or 0.05% formic acid (positive mode) or 0.1% TEA (negative mode). The relative unpolar organic solvents compete with analyte for binding positions, the acids or base added in the mobile phase were used as ion pair and enabled the ionization. The elution of compound was monitored with UV-detection. The HPLC was performed on an agilent 1100 system which is coupled with an eletrospray ionization mass detector which allows ionization and mass detection at atmospheric pressure.

4.4.3 HRMS and MS/MS-fragmentation analysis

High resolution mass spectroscopy and MS/MS-fragmentation were employed to analyze extracted natural products or products of enzymatic reactions. The compounds were purified via an Agilent 1100 HPLC system which was conneted with an LTQ-FT instrument (*Thermo Fisher Scientific*), allowing HRMS and MS/MS fragmentation analysis.

4.4.4 Protein mass fingerprinting

Protein mass fingerprinting was used to validate the recombinant proteins. The recombinant protein was analyzed via SDS-PAGE followed by excising of the protein band showing correct size. The excised gel was treated with 200 µL wash solution for 30 min at 56°C. After removing the supernatant, the gel was dried at 56°C for 30 min. In-band tryptic digestion was carried out by adding 20 µL trypsin solution to the dry gel. After initial incubation at 37°C for 45 min, the excess of trypsin solution was removed. The mixture was further incubated at 37°C overnight. Proteolytic cleaved peptide fragments were achieved by adding 25 µL diffusion solution and subsequent sonification for 45 min. The sample was analyzed using spray-HPLC-QTOF-MS system. Comparison with MASCOT database allows the identification of the protein.

Wash solution

NH_4HCO_3	200 mM
MeCN	50% v/v

Trypsin solution

trypsin	0.02 µL/µL
NH_4HCO_3	10%
MeCN	10%
pH 8.1	

Diffusion solution

TFA	1% v/v
MeCN	10% v/v
pH 8.1	

4.5 Spectroscopic methods

4.5.1 NMR-spectroscopy

^{1}H-NMR-spectroscopy was utilized for verifying the product of organic synthesis. 10 mg sample was dissolved in 0.7 mL $CDCl_3$. The ^{1}H one-dimensional spectrum was recorded at room temperature on Bruker AV 600 and processed with Bruker Topspin 2.1.

4.6 Biochemical assays

4.6.1 ATP/^{32}PPi-exchange assay

Analysis of the recombinant adenylation domains was carried out using ATP/^{32}PPi-exchange assay based on reversible aminoacyl-AMP formation catalyzed by certain A-domains. The reaction mixture with final volume of 200 µL containing 50 mM Tris, 10 mM $MgCl_2$, 5 µM recombinant enzyme, 5 mM amino acid, 2.5 mM $Na_4P_2O_7$ and

^{32}P-$Na_4P_2O_7$ was incubated at 30°C for 10 min. The reaction was initiated by addition of ATP to a final concentration of 2.5 mM. After incubation at 30°C for 30 min the reaction was quenched by adding 0.75 mL stop solution (100 mM $Na_4P_2O_7$, 35 mM $HClO_4$, 0.15 g/L activated charcoal). The activated charcoal was pelleted via centrifugation (13,000 rpm for 3 min), washed twice with 0.8 mL water and resuspended in 0.8 mL water. The resuspended mixture was added to 3.5 mL liquid scintillation fluid (*Roth,* Rotiszint® eco plus). The radioactivity was measured via liquid scintillation counting (*Packard,* Tri-carb 2100TR Liquid Scintillation Analyzer). Assays were performed in triplicate and the activities are given relative to the highest measured activity.

4.6.2 Thioesterase catalyzed macrocyclization assay

The reaction mixture with final volume of 50 µL containing 25 mM HEPES, 50 mM NaCl, 10 - 400 µM peptidyl-SPh, 2 µL DMSO and 1 µM enzyme was incubated at 25°C for a certain period of time (2 h for activity analysis, 30 s for determination of kinetic parameters). The reaction was quenched by adding 50 µL MeOH to precipitate the enzyme. After centrifugation at 13,000 rpm for 10 min, the resulting supernatant was analyzed via LC-MS (*Agilent/HP* 1100 series, column*: Macherey-Nagel* cc125/2 Nucleodur 100-3 c18 ec, column temperature: 45°C) with the following gradient: solvent A: water (0.1% trifluoroacetic acid), solvent B: acetonitrile (0.1% trifluoroacetic acid), flow rate: 0.3 mL/min, gradient: 0 min, 30% B; 20 min, 50% B; 23 min, 95% B; 30 min, 95% B; 33 min, 30% B; 40 min, 30% B). Experiments were performed in triplicate for the determination of kinetic parameters.

4.6.3 Fluoresceinyl-CoA phosphopentetheinylation assay

Apo-PCP is converted to active *holo*-PCP by transfer of the ppan group from coenzyme A onto the conserved serine under catalysis by PPtases. It was shown that PPtase from *Bacillus subtilis* (Sfp) exhibits a high degree of substrate tolerance which makes Sfp capable of catalyzing the transfer of CoA-derivatized substrates onto PCPs *in vitro*.

Fluoresceinyl-CoA phosphopentetheinylation assay was employed to validate the ability of PCP-domains to be loaded with CoA-substrates *in vitro*.

Fluoresceinyl-CoA phosphopentetheinylation assay was carried out by incubation of the reaction mixture with a final volume of 50 µL containing 50 µM recombinant PCP, 300 µM fluoresceinyl-CoA, 5 µM recombinant Sfp and 50 µM $MgCl_2$ in Tris buffer (20 mM Tris, 100 mM NaCl, pH 7.0) at 25°C for 30 min. The labeled protein was separated via SDS-PAGE and visualized under UV-light (λ = 312 nm). A subsequent staining with Coomassie Brilliant Blue R250 enabled the comparison of the labeled and unlabeled samples.

4.6.4 [^{14}C]-acetyl-CoA phosphopentetheinylation assay

Similar to the fluoresceinyl-CoA phosphopentetheinylation assay, [^{14}C]-acetyl-CoA phosphopentetheinylation assay also uses the high substrate tolerance of Sfp to radio-label PCP-domains. SrfA-A PCP_1 was heterologously produced and purified as described earlier[16]. The phosphopentetheinylation assay was carried out by incubating 80 µM PCP with 8 µM Sfp and 2 µM [1-^{14}C]-acetyl-CoA (*Amersham*) in assay buffer (50 mM Tris/HCl, 10 mM $MgCl_2$, pH 7.5) for 45 min at 37°C. The final volume was 50 µL. The reaction mixture was directly used in following deacylation assay or treated with 1 mL 10% TFA and incubated on ice for 30 min. After 30 min centrifugation at 13,000 rpm, the protein pellet was then washed twice with 10% TFA and dissolved in 500 µL acetic acid. The radio activity was measured via liquid scintillation counting (*Packard,* Tri-carb 2100TR Liquid Scintillation Analyzer).

4.6.5 TE II mediated deacylation assay

Different recombinant thioesterases were added to the reaction mixture of [^{14}C]-acetyl-CoA phosphopentetheinylation reaction mixture with final concentration of 2 µM followed by incubation at 37°C for 1h. The reactions were quenched and the proteins were pelleted by adding 1 mL 10% TFA. The mixture was incubated on ice for 30

min and centrifuged at 13,000 rpm for 30 min. The protein pellet was washed twice with 10% TFA and dissolved in 500 µL acetic acid. The radioactivity was measured via liquid scintillation counting (*Packard,* Tri-carb 2100TR Liquid Scintillation Analyzer).

4.6.6 Deacylation study

Deacylation studies were based on the cleavage of the [1-^{14}C]-acetyl group from the stand-alone SrfA-A PCP_1 via LybB TE-mediated hydrolysis. SrfA-A PCP_1 was heterologously produced and purified as described earlier.[135] Loading of the PCP with [1-^{14}C]-acetyl-CoA (*Amersham*) was accomplished by incubating 80 µM PCP with 8 µM Sfp and 2 µM [1-^{14}C]-acetyl-CoA in assay buffer (50 mM Tris/HCl, 10 mM $MgCl_2$, pH 7.5) for 45 min at 37°C. The final reaction volume was 50 µL. Deacylation reactions were initiated by addition of 2 µM LybB-PCP-TE_1, LybB-TE_1 or LybB-TE_2 to the reaction mixture. The reactions were incubated at 37°C for 1 h. After the reactions were stopped, the PCP was precipitated by addition of 1 mL TCA (10 % v/v). The mixture was incubated on ice for 30 min and centrifuged at 13,000 rpm for 30 min. The protein pellet was washed twice with 1 mL TCA (10% v/v) and dissolved in 500 µL formic acid. The radioactivity was measured via liquid scintillation counting (*Packard,* Tri-carb 2100TR Liquid Scintillation Analyzer). Control reactions included measurement of *holo*-PCP directly after the loading procedure and after 1h incubation at 37°C lacking thioesterases to investigate non-enzymatic hydrolysis.

4.7 Natural product isolation

Lysobacter sp. ATCC 53042 was purchased from *LGC Standards GmbH* as freeze dried powder. The microorganism was revived by plating on soy agar slants (15 g BBL™ Trypticase™ soy broth BD21176, 7.5 g Agar Roth 2266.2 in 500 mL water) following the manufacturer's protocol. Fermentations were carried out as described earlier.[82, 83] After centrifugation of the culture at 7,000 rpm for 30 min, the resulted supernatant was extracted with butanol. After combination of the extracts, the organic solvents were removed using rotary evaporator to yield a yellow solid, which was then triturated in methanol. The soluble portion was separated and analyzed via LC-MS (*Agilent/HP* 1100

series, column: *Macherey-Nagel* cc125/2 Nucleodur 100-3 c18 ec) with the following gradient: solvent A: water (0.1% TFA), solvent B: acetonitrile (0.1% TFA), flow rate : 0.3 mL/min, gradient: 0 min, 10% B; 30 min, 60% B; 33 min, 95% B; 38 min, 10% B; 45 min, 10% B). HRMS-analysis of lysobactin was performed via high resolution mass spectrometry on an LTQ-FT instrument (*Thermo Fisher Scientific*).

5. Results

5.1 Genome sequencing of *Lysobacter sp.* and bioinformatic identification of NRPS/PKS coding gene clusters

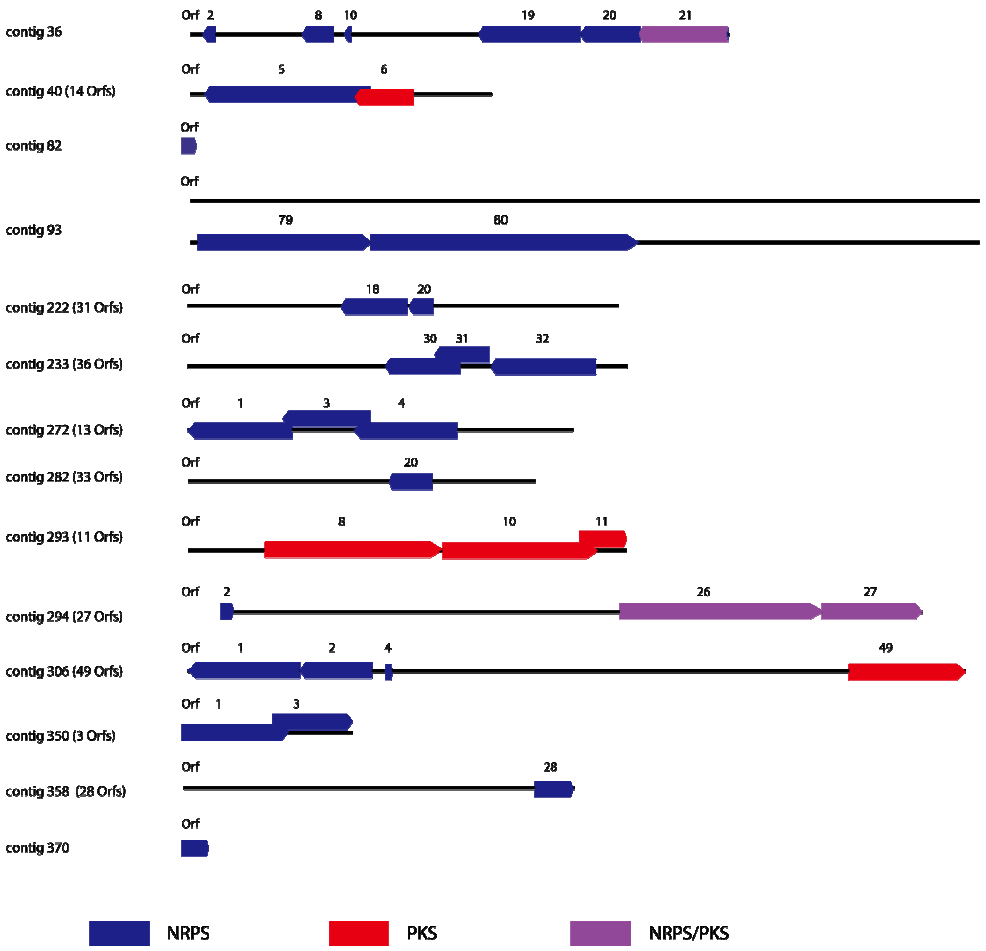


Figure 5.1: Contigs from sequencing result of genome of *Lysobacter sp.* ATCC 53042 containing NRPS/PKS coding genes. Red: PKS-coding genes. Blue: NRPS-coding genes, purple: hybrid NRPS/PKS coding genes.

The genomic DNA of *Lysobacter sp.* ATCC 53042 was isolated and sequenced using pyroseqeuncing methods on a GS FLX instrument. This genome sequencing was carried out by GATC-biotech, delivering 418 contigs ranging from 553 to 150,063 bp. To simplify the analysis, all contigs were connected head-to-tail, resulting in a sequence with a size of 6,011,895 bp. The order of the contigs and sequencing gaps were neglected. Analysis

of this sequence using GeneMark 2.4 annotation yielded 5,588 CDSs. The coding sequences crossing the contig borders (where two contigs were connected with each other) were treated carefully by deleting the part over the contig borders. CDSs containing the core motif of a PCP domain (GG(DH)SL)[136] were further analyzed using PKS/NRPS Analysis[137] resulting in the identification of NRPS or PKS/NRPS hybrid coding sequences. This analysis delivered 14 contigs containing NRPS or PKS/NRPS hybrid coding genes shown in Figure 5.1. Using the 4.6 kbp gene fragment reported by Bernhard, et al.[1], contig 93 covering a region of 150,063 bp with an average GC-content of 70.7% was identified to contain the full lysobactin biosynthesis machinery, which will be further discussed in chapter 5.2. Further analysis of the modular organization of the NRPS or PKS/NRPS hybrid CDSs in the other 13 contigs delivered 4 NRPS or PKS/NRPS CDSs showing logical organization. Other candidates are either illogical due to absence

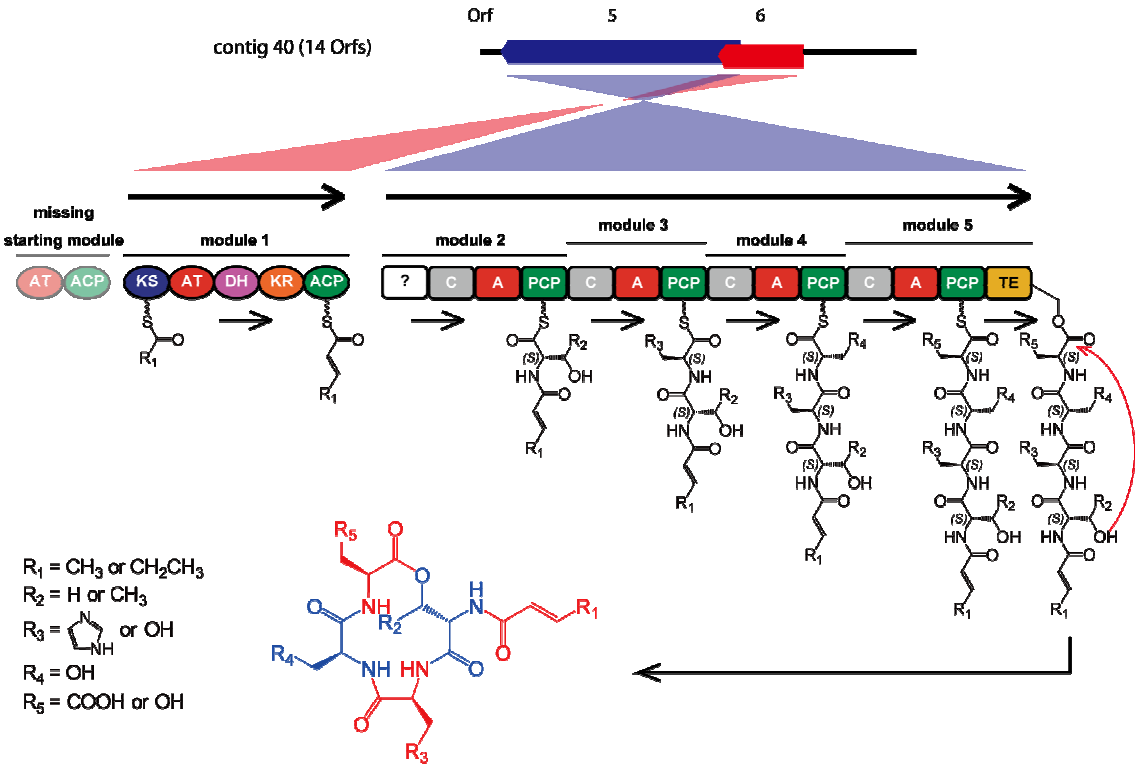


Figure 5.2: Potential hybrid NRPS/PKS encoding genes in contig 40 and the proposed corresponding product. *Orf* 5 encodes a tetramodular NRPS. Orf 6 encodes a monomodular PKS. The starting module is missing.

of essential domains, which is exemplified by NRPS/PKS hybrid CDSs in contig 294 lacking essential AT-domains, or small gene fragments possibly resulting from incomplete sequencing, which is exemplified by the short NRPS gene fragment found in contig 370. In the following chapters, these four NRPS and NRPS/PKS hybrid CDSs will be discussed in detail.

Contig 40

In contig 40, one NRPS coding gene with the size of 15,495 bp (*orf* 5) and one PKS coding gene with the size of 5,460 bp (*orf* 6) showing the same transcription direction were identified. The two *orf*s showed an overlapping region of 1,533 bp. The gene product of *orf* 5 displays 47% identity to arthrofactin synthetase from *Bradyrhizobium sp.* BTAi1 (accession number: YP 001242609.1) with an E-value of 0.0, while the gene product of *orf* 6 shows 46% identity to the polyketide synthase PksE from *Paenibacillus mucilaginosus* KNP414 (accession number: YP 004642886.1) with an E-value of 0.0. As shown in Figure 5.2, further analysis using PKS/NRPS analysis showed that *orf* 6 encodes a PKS containing 5 domains, including one keto-synthase, one acyltransferase, one dehydratase, one ketoreductase and one acyl carrier protein domain. A necessary starting module containing one AT- and one ACP-domain is missing. *Orf* 5 codes for a NRPS containing 4 modules (module 2 to 5), which can be further subdivided into 14 domains. Module 3 and 4 display a regular C-A-PCP organization. Module 2 contains an extra N-terminal unknown domain consisting of 554 amino acids, which could not be classified using PKS/NRPS analysis. Significant similarity could not be found via BLAST analysis. Module 5 contains at the C-terminus one thioesterase domain, which should be responsible for release of the final product via cyclization or hydrolysis. Comparison of the active-site residues determining the A-domain specificities with that of known A-domains[12, 138] suggested that the A-domain in module 2 activates L-Ser or L-Thr, the A-domain in module 3 L-Ser or L-His, the A-domain in module 4 L-Ser, and the A-domain in module 5 L-Ser or L-Asp. The hydroxyl groups in R_2, R_3 and R_4 could be used as

nucleophilic groups for the cyclization to build the final product. One of the possible structures of the final product is shown in Figure 5.2. The comparison of active-site residues is shown in Table 5.1.

