The surfactin biosynthetic complex of *Bacillus subtilis*: COM domain-mediated biocombinatorial synthesis, and single step purification of native multi-modular NRPSs and multi-enzyme complexes

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Claudia Chiocchini

aus Perugia/Italien

Marburg/Lahn 2006

Vom Fachbereich Biologie
der Philipps-Universität Marburg als Dissertation
am 14 Juni 2006 angenommen.

Erstgutachter : Prof. Dr. W. Buckel (Philipps-Universität, Marburg)
Zweitgutachter : Prof. Dr. M. A. Marahiel (Philipps-Universität, Marburg)

Tag der mündlichen Prüfung 29 Juni 2006

Dietro ogni problema c’ è un’ opportunità

Galileo Galilei

Summary

Most biosynthetic templates for the assembly of peptide natural products are composed of two or more nonribosomal peptide synthetases (NRPSs). For example, the surfactin biosynthetic complex consists of three NRPSs (SrfA-A, SrfA-B, and SrfA-C), which are encoded by the polycistronic *srfA* operon within the chromosome of the producer strain *Bacillus subtilis*. According to the molecular logic employed by NRPS assembly lines, the biosynthesis of a defined product relies on the proper, well-orchestrated interaction between partner-enzymes (i.e. SrfA-A/SrfA-B, and SrfA-B/SrfA-C), and the prevention of futile interactions between non-partner enzymes (i.e. SrfA-A/SrfA-C). Based on most recent *in vitro* studies, these selective interactions between NRPSs are controlled by the interplay of communication-mediating (COM) domains, located at the C- and N-termini of the corresponding donor and acceptor enzymes.

In the first part of this study, the potential of COM domains was exploited for the directed reprogramming of the surfactin biosynthetic complex, and the setting up of an *in vivo* system for the true biocombinatorial synthesis of lipopeptides. To this end, the first COM domain pair, facilitating the selective interaction between SrfA-A and SrfA-B, was substituted against various cognate, mis-cognate and non-cognate COM domain pairs. The consequences of these manipulations were then analyzed by means of HPLC and high-resolution MS. These experiments verified that COM domain pairs of the tyrocidine biosynthetic complex retain their functionality and selectivity even in the context of a heterologous host and NRPS system. Furthermore, utilization of a designated non-cognate COM domain pair allowed for an intended skipping of the second NRPS SrfA-B, the enforcement of a productive interaction between the natural non-partner enzymes SrfA-A and SrfA-C, and thus the directed synthesis of a shortened lipotetrapeptide product. In another experiment, all donor and acceptor enzymes of the biosynthetic complex were equipped with the same set of cognate COM domain pair. The resulting abrogation of the selectivity-barrier led to the establishment of an so-called universal communication system, and afforded the envisioned biocombinatorial synthesis of two lipopeptide products. All these experiments verified – for the first time *in vivo*, and within the context of a natural NRP assembly line – the decisive role of COM domains for the control of protein-protein communication between NRPSs.

The second objective of this work was the establishment of a gentle method for the purification of NRPSs and multi-enzymatic NRPS complexes. The approach taken was based on the utilization of polyol-responsive monoclonal antibodies (PR-mAb), which are able to release their bound antigen under gentle, non-denaturating conditions, in the presence of polyols. PR-mAbs were originally developed and used for the purification of the *E. coli* RNA polymerase holo-enzyme complex, including low-affinity bound σ-factors. Among others, these studies led to the identification of a antigen/antibody pair epitope/NT73. Within the scope of this work, the coding sequence of this epitope tag was fused to the 3'-end of the *srfA-A* gene within the chromosome of *B. subtilis*. Subsequently, the encoded SrfA-A-epi protein could be purified from cleared crude extracts of the resulting mutant using immunoaffininty chromatography. SDS-PAGE and MS/MS analyses, as well as biochemical characterizations unequivocally verified the purification of the epitop-tagged SrfA-A protein in active holo-form, as well as co-purification of SrfA-B (molecular weight of the dimeric complex: approx. 803 kDa). Under the conditions tested, the third NRPS, SrfA-C, as well as additional proteins, associated with the surfactin complex, could not be detected.

Zusammenfassung

Die meisten Biosynthese-Matrizen von Peptid-Naturstoffen sind aus mehreren nicht-ribosomalen Peptidsynthetasen (NRPS) aufgebaut. Der Biosynthese-Komplex des Lipopeptid-Antibiokums Surfactin z.B. besteht aus drei NRPSs (SrfA-A, SrfA-B und SrfA-C), welche vom polycistronischen *srfA* Operon des Produzentenstammes *B. subtilis* kodiert werden. Aufgrund der molekularen Logik der nichtribosomalen Peptidsynthese erfordert die Biosynthese eines definierten Produktes eine ausschließliche Interaktion von Partner-Enzymen (z.B. SrfA-A/SrfA-B und SrfA-B/SrfA-C). Anhand von *in vitro* Untersuchungen konnte gezeigt werden, dass diese selektiven Wechselwirkungen von Kommunikations-vermittelnden (*communication-mediating*, COM) Domänen kontrolliert werden, welche am C- bzw. N-Terminus der entsprechenden Donor- oder Akzeptor-Enzyme lokalisiert sind.
Im ersten Teil dieser Arbeit wurden das Potential von COM-Domänen zur Umprogrammierung des Surfactin-Biosynthese-Komplexes, sowie zur Generierung eines *in vivo* Systems zur biokombinatorischen Synthese von Lipopeptiden ausgenutzt. Hierfür wurde das erste COM-Domänenpaar (zw. SrfA-A und SrfA-B), gegen unterschiedliche verwandte, mis-verwandte und nicht-verwandte COM-Domänenpaare ausgetauscht. Die Auswirkungen dieser Manipulationen wurden anschließend mittels HPLC und MS analysiert. Hierbei konnte gezeigt werden, dass COM-Domänenpaare des Tyrocidine-Biosynthese-Komplexes auch im heterologen Wirt und NRPS-System ihre uneingeschränkte Funktionalität und Selektivität bewahren. Des weiteren konnte durch die Verwendung eines nicht-verwandten COM-Domänenpaares die tri-modulare Peptidsynthetase SrfA-B gezielt übersprungen und eine produktive Interaktion zwischen den nativen Nicht-Partner-Enzymen SrfA-A und SrfA-C etabliert werden. Dieses Übergehen von SrfA-B führte zur Bildung eines verkürzten Lipotetrapetid-Produktes. In einem weiteren Test wurden alle Donor- und Akzeptor-Enzyme mit demselben cognaten COM-Domänenpaar ausgestattet. Die hiermit verbundene Aufhebung der Selektivitätsbarriere führte zur Etablierung eines universellen Kommunikationssystems (*universal communication system*, UCS), sowie zur gleichzeitigen, biokombinatorische Synthese von zwei Lipopeptid-Produkten. All diese Ergebnisse verifizierten die entscheidende Bedeutung von COM-Domänen für die Kontrolle der Protein-Protein-Kommunikation zwischen NRPSs erstmals *in vivo* und im Kontext eines natürlichen NRP-Montagebandes.
Im zweiten Teil dieser Arbeit wurde eine schonende Methode zur Reinigung von NRPSs und multi-enzymatischen NRPS-Biosynthese-Komplexen entwickelt. Grundlage hierfür bildete die Verwendung von polyol-responsiven monoklonalen Antikörpern (PR-mAk), welche in der Lage sind ihr Antigen unter schonenden, nicht-denaturierenden Bedingun-gen – in Anwesenheit von Polyolen – freizusetzen. Diese Antikörper wurden erstmals für die Reinigung des *E. coli* RNA-Polymerase-Komplexes samt der locker gebundenen σ-Faktoren verwendet, was u.a. zur Identifizierung des Antigen/Antikörper-Paares Epitop/NT73 führte. Im Rahmen dieser Arbeit wurde der codierende Bereich dieses Epitops an das 3'-Ende des *srfA-A* Gens fusioniert. Anschließend konnte das kodierte Protein SrfA-A-epi aus Rohextrakten des entsprechenden *B. subtilis*-Stammes mittels Immunoaffinitäts-Chromatographie gereinigt werden. SDS-PAGE und MS/MS Analysen, sowie biochemische Charakterisierungen verifizierten hierbei die Reinigung von SrfA-A-epi in aktiver holo-Form, sowie im Komplex mit seinem natürlichen Partner-Enzym SrfA-B (Molekularmasse des dimeren Komplexes: ca. 803 kDa).