Contig 233

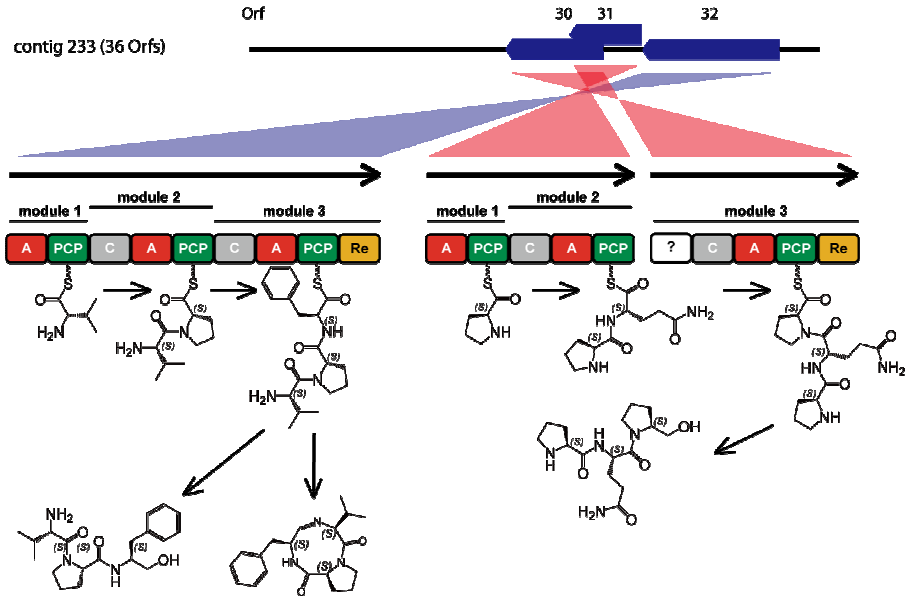


Figure 5.3: Potential NRPS encoding genes in contig 233 and the proposed corresponding products. *Orf* 32 encodes a trimodular NRPS. *Orf* 31 encodes a dimodular NRPS. *Orf* 32 encodes a monomodular NRPS with an unknown N-terminal domain

In contig 233, three consecutive NRPS coding sequences (*orf* 32, 31 and 30 in Figure 5.3) were found sharing the same direction of transcription. The gene product of *orf* 32 has a size of 9,822 bp and displays 51% identity to a non-ribosomal peptide synthetase module from *Burkholderia rhizoxinica* HKI 454 (accession number: YP 004021982.1) with an E-value of 0.0. Further analysis with PKS/NRPS analysis revealed that orf 32 consists of three NRPS modules. The first module consists of one A-domain and one PCP-domain. The second module displays a C-A-PCP tridomain structure. The third module has a C-A-PCP-R tetradomain architecture. Comparison of the active-site residues determining A-domain specificities with known A-domains suggested that the three A-domains activate L-Val, L-Pro and L-Phe, respectively. The comparison result is shown in Table 5.1.

The C-terminal R-domain displays 44% identity to MxaA from *Stigmatella aurantiaca* (accession number: AAK57184.1) with an E-value of 7e-70. *MxaA* encodes a tetradomain NRPS (C-A-PCP-Re) involved in the biosynthesis of myxalamid S, in which the C-terminal R-domain is responsible for NADPH-dependent reduction and release of the final product.[139]

Orf 31 covers a 5,094-bp region and *orf* 30 covers a 7,005-bp region. The two *orf*s show an overlapping region of 2,409 bp. The gene product of *orf* 31 displays 47% identity to syringopeptin synthetase C from *Photorhabdus asymbiotica* (accession number: CAQ84313.1) with an E-value of 0.0 and consists of 2 modules, which can be subdivided into 5 domains. The first module consists of one A-domain and one PCP-domain. The second module displays a C-A-PCP tridomain structure. Analysis of the active-site residues of the A-domains found in orf 31 suggests that the two A-domains activate L-Pro and L-Gln in turn.

Orf 30 consists of one module containing 5 domains. The three domains in middle (C-A-PCP) displays 50% identity to arthrofactin synthetase from *Burkholderia gladioli* (accession number: YP_004349031.1) with an E-value of 0.0. The N-terminal domain contains 849 amino acids and displays no significant identity to known protein (BLAST). The C-terminal R-domain displays 40% identity to polyketide synthase from *Nostoc sp.* PCC 7120 (accession number: NP 484397) with an E-value of 6e-71 and 36% identity to MxaA from *Stigmatella aurantiaca* (accession number: AAK57184.1) with an E-value of 1e-55. Analysis of the active-site residues of A-domain determining A-domain specificities suggests that the A-domain in orf 30 activate L-Pro.

Contig 306

In contig 306, 1 PKS coding gene and 3 NRPS coding genes were identified (Figure 5.4). *Orf* 49 encodes a type I PKS lacking the essential AT-domains (KS-?-KR-ACP-KS-KR-ACP-KS-DH). *Orf* 4 encodes a distinct thioesterase. *Orf* 1 and 2 have the same transcription direction opposite to that of *orf* 4. *Orf* 2 has a size of 6,708 bp

and its gene product shows 40% identity to a NRPS from *Nitrococcus mobilis* NB-231 (accession number: ZP 01127432.1) with an E-value of 0.0. Further analysis of orf 2 revealed that this NRPS consists of 2 modules. The first module displays a C-A-PCP-PCP tetradomain structure, where an extra PCP domain is found. The latter module shows a regular C-A-PCP tridomain organization. Comparison of the extracted active-site residues of the 2 A-domains with known A-domains suggests that the first A-domain activates L-Asp, while the second activates L-Val or L-Leu.

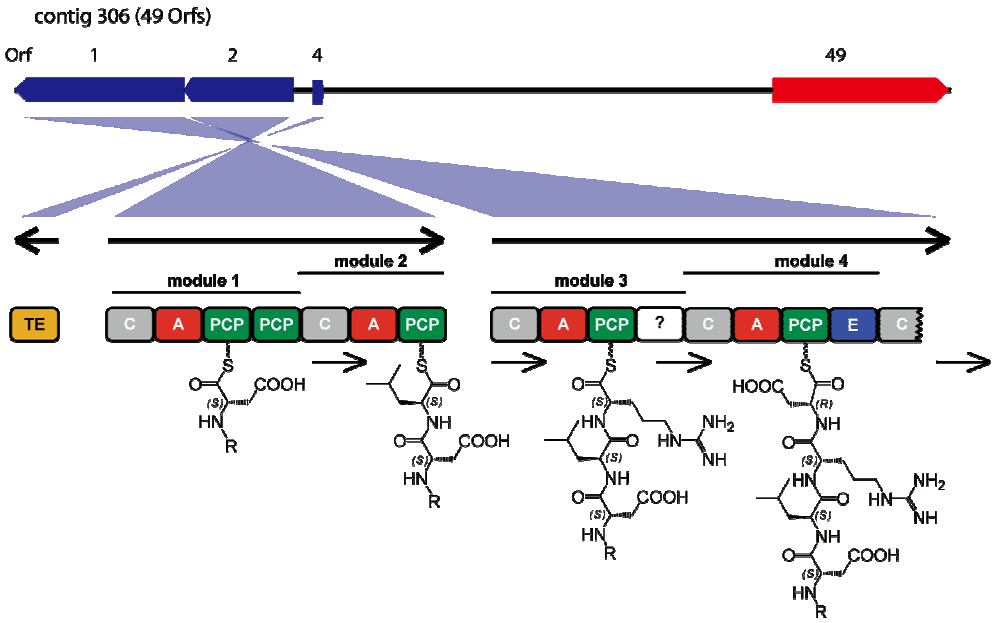


Figure 5.4: Potential NRPS encoding genes in contig 306 and the proposed corresponding substrates of the A-domains. Both *Orf* 1 and 2 encode NRPS consisting of 2 modules. Orf 4 encodes an alone-standing thioesterase. R: fatty acid chain.

Orf 1 has a size of 10,386 bp and its gene product shows 32% identity to an amino acid adenylation domain containing protein from *Nostoc punctiforme* PCC 73102 (accession number: YP 001866468.1) and 30% identity to a NRPS from *gamma proteobacterium* HdN1 (accession number: YP 003810749.1) with both E-values of 0.0. Further analysis using PKS/NRPS analysis revealed that orf 1 contains 2 complete NRPS modules and 1 additional incomplete NRPS module. The first module represents C-A-PCP organization followed by an unknown domain containing 474 amino acids. The N-terminal part of this

unknown domain (about 200 amino acids) contains conserved motifs of enzymes in the condensation superfamily, while the C-terminal part of this unknown domain contains conserved motifs of NRPS-para261 superfamily. The second module shows a C-A-PCP-E organization. And the third incomplete module contains only the 388 N-terminal amino acids of a C-domain. *Orf* 1 is not fully known because of the incomplete sequencing. Thus, *orf* 1 contains only the N-terminal part of a NRPS coding gene. Comparison of the active-site residues extracted from the A-domains in orf 1 with known A-domains suggests that these two A-domains to activate L-Arg and L-Asp in turn.

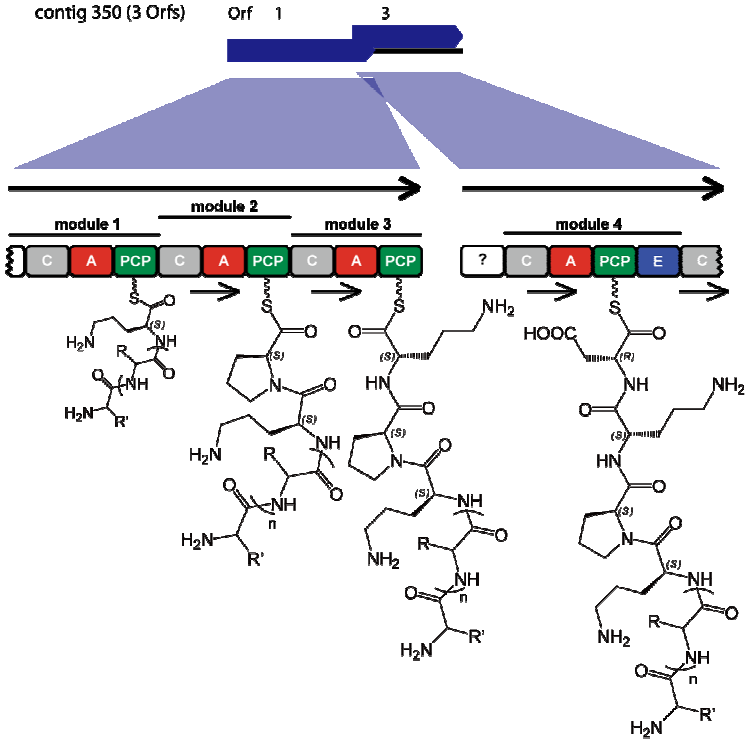


Figure 5.5: NRPS encoding genes in contig 350 and the proposed corresponding intermediates. *Orf 1* encodes an N-terminal incomplete trimodular NRPS. *Orf 3* encodes a C-terminal incomplete dimodular NRPS.

Contig 350

In contig 350, 2 NRPS coding genes were identified (Figure 5.5). *Orf* 1 covers a region of 10,080 bp and its gene product shows 35% identity to an amino acid adenylation protein (accession number: YP 322129.1) from *Anabaena variabilis* ATCC 29413 with an E-value

of 0.0. Further analysis revealed that orf 1 contains 3 complete NRPS modules, which could be further subdivided into 9 domains. Each module consists of three NRPS domains: C-A-PCP. At the N-terminal end of orf 1 there is one truncated domain because of the incomplete genome sequencing. The missing N-terminal part of *orf* 1 could be located in sequencing gaps or at the end of other contigs. Comparison of extracted active-site residues of A-domains found in orf 1 with known A-domains suggests that these A-domains activate L-Lys, L-Pro and L-Lys in turn. *Orf* 3 covers a region of 7,476 bp and its gene product displays 39% identity to linear gramicidin synthase subunit B (accession number: ADR59299.1) from *Pseudomonas putida* BIRD-1 with an E-value of 0.0. Further analysis of the amino acid sequence with PKS/NRPS analysis revealed that *orf* 3 encodes a NRPS containing one complete NRPS module (C-A-PCP-E) with an extra N-terminal unknown domain and an additional C-terminal incomplete C-domain. The C-terminal C-domain is truncated because of the incomplete genome sequencing and the missing part could be found in sequencing gaps or at the beginning of other contigs. Analysis of the N-terminal unknown domain via BLAST revealed no significant similarity to known proteins. Analysis of the extracted active-site residues of the A-domain found in orf 3 suggests that this A-domain activates L-Asp.

Table 5.1 Comparison of the extracted active-site residues determining the adenylation domain specificities of NRPSs found in contigs of the sequencing results of the *Lysobacter sp.* genome with known adenylation domains. Variations in the residue pattern are highlighted in red. The substrate prediction for each A-domain as well as the product of the NRPS is given. (The analysis was carried out using NRPSpredictor using comparison of 10 active-site residues. The comparison marked with * was done using PKS/NRPS analysis delivering comparison of 8 active-site residues)

A-domain	Active-site residues	Substrate	Product
Contig 40, orf 5			
A_1	D V W H V S L V D K		
BlmVI-A_1 [95]	D V W H V S L V D K	L-Ser	bleomycin
A_2	D S A L I A E V W K		

BacC-A3[140]	D S E L T A E V C K	L-His	bacitracin
A_3	D L W H I G L L D K		
SyrE-A_1[141]	D L W H L S L I D K	L-Ser	syringomycin
A4	D M W M L G V V I K		
ArfA-A_2	D S W K L G V V D K	L-Asp	arthrofactin
Contig 233, orf 32			
A_1	D A L L I G A V C K		
Q8NJX1-A_{14}	D A A L I G A V F K	L-Val	-
A_2	D V Q H V A H V A K		
SypA-A_2[142]	D V Q Y I A H V V K	L-Pro	syringopeptin
A_3	D A F T L G A V C K		
BacC-A_2[140]	D A F T V A A V C K	L-Phe	bacitracin
Contig 233, orf 31			
A_1	D V Q H V A H V V K		
SypA-A_2[142]	D V Q Y I A H V V K	L-Pro	syringopeptin
A_2	D A I Y L G V V L K		
LicA-A_1[143]	D A Q D L G V V D K	L-Gln	lichenysin
Contig 233, orf 30			
A_1	D V Q H A A H V A K		
SnbDE-A_1[144]	D V Q Y A A H V M K	L-Pro	pristinamycin
Contig 306, orf 2			
A_1	D L T K V G H V G K		
LchAB-A_2[145]	D L T K V G H I G K	L-Asp	lichenysin A
A_2	D M L V L G G V I K		
Tex1-A_4[146]	D M G F L G G V C K	L-Leu or L-Val	peptaibol
Contig 306, orf 1			
A_1	D T E D V G A I S K		

SyrE-A_5[141]	D V A D V G A I D K	L-Arg	syringomycin
A_2	D L T K V G H V G K		
LchAB-A_2[145]	D L T K V G H I G K	L-Asp	lichenysin A
Contig 350, orf 1			
*A_1	D I E S L G T V		
BacB-A_1[140]	D A E S I G S V	L-Lys	bacitracin
A_2	D V Q F A A H V V K		
ItuB-A_4[147]	D V Q F I A H V V K	L-Pro	iturin A
A_3	D I E S V G T V V K		
NpsA-A_1	D T E V V G T L V K	L-Lys	nourseothricin
Contig 350, orf 3			
A_1	D L T K V G H V G K		
LchAB-A_2[145]	D L T K V G H I G K	L-Asp	lichenysin A

5.2 Identification and characterization of lysobactin biosynthesis gene cluster

5.2.1 Confirmation of lysobactin production

To ensure that the *Lysobacter sp.* ATCC 53042 strain used in this study is a lysobactin producing strain, LC-MS analysis of culture extract was carried out. Under the conditions applied, lysobactin showed a retention time of t_R = 19.5 min (m/z = 1276.8 $[M+H]^+$ observed, m/z = 1276.7 $[M+H]^+$ calculated, Figure 5.6). Further analysis of the extracted compound using high-resolution MS (m/z = 1276.7278 $[M+H]^+$ observed, m/z = 1276.7260 $[M+H]^+$ calculated, Figure 5.6) confirmed the production of lysobactin under the fermentation conditions applied.

5.2.2 Identification and sequential study of the lysobactin biosynthetic gene cluster (*lyb*)

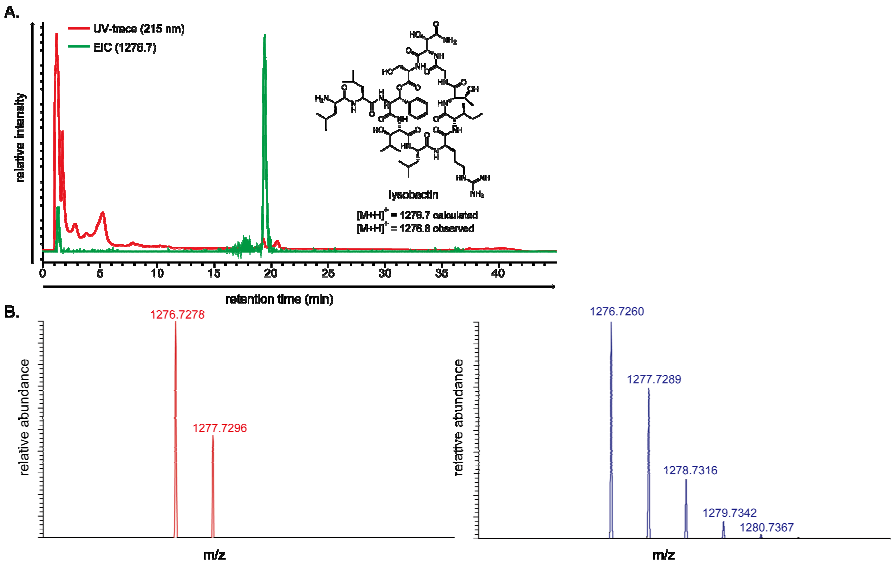


Figure 5.6: Extraction of native lysobactin. A. HPLC profile of a culture extract of *Lysobacter sp.* ATCC 53042. The red trace corresponds to the UV-signal at 215 nm. The green trace represents the extracted-ion-current using m/z = 1276.7 as query. Under the conditions applied, lysobactin shows a retention time of t_R = 19.5 min (m/z = 1276.8 $[M+H]^+$ observed, m/z = 1276.7 $[M+H]^+$ calculated). B. HRMS spectrum of lysobactin isolated from cultures of *Lysobacter sp.* ATCC 53042. The red signal corresponds to the observed ions. The spectrum accentuated in blue corresponds to the calculated spectrum for lysobactin (m/z = 1276.7278 $[M+H]^+$ observed, m/z = 1276.7670 $[M+H]^+$ calculated).

In the sequencing result, one contig (contig 93) was identified to contain the reported 4.6 kbp gene fragment.[1] This contig, covering a region of 150,063 bp with an average GC-content of 70.7%, was further analysed using GeneMark 2.4 annotation[148] and has been submitted to GeneBank (accession number JF412274). A detailed overview of the annotated genes, deduced protein functions and similarities to homologous is found in Table S1. The predicted lysobactin biosynthetic gene cluster encodes the two NRPSs LybA and LybB, three proteins conferring resistance to antibiotics (orf 78, 80 and 82) and one ABC-transporter permease/ATP-binding component (orf 79) putatively involved in the secretion of lysobactin (Figure 5.7).

The synthetase LybA displays 50% identity and 64% similarity with E-value of 0.0 to an NRPS from *Pseudomonas syringae pv. Syringae* B728a, while LybB shows 39% identity

and 59% similarity with E-value of 0.0 to a multimodular NRPS from *Herpetosiphon aurantiacus* ATCC 23779. Further bioinformatic analysis of the two synthetases using PKS/NRPS analysis revealed a multimodular organization (Figure 5.8)[137]. LybA consists of four modules and 11 domains, whereas LybB is constituted of 7 modules and 24 domains. Because the number of A-domains found within LybA and LybB directly

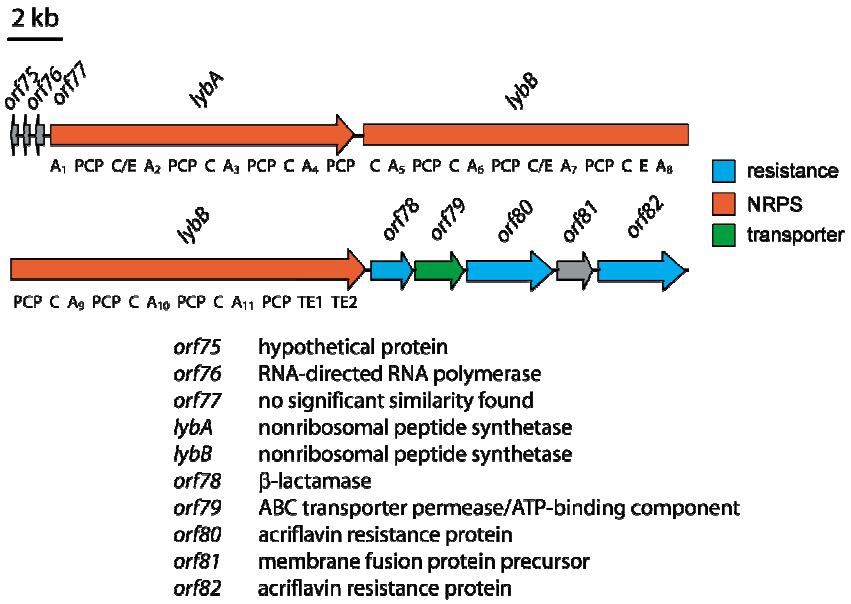


Figure 5.7: Schematic overview of the lysobactin biosynthetic gene cluster and the corresponding upstream and downstream regions. Functions of the proteins encoded within this region are based on BLAST-analysis and are given in the figure. Apart from the core components for the nonribosomal assembly of lysobactin by the synthetases LybA and LybB, genes coding for transporters and resistance-conferring proteins are found.

correlates with the primary sequence of lysobactin, a linear logic of lysobactin assembly is suggested. The reported 4.6 kbp gene fragment was found in *lybB* and codes for a $(C\text{-}A\text{-}PCP)_{10}\text{-}C_{11}$ tetradomain-region. The specificities of A-domains found in LybA and LybB were analysed using NRPSpredictor[138]. The predicted specificities of the A-domains nicely correlate with the primary sequence of lysobactin.

According to the prediction results, LybA-A_3 should activate L-Phe. The specificity determining residues of LybA-A_3 show only 80% identity comparing to BacC-A_2. This suggests activation of structurally similar building block[140]. Two further hydroxylated

amino acids could be found in lysobactin, namely β-OH-Leu$_4$ and β-OH-Asn$_{10}$. In contrast to LybA-A$_3$, the specificity conferring residues of LybA-A$_4$ are identical to those found in Leu-activating A-domain SrfA-A-A$_3$, involved in surfactin biosynthesis.[149] The extracted active-site residues of LybB-A$_{10}$ are identical to those found in Asn-activating A-domain NosC-A$_3$ responsible for activating and incorporating L-Asn into nostopeptolide A[150].

Lysobactin contains two D-configured amino acids, namely D-Leu$_1$ and D-Arg$_6$. The corresponding modules 1 and 6 are not equipped with epimerization domains (E-domain), which are normally required for stereoinversion of the α-stereocenter. Further examination of the downstream C-domains (C2 and C7) revealed that these two C-domains contain additional N-terminal sequences with the additional core-motif HHI/LXXXXGD. About 110 amino acids downstream of this additional core-motif, the histidine core-motif HHXXXD is observed.

Another unusual building block in lysobactin is the *allo*-Thr at position 8. The corresponding module contains an additional C-domain and shows a unique C-C-A-PCP organization. The incorporation of the *allo*-Thr will be further discussed in chapter 5.3. The termination module of LybB harbors an unusual tandem TE-domain architecture, similar to that of the arthrofactin synthetase ArfC[66]. Further analysis and discussion will be presented in detail in the next chapters.

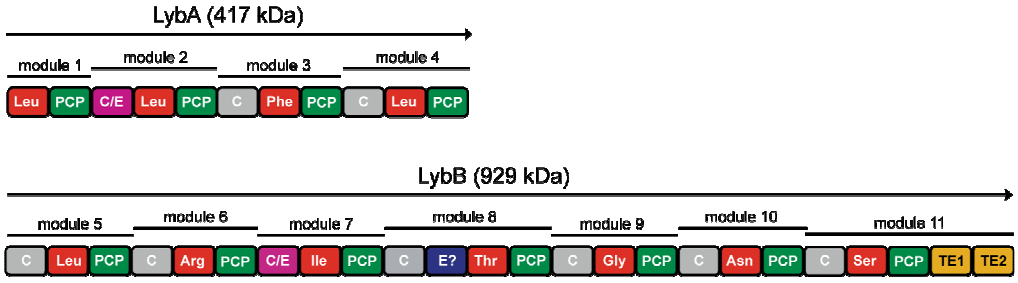


Figure 5.8: Modular organization of the lysobactin synthetase LybA and LybB. LybA consists of four modules and 11 domains respectively. LybB is constituted of 7 modules and 24 individual domains. The termination module contains two distinct thioesterase domains, responsible for the release of the PCP-bound intermediate via macrolactonization.

The upstream flanking region of the *lyb* gene cluster is defined by *orf*s 75-77 which encode hypothetical proteins and a RNA-directed RNA polymerase in the reverse direction. The downstream region of the *lyb* gene cluster comprises *orf* 78, encoding a β-lactamase-type protein, *orf* 80 and *orf* 82, encoding RND-family acriflavin resistance proteins, and *orf* 79, encoding an ABC transporter permease/ATP-binding component. The direction of transcription for *lybA*, *lybB* and *orf*s 78-82 is identical.

5.2.3 Substrate specificity studies of the A-domains

Overproduction of LybA-A_1, LybB-A_6 and LybB-A_9

The gene fragments *LybA-A$_1$*, *LybB-A$_6$* and *LybB-A$_9$* were cloned in the expression vector pET28a (*Novagen*) and overexpressed in *E. coli* Rosetta 1 DE 3 / *E. coli* BL21 as described in the Methods section. The recombinant His-tagged proteins were purified by Ni^{2+}-NTA affinity chromatography. The fractions were analysed using SDS-PAGE and visualized by Coomassie stain (Figure 5.9). Concentration using centrifugal filter (*Millipore*) afforded 1.44 (LybA-A_1), 0.85 (LybB-A_6) and 2.09 (LybB-A_9) mg/L media of recombinant protein.

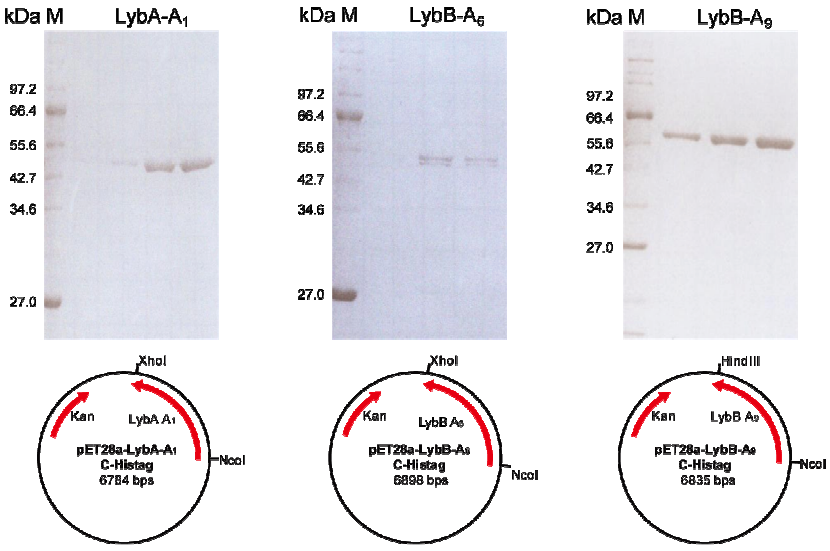


Figure 5.9: Plasmid maps and Coomassie-stained SDS-PAGE of purified recombinant A-domains of lysobactin synthetase LybA-A_1 (56.9 kDa), LybB-A_6 (60.4 kDa) and LybB-A_9 (58.2 kDa). Lane M: Protein Marker, Broad Range (2-212 kDa) P7702 (*NEB*)

In vitro characterization of A-domain specificities

The specificities of recombinant A-domains were studied via ATP/PPi-exchange assays. The assays were carried out as described in the methods section. Different amino acids were used for the characterization of LybA-A_1. A negative control was carried out using the same assay in the absence of substrate. As shown in Figure 5.10, D-Leu showed the highest activity. The predicted preferred substrate L-Leu showed a reduced activity of 0.94 compared to that of D-Leu. Structurally related amino acids such as Ala, Ile and Val showed reduced activities compared to D-Leu ranging from 0.11 (D-Val) to 0.76 (L-Ile). Other studied amino acids such L-Phe, L-Ser and L-Lys showed reduced activities compared to that of D-Leu ranging from 0.24 (L-Lys) to 0.56 (L-Ser). The negative control showed a background activity of 0.002 compared to D-Leu.

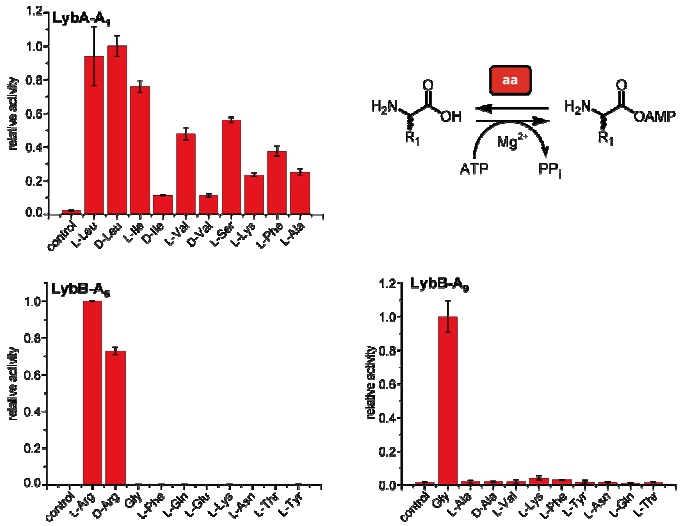


Figure 5.10: Analysis of the recombinant A-domains from lysobactin synthetase via ATP/PPi-exchange assay.

In the study of LybB-A_6, the predicted preferred substrate L-Arg showed the highest activity. The negative control without substrate showed a background activity of 0.003 compared to that of L-Arg. D-Arg showed reduced activity related to that of L-Arg of 0.73. Activities of other studied amino acids such as Gly, L-Phe and L-Tyr did not exceed 0.006.

In the study of LybB-A_9, the predicted preferred substrate Gly showed the highest activity.

The negative control in the absence of substrate showed a background activity of 0.02. Activities of other tested substrates such as L-Ala, L-Phe and L-Tyr did not exceed 0.04.

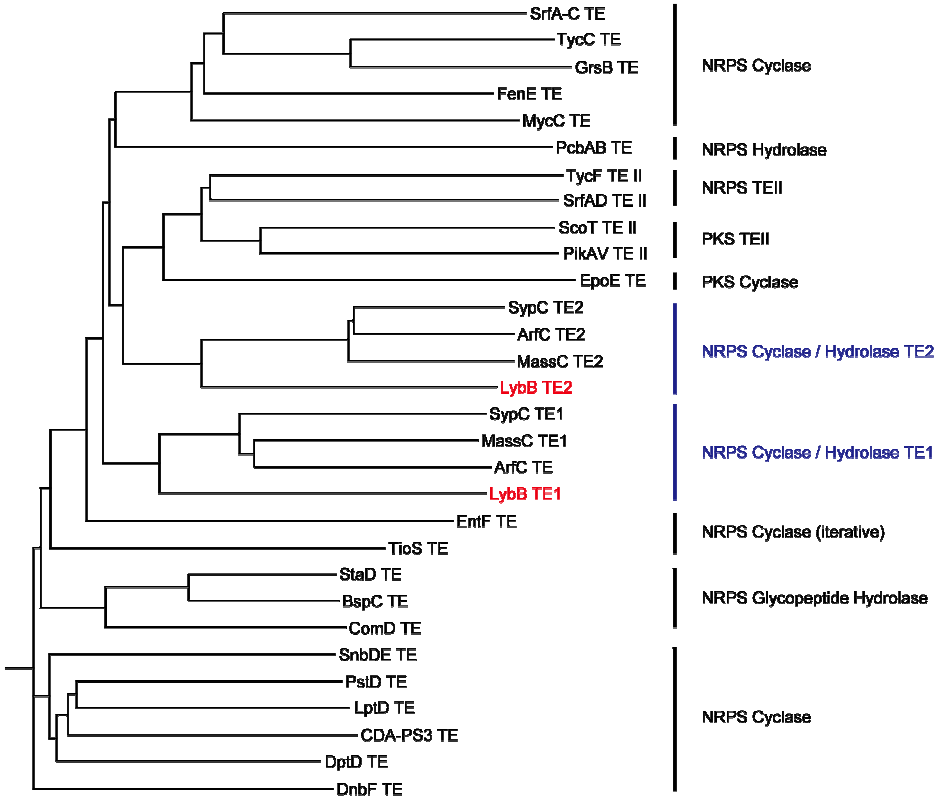


Figure 5.11: Phylogenetic analysis of TE domains derived from various microorganisms. LybB-TE_1 and TE_2 are accentuated in red. The TE domains employed are: FenE-TE from *Bacillus amyloliquefaciens* DSM7, TycC-TEand TycF-TE II from *Bacillus brevis*, Grs-TE from *Bacillus subtilis*, SrfA-C-TE and SrfAD-TE II from *Bacillus amyloliquefaciens* FZB 42, MycC-TE from *Bacillus subtilis subsp. Spizizenii* str. W23, PcbAB-TE from *Aspergillus flavus* NRRL 3357, EpoE-TE from *Sorangium cellulosum*, ScoT-TE II from *Photorhabdus asymbiotica*, PikA-TE II from *Streptomyces venezuelae*, MassC-TE_1 and MassC-TE_2 from *Pseudomonas fluorescens*, ArfC-TE_1 and ArfC-TE_2 from *Pseudomonas* sp. MIS38, LybB-TE_1 and LybB-TE_2 from *Lysobacter* sp. ATCC 53042, SypC-TE_1 and TE_2 from *Pseudomonas syringae*, EntF-TE from *Escherichia coli*, TioS-TE from *Micromonospora* sp. ML1, BspC-TE from *Amycolatopsis balhimycina*, StaD-TE from *Streptomyces toyocaensis*, ComD-TE from *Streptomyces lavendulae*, LptD-TE from *Streptomyces fradiae*, PstD-TE from *Actinoplanes friuliensis*, CDA-PS3-TE from *Streptomyces coelicolor* A3(2), DptD-TE from *Streptomyces roseosporus* NRRL 11379, SnbDE-TE from *Streptomyces pristinaespiralis*, DnbF-TE from *Bacillus subtilis*.

5.2.4 Characterization of the LybB thioesterases

Bioinformatic analysis of the TE-domains in LybB

Phylogenetic analysis of LybB-TE_1 and LybB-TE_2 compared to thioesterases from NRPSs or PKSs was carried out. The results shown in Figure 5.11 revealed the LybB-TEs to cluster with several other tandem TEs. LybB-TE_1 clusters with the TE_1 domains found in other tandem TE systems as arthrofactin (ArfC-TE_1), massetolide (MassC-TE_1) and syringopeptin (SypC-TE_1) biosynthetic systems, while LybB-TE_2 clusters with TE_2 domains from these biosynthetic systems.

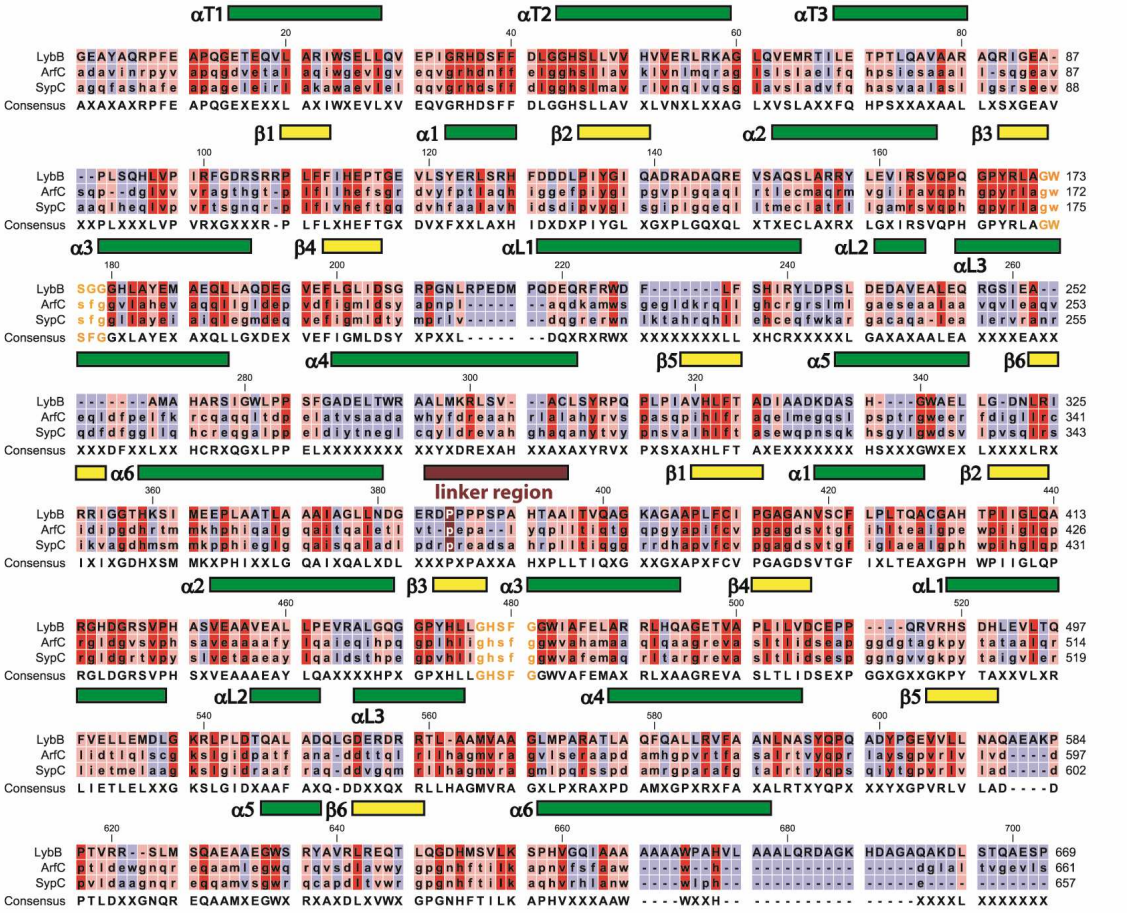


Figure 5.12: Multiple sequence alignment of tandem TE domains. Proteins are LybB-PCP-TE_1-TE_2, ArfC-PCP-TE_1-TE_2 and SypC-PCP-TE_1-TE_2. The degree of conservation is indicated by color: Red indicates complete and blue no agreement. The GXSXG core motif of both TEs is displayed in orange, α-helices are accentuated in green and β-sheets are given in yellow.

Further sequential analysis of LybB-TE_1 and TE_2 based on core motifs, secondary structure elements and multiple sequence alignments with tandem TE-domains, under application of ClustalW algorithm, JPred3 and XtalPred servers, were carried out (Figure 5.12). Both LybB-TEs share sequence homology to α/β hydrolase fold proteins and contain a catalytic triad consisting of Ser-His-Asp and the core-motif GXSXG. A region between α6 of TE_1 and β1 of TE_2 lacking a secondary structure was observed.