Ein Teil der währen der Promotion erzielten Ergebnisse wurde in folgender Originalpublikation veröffentlicht

Claudia Chiocchini, Uwe Linne and Torsten Stachelhaus, *In vivo* biocombinatorial synthesis of lipopeptides by COM domainmediated reprogramming of the surfactin biosynthetic complex, *Chem. Biol*., under revision.

Table of contents

Summary .....

Zusammenfassung .....

Table of contents .....

1 Abbreviations 1

2 Introduction 4

2.2 Natural products and combinatorial chemistry 4

2.3 Nonribosomal peptides 6

2.4 Organization and function of NRPS domains. 8

2.5 The lipoheptapeptide surfactin 11

2.6 Surfactin biosynthesis 12

2.7 Targeted modifications of non ribosomal peptides 13

2.7.1 Rational manipulations of biosynthetic templates 14

2.7.2 The chemoenzymatic approach 17

2.8 Combinatorial libraries of natural products 18

2.9 The first steps towards combinatorial biosynthesis in polyketide synthases. 19

2.10 Protein-protein communication in NRPS. 22

2.11 Heterologous expression of NRPS genes. 25

2.12 Post translational modification of NRPS 26

2.13 Purification of NRPSs from the natural producer strain 27

2.13.1 Immobilized metal ion affinity chromaptography (IMAC) 28

2.13.2 Immunoaffinity chromatography, using polyol-responsive monoclonal antibodies (PR-mAB) 28

2.14 Goal 30

3 Material 31

3.1 Microorganisms 31

3.2 Plasmids 31

3.2.1 pQE60 31

3.2.2 pDG646 /pDG782 32

3.2.3 pCm::Tc 32

3.2.4 pTOPO-XL 33

3.2.5 pKE27 33

3.3 Media 33
3.4 Chemicals and laboratory products 34
3.5 Equipment 35
4 Methods 37
4.1 Molecular Biology 37
4.1.1 Construction of plasmids 37
4.1.2 *B. subtilis* transformation 42
4.1.2.1 Congression 43
4.1.3 Southern blot analysis. 43
4.2 Biochemical methods 44
4.2.1 Preparation of *B. subtilis* cell extracts 44
4.2.2 Western blot 44
4.2.3 Immunodetection 44
4.2.4 Protein purification by Ni-NTA affinity chromatography 45
4.2.5 Protein purification by immunoaffinity chromatography 46
4.2.6 ATP-PPi exchange 47
4.2.7 Covalent loading of NRPSs with radioactively labelled aminoacids 49
4.2.8 Surfactin and lipo-tetrapeptide preparation 49
4.2.9 Surfactin analysis on blood agar plates. 50
4.2.10 Surfactin and lipo-tetrapeptide analysis by HPLC/MS 50
4.2.11 Identification of NRPSs by tryptic digestion and ESI/MS-MS analysis 51
5 Results (I) 52
5.1 *In vivo* investigation of protein-protein communication in the surfactin biosynthetic complex 52
5.2 Maintenance of *B. subtilis* genetic competence 54
5.3 Construction of the disruption strain CC64 54
5.4 COM domain swapping 57
5.4.1 Elaboration of a cloning strategy for COM domain swapping 57
5.4.2 Swapping of $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}B1}$: a proof of principle 58
5.4.3 Activity of heterologous COM domains: $COM^{D}_{TycB}/COM^{A}_{TycC.}$ 62
5.4.4 Skipping of SrfA-B within the surfactin biosynthetic complex: $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}B1}$ vs. $COM^{D}_{TycA}/COM^{A}_{TycC.}$ 63
5.4.5 Combinatorial *in vivo* biosynthesis; $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C}$ 67

5.4.6 Investigation of the proposed non-cognate COM domain pair $COM^{D}_{SrfA-A3}$/COM^{A}_{SrfA-C} 69
6 Results (II) 72
6.1 Method development for the gentle purification of NRPSs and entire NRPS multi-enzyme complexes 72
6.2 Purification by Ni-NTA affinity chromatography 73
6.2.1 Integration of a second copy of *comS* in the *amyE*-site of the surfactin producer *B. subtilis* OKB105 73
6.2.2 Integration of a resistance marker in the transition site of *srfA-A-srfA-B* 75
6.2.3 Fusion of a Hexahistidine-tag to the C-terminus of SrfA-A 76
6.2.4 SrfA-A-His purification. 78
6.3 Immunoaffinity chromatography using polyol-responsive monoclonal antibodies. 79
6.3.1 Fusion of the epitope tag at the C-terminus of SrfA-A 80
6.3.2 SrfA-A-epi purification by immunoaffinity chromatography 82
6.3.3 Reconstitution of the surfactin biosynthetic complex 84
7 Discussion 89
7.1 Elaboration of a strategy for COM domain swapping 90
7.2 Reconstitution of surfactin biosynthesis using heterologous COM domains 94
7.3 COM domains as a tool for combinatorial biosynthesis 94
7.4 Reprogramming of the surfactin biosynthetic assembly line: skipping of SrfA-B 96
7.5 The 5-residue model: identification of the selectivity-conferring residues of COM domains 98
7.6 COM domains and docking domains: differences and homologies. 101
7.7 Outlook 103
7.8 Purification of large NRPSs and multi-enzymatic NRPS complexes 106
7.9 Method optimization for purification of large NRPSs and multi-enzymatic NRPS complexes 108
7.9.1 Optimization of IMAC 108
7.9.2 Optimization of immunoaffinity chromatography 109
7.10 Evaluation of different methods for the genetic manipulation of *B. subtilis* chromosome 111
7.11 Outlook 114
8 Appendix 116

8.1 Strains and plasmids 116
8.2 Oligonucleotides 118
8.3 The conserved core motifs of NRPS domains 119
8.4 Biosynthetic systems 121
Literature 122

1 Abbreviations

A	adenine
A domain	adenylation domain
aa	amino acid
Ab	antibody
ACP	acyl carrier protein
Ag	antigen
amp	ampicillin
AMP	adenosyne-5'-monophosphate
AT	acyltransferase
ATCC	strain collection (american type culture collection)
AU	arbitrary units
bla	ampicilin resistance cassette (β-lactamase coding gene)
bp	base pairs
BSA	bovine serum albumine
C	cytosine
C domain	condensation domain
cat	chloramepenicol resistance
Cm	chloramphenicol
CoA	coenzyme A
cpm	counts per minute
Da	dalton
dATP	2'-deoxyadenosine-5'-triphosphate
DEBS	6-deoxyerythroenolide-B-synthase
DKP	D-Phe-L-Pro diketopiperazine
dpm	disintegrations per minute
DH	dehydrogenase
E domain	epimerization domain
EDTA	ethylene diamino tetra ecetic acid
ESI	electrospray ionization
ER	enoyl reductase
FA	β-hydroxyl fatty acid
FT-ICR MS	Fourier transform-ion cyclotron resonance mass spectrometry
G	guanine
HEPES	2[4-(2-Hydroxyethyl)piperazine]-1-ethanesulphonic acid
HPLC	high performance liquid chromatography
IMAC	immobilized metal ion affinity chromatography
kan	kanamycin resistance cassette
kb	kilobase pairs
Km	kanamycin
KR	ketoreductase