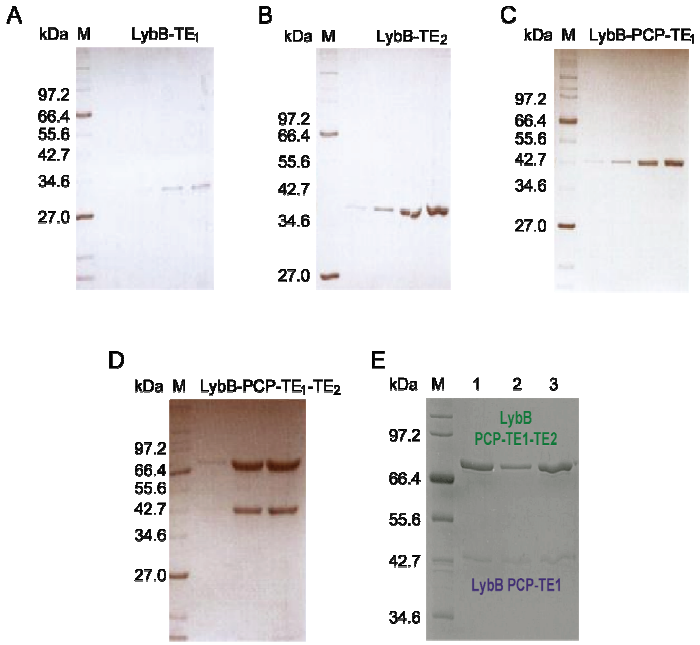
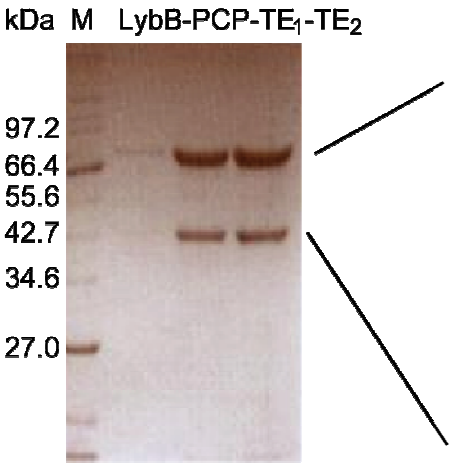


Figure 5.13: Coomassie-stained SDS-PAGE of purified recombinant thioesterases of lysobactin synthetase. Protein Marker (M) was Broad Range Protein Marker P7702 (*NEB*). A. LybB-TE_1 (34.4 kDa), B. LybB-TE_2 (34.3 kDa), C. LybB-PCP-TE_1 (46.1 kDa), D. LybB-PCP-TE_1-TE_2 (76.7 kDa), E. SDS-PAGE analysis of LybB-PCP-TE_1-TE_2 purified in the presence and absence of protease inhibitors (lane 1: Purification without inhibitors; lane 2: Purification with 1-fold concentration of protease inhibitors; lane 3: Purification with 8-fold concentration of protease inhibitors).

Overproduction of LybB-TE_1, LybB-TE_2, LybB-PCP-TE_1 and LybB-PCP-TE_1-TE_2

The gene fragments *LybB-TE_1, LybB-TE_2, LybB-PCP-TE_1 and LybB-PCP-TE_1-TE_2* were cloned in the expression vector pET28a (*Novagen*) and overexpressed in *E. coli* BL21 as described in the methods section. The recombinant His-tagged proteins were purified by

Ni^{2+}-NTA affinity chromatography. The fractions were analysed using SDS-PAGE and visualized by Coomassie stain (Figure 5.13). Concentration using centrifugal filter (*Millipore*) afforded 7.4 (LybB-PCP-TE_1-TE_2), 3.0 (LybB-PCP-TE_1), 1.6 (LybB-TE_1) and 6.4 (LybB-TE_2) mg/L media of recombinant proteins. Purification of heterologous produced LybB-PCP-TE_1-TE_2 showed a co-eluted protein band with a size of approximately 42 kDa. The two protein bands were excised from the gel and analysed via peptide mass fingerprinting. Peptide fragments obtained from tryptic digestion were identified via HRMS-analysis and are shown in Figure 5.14. To investigate if the cleavage of the full length peptide is an autocatalytic event, the mixture of PCP-TE_1-TE_2 and PCP-TE_1 was incubated for two days at various pH-values. Subsequent SDS-PAGE analysis of the reactions did not reveal an increased formation of LybB-PCP-TE_1. LybB-PCP-TE_1-TE_2 was also purified in the presence of protease inhibitors to investigate the influence of




```
  1 MGSSHHHHHH SSGLVPRGSH MAYAQRPFEA PQGETEQVLA RIWSELLQVE
 51 PIGRHDSFFD LGGHSLLVVH VVERLRKAGL QVEMRTILET PTLQAVAARA
101 QRIGEAPLSQ HLVPIRFGDR SRRPLFFIHE PTGEVLSYER LSRHFDDDLP
151 IYGIQADRAD AQREVSAQSL ARRYLEVIRS VQPQGPYRLA GWSGGGHLAY
201 EMAEQLLAQD EGVEFLGLID SGRPGNLRPE DMPQDEQRFR WDFLFSHIRY
251 LDPSLDEDAV EALEQRGSIE AAMAHARSIG WLPPSFGADE LTWRAALMKR
301 LSVACLSYRP QPLPIAVHLF TADIAADKDA SHGWAELLGD NLRIRRIGGT
351 HKSIMEEPLA ATLAAAIAGL LNDGERDPPP PSPAHTAAIT VQAGKAGAAP
401 LFCIPGAGAN VSCFLPLTQA CGAHTPIIGL QARGHDGRSV PHASVEAAVE
451 ALLPEVRALG QGGPYHLLGH SFGGWIAFEL ARRLHQAGET VAPLILVDCE
501 PPQRVRHSDH LEVLTQFVEL LEMDLGKRLP LDTQALADQL GDERDRRTLA
551 AMVAAGLMPA RATLAQFQAL LRVFAANLNA SYQPQADYPG EVVLLNAQAE
601 AKPPTVRRSL MSQAEAAEGW SRYAVRLREQ TLQGDHMSVL KSPHVGQIAA
651 AAAAAWPAHV LAAALQRDAG KHDAGAQAKD LSTQAESP

  1 MGSSHHHHHH SSGLVPRGSH MAYAQRPFEA PQGETEQVLA RIWSELLQVE
 51 PIGRHDSFFD LGGHSLLVVH VVERLRKAGL QVEMRTILET PTLQAVAARA
101 QRIGEAPLSQ HLVPIRFGDR SRRPLFFIHE PTGEVLSYER LSRHFDDDLP
151 IYGIQADRAD AQREVSAQSL ARRYLEVIRS VQPQGPYRLA GWSGGGHLAY
201 EMAEQLLAQD EGVEFLGLID SGRPGNLRPE DMPQDEQRFR WDFLFSHIRY
251 LDPSLDEDAV EALEQRGSIE AAMAHARSIG WLPPSFGADE LTWRAALMKR
301 LSVACLSYRP QPLPIAVHLF TADIAADKDA SHGWAELLGD NLRIRRIGGT
351 HKSIMEEPLA ATLAAAIAGL LNDGERDPPP PSPAHTAAIT VQAGKAGAAP
401 LFCIPGAGAN VSCFLPLTQA CGAHTPIIGL QARGHDGRSV PHASVEAAVE
451 ALLPEVRALG QGGPYHLLGH SFGGWIAFEL ARRLHQAGET VAPLILVDCE
501 PPQRVRHSDH LEVLTQFVEL LEMDLGKRLP LDTQALADQL GDERDRRTLA
551 AMVAAGLMPA RATLAQFQAL LRVFAANLNA SYQPQADYPG EVVLLNAQAE
601 AKPPTVRRSL MSQAEAAEGW SRYAVRLREQ TLQGDHMSVL KSPHVGQIAA
651 AAAAAWPAHV LAAALQRDAG KHDAGAQAKD LSTQAESP
```

Figure 5.14: Peptide mass fingerprinting analysis of the two bands in heterologous production of LybB-PCP-TE_1-TE_2. The fragments observed in peptide mass fingerprinting analysis are accentuated in red.

protease inhibition on protein cleavage. Addition of inhibitors (EDTA, AEBST, leupeptin, bestatin, aprotinin and E-64 for the inhibition of serine, cysteine and metalloproteases) did not reduce the cleavage of LybB-PCP-TE_1-TE_2[10] (Figure 5.14).

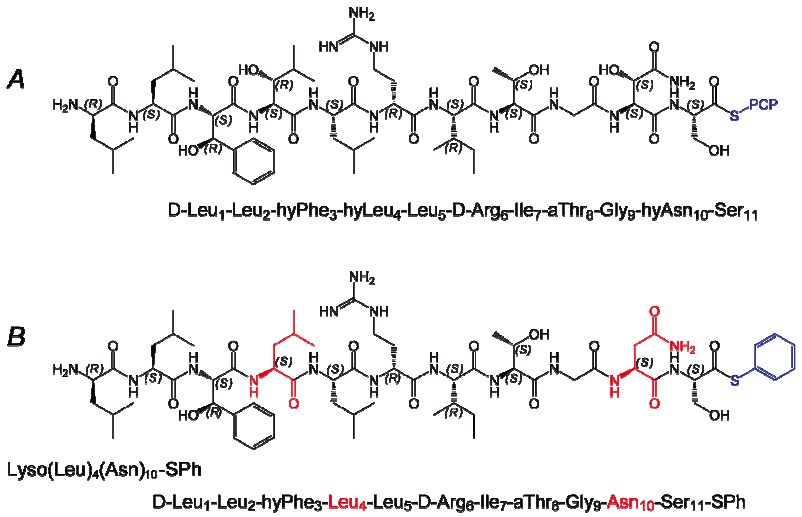


Figure 5.15: Chemical structures of linear lysobactin and corresponding artificial analog. A. Chemical structure of the linear, cognate lysobactin precursor as a PCP-bound substrate for TE-mediated macrolactonization *in vivo*. B: Chemical structure of the substrate employed in macrocyclization studies. The C-terminal thiophenol activation (blue) mimics the naturally occurring Ppant-cofactor of the PCP-bound substrate. Conservative substitutions of the commercially non-available building blocks are accentuated in red.

In vitro cyclization/hydrolyzation activity of LybB thioesterases

For biochemical characterization of LybB thioesterases, a linear peptide substrate was synthesized via solid phase peptide synthesis (as described in the methods section) and C-terminally activated with thiophenol (SPh). The chemical structure of the synthesized substrate and the linear cognate lysobactin precursor are shown in Figure 5.15. For synthetic reasons, $hyLeu_4$ and $hyAsn_{10}$ were replaced with L-Leu and L-Asn, which is why the substrate shows minor modifications of the peptide backbone. *In vitro* macrocyclization assays (as described in the methods section) were carried out to characterize the recombinant thioesterases. All reactions were monitored via liquid chromatography-mass spectrometry (LC-MS) and the results are shown in Figure 5.16. In the control reaction in absence of the enzymes, slight hydrolysis (t_R = 4.2 min) and

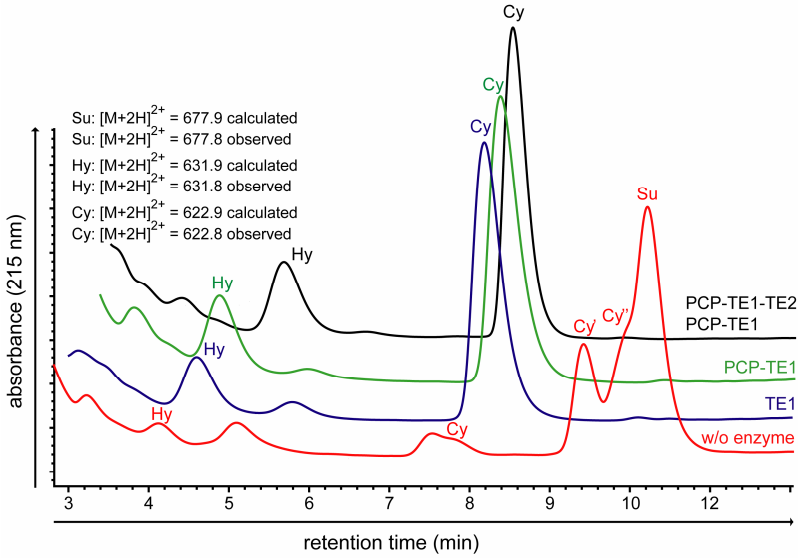


Figure 5.16: Cyclization of the peptidyl-SPh substrate Lyso(Leu)$_4$(Asn)$_{10}$-SPh (B in Figure 5.15) by LybB thioesterases, monitored by LC-MS. The red trace shows the result of the negative control reaction in absence of enzyme. The blue trace shows the result of the assay using LybB-TE$_1$. The green trace corresponds to the result of the assay using LybB-PCP-TE$_1$. The black trace shows the result of the assay using LybB-PCP-TE$_1$-TE$_2$/PCP-TE$_1$.

non-enzymatic cyclization (Cy, Cy' and Cy'', t_R = 7.5, 9.4 and 9.9 min) were observed. The macrocyclization assay with LybB-PCP-TE$_1$-TE$_2$/PCP-TE$_1$ (black trace in Figure 5.16) resulted in the complete conversion of the undecapeptidyl-thioester to the macrocyclic product (Cy, t_R = 8.28 min) and the hydrolyzed linear peptide (Hy, t_R = 5.27 min). In the negative control reaction the other two cyclization products Cy' and Cy'' were not detected. To confirm that Cy is the expected cyclization product, in which the β-hydroxyl group of hyPhe$_3$ is used as nucleophlic group, additional MS^2-analysis of Cy was carried out. The analysis revealed the formation of b-series fragment ions carrying a dehydro-phenylalanine species, which resulted from a loss of one molecule of water and is characteristic for the fragmentation of lactones[151]. The fragment ions of the y- and b-series are shown in Figure 5.17. Determination of kinetic parameters for LybB-PCP-TE$_1$-TE$_2$- / LybB-PCP-TE$_1$-mediated macrocyclization revealed the enzymes to follow Michaelis-Menten-kinetics. LybB-PCP-TE$_1$-TE$_2$ showed K_M = 1.03 mM, k_{cat} = 11.1 s^{-1} and k_{cat}/K_M = 10.8 s^{-1}·mM^{-1}, while LybB-PCP-TE$_1$ showed K_M = 0.86 mM, k_{cat} = 15.8 s^{-1} and

k_{cat}/K_M = 18.4 $s^{-1} \cdot mM^{-1}$. LybB-TE_1 was also investigated for its macrocyclization capabilities and proofed to be able to catalyze the predominant formation of the macrolactone (blue trace in Figure 5.16). The reactions catalyzed by LybB-PCP-TE_1-TE_2,

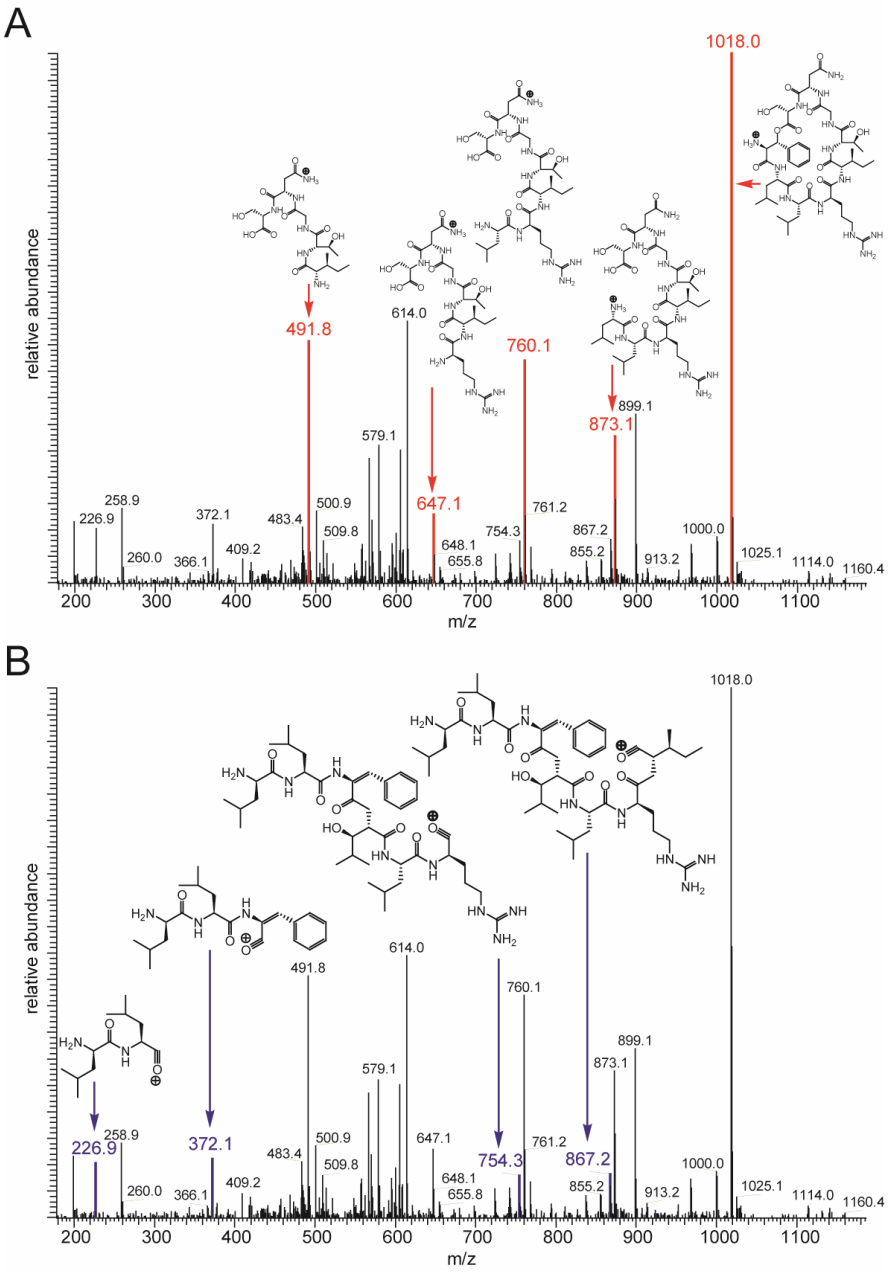


Figure 5.17: MS^2-spectra of the macrocyclic product Cy. A. The observerd y-series fragment ions (red). B. The observed b-series fragment ions (blue). The formation of dehydrophenylalanine species strongly supports the identity of Cy to be the expected macrolactone, cyclized via the hydroxyl group of $hyPhe_3$.

LybB-PCP-TE_1 and LybB-TE_1 showed cyclization-to-hydrolysis ratios of 5.7, 4.6 and 4.9 after 2h incubation at 25°C. To identify the building blocks in the lysobactin peptide sequence that are important for substrate recognition, several substrates based on Lyso$(Leu)_4(Asn)_{10}$-SPh (shown in Figure 5.15) were synthesized via SPPS. In each substrate, one single amino acid with polar sidechain at positions 6, 8, 10 and 11 was substituted by L-alanine (Figure 5.18). *In vitro* macrocyclization assays catalyzed by LybB-PCP-TE_1 showed that all substrates were cyclized with reduced yields and cyclization-to-hydrolysis ratios of about 1.

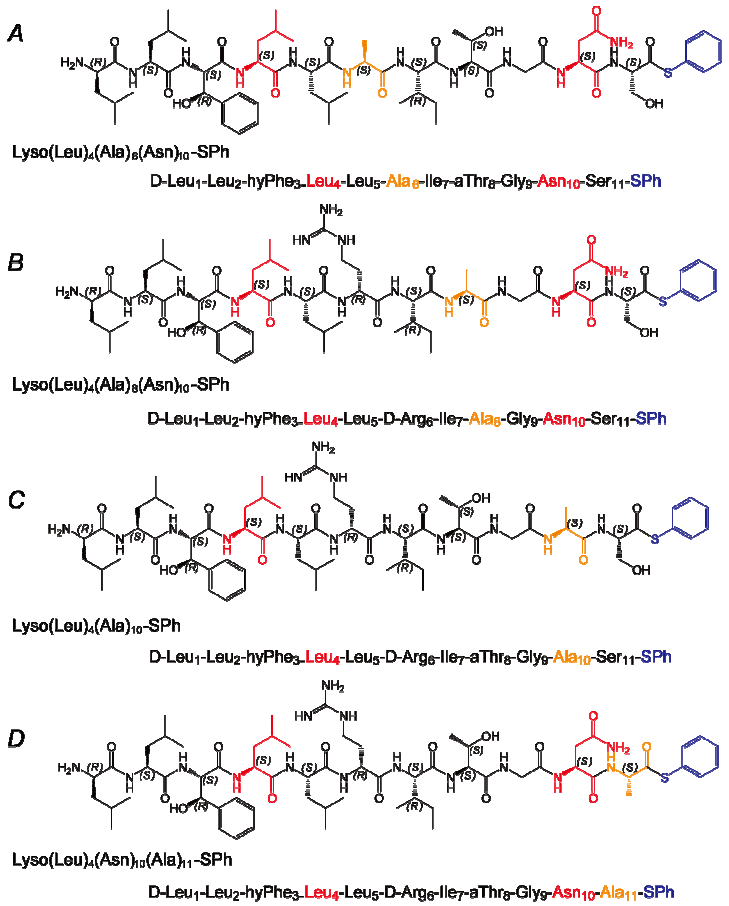


Figure 5.18: Chemical structures of peptide substrates used in alanine-scan experiment. Conservative substitutions of the commercially non-available building blocks are accentuated in red. The C-terminal thiophenol activation (blue) mimics the naturally occurring Ppant-cofactor of the PCP-bound substrate. Amino acid substitutions by L-alanine are highlighted in orange.

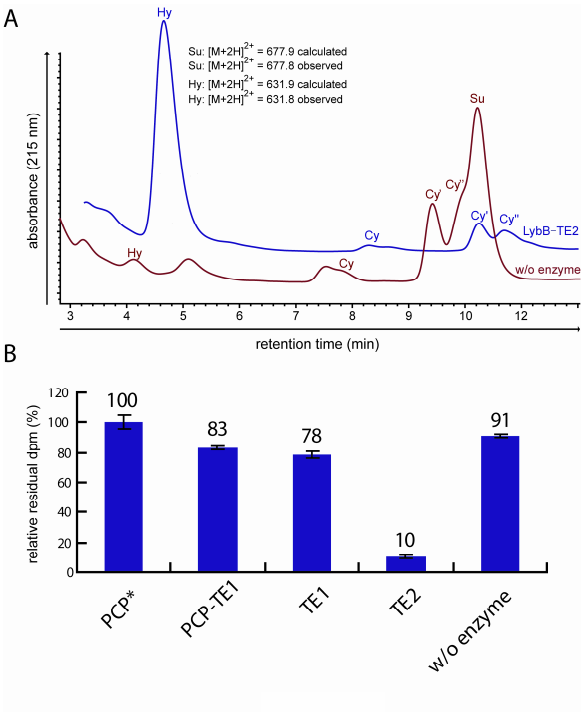


Figure 5.19: Hydrolysis of the thiophenol-activated substrate Lyso(Leu)$_4$(Asn)$_{10}$-SPh mediated by LybB-TE$_2$. A. The blue HPLC trace corresponds to the incubation of the substrate with the recombinant enzyme. The red HPLC trace corresponds to the negative control lacking the enzyme. The formation of non-enzymatically cyclized peptides Cy' and Cy'' is observed in the absence and presence of the thioesterase. Enzymatic conversion of the substrate (Su) leads to the predominant formation of the hydrolyzed peptide Hy at t_R = 4.2 min. B. Results of the deacylation studies for LybB-TEs on the stand-alone *holo*-PCP SrfA-A PCP$_1$.

The role of LybB-TE$_2$ was also investigated *in vitro* by incubating the recombinant enzyme with the substrate Lyso(Leu)$_4$(Asn)$_{10}$-SPh (Figure 5.15). A substantial conversion of the peptidyl-thioester to the corresponding hydrolysis product (Hy, t_R = 4.2 min, Figure 5.19) was observed. In addition, the three cyclization products (Cy, Cy' and Cy'') were detected in comparable ratios. LybB-TE$_2$ displayed a cyclization-to-hydrolysis ratio of 0.04. To further study the function of LybB-TE$_2$, deacylation studies employing LybB-TEs and the stand-alone SrfA-A-PCP$_1$ were carried out.[16] SrfA-A-PCP$_1$ was artificially misprimed *in vitro* utilizing Sfp and ^{14}C-acetyl-CoA. Hydrolytic cleavage of the acetyl-group was achieved by incubation of *holo*-PCP$_1$ with recombinant LybB-TEs and residual radioactivity was quantified after precipitation of the enzymes.

5.3 Initial study of the putative C_{β}-epimerase

Bioinformatic analysis of the C-domains in module 8 of LybB

Phylogenetic analysis of the two C-domains in module 8 of LybB compared to other C-domains, E-domain and C/E-dual domains from NRPSs was carried out in order to get a first insight into the possible biological functions of the two C-domains. As shown in Figure 5.20, domains with epimerization activity (E-domains, C/E-dual domains) cluster together, while domains without epimerization activity (normal C-domains) cluster together. LybB-$C_{8\text{-}1}$ and LybB-$C_{8\text{-}2}$ locate between the two groups.

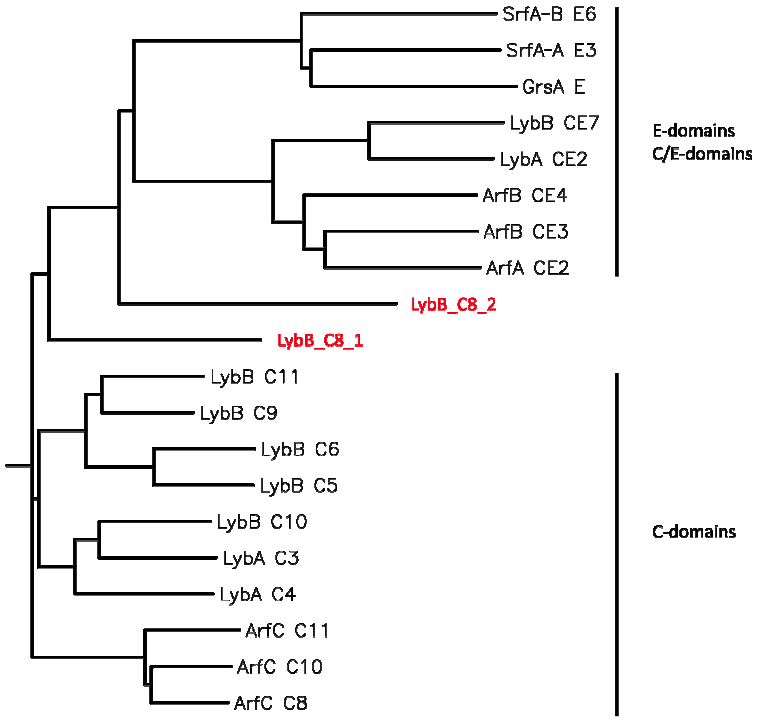


Figure 5.20: Phylogenetic analysis of LybB-$C_{8\text{-}1}$ and LybB-$C_{8\text{-}2}$ compared to other C-domains, E-domains and C/E-dual domains from NRPSs. SrfA-A, SrfA-B: surfactin synthetase, GrsA: gramicidin synthetase, ArfA, ArfB, ArfC: arthrofactin synthetase, LybA, LybB: lysobactin synthetase.

Overproduction of LybB-A_8, and LybB-C-A-PCP_8

The gene fragments *lybB-A*$_8$ and *lybB-C-A-PCP*$_8$ were cloned in the expression vector pET28a (*Novagen*) and overexpressed in *E. coli* Rosetta 1 DE3 as described in the methods section. The recombinant His-tagged proteins were purified by Ni^{2+}-NTA affinity chromatography. The fractions were analysed using SDS-PAGE and visualized by Coomassie stain (Figure 5.21). Concentration using centrifugal filter (*Millipore*) afforded 0.9 (LybB-A_8) and 1.4 (LybB-C-A-PCP_8) mg/L media of recombinant protein.

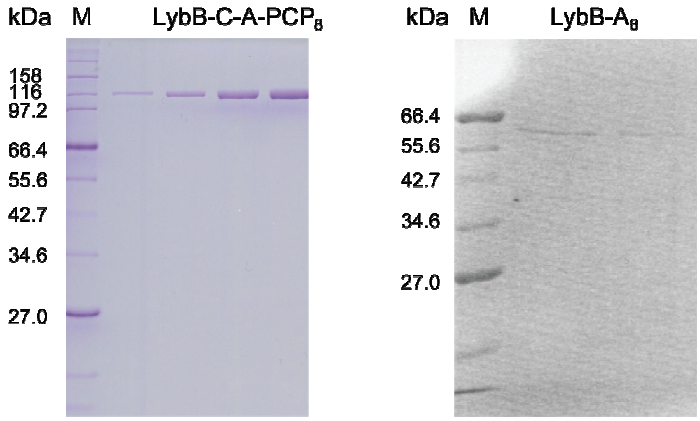


Figure 5.21: Coomassie-stained SDS-PAGE of recombinant LybB-C-A-PCP_8 and LybB-A_8. Lane M: Protein Marker, Broad Range (2-212 kDa) P7702 (*NEB*).

In vitro characterization of LybB-A_8 and LybB-C-A-PCP_8

The activity and substrate specificity of LybB-A_8 was characterized via ATP/^{32}PPi exchange assay (as described in the methodes section) using both recombinant LybB-A_8 and LybB-C-A-PCP_8. For characterization of LybB-C-A-PCP_8, different amino acids were used, including L-Thr, D-Thr, *allo*-Thr and L-Ser. The negative control was carried out using the same assay in absence of substrate. As shown in Figure 5.22, L-Thr showed the highest activity. The structurally similar amino acids D-Thr, *a*Thr and Ser showed only reduced activities of 0.03, 0.04 and 0.11 compared to L-Thr. The negative control without

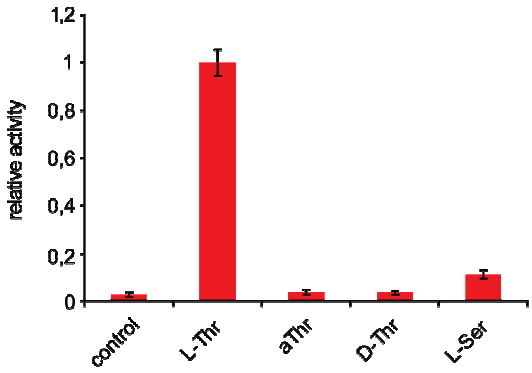


Figure 5.22: Substrate specificity analysis of recombinant LybB-C-A-PCP$_8$

substrate showed a background activity of 0.03 compared to L-Thr. In the experiment of recombinant LybB-A$_8$, no activity was observed.

To confirm that the PCP-domain in the construct LybB-C-A-PCP$_8$ was active, a fluoresceinyl-CoA phosphopentetheinylation assay was carried out as described in the methods section. After incubating with Sfp and fluoresceinyl-CoA, the reaction mixture was analyzed with SDS-PAGE. The fluorescence labeled LybB-C-A-PCP$_8$ was visualized under UV-light (λ = 312 nm) and compared with subsequent Coomassie blue staining results. Figure 5.23 clearly shows that LybB-C-A-PCP$_8$ was able to be labeled after treatment with Sfp and fluoresceinyl-CoA.

Epimerization experiment of LybB-C-A-PCP$_8$

To study if the second C-domain in module 8 has the supposed C$_\beta$-epimerization activity, the following experiment was carried out. In the first step, the purified LybB-C-A-PCP$_8$ was incubated with Sfp and CoA at 25°C for 30 min followed by addition of ATP and [^{14}C]-Thr. The control reaction was carried out by adding [^{14}C]-acetyl-CoA instead of ATP and [^{14}C]-Thr. The reactions were either quenched by adding 10% TFA and the radioactivity of precipitated protein was analyzed via liquid scintillation counting (*Packard*, Tri-carb 2100TR Liquid Scintillation Analyzer), or further treated with LybB-TE$_2$ or KOH to cleave the amino acid tethered to the PCP-domain, of which the chirality was

further analyzed via chiral TLC to reveal, if the chirality of β-C-atom of Thr was changed. The liquid scintillation counting showed that [^{14}C]-Thr was not able to be loaded onto the PCP-domain, although the apo-LybB-C-A-PCP8 was converted to holo form efficiently.

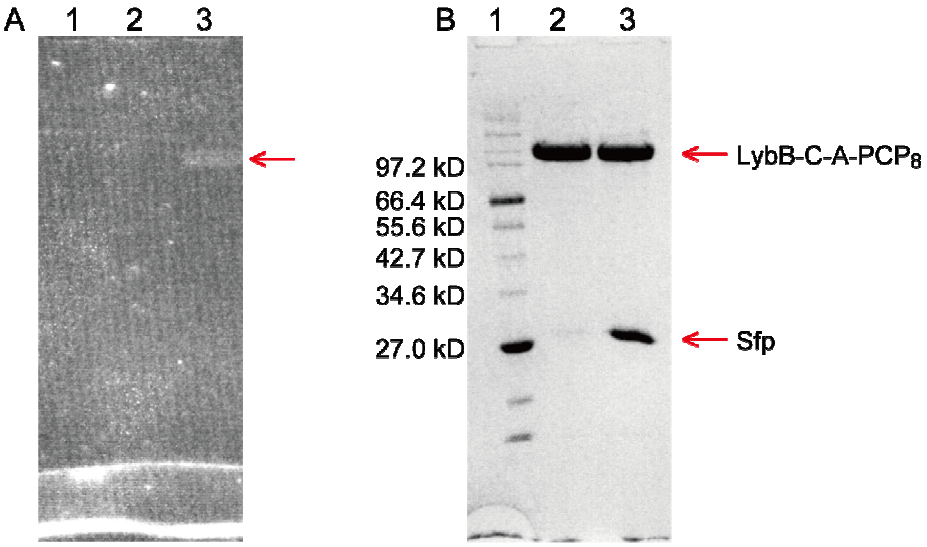


Figure 5.23: Fluoresceinyl-CoA loading assay of LybB-C-A-PCP$_8$. A. SDS-PAGE gel of LybB-C-A-PCP$_8$ after fluoresceinyl-CoA labelling assay under UV-light (λ = 312 nm). B. SDS-PAGE gel of LybB-C-A-PCP$_8$ after Coomassie blue staining. Lane 1: Protein Marker Broad Range P7702 (*NEB*). Lane 2: LybB-C-A-PCP$_8$. Lane 3: LybB-C-A-PCP8 after fluoresceinyl-CoA labelling assay.

6. Discussion

6.1 Genome sequencing of *Lysobacter sp*. ATCC 53042 and bioinformatic identification of NRPS/PKS gene clusters.

Based on the DNA-sequence of the contigs resulted from the genome sequencing of *Lysobacter sp.* ATCC 53042, searching of the potential NRP producing gene cluster was carried out. NRPS- and PKS/NRPS hybrid CDSs were discovered in 14 contigs. 4 groups of the NRPS- and PKS/NRPS CDSs seem logical and were further analysed.

6.1.1 PKS/NRPS hybrid coding genes in contig 40

In contig 40, two overlapping genes encoding one PKS and one NRPS were identified. The PKS encoded by *orf* 6 consists of 5 domains (KS, AT, DH, KR and ACP). According to its organization, this module is supposed to be responsible for incorporating an α-β-unsaturated building block. A necessary starting module, normally consisting of one AT-domain and one ACP-domain responsible for initiating the polyketide synthesis, is missing. No genes encoding a starting module could be found in the flanking area. This could be compensated by a distant gene encoding starting module located in another contig or in the sequencing gaps. Another possibility is, that the AT and ACP domain in the first module could be responsible for inititation of the polyketide synthesis. The α-β-unsaturated building block tethered on the ACP-domain is supposed to condense with the amino acid activated by the A-domain in module 2 and therefore the elongating chain is transferred from PKS to NRPS encoded by *orf* 5. The NRPS features an N-terminal domain with a size of 554 amino acids, which shows no similiarity to known protein. Although it could be predicted that one function of this domain could be involved in the communication between the PKS and NRPS, enabling the transfer of the growing chain from PKS to NRPS, the exact function of the domain could not be proposed, unless the structure of the natural product is resolved. The rest part of the NPRS (orf 5) consists of four regular modules, responsible for incorporating four further building blocks.

Bioinformatic analysis of the extracted residues of the A-domains determining the A-domain specificity suggested the primary sequence of the NRP to be L-(Thr/Ser)-L-(His/Ser)-L-Ser-L-(Ser/Asp). The C-terminal TE-domain should be responsible for release of the final product in a macrocyclized or hydrolysed form. Using these A-domain specificity predictions, product structure of this PKS/NRPS hybrid biosynthetic machinery could be predicted and is shown in Figure 5.2. In the presented structure, the hydroxyl group in the second building block was used as the nucleophile group for the macrocyclization catalyzed by C-terminal TE-domain. The hydroxyl groups in R_3 and R_4 are also possible nucleophile groups for the macrocyclization.

6.1.2 PKS/NRPS hybrid coding genes in contig 233

In contig 233, three NRPS-encoding genes (*orf* 30-32) showing the same direction of transcription were identified. *Orf* 32 encodes a NRPS containing 3 modules and 9 domains. The modules in orf 32 exhibit regular organization (C-A-PCP in the extending module, A-PCP in the initial module). Instead of a TE-domain, an R-domain was found at the C-terminal end of the synthetase. Generally, R-domains are known to catalyse either reduction of heterocyclen derived from Thr/Ser/Cys, or reductive release of the final product in form of an aldehyde or a primary alcohol. Due to the absence of Thr/Ser/Cys in the predicted primary structure of the natural product, the C-terminal R-domain could rather be responsible for NADPH-dependent reduction of the C-termnal carboxylic thioester to an alcohol or aldehyde and the release of the final product. The R-domain showed 44% identity to the C-terminal R-domain in MxaA from *Stigmatella aurantiaca* (accession number: AAK57184.1) with E-value of 7e-70. MxaA is a tetradomain NRPS (C-A-PCP-Re) involved in the biosynthesis of myxalamid S, in which the C-terminal R-domain is responsible of NADPH-dependent reduction and release of the final product.[139] This BLAST result supports also the prediction, that the orf 32 could be responsible for biosynthesis of a primary alcohol or aldehyde originated from a tripeptide L-Val-L-Pro-L-Phe.

Orf 31 consists of 2 modules and 5 domains, while orf 30 consists of 1 module and 5 domains. The first module in orf 32 is an initial module (A-PCP). The absence of an N-terminal C-domain excludes the possibility that orf 32 and 31 belong to the same NRPS system. The tripeptide synthesized by orf 32 could not be handed over to orf 31 and undergoes further extension steps. Orf 30 exhibits at N-terminal an unknown domain and a C-terminal R-domain instead of a TE-domain. The biological function of the N-terminal unknown domain can be only predicted after the structure of the natural product is solved. Due to the absence of Thr/Ser/Cys in the predicted primary structure of the natual product and the high similarity of the C-terminal R-domain to the C-terminal R-domain in MxaA, the C-terminal R-domain in orf 30 could be also responsible for NADPH-dependent reduction of the C-terminal carboxylic thioester of the peptide chain tethered on the upstream neighbouring PCP-domain to aldehyde or primary alcohol and the release of the final product.

By summarizing the information of all three CDSs, the following suggestion could be made. *Orf* 32 encodes a NRPS, producing a primary alcohol derived from a tripeptide (Trp-Pro-Trp), while *orf* 31 and 30 encode two NRPSs, producing another primary alcohol originated from a tripeptide (Pro-Val-Pro). The predicted structures are shown in Figure 5.3.

6.1.3 NRPS coding genes in contig 306

In contig 306, one PKS coding gene (*orf* 49) and 3 NRPS coding genes (*orf* 1, 2 and 4) were identified. Analysis via PKS/NRPS analysis showed that *orf* 49 encodes a type I PKS lacking the essential AT-domains. This PKS could be an inactive gene and will not be further discussed. *Orf* 4 encodes a distinct TE-domain, which could serve as an editing typ II TE for PKS- or NRPS systems. *Orf* 2 encodes a NRPS with 2 modules and 7 domains. The first module starts with a C-domain, suggesting condensation of an N-terminal acyl chain resulting in a lipo-peptide-like final product. An additional PCP-domain was observed in the first module. Which PCP-domain is essential for the biosynthesis and the

roll of the other PCP-domain is entirely unknown. After the incorporation of the second amino acid, the peptide chain is supposed to be transfered to the NRPS encoded by *orf* 1. After the first PCP-domain in orf 1 is an unknown domain, which shows no similarity to known protein. The biological function of this unknown domain can only be predicted after the structure of the natural product is solved. The next module contains an E-domain after the PCP-domain, suggesting incorporation of a D-configured amino acid. In the next module, only the sequence information of the N-terminal part of a C-domain (388 amino acids) is known. The sequence of the rest of the synthetase was missing because of the incomplete sequencing. The missing part could locate in the sequencing gaps or on the beginning of other contigs. One possibility is *orf* 2 in contig 294, which encodes a PCP-TE didomain. The N-terminus of this didomain is not fully known because of incomplete sequencing. PCR experiment could be carried out to verify if the orf 2 in contig 294 and orf 1 and 2 in contig 306 belong to the same NRPS assembly.

In summary, the *orf* 1 and 2 in contig 306 could encode a part of an NRPS machinery producing a lipo-peptide-like product. From the known DNA-sequence, only the N-terminal part of the primary structure of this lipo peptide could be proposed to be R-Asp_1-Val/Leu_2-Arg_3-Asp_4-, in which R stands for a fatty acid.

6.1.4 NRPS coding genes in contig 350

In contig 350, two NRPS coding genes (*orf* 1 and 3) were identified. These two *orf* encode part of an NRPS machinery. A part of the primary structure of this peptide can be proposed from the known NRPS sequence and could be -L-Lys-L-Pro-L-Lys-D-Asp-.