KS	ketosynthase
l	liter
M	molar (moles/l)
µ	micro (10^{-6})
MCS	multiple cloning site
min	minutes
ml	milliliter
mls	macrolide, lincosamide, streptogramine resistance cassette
MLS	macrolide, lincosamide, streptogramine
MS	mass spectrometry
MW	molecular weight
n	nano (10^{-9})
n.d.	not detected
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthetase
NTA	nitrilotriacetate
OD_{600}	optical density at 600 nm
ori	origin of replication
P_{Srf}	surfactin promoter
PAGE	polyacrylamide gelelectrophoresis
PCP	peptidyl carrier protein or thiolation domain
PCR	polymerase chain reaction
PEGA	polyethilene glycol amide
PK	poliketide
PKS	polyketide synthase
PMSF	phenyl methyl sulfonyl fluoride
PR-mAb	polyol-responsive monoclonal antibody
Ppant	4'-phosphopantetheine
PP_i	inorganic pyrophosphate
PPTase	4'-phosphopantetheine transferase
RBS	ribosomal binding site
rpm	rounds per minute
RT	room temperature
SDS	sodiumdodecylsulfate
SNAC	N-acetylcysteamine
SPPS	solid phase peptide synthesis
srf	surfactin
Srf TE	TE domain from the surfactin NRPS
T	tymine
TE domain	thioesterase domain
TE II	type II thioesterase
tet	tetracycline resistance cassette

Tet	tetracycline
TCA	tricloroacetic acid
TFA	trifluoroacetic acid
UCS	universal communication system
v/v	volume per volume
w/v	weight per volume
wt	wild type

Tab. 1.1 Amino acids: Abbreviations and molecular weights

Amino acid	3-letter code	1-letter code	MW [g/mol]
Alanine	Ala	A	89
Arginine	Arg	R	174
Asparagine	Asn	N	132
Aspartate	Asp	D	133
Cysteine	Cys	C	121
Glutamate	Glu	E	147
Glutamine	Gln	Q	146
Glycine	Gly	G	75
Histidine	His	H	155
Isoleucine	Ile	I	131
Leucine	Leu	L	131
Lysine	Lys	K	146
Methionine	Met	M	149
Ornithine	Orn	O	132
Phenylalanine	Phe	F	165
Proline	Pro	P	115
Serine	Ser	S	105
Threonine	Thr	T	119
Tryptophane	Trp	W	204
Tyrosine	Tyr	Y	181
Valine	Val	V	117

2 Introduction

2.2 Natural products and combinatorial chemistry

The last decades revealed a growing need for new antibiotics, due to the emergence of resistant, pathogenic microorganisms. Prominent examples are represented by methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant enterococci (VREs). In USA, about 12% of *Enterococci faecalis* is resistant to vancomycin, and 35% of *Staphylococcus aureus* is resistant to all available β-lactams antibiotics. Multi-resistant Gram-positive bacteria also represent an increasing threat, because of their ability to induce severe infections in hospitals [Page M.GP, 2005] The necessity of improving the efficiency of the drug discovery process induced pharmaceutical companies to the search for new possible sources [Ortholand, 2004].

The vast majority of commonly used antimicrobial agents are either natural products or derivatives of natural products. Traditionally, most of these compounds had been identified by screening of natural product libraries, which per se represents a very time-consuming process that suffers of several disadvantages: i) maintaining a high quality strain collection is expensive, ii) the identification of a hit, which is isolated from an extract has to be followed by bioassay guided fractionation, in order to identify the active component, and iii) the novelty of the isolated molecule is not predictable and requires further characterization. In the worst case, comparison with bioactive molecules available in data bases can reveal that the identified lead is an already well-known natural product.

The introduction of high throughput screening and combinatorial chemistry initially seemed to offer a valuable alternative to the traditional methods for discovering new biologically active compounds. Combinatorial libraries began to compete with natural product extracts for screening resources. The possibility to produce large libraries and screen a high number of molecules in a short time seemed to be a promising tool to find new hits. However, screening of the first combinatorial libraries was pretty upsetting; the most probable reason of this failure being the composition of those libraries, which contained mostly oligomeric molecules that were easy to synthesize, like peptides or nucleotides, but structurally too simple to resemble small drugs. Successive studies aimed to create smaller libraries with improved quality, using filter for lead-like [Hann et al., 2001; Oprea et al., 2001] or drug-like properties [Egan et al., 2002; Lipinski et al., 2001]. As a consequence the quality of the leads discovered from combinatorial libraries started to increase.

These results suggested that also known natural products could be exploited in order to obtain new bioactive molecules. Therefore further approaches aimed to combine the interesting features of natural products with the potential of combinatorial chemistry to produce natural product-like libraries. For example, natural products were used as source of building block. Affimax reported the degradation of the thiazole antibiotic GE2270 A, which was followed by combinatorial synthesis of different A-ring modifications [Clough et al., 2003]. Another promising approach is the derivatization of natural products. A significant example was the preparation of a library of synthetic vancomycin dimers, which were tested for their ability to bind D-Ala-D-Ala, as well as D-Ala-D-Lac. The glycopeptide vancomycin [van Wageningen et al., 1998] sequesters the peptidoglycan substrate D-Ala-D-Ala by forming five hydrogen bonds, this way preventing the transpeptidation of the cell wall building blocks. Vancomycin-resistant bacteria use the alternative substrate D-Ala-D-Lac, which, due to the loss of one hydrogen bond, is bound by vancomycin with a 1000-times lower affinity. Therefore binding of vancomycin derivatives to D-Ala-D-Lac was initially thought to be required to contrast the infections due to vancomycin resistant bacteria. Utilizing solid-phase peptide synthesis, a large library of dimeric vancomycin analogues was prepared and several vancomycin variants were detected, which exhibited a higher affinity for the target and were active against vancomycin-resistant bacteria [Nicolaou et al., 2001]. Natural product scaffolds were also used as starting point for combinatorial approaches. A so called split and mix library of 39.300 molecules was synthesized by Ellman and co-workers using the biaryl template of vancomycin in order to detect analogues with increased selectivity for vancomycin-resistant bacterial strains [Jain et al., 2003; Xu, 1999; Xu Ruo, 1999].

Given the structural complexity of the majority of antibiotics and pharmacologically important compounds, organic chemistry is not always a suitable tool for introducing structural modifications. For example, chemical modification of certain portions of molecule backbone may not always be feasible, due to the lack of suitable carbonyl, amide of hydroxyl moieties.

A valuable alternative approach to produce novel natural products is the manipulation of the correspondent biosynthetic pathways. Several successes were already obtained within the groups of nonribosomal peptides (NRPs) and polyketides (PKs). Here, the modular organization of the biosynthetic templates - nonribosomal peptide synthetases (NRPSs) or polyketide synthetases (PKSs) - render both enzyme families attractive targets for the genetic engineering. The directed modification of NRPS and PKS assembly line were carried out, and

aimed on synthesizing novel natural products. Several examples for the successful application of corresponding approaches are reported in section 2.7.

2.3 Nonribosomal peptides

Nonribosomal peptides (NRPs) represent a diverse group of pharmacologically important natural products synthesized by numerous microorganisms, including filamentous fungi, Gram-positive and - to minor extent - Gram-negative bacteria. Among the Gram-positive bacteria, the group of *Actinomycetes,* as well as members of the genus *Bacillus* are the primary producers of NRPs.

Members of this heterogeneous class of natural compounds comprise the antibiotics bacitracin A (1) [Konz et al., 1997], tyrocidine A (2) [Mootz and Marahiel, 1997], vancomycin (3) [van Wageningen et al., 1998] and daptomycin (4) [Raja et al., 2003]. Of non-ribosomal origin are also the immuno-suppressant cyclosporine A (5) [Weber et al., 1994], the anti-tumor drugs bleomycin (6) [Du et al., 2000] and epothilone (7) [Molnar et al., 2000], as well as the biotenside surfactin (8) [Peypoux et al., 1999](see Fig. 1). The structural complexity of NRPs reflects their broad activity spectrum. The precursors for the biosynthesis of NRPs are not restricted to the 20 proteinogenic amino acids, but comprise over 300 building blocks like nonproteinogenic amino acids, carboxy- and hydroxyl acids. Moreover, modifications like N-, C- and O-methylation, glycosylation, acylation, Cα epimerization and heterocyclic ring formation are often observed. Additional constraints such as cyclic or branched-cyclic backbones, as well as oxidative cross-linking confer to the molecules rigid structures, which are fundamental for their bioactivity.