To verify if these 4 NRPS or PKS-NRPS hybrid assemblies discussed above are active, radioactive precursor feeding and gene inactivation/compensation could be performed. To study the structure of the products of these NRPS assemblies, natural product isolation and NMR studies could be carried out. A-domain specificity test *in vitro* could be also carried out to verify the correlation between the NRPS assembly and the natural product.

6.2 Identification and characterization of lysobactin biosynthesis gene cluster

Lysobactin, first isolated at Squibb Institute of Medical Research from *Lysobacter sp.* ATCC 53042[82, 83], is a potential antibiotic agent with high activity against MRSA and VRE. Efforts have been made for derivatization of this compound.[84, 86, 88] In 1996, Bernhard et al. have reported the identification of a 4.6 kb gene segment involved in the biosynthesis of lysobactin in *Lysobacter sp.*.[1] In this work, we have sequenced the chromosome of the lysobactin producing strain *Lysobacter sp.* ATCC 53042 and used the reported 4.6 kbp gene fragment to identify the biosynthetic machinery of lysobactin.

6.2.1 Characterization of A-domains in lysobactin synthetase.

In the primary sequence of lysobactin, there are 3 β-hydroxylated amino acids, $hyPhe_3$, $hyLeu_4$ and $hyAsn_{10}$. The extracted specificities determining residues of lybA-A_3, which is responsible of incorporation of the $hyPhe_3$ in the final product, showed 80% identity comparing to BacC-A_2. This variation in the specificity determining motif might indicate the direct activation and incorporaton of β-OH-Phe into lysobactin instead of the activation of L-Phe followed by hydroxylation *in trans* as reported in kutzneride biosynthesis.[47] The suggested model for lysobactin biosynthesis would require hydroxylation of free amino acid substrate, as reported in the viomycin biosynthesis.[152] In contrast to LybA-A_3, the specificity conferring residues of LybA-A_4 and LybB-A_{10} are identical to those found in Leu-activating A-domain SrfA-A-A_3 and Asn-activating A-domain NosC-A3, which are involved in biosynthesis of surfactin[149] and nostopeptolide A.[150].

Based on the sequence information of *lybA* and *lybB*, several A-domains (A_1, A_6 and A_9) were expressed and purified. The characterization was carried out using the ATP/^{32}PPi exchange assay with different amino acids as substrates. As shown in Figure 5.10, LybA-A_1 showed highest activity towards L-Leu and D-Leu. Besides the proposed cognate substrate, LybA-A_1 showed also activities against other structurally related amino acids

such as L-/D-Ile or even structurally distinct amino acids such as L-Ser and L-Phe. This relaxed substrate specificity was also reported by TycA-A_1, involved in tyrocidine biosynthesis.[9] Generally, the substrate specificity of A-domains are considered to be the gate-keeper ensuring the incorporation of correct building blocks into the final product.[10] The flexibility in the substrate recognition of the A-domain could facilitate the production of derivatized natural product[9]. In contrast, LybB-A_6 showed strict substrate specificity towards L-Arg and D-Arg, which is consistent with the prediction that L-Arg is the cognate substrate for LybB-A_6. The result of LybB-A_9 showed clearly that sorely Gly is accepted by LybB-A_9 *in vitro*, which is consistent with the predicted biosynthesis model. Analysis of A-domain specificities matchs the prediction in full agreement and confirms that LybA and LybB to mediate the biosynthesis of lysobactin following a linear logic. Intriguingly, both LybA-A_1 and LybB-A_9 showed almost equal activities towards L- and D-configured substrates. The incorporation of correct L-enantiomer could be ensured due to the relative low cytoplasmic abundance of D-configured substrates. The acceptor site specificity of the upstream C-domain could also contribute in preventing the incorrect incorporation of D-configured building block.[15]

6.2.2 Characterization of LybB thioesterase activities

In contrast to the well-studied thioesterases of the type I and II introduced above, an unusual type of tandem thioesterase architecture locating at the C-terminal end of the assembly line was found in the lysobactin synthetase, which was only reported in other few systems such as arthrofactin, massetolide and syringopeptin.[67, 142, 153] *In vivo* studies of arthrofactin tandem TE showed that the inactivation of the first TE-domain via Ser/Ala mutation led to total abolishment of the arthrofactin production, while inactivation or deletion of the ArfC-TE_2 drastically decreased the arthrofactin production.[66] These results confirmed both TEs in arthrofactin assembly to be directly involved in the efficient production of the natural product.

The phylogenetic study of LybB-TE_1 and LybB-TE_2 compared with other thioesterases from NRPSs revealed the TEs from tandem TE organizations (synthetases of arthrofactin, massetolide, syringopeptin and lysobactin) cluster together. It was speculated that the TE_1s from these tandem TEs represents a novel subclass of thioesterase as the natural products of these group are all macrolactones. The TE_2s from these tandem TEs constitute another discrete subclass other than that of TE II. This leds to the suggestion that the TE_2 should be not a variant of external TE II. As the TEs from the tandem TE organizations cluster together, it was suggested that the tandem TEs could share a common mode of action. Further secondary structure prediction of LybB-TE_1 and LybB-TE_2 suggested that both TEs represent the typical α-β-hydrolase fold and contain the core motif GXSXG and the catalytic triad Ser-His-Asp. These bioinformatic analysis results suggested that both LybB-TEs to exhibit either cyclization or hydrolyzation activity.

To investigate the role of the two TEs in the cyclization process of lysobactin biosynthesis, LybB-TE_1, LybB-TE_2, LybB-PCP-TE_1-TE_2 and LybB-PCP-TE_1 were heterologously produced in *E. coli* BL21. The heterologous production of LybB-PCP-TE_1-TE_2 gave rise to a protein with size of about 46 kD. Peptide mass fingerprinting analysis confirmed this protein to be LybB-PCP-TE_1^*. The primary sequence of the protein is longer than that of the constructed LybB-PCP-TE_1 in the C-terminal. So that the contamination from the constructed LybB-PCP-TE_1 can be excluded. This observed LybB-PCP-TE_1^* is supposed to be resulted from degradation of LybB-PCP-TE_1-TE_2. The protein mixture of LybB-PCP-TE_1-TE_2 and LybB-PCP-TE_1^* was incubated at room temperature at different pH-value for two days. Subsequent SDS-PAGE did not reveal increase of degradation product. Thus, an autocatalytic degradation could be ruled out. Purification of LybB-PCP-TE_1-TE_2 in presence of various protease inhibitors (EDTA, AEBSF, leupeptin, bestatin, aprotinin and E-64 for the inhibition of serine proteases, cysteine proteases and metalloproteases) was carried out,

which did not reduce the cleavage of the LybB-PCP-TE_1^*. Taking these results together, it could be suggested that an intracellular *E. coli* protease could be involved in the posttranslational cleavage of the enzyme during the heterologous production. Further suggestion could be made that the LybB-TE_2 is also cleaved off posttranslationally from the native LybB prior to the biosynthesis of lysobactin. The resulted stand-alone LybB-TE_2 works as a TE II and is responsible for regenerating the mis-primed PCP-domains on the lysobactin assembly, although it could be only speculated which specific protease is involved in this cleavage process. This suggestion is further supported by the result that LybB-TE_2 could cleave acetyl groups from the stand-alone holo-PCP *in vitro*, which will be discussed in detail in the next chapter. A sequential anaylsis of both LybB-TEs, secondary structure elements and multiple sequence alignments with other tandem TEs revealed a proline-rich region between α6 of TE_1 and β1 of TE_2 lacking secondary structure motifs. This region is considered to be the linker between TE_1 and TE_2 and could contain a putative protease site, although it can only be speculated if a specific cleavage site is involved in the cleavage process. The conserved proline residue which was considered to be the N-terminus of ArfC TE_2 is also located within this region.[10, 66]

To study the function of the individual thioesterases in lysobactin assembly, the four heterologously produced LybB thioesterases (LybB-PCP-TE_1-TE_2, LybB-PCP-TE_1, LybB-TE_1 and LybB-TE_2) were biochemically characterized *in vitro* using the artificial peptidyl substrate Lyso$(Leu)_4(Asn)_{10}$-SPh (Figure 5.15). This *in vitro* characterization bases on the fact that excised TE-domains remain catalytically active and can be incubated with synthetic substrates for the evaluation of their inherent catalytic potential.[62] The substrate was synthesized via SPPS and C-terminally activated with thiophenol to imitate the naturally occurring PCP-ppant-bound substrate.[154] Due to synthetic reasons, minor changes were made on the undecapeptide backbone by omitting the β-hydroxygroup at building blocks 4 and 10. The stereochemistry was conserved throughout the peptide. To our knowledge, these minor modifications can influence the *in vitro* characterization of

thioesterases only slightly.

In the control reaction without enzyme, slight hydrolysis (Hy, t_R = 4.2 min) and three different peaks showing the mass of cyclization products (Cy, Cy' and Cy'' with t_R = 7.5, 9.4, 9.9 min) were observed, suggesting three different cyclization products were formed. In the structure of the applied substrate Lyso(Leu)$_4$(Asn)$_{10}$-SPh (Figure 5.15), three potential nucleophile groups are present, namely the primary amino group of D-Leu$_1$ and the secondary hydroxyl groups of hyPhe$_3$ and *a*Thr$_8$. The three different cyclization products could be formed via non-catalytic intramolecular nucleophilic attack of these nucleophile groups onto the C-terminal thioester of the substrate.

Incubating of the substrate with recombinante enzyme mixture LybB-PCP-TE$_1$-TE$_2$/ LybB-PCP-TE$_1$* resulted in completely conversion of peptidyl-thioester substrate into primary cyclization product Cy and minor hydrolysis product Hy. The other two cyclzation products Cy' and Cy'' observed in the control reaction were not at present, indicating that Cy is the expected cyclization product catalyzed by thioesterase with the secondary hydroxyl group of hyPhe$_3$ as nucleophile group. This was further confirmed by the MS^2-analysis of the cyclization product Cy. Analogously, the incubation of the same substrate with the other two recombinant enzymes LybB-PCP-TE$_1$ and LybB-TE$_1$ also resulted in fully conversion of the substrate into primary cyclization product Cy and minor hydrolysis product Hy. In contrast, the incubation of the substrate with recombinant enzyme LybB-TE$_2$ resulted in primary hydrolysis product Hy, all the three cyclization products (Cy, Cy' and Cy'') were also observed. This suggests that LybB-TE$_2$ mainly catalyzes the hydrolysis of the substrate. However the reaction is much slower than the cyclization reaction catalyzed by LybB-TE$_1$, due to the fact that the non-catalytic reactions were not totally oppressed.

In the further kinetic studies of the macrocyclization reaction catalyzed by LybB-PCP-TE$_1$-TE$_2$ and LybB-PCP-TE$_1$, both recombinant enzymes were found to follow the Michaelis-Menton-kinetics. The kinetic parameters of both enzymes showed only

minimal differences (the data are shown in section 5.2.4). The *in vitro* cyclization assay mediated by these two recombinant enzymes discussed above did not show noteworthy difference either. These results suggest that the additional internal thioesterase domain LybB-TE_2 does not influence the product formation. The comparison of the kinetic parameters of LybB-PCP-TE_1-TE_2 and LybB-PCP-TE_1 with that of recombinant thioesterase derived from other NRPS systems (DptD-PCP-TE involved in daptomycin biosynthesis and A54145-PCP-TE involved in A54145 biosynthesis) revealed that the lysobactin thioesterases display lower substrate affinity and higher turnover numbers (s. Table 6.1).

Table 6.1: The kinetic parameters of recombinant thioesterases

	LybB-PCP-TE_1-TE_2	**LybB-PCP-TE_1**	**DptD-PCP-TE[155]**	**A54145-PCP-TE[156]**
K_M	1.03 mM	0.86 mM	50.1 µM	80.2 µM
k_{cat}	11.1 s^{-1}	15.8 s^{-1}	0.003 s^{-1}	0.0036 s^{-1}

In this study, the function of LybB-TE_2 was investigated *in vitro*. Incubation of recombinant LybB-TE_2 with the substrate Lyso$(Leu)_4(Asn)_{10}$-SPh led to substantial hydrolysis. In addition, all the three cyclization products (Cy, Cy' and Cy'') were also observed with comparable ratio as observed in the negative control reaction without enzyme. The cyclization-to-hydrolysis ratio of the reaction catalyzed by TE_2 was 0.04. Therefore, LybB-TE_2 is considered to catalyze solely hydrolysis of the linear lysobactin precursor. We have suggested that the LybB-TE_2 could be posttranslationally cleaved and act as a TE II editing domain. The deacylation study (s. section 5.2.4) revealed that LybB-TE_2 hydrolyzes the PCP-bound acetyl group efficiently. In contrast, LybB-TE_1 and LybB-PCP-TE_1 did not result in substantial formation of deacylated PCP. The results of this deacylation study contradicts the postulated mechanism of ArfC thioesterase reported by Roongsawang et al., which suggests that the peptidyl chain is transferred onto the active serine of ArfC-TE_1 and then onto the active serine of ArfC-TE_2, before it undergoes

an intramolecular nucleophilic attack catalyzed by ArfC-TE_2, which leads to macrolactonization and release of the final product.

The results achieved in this study assign different functions to the individual LybB thioesterases. The *in vitro* characterization of excised LybB-TE_1 confirmed clearly that the first thioesterase of LybB conducts solely the cyclization and the subsequent release of the product. The presence of TE_2 did not influence the yield or the cyclization-to-hydrolysis ratio of the macrocyclization reaction. This suggests the macrocyclization to be the preferred reaction. In contrast to that of LybB-TE_1, LybB-TE_2 mediated exclusively the hydrolysis of the peptidyl thioester substrate and the mis-primed PCP domain *in vitro*. Combined with the proteolytic cleavage of LybB-PCP-TE_1-TE_2 observed in the heterologous production, we suggested that the LybB-TE_2 could be posttranslationally cleaved and acts as a stand-alone TE II. This hypothesis could also explain the observation that inactivation of ArfC-TE_2 drastically decrease the production yield of arthrofactin, as disruption of TE II-encoding genes in NRPS- or PKS-gene clusters commonly leads to drastically decrease of the corresponding natural product, which can be complemented by a heterologous TE II.[157, 158] Taking the results together, it is concluded that LybB-TE_1 mediates the macrocyclization and release of the final product, while LybB-TE_2 deacylates the mis-primed PCPs *in cis* or *in trans*, ensuring a continuous production.

6.2.3 Initial study of the putative C_β-epimerase

Besides the D-configured and β-hydroxylated amino acids, lysobactin contains a further unusual building block, *allo*-Thr, at position 8. *Allo*-configured amino acids were observed in several NRPs exemplified by enduracidin A and coronatine. Enduracidin A, produced by *Streptomyces fungicidicus* ATCC 21013, is active against Gram-positive bacteria such as MRSA and VRE. This branched cyclic peptide disrupts the biosynthesis of the bacteria cell wall by blocking the elongation step of peptidoglycan biosynthesis.[159, 160] Enduracidine contains one D-*allo*-Thr and one L-*allo*-Thr at position 5 and 8. The

corresponding A-domains of which were reported to activate *allo*-configured amino acids specifically.[29] Phytotoxin coronatine is produced by phytopathogenic bacterium *Pseudomonas syringae* containing in the structure one coronamic acid building block, which is derived from L-*allo*-Thr. An *allo*-Thr activating A-domain was reported to be involved in the biosynthesis of coronatine.[30] The source of *allo*-configured amino acids in the biosynthesis of these natural products is so far unknown.

Module 8 in the lysobactin biosynthesis machinery, which is responsible for incorporating *allo*-Thr into the final product, contains an additional C-domain. The *in vitro* ATP/^{32}PPi exchange assay revealed the A-domain in this module to activate L-Thr specifically. The activities of *allo*-Thr and D-Thr relative to that of L-Thr did not exceed 0.04. This result clearly suggests that the chirality conversion of the β-carbon of the threonine should occur after Thr is activated and loaded onto the corresponding PCP-domain. Interestingly, it was observed that among the three recombinant proteins produced heterologously, namely LybB-C-A-PCP$_8$, LybB-A-PCP$_8$ and LybB-A$_8$, only the activity of LybB-C-A-PCP$_8$ could be observed in the ATP/^{32}PPi exchange assay *in vitro*. A possible reason for this could be that the neighbouring flanking domains ensure the expression of intact A-domain with proper folding.

Because of the similarity between E- and C-domains, it was postulated that one C-domain in module 8 could mediate the epimerization of the C$_\beta$-carbon, while the other C-domain catalyzes the peptide bond formation and meanwhile transfers the growing peptide chain from the upstream PCP- onto the PCP-domain in the current module. A phylogenetic analysis of the both C-domains in module 8 compared with C-domains, E-domains and C/E dual domains from other NRPS systems was carried out, trying to distinguish the both C-domains observed in module 8. As shown in Figure 5.20, the C-domains from different NRPS systems cluster together in the lower part, whereas the proteins with epimerization activity, including E-domains and C/E dual domains, cluster together in the upper part of the phylogenetic tree. The two C-domains from module 8

of LybB locate between these two groups. This phylogenetic study did not indicate which C-domain could mediate the epimerization of the β-carbon of the PCP-tethered threonine.

To investigate which C-domain has the epimerization activity, an *in vitro* epimerization assay was planed using organically synthesized Thr-SNAC as substrate. Heterologous expression of the both C-domains was carried out in *E. coli strains* (BL21 and Rosetta 1). The overexpression of LybB-$C_{8\text{-}1}$ was not detectable, while recombinant LybB-$C_{8\text{-}2}$ was achieved in good yield. The recombinant LybB-$C_{8\text{-}2}$ was incubated with Thr-SNAC and analysed via HPLC-MS and compared with the standard Thr-SNAC/*a*Thr-SNAC synthesized organically. It was not very surprising, that the analysis result did not give any evidence of epimerization. It was reported, that *in vitro* epimerization assay of E-domains incubated with corresponding substrate C-terminally activated with SNAC did not yield any epimerization product.[161] One possible reason of the negative result achieved in this study could be that the SNAC-activated substrata could be not accepted by the LybB-$C_{8\text{-}2}$ because SNAC only mimics a short part of the natural substrate. It could not be ruled out that LybB-$C_{8\text{-}2}$ does not have an epimerization activity.

The C-domains in module 8 was further investigated. After the *apo*-LybB-C-A-PCP_8 was pre-treated with $MgCl_2$, CoA and Sfp, the resulted *holo*-LybB-C-A-PCP_8 was then incubated *in situ* with [^{14}C]-Thr and ATP, the enzyme was precipitated by adding 10% TCA. The resulted enzyme pellet was treated with KOH followed by centrifugation. The supernatant was analyzed with chiral-TLC. It was expected that the [^{14}C]-Thr could be activated and loaded onto the PCP-domain. The β-carbon of PCP-tethered [^{14}C]-Thr could be epimerized by the C-domain and detected with chiral-TLC after KOH-cleavage. But the chiral-TLC result did not show any trace of radioactive substance despite considerable efforts. This result indicates that Thr could not be loaded onto the PCP-domain under the conditions applied, although it could be activated by the A-domain in the LybB-C-A-PCP_8 construct, which was proved by the ATP/PPi-exchange assay.

To circumvent the loading step catalyzed by the A-domain, which could be the main obstacle of the experiment discussed above, it was attempted to load the recombinant LybB-C-A-PCP_8 with synthesized Thr-CoA and Sfp. The Thr was subsequently cleaved by KOH or LybB-TE_2 and analyzed via LC-MS. One possible hindrance of this experiment is that the turnover number of this reaction is limited to one, which could make the subsequent detection via HPLC-MS difficult. Parallel reactions were pooled together and screened via HPLC-MS. No formation of *a*Thr could be observed.

In the initial study of the putative β-carbon epimerization domain, we have proved that the A-domain in module 8 accepts specifically L-Thr, which indicates an on-line epimerization of the β-carbon. Different epimerization possibilities could be put forth based on the well-known epimerization mechanismen of the C_α-carbon. The β-carbon of Thr could be epimerized before or after the condensation with the upstream peptidyl-chain.[74] Despite considerable efforts, no epimerization could be observed in this study. Further study with synthesized peptidyl-CoA could be applied to test if the epimerization occurs after condensation with the upstream peptide chain.