Despite their structural heterogeneity, NRPs share a common mode of synthesis, namely a stepwise assembly of the peptide backbone directed by large, multi-modular assembly lines, termed nonribosomal peptide synthetases (NRPSs). These enzymes contemporary act as template and biosynthetic machinery [Finking and Marahiel, 2004; Marahiel et al., 1997], and possess a modular organization. Each module is responsible for recognition, activation and incorporation of one specific substrate in the growing peptide chain, and can be subdivided in single catalytic units called domains.

In most NRPS assembly lines, the number and order of modules in the enzyme reflects the number and order of the amino acid substrate in the peptide product [Mootz et al., 2002c]. However, this colinearity principle finds some exceptions e.g. in the assembly lines of bacillibactin, yersiniabactin [Gehring et al., 1998; Miller et al., 2002] (9) and bleomycin.

There are also examples of mixed NRPS-PKS hybrid assembly lines, like in the case of the anti-tumor agent epothilone A (7).

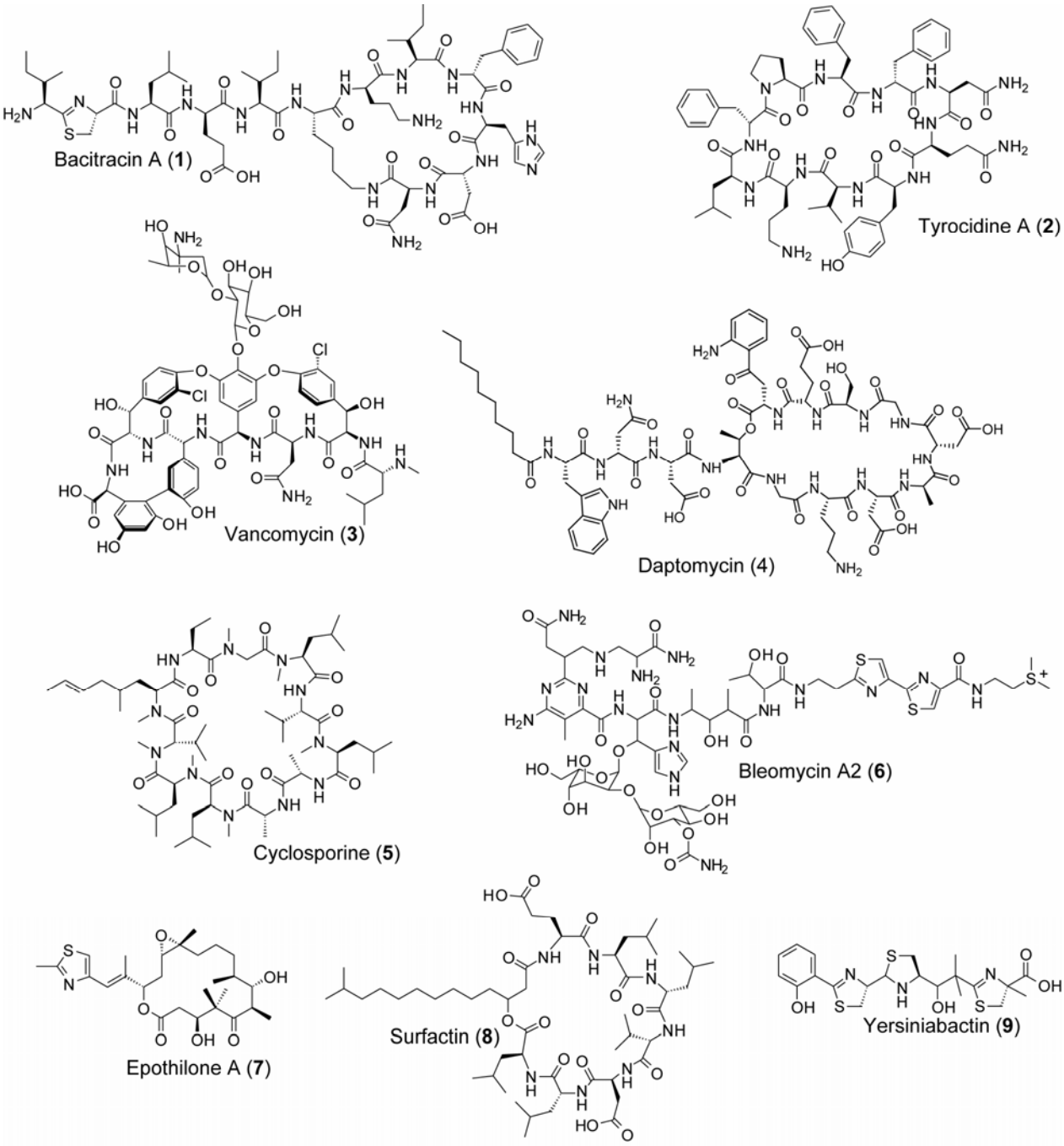


Fig. 1
A selection of nonribosomal peptides produced by bacteria and filamentous fungi is shown above. **(1)** Bacitracin A, produced by *Bacillus licheniformis*; **(2)** Tyrocidine A, produced by *Bacillus brevis*; **(3)** Vancomycin produced by *Streptomyces orientalis*; **(4)** Daptomycin, produced by *Streptomyces roseosporus*. **(5)** Cyclosporine, produced by *Tolypocladium niveum*; **(6)** Bleomycin A2, produced by *Streptomyces verticillus*; **(7)** Epothilone A produced by *Sorangium cellulosum*; **(8)** Surfactin, produced by *Bacillus subtilis*; **(9)** Yersiniabactin produced by *Yersinia pestis*.

2.4 Organization and function of NRPS domains.

According to the so-called multiple carrier thio-template mechanism, the substrate amino acids and the nascent peptide chain are tethered to the NRPSs via 4´-phosphopantetheine (Ppant) moieties. The prosthetic group represents an integral part of a module's peptidyl carrier (PCP) domain, which has to be converted into its active holo-form by phosphopantetheinylation (Fig. 2A). This priming reaction requires specialized phosphopantetheinyl transferases (PPTase), which catalyze the nucleophilic attack of an invariant serine residue of the PCP domain onto the β-3'-phosphate group of CoA [Lambalot et al., 1996]. In analogy to aminoacyl-tRNA synthetases, NRPS adenylation (A) domains select their cognate amino acid from the pool of available substrates and activate it as aminoacyl adenylate at the expense of ATP and Mg^{2+} [Stachelhaus and Marahiel, 1995b]. The activated aminoacyl-O-AMP oxoesters are subsequently transferred to the adjacent holo-PCP and bound as thioester to the 4´-Ppant cofactor [Stein et al., 1996; Stein et al., 1994] (Fig. 2B,C).

The peptide-bond formation between aminoacyl or peptidyl-S-PCP donor and an aminoacyl-S-PCP acceptor substrates is catalyzed by the C domain of the acceptor module. The nucleophilic attack of the amino group of downstream located aminoacyl-S-PCP acceptor onto the up-stream localized peptidyl-S-PCP thioester eventually yields the translocation of the thioester-bound peptide chain to the PCP domain of the acceptor-module [Stachelhaus et al., 1998] (Fig. 2 D). Based on mis-priming experiments, a low selectivity for the upstream donor substrate and a high selectivity for the downstream acceptor residue could be determined for the C domain [Belshaw et al., 1999]. This property of the C domains probably prevents mis-initation of the biosynthetic process, since this way, the activated aminoacyl-S-PCP is prevented from directly being transferred to the downstream C domain and accepted as donor acyl-thioester. The iteration of substrate activation, binding as thioester to the PCP domain, peptide chain elongation and translocation to the downstream module (Fig. 3) leads to the formation of the full length peptide chain, bound to the 4'Ppant-cofactor of the last (termination) module. The final product release is then catalyzed by a TE domain, located at the C-terminus of the termination module (Fig. 2E). TE domains, which catalyze product release in NRPS as well as in fatty acid synthase (FS) and PKS assembly lines, belong to the hydrolase family that utilizes the catalytical triade, consisting of acidic residue-His-Ser (or Cys). In a first step, the peptidyl intermediate is transferred from the downstream PCP domain to a highly conserved serine residue, which is part of the core motif GxSxG and the catalytic triade [Shaw-Reid et al., 1999].

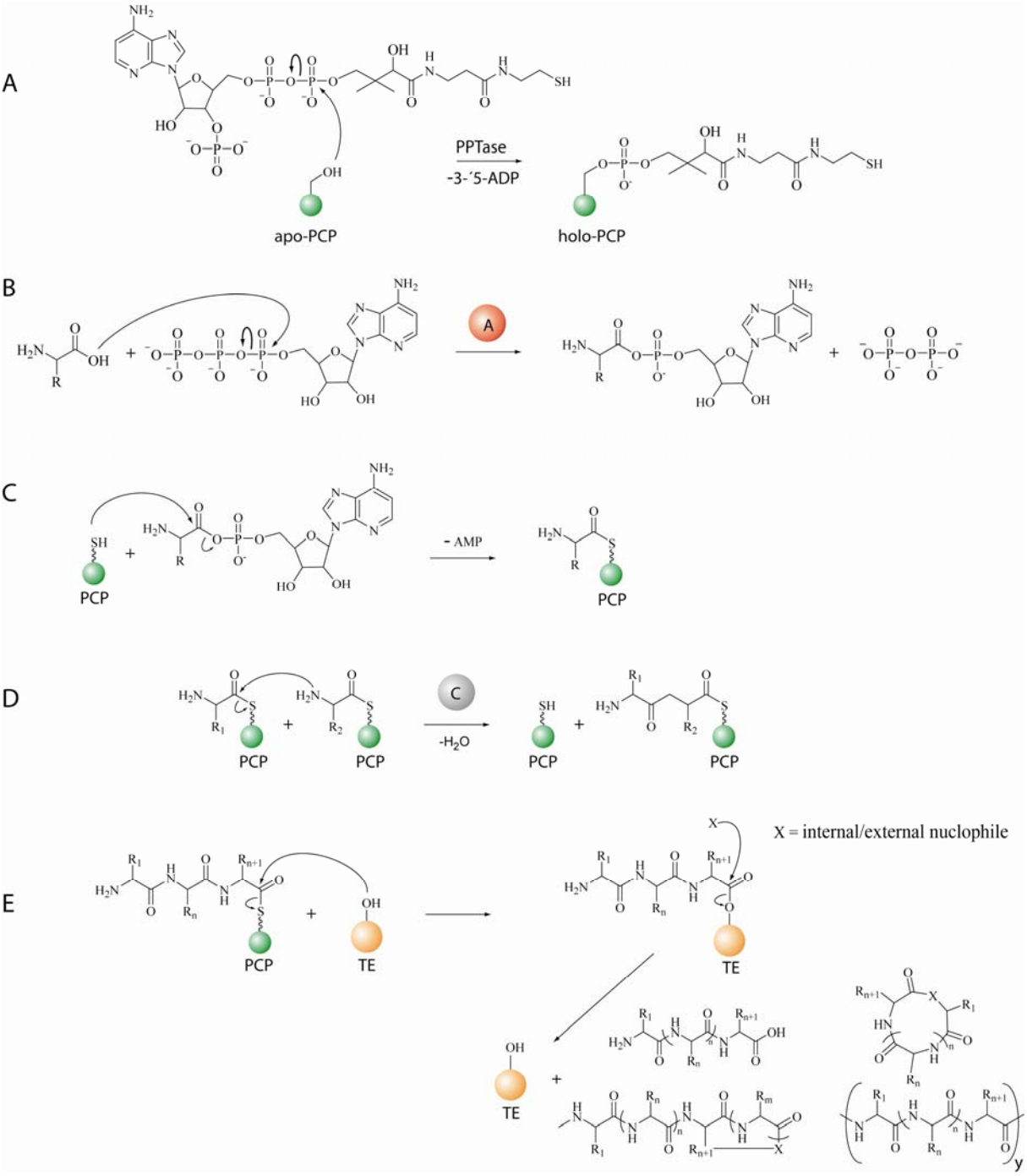


Fig. 2 The five essential reactions of non ribosomal peptide biosynthesis
(A) The post-translational modification of the PCP domain (in green) is carried out by a specific PPTase, which catalyze the nucleophilic attack of the hydroxyl-group of an invariant serine residue of the PCP domain onto the 3'-phosphate group of CoA, resulting in the covalent loading of the Ppant cofactor to PCP. **(B)** An adenylation (A) domain (in red) selectively recognizes a substrate amino acid and activates it as aminoacyl adenylate at expense of ATP and Mg^{2+}. **(C)** The activated amino acid is bound to the Ppant cofactor of the associated PCP as aminoacyl thioester. **(D)** The C domain catalyzes the peptide bond formation between two intermediates covalently bound to neighbouring PCP domains. **(E)** The thioesterase domain, localized at the C-terminus of the last module, forms an oxoester intermediate with the peptide chain and finally catalyzes the product release either in linear, circular, branched circular or multi-meric form.

The subsequent deacylation of the resulting oxoester intermediate occurs by nucleophilic attack of either (i) an internal nucleophile, which – depending on its position – leads to

intramolecolar cyclization and the generation of cyclic or branched-cyclid macrolactones or macrolactames, (ii) water, which causes the release of linear peptides, or (iii) a N-terminal nucleophile of another peptide chain, leading to multi-meric compounds (Fig. 2E).

The described domains are the so-called core domains. A minimal elongation module consists of the three domains: C-A-PCP. An A-PCP di-domain pair forms an initiation module, and in fact - in absence of an up-stream C domain - an amino acid can be activated, tethered to PCP and directly translocated to the downstream module. Finally, the C-terminal termination module (C-A-PCP-TE) has an additional TE for peptide product release.

Many NRPS modules contain additional domains, which modify the activated aminoacyl or peptidyl-thioester intermediates. Examples for those optional domains are the epimerization (E) domain, N-methylation, (M) domain, oxidation (Ox) domain, formylation (F) domain, cyclization (Cy) domain. Many pharmacologically important non-ribosomal peptides undergo one ore more of the mentioned modifications, which were shown to be fundamental for their activity.

In most NRPS assembly lines, the number and order of modules in the enzyme reflect the number and order of amino acid constituents of the peptide product [Mootz et al., 2002c].

This so-called co-linearity principle is illustrated in more detail in the next section, on the basis of the nonribosomal biosynthesis of the lipoheptapeptide surfactin.

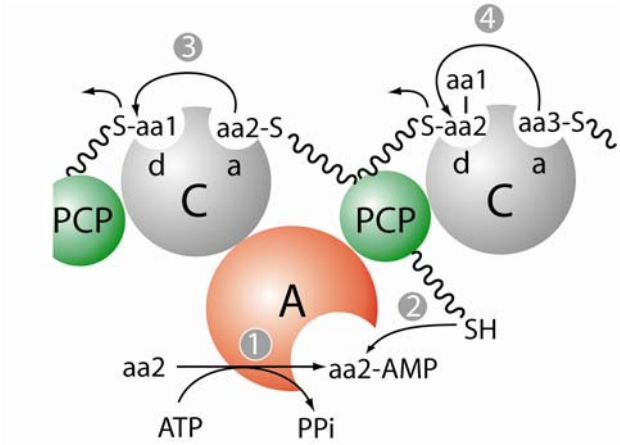


Fig. 3 Domain organization of a NRPS elongation module

An elongation module has the organization C-A-PCP. **(1)** the A domain is responsible for recognition and activation of the amino acid substrate as aminoacyl adenylate. **(2)** The holo-PCP binds the activated amino acid as aminoacyl-thioester through its phosphopantetheine cofactor. **(3)** The aminoacyl thioester is translocated to the acceptor site of the C domain, which catalyzes the nucleophylic attack of aa 2-S-PCP onto aa 1-S-PCP with consequent peptide bond formation between the aa1-aa 2. **(4)** Finally the peptidyl-S-Ppant is transferred to the donor site of the downstream C domain of the following module, where a second peptide bond formation takes place between aa 3-S-Ppant and peptidyl-S-Ppant.