7. References

1. Bernhard, F., et al., *Identification of genes encoding for peptide synthetases in the gram-negative bacterium Lysobacter sp. ATCC 53042 and the fungus Cylindrotrichum oligospermum.* DNA Seq, 1996. **6**(6): p. 319-30.
2. Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs over the last 25 years.* J Nat Prod, 2007. **70**(3): p. 461-77.
3. Sieber, S.A. and M.A. Marahiel, *Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides.* J Bacteriol, 2003. **185**(24): p. 7036-43.
4. Sieber, S.A. and M.A. Marahiel, *Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics.* Chem Rev, 2005. **105**(2): p. 715-38.
5. Felnagle, E.A., et al., *Nonribosomal peptide synthetases involved in the production of medically relevant natural products.* Mol Pharm, 2008. **5**(2): p. 191-211.
6. Walsh, C.T., *Polyketide and nonribosomal peptide antibiotics: modularity and versatility.* Science, 2004. **303**(5665): p. 1805-10.
7. Eriani, G., et al., *Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs.* Nature, 1990. **347**(6289): p. 203-6.
8. Francklyn, C.S., *DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression.* Biochemistry, 2008. **47**(45): p. 11695-703.
9. Schaffer, M.L. and L.G. Otten, *Substrate flexibility of the adenylation reaction in the Tyrocidine non-ribosomal peptide synthetase.* Journal of Molecular Catalysis B: Enzymatic, 2009. **59**(1-3): p. 140-4.
10. Hou, J., L. Robbel, and M.A. Marahiel, *Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture.* Chem Biol, 2011. **18**(5): p. 655-64.
11. Conti, E., et al., *Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.* EMBO J, 1997. **16**(14): p. 4174-83.
12. Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chem Biol, 1999. **6**(8): p. 493-505.
13. Samel, S.A., et al., *Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase.* Structure, 2007. **15**(7): p. 781-92.
14. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.* Nat Struct Biol, 2002. **9**(7): p. 522-6.
15. Linne, U. and M.A. Marahiel, *Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization.* Biochemistry, 2000. **39**(34): p. 10439-47.
16. Kraas, F.I., et al., *Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation.* Chem Biol. **17**(8): p. 872-80.

17. Duerfahrt, T., et al., *Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis.* Chem Biol, 2004. **11**(2): p. 261-71.
18. Rausch, C., et al., *Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution.* BMC Evol Biol, 2007. **7**: p. 78.
19. Koglin, A., et al., *Conformational switches modulate protein interactions in peptide antibiotic synthetases.* Science, 2006. **312**(5771): p. 273-6.
20. Yanai, K., et al., *Para-position derivatives of fungal anthelmintic cyclodepsipeptides engineered with Streptomyces venezuelae antibiotic biosynthetic genes.* Nat Biotechnol, 2004. **22**(7): p. 848-55.
21. Feifel, S.C., et al., *In vitro synthesis of new enniatins: probing the alpha-D-hydroxy carboxylic acid binding pocket of the multienzyme enniatin synthetase.* Chembiochem, 2007. **8**(15): p. 1767-70.
22. Haese, A., et al., *Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in Fusarium scirpi.* Mol Microbiol, 1993. **7**(6): p. 905-14.
23. Haese, A., et al., *Bacterial expression of catalytically active fragments of the multifunctional enzyme enniatin synthetase.* J Mol Biol, 1994. **243**(1): p. 116-22.
24. Worthington, A.S. and M.D. Burkart, *The old is new again: asparagine oxidation in calcium-dependent antibiotic biosynthesis.* ACS Chem Biol, 2007. **2**(3): p. 152-4.
25. Strieker, M., et al., *Mechanistic and structural basis of stereospecific Cbeta-hydroxylation in calcium-dependent antibiotic, a daptomycin-type lipopeptide.* ACS Chem Biol, 2007. **2**(3): p. 187-96.
26. Muller, C., et al., *Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic Friulimicin in Actinoplanes friuliensis.* Antimicrob Agents Chemother, 2007. **51**(3): p. 1028-37.
27. Nelson, J.T., et al., *Characterization of SafC, a catechol 4-O-methyltransferase involved in saframycin biosynthesis.* Appl Environ Microbiol, 2007. **73**(11): p. 3575-80.
28. Heinzelmann, E., et al., *A glutamate mutase is involved in the biosynthesis of the lipopeptide antibiotic friulimicin in Actinoplanes friuliensis.* Antimicrob Agents Chemother, 2003. **47**(2): p. 447-57.
29. Yin, X. and T.M. Zabriskie, *The enduracidin biosynthetic gene cluster from Streptomyces fungicidicus.* Microbiology, 2006. **152**(Pt 10): p. 2969-83.
30. Vaillancourt, F.H., et al., *Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis.* Nature, 2005. **436**(7054): p. 1191-4.
31. Hoffmann, K., et al., *Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis.* J Biol Chem, 1994. **269**(17): p. 12710-4.
32. Stachelhaus, T. and C.T. Walsh, *Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase.* Biochemistry, 2000. **39**(19): p. 5775-87.
33. Kagan, R.M. and S. Clarke, *Widespread occurrence of three sequence motifs in diverse*

S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch Biochem Biophys, 1994. **310**(2): p. 417-27.

34. Katz, J.E., M. Dlakic, and S. Clarke, *Automated identification of putative methyltransferases from genomic open reading frames.* Mol Cell Proteomics, 2003. **2**(8): p. 525-40.
35. Weber, G., et al., *The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame.* Curr Genet, 1994. **26**(2): p. 120-5.
36. Broberg, A., A. Menkis, and R. Vasiliauskas, *Kutznerides 1-4, depsipeptides from the actinomycete Kutzneria sp. 744 inhabiting mycorrhizal roots of Picea abies seedlings.* J Nat Prod, 2006. **69**(1): p. 97-102.
37. Gulavita, N.K., et al.,, *Isolation and structure elucidation of perthamide B, a novel peptide from the sponge Theonella sp.* Tetrahedron Letters, 1994. **35**(37): p. 6815-18.
38. Miller, D.A., et al.,, *C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase.* J Am Chem Soc, 2001. **123**: p. 8434-5.
39. Weinig, S., et al., *Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases.* Chem Biol, 2003. **10**(10): p. 939-52.
40. Schoenafinger, G., et al., *Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin.* J Am Chem Soc, 2006. **128**(23): p. 7406-7.
41. Rouhiainen, L., et al., *Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90.* Mol Microbiol, 2000. **37**(1): p. 156-67.
42. Roy, R.S., et al., *Thiazole and oxazole peptides: biosynthesis and molecular machinery.* Nat Prod Rep, 1999. **16**(2): p. 249-63.
43. Du, L., et al., *An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003.* FEMS Microbiol Lett, 2000. **189**(2): p. 171-5.
44. Schneider, T.L., B. Shen, and C.T. Walsh, *Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation.* Biochemistry, 2003. **42**(32): p. 9722-30.
45. Silakowski, B., et al., *New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1.* J Biol Chem, 1999. **274**(52): p. 37391-9.
46. Reimmann, C., et al., *Essential PchG-dependent reduction in pyochelin biosynthesis of Pseudomonas aeruginosa.* J Bacteriol, 2001. **183**(3): p. 813-20.
47. Strieker, M., et al., *Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis.* J Am Chem Soc, 2009. **131**(37): p. 13523-30.
48. Vaillancourt, F.H., J. Yin, and C.T. Walsh, *SyrB2 in syringomycin E biosynthesis is a*

nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. Proc Natl Acad Sci U S A, 2005. **102**(29): p. 10111-6.

49. Neumann, C.S. and C.T. Walsh, *Biosynthesis of (-)-(1S,2R)-allocoronamic acyl thioester by an Fe(II)-dependent halogenase and a cyclopropane-forming flavoprotein.* J Am Chem Soc, 2008. **130**(43): p. 14022-3.
50. Walsh, C.T., et al., *Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines.* Curr Opin Chem Biol, 2001. **5**(5): p. 525-34.
51. Walsh, C.T., et al., *Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story.* Chem Biol, 1996. **3**(1): p. 21-8.
52. Walsh, C.T. and M.A. Fischbach, *Natural products version 2.0: connecting genes to molecules.* J Am Chem Soc. **132**(8): p. 2469-93.
53. He, X.M. and H.W. Liu, *Formation of unusual sugars: mechanistic studies and biosynthetic applications.* Annu Rev Biochem, 2002. **71**: p. 701-54.
54. Thibodeaux, C.J., C.E. Melancon, 3rd, and H.W. Liu, *Natural-product sugar biosynthesis and enzymatic glycodiversification.* Angew Chem Int Ed Engl, 2008. **47**(51): p. 9814-59.
55. Keating, T.A., et al., *Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis.* Chembiochem, 2001. **2**(2): p. 99-107.
56. Tseng, C.C., et al., *Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase.* Biochemistry, 2002. **41**(45): p. 13350-9.
57. Kopp, F. and M.A. Marahiel, *Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis.* Nat Prod Rep, 2007. **24**(4): p. 735-49.
58. Bruner, S.D., et al., *Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE.* Structure, 2002. **10**(3): p. 301-10.
59. Samel, S.A., et al., *The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide.* J Mol Biol, 2006. **359**(4): p. 876-89.
60. Trauger, J.W., et al., *Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase.* Nature, 2000. **407**(6801): p. 215-8.
61. Sieber, S.A., et al., *Peptidyl thiophenols as substrates for nonribosomal peptide cyclases.* Angew. Chem. Int. Ed., 2004. **43**: p. 493-8.
62. Kopp, F. and M.A. Marahiel, *Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides.* Curr Opin Biotechnol, 2007. **18**(6): p. 513-20.
63. Grunewald, J. and M.A. Marahiel, *Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides.* Microbiol Mol Biol Rev, 2006. **70**(1): p. 121-46.
64. Lin, H., et al., *Macrolactamization of glycosylated peptide thioesters by the thioesterase domain of tyrocidine synthetase.* Chem Biol, 2004. **11**(12): p. 1635-42.
65. Kohli, R.M., J. Takagi, and C.T. Walsh, *The thioesterase domain from a nonribosomal peptide synthetase as a cyclization catalyst for integrin binding peptides.* Proc Natl Acad Sci U S A, 2002. **99**(3): p. 1247-52.

66. Roongsawang, N., K. Washio, and M. Morikawa, *In vivo characterization of tandem C-terminal thioesterase domains in arthrofactin synthetase.* Chembiochem, 2007. **8**(5): p. 501-12.
67. de Bruijn, I., et al., *Massetolide A biosynthesis in Pseudomonas fluorescens.* J Bacteriol, 2008. **190**(8): p. 2777-89.
68. Kessler, N., et al., *The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains.* J Biol Chem, 2004. **279**(9): p. 7413-9.
69. Gaitatzis, N., B. Kunze, and R. Muller, *In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11136-41.
70. Weber, G. and E. Leitner, *Disruption of the cyclosporin synthetase gene of Tolypocladium niveum.* Curr Genet, 1994. **26**(5-6): p. 461-7.
71. Reuter, K., et al., *Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily.* EMBO J, 1999. **18**(23): p. 6823-31.
72. Mofid, M.R., R. Finking, and M.A. Marahiel, *Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp.* J Biol Chem, 2002. **277**(19): p. 17023-31.
73. Balibar, C.J., F.H. Vaillancourt, and C.T. Walsh, *Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains.* Chem Biol, 2005. **12**(11): p. 1189-200.
74. Stein, D.B., U. Linne, and M.A. Marahiel, *Utility of epimerization domains for the redesign of nonribosomal peptide synthetases.* FEBS J, 2005. **272**(17): p. 4506-20.
75. Belshaw, P.J., C.T. Walsh, and T. Stachelhaus, *Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis.* Science, 1999. **284**(5413): p. 486-9.
76. Schwarzer, D., et al., *Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14083-8.
77. Kim, B.S., et al., *Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin.* J Biol Chem, 2002. **277**(50): p. 48028-34.
78. Debono, M., et al., *Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032).* J Antibiot (Tokyo), 1988. **41**(8): p. 1093-105.
79. Mahlert, C., et al., *Chemoenzymatic approach to enantiopure streptogramin B variants: characterization of stereoselective pristinamycin I cyclase from Streptomyces pristinaespiralis.* J Am Chem Soc, 2005. **127**(26): p. 9571-80.
80. Nguyen, K.T., et al., *Genetically engineered lipopeptide antibiotics related to A54145 and daptomycin with improved properties.* Antimicrob Agents Chemother. **54**(4): p. 1404-13.

81. Weist, S., et al., *Mutasynthesis of glycopeptide antibiotics: variations of vancomycin's AB-ring amino acid 3,5-dihydroxyphenylglycine.* J Am Chem Soc, 2004. **126**(19): p. 5942-3.
82. Bonner, D.P., et al., *Lysobactin, a novel antibacterial agent produced by Lysobacter sp. II. Biological properties.* J Antibiot (Tokyo), 1988. **41**(12): p. 1745-51.
83. O'Sullivan, J., et al., *Lysobactin, a novel antibacterial agent produced by Lysobacter sp. I. Taxonomy, isolation and partial characterization.* J Antibiot (Tokyo), 1988. **41**(12): p. 1740-4.
84. von Nussbaum, F., et al., *Structure and total synthesis of lysobactin (katanosin B).* Angew Chem Int Ed Engl, 2007. **46**(12): p. 2039-42.
85. Maki, H., K. Miura, and Y. Yamano, *Katanosin B and plusbacin A(3), inhibitors of peptidoglycan synthesis in methicillin-resistant Staphylococcus aureus.* Antimicrob Agents Chemother, 2001. **45**(6): p. 1823-7.
86. Guzman-Martinez, A., R. Lamer, and M.S. VanNieuwenhze, *Total synthesis of lysobactin.* J Am Chem Soc, 2007. **129**(18): p. 6017-21.
87. Tymiak, A.A., et al,, *Structure determination of lysobactin, a macrocyclic peptide lactone antibiotic.* J. Org. Chem., 1989. **54**: p. 1149-57.
88. von Nussbaum, F., Beck H., Brunner, N., Endermann, R., Koebberling, J., Ragot, J., Telser, J., Schuhmacher J., Anlauf, S., Grande, Y.C., Greschat S., Militzer, H.-C., Schiffer, G., *Lysobactin amides*. 2009., IPC8 Class: AA61K3812FI.
89. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**(4): p. 380-416.
90. Moran, M.A., L.T. Rutherford, and R.E. Hodson, *Evidence for indigenous Streptomyces populations in a marine environment determined with a 16S rRNA probe.* Appl Environ Microbiol, 1995. **61**(10): p. 3695-700.
91. Schnappinger, D. and W. Hillen, *Tetracyclines: antibiotic action, uptake, and resistance mechanisms.* Arch Microbiol, 1996. **165**(6): p. 359-69.
92. Weber, T., et al., *Exploiting the genetic potential of polyketide producing streptomycetes.* J Biotechnol, 2003. **106**(2-3): p. 221-32.
93. Couzinet, S., et al., *In vitro activity of the polyether ionophorous antibiotic monensin against the cyst form of Toxoplasma gondii.* Parasitology, 2000. **121 (Pt 4)**: p. 359-65.
94. Hopwood, D.A., K.F. Chater, and M.J. Bibb, *Genetics of antibiotic production in Streptomyces coelicolor A3(2), a model streptomycete.* Biotechnology, 1995. **28**: p. 65-102.
95. Du, L., et al., *The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase.* Chem Biol, 2000. **7**(8): p. 623-42.
96. Hartmann, T., *From waste products to ecochemicals: fifty years research of plant secondary metabolism.* Phytochemistry, 2007. **68**(22-24): p. 2831-46.
97. Hopwood, D.A., *Forty years of genetics with Streptomyces: from in vivo through in vitro*

to in silico. Microbiology, 1999. **145 (Pt 9)**: p. 2183-202.

98. Mailman, T., M. Hariharan, and B. Karten, *Inhibition of neuronal cholesterol biosynthesis with lovastatin leads to impaired synaptic vesicle release even in the presence of lipoproteins or geranylgeraniol.* J Neurochem.
99. Hertweck, C., *The biosynthetic logic of polyketide diversity.* Angew Chem Int Ed Engl, 2009. **48**(26): p. 4688-716.
100. Collie, J.N., *Derivatives of the multiple ketene group.* Proc. Chem. Soc., 1907. **23**: p. 230-231.
101. Birch, A.J.a.D., F. W., *Studie in relation to biosynthesis I. Some possible routes to derivatives of orcinol and phloroglucinol.* Aust. J. Chem., 1953. **6**: p. 360-368.
102. Hopwood, D.A., *Genetic Contributions to Understanding Polyketide Synthases.* Chem Rev, 1997. **97**(7): p. 2465-2498.
103. Rawlings, B.J., *Type I polyketide biosynthesis in bacteria (Part A--erythromycin biosynthesis).* Nat Prod Rep, 2001. **18**(2): p. 190-227.
104. Cane, D.E. and C.T. Walsh, *The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases.* Chem Biol, 1999. **6**(12): p. R319-25.
105. Bizukojc, M. and S. Ledakowicz, *Physiological, morphological and kinetic aspects of lovastatin biosynthesis by Aspergillus terreus.* Biotechnol J, 2009. **4**(5): p. 647-64.
106. Malpartida, F. and D.A. Hopwood, *Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host.* Nature, 1984. **309**(5967): p. 462-4.
107. Fernandez-Moreno, M.A., et al., *Nucleotide sequence and deduced functions of a set of cotranscribed genes of Streptomyces coelicolor A3(2) including the polyketide synthase for the antibiotic actinorhodin.* J Biol Chem, 1992. **267**(27): p. 19278-90.
108. Tropf, S., et al., *Reaction mechanisms of homodimeric plant polyketide synthase (stilbenes and chalcone synthase). A single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones, and 6'-deoxychalcones.* J Biol Chem, 1995. **270**(14): p. 7922-8.
109. Austin, M.B. and J.P. Noel, *The chalcone synthase superfamily of type III polyketide synthases.* Nat Prod Rep, 2003. **20**(1): p. 79-110.
110. Moore, B.S. and J.N. Hopke, *Discovery of a new bacterial polyketide biosynthetic pathway.* Chembiochem, 2001. **2**(1): p. 35-8.
111. Pfeifer, V., et al., *A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine.* J Biol Chem, 2001. **276**(42): p. 38370-7.
112. Seshime, Y., et al., *Discovery of a novel superfamily of type III polyketide synthases in Aspergillus oryzae.* Biochem Biophys Res Commun, 2005. **331**(1): p. 253-60.
113. Ikeda, H., et al., *Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in Streptomyces avermitilis.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9509-14.
114. Palaniappan, N., et al., *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, 2003. **278**(37): p. 35552-7.

115. Lowden, P.A., et al., *Origin and True Nature of the Starter Unit for the Rapamycin Polyketide Synthase We thank Dr. Bradley S. Moore for help with the deuterium NMR experiments. This work was supported by grants from The Wellcome Trust (to J.S. and P.F.L.) and from the NIH (AI20264 to H.G.F.).* Angew Chem Int Ed Engl, 2001. **40**(4): p. 777-779.
116. Floss, H.G. and T.W. Yu, *Rifamycin-mode of action, resistance, and biosynthesis.* Chem Rev, 2005. **105**(2): p. 621-32.
117. Campelo, A.B. and J.A. Gil, *The candicidin gene cluster from Streptomyces griseus IMRU 3570.* Microbiology, 2002. **148**(Pt 1): p. 51-9.
118. Dewick, P.M., *Medicinal Natural Products: a Biosynthetic Approach.* Wiley, Chichester, 1997.
119. Horinouchi, S., *Combinatorial biosynthesis of non-bacterial and unnatural flavonoids, stilbenoids and curcuminoids by microorganisms.* J Antibiot (Tokyo), 2008. **61**(12): p. 709-28.
120. Austin, M.B., et al., *Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase.* Nat Chem Biol, 2006. **2**(9): p. 494-502.
121. Miyanaga, A., et al., *Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis.* Proc Natl Acad Sci U S A, 2008. **105**(3): p. 871-6.
122. Stassi, D.L., et al., *Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering.* Proc Natl Acad Sci U S A, 1998. **95**(13): p. 7305-9.
123. Weber, T., et al., *Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety.* Chem Biol, 2008. **15**(2): p. 175-88.
124. Wenzel, S.C., et al., *On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of "glycolate" units into ansamitocin and soraphen A.* J Am Chem Soc, 2006. **128**(44): p. 14325-36.
125. Reeves, C.D., et al., *A new substrate specificity for acyl transferase domains of the ascomycin polyketide synthase in Streptomyces hygroscopicus.* J Biol Chem, 2002. **277**(11): p. 9155-9.
126. Haydock, S.F., et al., *Organization of the biosynthetic gene cluster for the macrolide concanamycin A in Streptomyces neyagawaensis ATCC 27449.* Microbiology, 2005. **151**(Pt 10): p. 3161-9.
127. Emmert, E.A., et al., *Genetics of zwittermicin a production by Bacillus cereus.* Appl Environ Microbiol, 2004. **70**(1): p. 104-13.
128. Morita, H., et al., *Structural insight into chain-length control and product specificity of pentaketide chromone synthase from Aloe arborescens.* Chem Biol, 2007. **14**(4): p. 359-69.
129. Wu, K., et al., *The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus*

(ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene, 2000. **251**(1): p. 81-90.