2.5 The lipoheptapeptide surfactin

The cyclic lipoheptapeptide surfactin is produced by different *B. subtilis* strains, and consists of a β- hydroxy fatty acid chain of variable length (12 to 15 carbon atoms). As shown in Fig. 4, the fatty acid is linked to both ends of the heptapeptide chain LGlu-LLeu-DLeu-LVal-LLeu-LAsp-DLeu-LLeu, via an amide (N-terminal end) and a lactone (C-terminal end) bond, respectively [Arima et al., 1968]. Surfactin derivatives were isolated, varying not only in the length of the lipo-moiety, but also in the peptide composition, with either valine or isoleucine substituting the leucine moietiy in position 7.

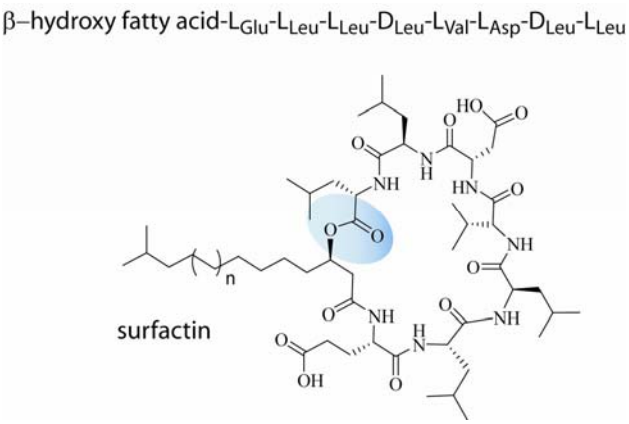


Fig. 4 The lipoheptapeptide surfactin
The lipoheptapeptide surfactin is synthesized as a mixture of compounds differing in the length of the β-hydroxy-fatty acid chain (n = 1-3) as well as in the amino acid composition (leucine can be substituted by isoleucine or valine). In blue, the bond between leucine at position 7 and the β-hydroxyl group of the FA chain is marked, resulting in macrolactonization.

At a concentration of 20 µM, surfactin is able to reduce the surface tension of water from 72 mJ/m^2 to 27 mJ/m^2. This makes surfactin one of the most powerful biotensides known so far, and an interesting for industrial application. In particular, surfactin represents an alternative (or supplement) to chemical surfactants, which have a detrimental effect [Peypoux et al., 1999]. Because of its ampholytic nature, surfactin is soluble in polar and non-polar solvents. In aqueous phase as well as at the water/air interface, surfactin assumes a characteristic horse-saddle conformation, which is probably responsible for its broad spectrum of biological activity [Bonmatin JM, 1994]. Surfactin is able to inhibit the clot formation [Arima et al., 1968], and has been described as antibacterial, antitumoral and hypercholesterolemic agent [Imai, 1971; Kameda et al., 1972; Tsukagoshi et al., 1970]. More recent are the observations of an emulsification [Razafindralambo et al., 1998], as well as an antiviral and an antimycoplasma activity of surfactin [Vollenbroich et al., 1997].

The antibiotic activity of surfactin is based on in its ability to alter the cell membrane [Bernheimer and Avigad, 1970]. The interaction with divalent cations, which bind to the "claw" formed by aspartate and glutamate, neutralizes their negative charges and facilitates the penetration of biological membranes. At low concentration, surfactin is miscible with

phospholipids; at intermediate concentration it forms small micelles within the bilayer, whereas at high concentration it behaves like a detergent leading to cell disruption**.**

2.6 Surfactin biosynthesis

The surfactin biosynthetic assembly line consists of three large NRPSs: SrfA-A (402 kDa), SrfA-B (401 kDa) and SrfA-C (144 kDa), comprising of a total of seven modules. The corresponding NRPS genes are organized in the *srfA* operon. The first step of surfactin biosynthesis is the loading of the fatty acid chain on the L-glutamate residue, previously activated by the first module. In contrast to the majority of NRPSs assembly lines that begin with a initiation module the surfactin protein template starts with an elongation module C-A-PCP.

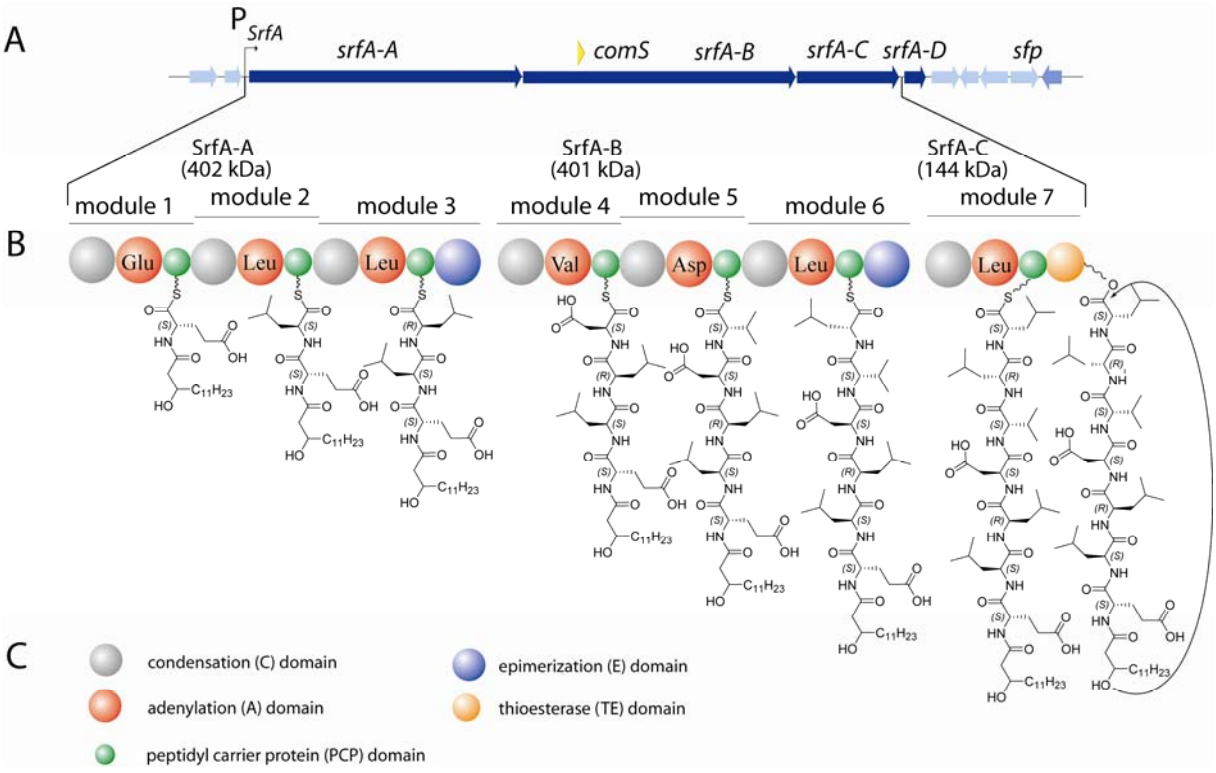


Fig. 5 The surfactin biosynthetic assembly line
(A) The surfactin biosynthetic gene cluster of *Bacillus subtilis* encodes for the non-ribosomal protein template for the synthesis of the lipoheptapetide surfactin. **(B)** This biosynthetic complex consists of three surfactin synthetases: SrfA-A,-B,-C, consisting of seven distinct modules, each responsible for recognition, activation and loading of a single amino acid substrate. Two epimerization domains are found in modules 3 and 6, converting L-Leu into the stereoisomer D-Leu, respectively. The cyclization and release of the final heptapeptide as macrolactone is catalyzed by the TE domain. **(C)** In different colors the single domains are represented.

The first C domain catalyzes the formation of a covalent peptide-bond between the α-amino group of L-glutamate and the α-carboxyl group of the fatty acid. The acyl-chain donor could be a CoA-dependent acyl transferase, observed for the first time in the crude cell extract of *B. subtilis* ATCC 21332 [Menkhaus et al., 1993]. However, this 40 kDa-protein could not be

isolated in subsequent studies and the mechanism of surfactin lipo-initiation has not been completely clarified yet. Recently, it was proposed that the protein TE II, encoded by the gene *srfA-D* and involved in the proof-reading of peptide synthesis [Schwarzer et al., 2002] may also act as acyl transferase, and possibly be required for in the lipo-initiation of surfactin synthesis [Steller et al., 2004].

Surfactin biosynthesis proceeds according to the multiple carrier thiotemplate mechanism, by a stepwise activation and loading of the single amino acid substrates. Two epimerization domains, located at the C-terminus of SrfA-A and SrfA-B, respectively, catalyze the conversion of the L-Leu into D-Leu at position 3 and 6.

The release of the peptide product from the protein template as macrolactone is catalyzed by the TE domain and is the result of the nucleophilic attack of the β-hydroxy moiety of the fatty acid onto the α-carboxy group of the Leu 7. The resolution of the x-ray structure of Srf-TE provided more insight into the structural basis of the cyclization reaction. [Bruner et al., 2002].

Immediately downstream of the coding region of *srfA-C,* the *srfA-D* gene is localized, encoding for the type II thioesterase (TE II), which was shown to have an editing activity. According to the so called “cleaning model”, TE IIs restore the peptide synthesis after mis-loading of PCP domains by catalyzing the hydrolysis of the unprocessed aminoacyl-S-PCP substrates [Schneider and Marahiel, 1998; Schwarzer et al., 2002; Yeh et al., 2004].

About 4 kb downstream the *srfA* operon, the *sfp* gene is located, encoding for the PPTase Sfp, which is required for the conversion of the PCP domains of the surfactin biosynthetic complex from their inactive apo- to the active holo-form [Lambalot et al., 1996]. Due to its high substrate tolerance, Sfp is widely exploited *in vitro* as well as *in vivo* for modification of different peptidyl and acyl carrier proteins either with their natural cofactors or with cofactor mimicries (see 2.7.2).

2.7 Targeted modifications of non ribosomal peptides

The described modular organization of NRPSs and the colinearity between protein template and final peptide product render NRPSs a promising tool for the rational design of novel bioactive compounds. Many questions addressed in the past, concerning the tolerance of NRPSs to genetic modifications, could be answered in the last decade. The strategies described for introducing targeted modifications in NRPs can be divided into two categories: the rational manipulation of biosynthetic NRPS template, and the chemoenzymatic approach.

2.7.1 Rational manipulations of biosynthetic templates

The first relevant examples for the manipulation of NRPS pathways were based on genetic engineering of the surfactin biosynthetic operon and consisted of the exchange of the encoding gene fragment of the A-PCP didomain of module 7 (SrfA-C, see Fig. 5) against corresponding gene fragments derived from different NRPS systems. This so-called swapping of a minimal modules led to the intended biosynthesis of surfactin derivatives, in which the L-Leu moiety at position 7 was substituted against L-Cys; L-Phe, L-Orn and L-Val, respectively [Stachelhaus and Marahiel, 1995b]. In successive experiments L-Leu at position 2 of SrfA-A was substituted with L-Orn [Schneider et al., 1998; Stachelhaus et al., 1995] (Fig. 6A).

Another interesting example of manipulation of the surfactin assembly line was the translocation of the terminal TE domain. The fusion of the corresponding DNA coding sequence to the upstream located coding fragments of surfactin modules 4 and 5 led to the expected production of the truncated lipotetra and lipopentapeptide derivatives of surfactin [de Ferra et al., 1997]. These experiments also demonstrated the *in vivo* portability of the TE domain (Fig. 6B).

Although the mentioned *in vivo* strategies resulted in the establishment of modified NRPS pathways, which were able to produce the expected peptide products, all mentioned approaches suffered from an extensive reduction of the product titers when compared with the natural systems. Based on the current understanding of NRPS assembly-lines, these reductions in productivity can be mostly attributed to the incorrect definition of the domain boundaries used for the minimal module swaps.

Notably, no dramatic drop in productivity was observed in case of the exchange of the first module of SrfA-A (organization C-A_{Glu}-PCP) with the first module of LicA (C-A_{Gln}-PCP, lichenysin synthetase 1) [Yakimov et al., 2000]. This success, however, was probably only due to the high sequence homology shared by NRPS modules (> 60%), rather than to an improved understanding of domain boundaries.

Resolution of the crystal and/or NMR structures of A, C and PCP domains provided interesting insights in the mechanism and architecture of NRPS assembly lines [Bruner et al., 2002; Conti et al., 1997; Keating et al., 2002; Weber et al., 2000] and also allowed for a more precise determination of domain boundaries . This knowledge was exploited by two different approaches for the construction of hybrid NRPSs : i) the fusion at core sequences, and ii) the fusion at linker regions between catalytic domains. Interestingly, it was observed that fusions of two different A domains were only functional when the fusion site was located at the postulated hinge region, corresponding to the conserved sequence GRIDxQ. Active hybrid

enzymes were also obtained by fusions at the core motif of PCP (LGG(DH)SL) and at the highly conserved core 3 of the C domain (MHHxISDG(WV)S) [Symmank et al., 1999] (Fig. 6C).

The construction of functional hybrid NRPSs was mostly done by linker fusion of individual modules, and relied – apart from the definition of the boundaries between domains - on the understanding of the substrate selectivity and tolerance of the certain catalytic domains. For instance, it was observed that C domains exhibit high substrate and enantio-selectivity for the acceptor aminoacyl-PCP, while it has only enantio-selectivity for the donor aminoacyl/ peptidyl-PCP [Belshaw et al., 1999; Ehmann et al., 2000]. These results actually suggested that adjacent C and A domains must not be separated in course of module swapping experiments.

The substrate tolerance of the E domain has also been investigated. This optional domain catalyzes the conversion of L-configured amino acid in the corresponding D-isomer (and *vice versa*), and acts on the thioester-stage after the substrate has been tethered to the module's PCP domain. In the case of C-A-PCP-E elongation modules, the activated L-aminoacyl-S-PCP is first condensed with the nascent peptidyl chain – rather than being directly epimerised. This, however, appears to be mainly due to the immediate trapping of the aminoacyl-S-PCP in the acceptor binding pocket of the adjacent C domain. E domains are able to epimerise both, L-aminoacyl-PCP and peptidyl-S-PCP, although E domains of elongation modules revealed a clear preference for the latter substrate [Belshaw et al., 1999; Linne et al., 2001]. Fusion experiments also revealed that the E domain's substrate has to be presented by a designated PCP domain, and that PCP-E didomain pairs – not unlike the already mentioned C-A didomain pairs must not separated in the course of domain or module swapping experiments. Further investigations showed signature differences in sequence between PCP domains, which usually do not interact with E domains (PCP^C; core motif GGHSL), and those who does (PCP^E; core motif GGDSI). The observed differences in the amino acid sequence are likely to play a key role in the interaction between PCP and E domains. This hypothesis was supported by means of mutational analysis of a PCP^E domain [Linne et al., 2001].

All this information was exploited for the generation hybrid-NRPSs *in vitro* and *in vivo*. All these approaches also used highly variable regions (so-called linker), located at the transition between domains, as fusion sites for the construction of the hybrid NRPSs [Doekel and Marahiel, 2001; Duerfahrt et al., 2004; Mootz et al., 2000]. Fusion at those poorly conserved regions was expected to least affect the communication between modules. Hybrid dimodular enzymes were generated by fusion between both A and PCP domains [Doekel and Marahiel,

2001] (Fig. 6D), whereas in an advance study the junction between PCP and C was used for module fusions . The resulting tri-modular NRPS systems were successfully tested *in vitro* for the production of the predicted tri-peptide products [Mootz et al., 2000] (Fig. 6E). In both cases a TE domain was fused to an elongation module for the catalytic release of the products. The effect of a module deletion was investigated *in vivo* in the surfactin system. The second module of the tri-modular SrfA-A, responsible for activation and loading of L-Leu, was deleted and the resulting *B. subilis* mutant was analysed for product formation [Mootz et al., 2002a]. As predicted a hexapaptide surfactin derivative was produced, lacking L-Leu at position 2, however in a significantly lower amount in comparison to the wild type (Fig. 6F).