130. Reynolds, K.A., et al., *Biosynthesis of the shikimate-derived starter unit of the immunosuppressant ascomycin: stereochemistry of the 1,4-conjugate elimination.* J Antibiot (Tokyo), 1997. **50**(8): p. 701-3.
131. Fischbach, M.A. and C.T. Walsh, *Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms.* Chem Rev, 2006. **106**(8): p. 3468-96.
132. Miller, D.A., et al., *Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis.* Chem Biol, 2002. **9**(3): p. 333-44.
133. Suo, Z., C.C. Tseng, and C.T. Walsh, *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, 2001. **98**(1): p. 99-104.
134. Miller, D.A., C.T. Walsh, and L. Luo, *C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase.* J Am Chem Soc, 2001. **123**(34): p. 8434-5.
135. Kraas, F.I., et al., *Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation.* Chemistry & biology, 2010. **17**(8): p. 872-80.
136. Stachelhaus, T., A. Huser, and M.A. Marahiel, *Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases.* Chem Biol, 1996. **3**(11): p. 913-21.
137. Bachmann, B.O. and J. Ravel, *Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data.* Methods Enzymol, 2009. **458**: p. 181-217.
138. Rausch, C., et al., *Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).* Nucleic Acids Res, 2005. **33**(18): p. 5799-808.
139. Silakowski, B., et al., *Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium Stigmatella aurantiaca Sga15.* Chem Biol, 2001. **8**(1): p. 59-69.
140. Konz, D., et al., *The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases.* Chem Biol, 1997. **4**(12): p. 927-37.
141. Guenzi, E., et al., *Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases.* J Biol Chem, 1998. **273**(49): p. 32857-63.
142. Scholz-Schroeder, B.K., J.D. Soule, and D.C. Gross, *The sypA, sypS, and sypC synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by Pseudomonas syringae pv. syringae B301D.* Mol Plant Microbe Interact,

2003. **16**(4): p. 271-80.

143. Konz, D., S. Doekel, and M.A. Marahiel, *Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin.* J Bacteriol, 1999. **181**(1): p. 133-40.
144. de Crecy-Lagard, V., et al., *Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Streptomyces virginiae: molecular characterization of the last structural peptide synthetase gene.* Antimicrob Agents Chemother, 1997. **41**(9): p. 1904-9.
145. Yakimov, M.M., et al., *A putative lichenysin A synthetase operon in Bacillus licheniformis: initial characterization.* Biochim Biophys Acta, 1998. **1399**(2-3): p. 141-53.
146. Wiest, A., et al., *Identification of peptaibols from Trichoderma virens and cloning of a peptaibol synthetase.* J Biol Chem, 2002. **277**(23): p. 20862-8.
147. Tsuge, K., T. Akiyama, and M. Shoda, *Cloning, sequencing, and characterization of the iturin A operon.* J Bacteriol, 2001. **183**(21): p. 6265-73.
148. Borodovsky, M., et al., *Prokaryotic gene prediction using GeneMark and GeneMark.hmm.* Curr Protoc Bioinformatics, 2003. **Chapter 4**: p. Unit4 5.
149. Cosmina, P., et al., *Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis.* Mol Microbiol, 1993. **8**(5): p. 821-31.
150. Hoffmann, D., et al., *Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from Nostoc sp. GSV224.* Gene, 2003. **311**: p. 171-80.
151. Crotti, A.E.M., Fonseca, T., Hong, H., Staunton, J., Galembek, S. E., Lopes, N. P., *The fragmentation mechanism of five-membered lactones by electrospray ionisation tandem mass spectrometry.* Int. J. Mass Spectrometry, 2004. **232**(1): p. 271-276.
152. Helmetag, V., et al., *Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis.* FEBS J, 2009. **276**(13): p. 3669-82.
153. Roongsawang, N., et al., *Cloning and characterization of the gene cluster encoding arthrofactin synthetase from Pseudomonas sp. MIS38.* Chem Biol, 2003. **10**(9): p. 869-80.
154. Kohli, R.M., C.T. Walsh, and M.D. Burkart, *Biomimetic synthesis and optimization of cyclic peptide antibiotics.* Nature, 2002. **418**(6898): p. 658-61.
155. Grunewald, J., et al., *Synthesis and derivatization of daptomycin: a chemoenzymatic route to acidic lipopeptide antibiotics.* J Am Chem Soc, 2004. **126**(51): p. 17025-31.
156. Kopp, F., et al., *Chemoenzymatic design of acidic lipopeptide hybrids: new insights into the structure-activity relationship of daptomycin and A54145.* Biochemistry, 2006. **45**(35): p. 10474-81.
157. Kotowska, M., et al., *Type II thioesterase from Streptomyces coelicolor A3(2).* Microbiology, 2002. **148**(Pt 6): p. 1777-83.
158. Schneider, A. and M.A. Marahiel, *Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis.* Arch Microbiol, 1998. **169**(5): p. 404-10.
159. Cudic, P., et al., *Complexation of peptidoglycan intermediates by the lipoglycodepsipeptide antibiotic ramoplanin: minimal structural requirements for*

intermolecular complexation and fibril formation. Proc Natl Acad Sci U S A, 2002. **99**(11): p. 7384-9.

160. Fang, X., et al., *The mechanism of action of ramoplanin and enduracidin.* Mol Biosyst, 2006. **2**(1): p. 69-76.

161. Linne, U., *Epimerisierungsdomänen aus Peptidsynthetasen - Untersuchungen zur Substratespezifität und zur Interaktion mit anderen Domänen.* 2001.

8. Supplementry section

Table S1. Bioinformatic analysis of the contig harboring the lysobactin synthetase genes *lybA* and *lybB*, responsible for lysobactin biosynthesis. The length of each gene is given in basepairs, whereas the length of the gene product is given in the number of amino acids. Lysobactin synthetases are accentuated in red. The contig, covering a region of 150063 bp with an average GC-content of 70.7%, encompasses 91 CDSs based on GeneMark2.4 annotation. The proposed functions of the encoded proteins and the corresponding homology to related proteins are based on BLAST-analysis.

gene	length (bp)	length (aa)	proposed function	similar protein, organism	identity / similarity (%)	E-value
orf1	1083	361	alkaline phosphatase	ZP-02241451.1, *Xanthomonas oryzae pv. oryzae* PXO99A	71/78	$2e^{-141}$
orf2	510	170	no significant similarity found	/	/	/
orf3	225	75	no significant similarity found	/	/	/
orf4	1317	439	RND family efflux transporter MFP subunit	YP_001684934.1, *Caulobacter* sp. K31	55/70	$3e^{-107}$
orf5	3105	1035	acriflavin resistance protein	YP_001684933.1, *Caulobacter* sp. K31	72/83	0.0
orf6	3198	1066	acriflavin resistance protein	YP_549641.1, *Polaromonas* sp. JS666	64/79	0.0
orf7	372	124	putative exported protein	CBW26876.1, *Bacteriovorax marinus* SJ	27/40	0.38
orf8	330	110	hypothetical protein	ZP_03228483.1, *Bacillus coahuilensis* m4-4	32/64	6.9
orf9	1131	377	peptidase	YP_002129213.1, *Phenylobacterium zucineum* HLK1	40/52	$2e^{-47}$
orf10	1335	445	hypothetical protein	YP_001851229.1, *Mycobacterium marinum* M	35/49	$2e^{-42}$
orf11	123	41	no significant similarity found	/	/	/
orf12	1017	339	Mg^{2+} transporter protein CorA family protein	ZP_06464109.1, *Burkholderia* sp. CCGE1003	51/72	$3e^{-98}$
orf13	315	105	no significant similarity found	/	/	/
orf14	1230	410	benzoate transporter	YP_003167664.1, *Candidatus Accumulibacter phosphatis* clade IIA str. UW-1	64/80	$4e^{-134}$
orf15	231	77	No significant similarity	/	/	/
orf16	1503	501	transcriptional regulator	ZP_01052708.1, *Polaribacter* sp. MED152	21/40	0.44
orf17	1224	408	hypothetical protein	NP_821723.1, *Streptomyces avermitilis* MA-4680	26/40	$5e^{-15}$
orf18	624	208	HupE / UreJ protein	ZP_05112692.1, *Labrenzia alexandrii* DFL-11	50/64	$3e^{-38}$
orf19	489	163	no significant similarity found	/	/	/
orf20	837	279	hypothetical protein/Co^{2+} ABC-type transporter system	ZP_04715208.1, *Alteromonas macleodii* ATCC 27126	34/51	$3e^{-35}$
orf21	450	150	hypothetical protein	ZP_07044210.1, *Comamonas testosteroni* S44	67/79	$5e^{-22}$
orf22	555	185	putative transmembrane protein	YP_001974093.1, *Stenotrophomonas maltophilia* K279a	47/67	$1e^{-36}$
orf23	63	21	multi-sensor hybrid histidine kinase	YP_003444812.1, *Allochromatium vinosum* DSM 180	73/86	9.0
orf24	924	308	glycerophosphoryl diester phosphodiesterase family protein	ZP_02364269.1, *Burkholderia oklahomensis* C6786	56/67	$8e^{-72}$
orf25	891	297	endonuclease	ZP_06490315.1, *Xanthomonas campestris pv. musacearum* NCPPB4381	44/60	$4e^{-51}$
orf26	1203	401	glycoside hydrolase family protein	YP_001269345.1, *Pseudomonas putida* F1	65/75	$3e^{-144}$
orf27	459	153	transcriptional regulator, BadM/Rrf2 family	YP_003693117.1, *Starkeya novella* DSM 506	71/85	$5e^{-51}$
orf28	369	123	sec-independent protein translocase protein TatC	YP_001260172.1, *Sphingomonas wittichii* RW1	50/65	$9e^{-26}$
orf29	1227	409	oxidoreductase FAD/NAD(P)-binding domain protein	YP_001682485.1, *Caulobacter* sp. K31	59/71	$1e^{-133}$
orf30	1224	408	OmpA family outer membrane lipoprotein	YP_001905001.1, *Xanthomonas campestris pv. campestris* str. B100	70/81	$2e^{-135}$
orf31	405	135	hypothetical protein	YP_917345.1, *Paracoccus denitrificans* PD1222	36/50	$6e^{-6}$
orf32	2184	728	glycosyl hydrolase family 3 N terminal domain protein	ZP_05135635.1, *Stenotrophomonas* sp. SKA14	64/77	0.0
orf33	2208	736	putative ferric siderophore receptor protein	YP_001971588.1, *Stenotrophomonas maltophilia* K279a	68/80	0.0
orf34	930	310	possible beta-lactamase type II	ZP_00946799.1, *Ralstonia solanacearum* IPO1609	70/81	$1e^{-115}$
orf35	1806	602	YvcC, ABC transporter	YP_001422767.1, *Bacillus amyloliquefaciens* FZB42	42/62	$2e^{-136}$
orf36	297	99	hypothetical protein	YP_002433262.1, *Desulfatibacillum alkenivorans* AK-01	28/45	6.2
orf37	168	56	no significant similarity found	/	/	/

Supplementry section

orf38	1695	565	malate:quinone oxidoreductase	YP_002027471.1, *Stenotrophomonas maltophilia* R551-3	79/87	0.0
orf39	54	18	no significant similarity found	/	/	/
orf40	906	302	hypothetical protein, fatty acid hydroxylase	YP_001350148.1, *Pseudomonas aeruginosa* PA7	61/70	$2e^{-87}$
orf41	684	228	AraC family transcriptional regulator	YP_001669525.1, *Pseudomonas putida* GB-1	40/55	$3e^{-30}$
orf42	1290	430	cytochrome-c peroxidase	ZP_03657868.1, *Helicobacter cinaedi* CCUG 18818	52/68	$1e^{-105}$
orf43	2556	852	TonB-dependent outer membrane receptor	ZP_06702903.1, *Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 11122	53/70	0.0
orf44	1410	470	putative oxidoreductase/FAD-dependent	YP_001974019.1, *Stenotrophomonas maltophilia* K279a	67/78	$8e^{-144}$
orf45	630	210	nucleoside-diphosphate-sugar epimerase-like protein	YP_001764517.1, *Burkholderia cenocepacia* MC0-3	66/77	$4e^{-73}$
orf46	897	299	putative transcriptional regulator	YP_002638715.1, *Salmonella enterica* subsp. *enterica serovar* RKS4594	52/70	$1e^{-78}$
orf47	1395	465	major facilitator family transporter	ZP_03572459.1, *Burkholderia multivorans* CGD2M	62/71	$2e^{-108}$
orf48	711	237	alpha/beta hydrolase fold protein	YP_002501846.1, *Methylobacterium nodulans* ORS 2060	62/71	$2e^{-65}$
orf49	462	154	hypothetical protein	ZP_01450738.1, *alpha proteobacterium* HTCC2255	47/59	$5e^{-19}$
orf50	447	149	conserved hypothetical protein	ZP_06705519.1, *Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 11122	64/72	$9e^{-45}$
orf51	504	168	putative Mg(2+) transporter	YP_001972661.1, *Stenotrophomonas maltophilia* K279a	44/66	$4e^{-24}$
orf52	1062	354	putative DNA topoisomerase	YP_001901445.1, *Xanthomonas campestris pv. campestris* str. B100	61/74	$5e^{-111}$
orf53	2298	766	TonB-dependent receptor	ZP_05136801.1, *Stenotrophomonas* sp. SKA14	71/82	0.0
orf54	111	37	no significant similarity found	/	/	/
orf55	180	90	hypothetical protein	ZP_05136538.1, *Stenotrophomonas* sp. SKA14	56/69	$2e^{-13}$
orf56	303	101	hypothetical protein	YP_739616.1, *Shewanella* sp. MR-7	51/67	$3e^{-16}$
orf57	1620	540	PepSY-associated TM helix	ZP_05134261.1, *Stenotrophomonas* sp. SKA14	65/76	0.0
orf58	1023	341	putative esterase	YP_001972145.1, *Stenotrophomonas maltophilia* K279a	50/62	$6e^{-65}$
orf59	984	328	beta-lactamase domain-containing protein	YP_001672104.1, *Caulobacter* sp. K31	50/65	$2e^{-75}$
orf60	468	156	hypothetical protein	ZP_06704248.1, *Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 11122	62/79	$4e^{-43}$
orf61	945	315	hypothetical protein	YP_366148.1, *Xanthomonas campestris pv. vesicatoria* str. 85-10	70/80	$1e^{-124}$
orf62	375	125	hypothetical protein	YP_002029693.1, *Stenotrophomonas maltophilia* R551-3	81/88	$4e^{-51}$
orf63	750	250	heme oxygenase	ZP_05135978.1, *Stenotrophomonas* sp. SKA14	70/77	$5e^{-67}$
orf64	258	86	hypothetical protein	YP_002029695.1, *Stenotrophomonas maltophilia* R551-3	71/78	$1e^{-26}$
orf65	153	51	no significant similarity found	/	/	/
orf66	1275	425	membrane fusion protein	ZP_05137116.1, *Stenotrophomonas* sp. SKA14	75/87	$6e^{-150}$
orf67	3321	1107	acriflavin resistance protein	YP_002030072.1, *Stenotrophomonas maltophilia* R551-3	87/93	0.0
orf68	3327	1109	acriflavin resistance protein	YP_002030073.1, *Stenotrophomonas maltophilia* R551-3	46/65	0.0
orf69	369	123	hypothetical protein, NIPSNAP family protein	YP_003211797, *Cronobacter turicensis* z3032	42/64	$3e^{-18}$
orf70	204	68	no significant similarity found	/	/	/
orf71	213	71	no significant similarity found	/	/	/
orf72	984	328	TesB-like acyl-CoA thioesterase II	YP_003275253.1, *Gordonia bronchialis* DSM 43247	48/61	$5e^{-52}$
orf73	186	62	hypothetical protein	YP_003766914.1, *Amycolatopsis mediterranei* U32	37/58	0.026
orf74	2868	956	lanthionine synthetase C-like protein	ZP_05136619, *Stenotrophomonas* sp. SKA14	57/69	0.0
orf75	306	102	hypothetical protein	YP_003382479.1, *Kribbella flavida* DSM 17836	29/44	1.9
orf76	93	31	RNA-directed RNA polymerase	Q3KSM3.1, *Pieris rapae* virus	66/66	2.0
orf77	192	64	no significant similarity found	/	/	/
lybA	**11490**	**3830**	**amino acid adenylation/NRPS**	**YP_236792.1, *Pseudomonas syringae pv. syringae* B728a**	**50/64**	**0.0**
lybB	**25611**	**8537**	**amino acid adenylation domain-containing protein/NRPS**	**YP_001546733.1, *Herpetosiphon aurantiacus* ATCC 23779**	**39/53**	**0.0**
orf78	1593	531	lactamase B	YP_003375569.1, *Xanthomonas albilineans*	60/73	0.0
orf79	1809	603	ABC transporter, permease/ATP-binding protein	ZP_03109194.1, *Bacillus cereus* NVH0597-99	41/62	$6e^{-128}$
orf80	3222	1074	RND superfamily protein, acriflavin resistance protein	ZP_06729535, *Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 10535	47/64	0.0
orf81	1278	426	membrane fusion protein precursor	ZP_0672953, *Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 10535	62/77	$6e^{-114}$
orf82	3225	1075	RND superfamily protein, acriflavin resistance protein	YP_365697.1, *Xanthomonas campestris pv. vesicatoria* str. 85-10	74/86	0.0
orf83	1377	459	hypothetical protein, alpha/beta fold family hydrolase	YP_001265258.1, *Sphingomonas wittichii* RW1	54/66	$3e^{-112}$
orf84	153	51	no significant similarity found	/	/	/
orf85	159	53	no significant similarity found	/	/	/

Supplementry section

orf86	11271	3757	keto-hydroxyglutarate-aldolase/polyketide synthase	ZP_02164195.1, *Kordia algicida* OT-1	48/65	0.0
orf87	3693	1231	polyketide synthase of type I	ZP_02163855.1, *Kordia algicida* OT-1	38/55	0.0
orf88	2607	869	exporters of the RND superfamily-like protein	ZP_02163857.1*Kordia algicida* OT-1	43/66	0.0
orf89	915	305	hypothetical protein	GB_ACX30634.1, uncultured bacterium ARCTIC96BD-19	38/60	$2e^{-55}$

Aknowledgement

First of all, I would like to thank Prof. Dr. M. A. Marahiel for being an excellent supervisor, his scientific guidance and generous support during my PhD thesis. I gratefully acknowledge his scientific advices, support with new ideas and willingness for helpful discussions. With his enthusiasm about science, he generated an inspiring atmosphere in his laboratory, which made work there very exciting.

I gratefully thank Prof. Dr. A. Geyer for reviewing of this thesis. I would like to thank Prof. Dr. A. Seubert for being my thesis committee.

I would like to thank Dr. Lars Robbel for the cooperation, many helpful advices of the lysobactin project. His profound experience of natural product offered an excellent acceleration.

I would like to thank Alan Tanovic for imparting the experience of protein crystallization and his willingness for helpful discussions. I acknowledge Dr. Marcus Miethke for his scientific advises.

Especially, I would like to thank my friend, Dr. Dong Sun, for his expert advices in the organic chemistry.

I gratefully thank Dr. Uwe Linne for his expert opinions concerning bioanalytical questions and his helpful advises. I acknowledge Antje Schäfer, Gabriele Schimpff-Weiland and Christiane Bomm for their technical assistance.

I acknowledge the whole Marahiel group for the nice working atmosphere. I really enjoyed the time in the lab.

I thank Julian Hegemann, Femke Kraas, Tobias Gießen and Leif Flühe for proof-reading this thesis.

Special thanks to my parents, for their financial and moral support throughout my whole life. I would like to thank my wife Lei for her support, for the wonderful time we spent together and that will come. I wish to dedicate this thesis to my family.