A less invasive approach for the manipulations of NRPS pathways is based on the introduction of point mutations to alter the specificity of the A domain. This approach benefited from the elucidation of the crystal structure of the Phe-activating A domain of GrsA [Conti et al., 1997], which - in turn - allowed for the identification of the key amino acid residues for recognition and binding of the cognate amino acid substrates (Fig. 6G) Subsequent mutational analysis, validated the specificity-conferring code and confirmed the results of the *in silico* investigations [Stachelhaus et al., 1999]. In a more recent work, the knowledge of the nonribosomal code was exploited *in vivo* to change the specificity of the Asp-activating A domain of Srf-A-B from Asp to Asn. Interestingly the resulting surfactin analogue was biosynthesised, although at significantly reduced rates [Eppelmann et al., 2002].

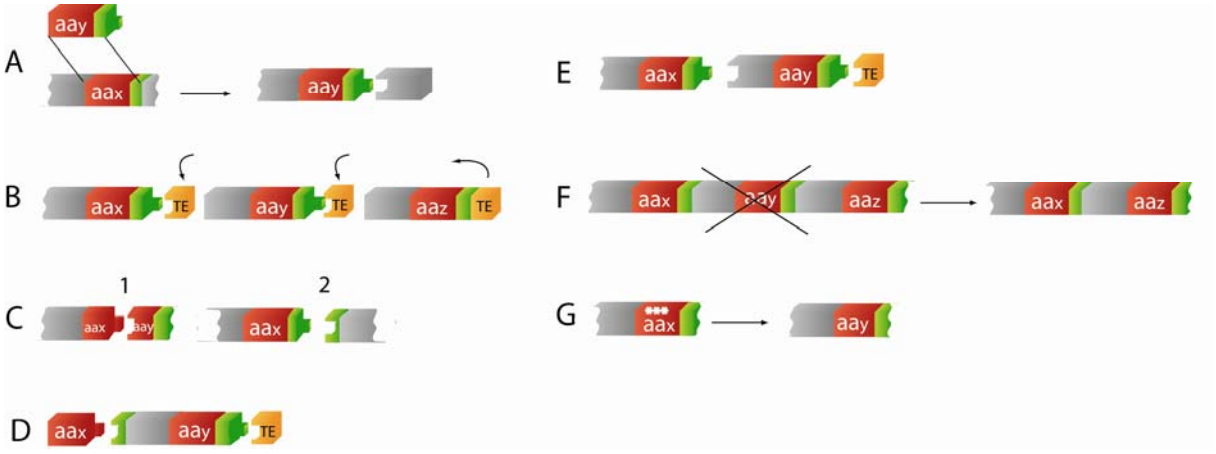


Fig. 6 Manipulation of NRPS biosynthetic pathways
(A) *In vivo* swapping of A-PCP didomains (minimal modules). **(B)** Translocation of TE domain to generate truncated peptides **(C)** NRPS fusion within A and PCP domains (1 and 2 respectively) **(D)** Construction of di-modular hybrid NRPSs by fusion at the linker region between A and PCP **(E)** Construction of di-modular hybrid NRPS by fusion of the elongation module C-A-PCP, using as fusion site the linker region between PCP and C **(F)** Deletion of an elongation module (C-A-PCP) to obtained a shortened peptide **(G)** Modification of the specificity of the A domain by introduction of point mutations

2.7.2 The chemoenzymatic approach

The increasing knowledge about NRPS structure and function provided the basis for the genetic manipulation of NRPS clusters. However, most of these attempts were still connected with significant reductions in productivity of the newly engineered biosynthetic assembly lines.

Recently, chemoenzymatic approaches, which combine the advantages of natural biosynthesis and solid-phase peptide synthesis, were shown to be powerful alternatives for the introduction of targeted modifications in the NRP structures. In this connection, the template-driven synthesis of NRPs was replaced by solid-phase peptide synthesis (SPPS), which – among others – allowed for the incorporation of a variety of substrate analogues in a given NRP product, and even variation of the chain length. On the other hand, macrocyclization of the final product, which is usually difficult to achieve by means of organic chemistry, e.g. due to problems associated with steric repulsion, mis-alignment of ring residues, and the requirement of protecting groups, is achieved by enzymatic cyclization as catalyzed by the excised, stand-alone TE domains of the corresponding biosynthetic complex. Given their regio and stereoselectivity, TE domains permitted the specific formation of the desired macrocyclic structures of the natural product and its derivatives.

In 2002, the concept of chemoenzymatic NR synthesis was for the first time proven on the example of the TycC TE domain, derived from the tyrocidine biosynthetic complex. Kohli et al. demonstrated that the isolated TycC TE domain is able to catalyze the cyclization of a variety of tyrocidine-precursor molecules, which had been synthesized by SPPS, and were bound to a solid support, mimicking the natural Ppant linker [Kohli et al., 2002]. This property of TycC TE was exploited to construct a library of 192 tyrocidine derivatives, varying in the amino-acid composition at positions 1 and 4 of the cyclic decapeptide. The generated combinatorial library was subsequently screened for natural product analogues with potential therapeutic utility, i.e. compounds that have an increased preference for bacterial over eukaryotic membranes, and an improved spectrum of activity against common bacterial pathogens. Interestingly, most of the synthesized variants showed improved therapeutical properties compared to the natural decapeptide.

Notably, NRPS TE domains not only exhibit a relatively large promiscuity for the composition of the linear precursor peptide, but also towards the nature of the corresponding leaving group, activating the terminal carboxy moiety. Initial approaches e.g. used N-acetylcysteamine (SNAC), representing a mimicry of the naturally occurring Ppant co-factor of NRPS PCP domains [Kohli et al., 2001; Trauger et al., 2000; Trauger et al., 2001], while in

the further studies Kohli et al. were directly using the polyethilene glycol amide (PEGA) support during SPPS [Kohli et al., 2002]. More recently, Sieber et al., reported about the superior performance of thiophenol leaving groups [Sieber et al., 2004], which was used e.g. for derivatization of the antibiotics daptomycin [Grünewald et al., 2004], and pristinamycin [Mahlert et al., 2005].

The described chemoenzymatic approach represents a valuable method to introduce specific modifications in a certain peptide structure, and in a few cases it has actually been shown to give rise to variants with an improved pharmacological activity.

Given the huge availability of amino acid analogues, chemo-enzymatic synthesis provides countless possibilities for the generation of natural product derivatives. On the other hand, the approach is limited by the mentioned stereo and regioselectivity of TE domains, excluding certain – especially terminal – positions from being modified. Moreover, once an interesting variant has been identified, preparative-scale chemoenzymatic synthesis may be the only method to making available higher quantities of a given compound. However, the up-scalability of this technique has still to be proven.

2.8 Combinatorial libraries of natural products

Although the rational design of NRPs via the genetic manipulation of the protein template or chemoenzymatic synthesis may lead to the formation of a desired peptide, every single modification to be introduced requires an individual experiment, making – especially the genetic manipulation – laborious and time-consuming.

In contrast to the rational design of NRP biosynthetic assembly lines, a so-called combinatorial approach would allow to easily interrogate the effect of individual modifications, and combine them in all possible permutations, to generate a large library of analogues. The advantage towards rational design would be the possibility to create numerous biosynthetic pathways, and potentially a corresponding number of novel products with less effort. The approach may, however, be limited by the chance that some of the random combinations could be incompatible, causing an impeding or even abortion of product formation.

Up to now, the prospects of combinatorial biosynthesis have only been studied for polyketide synthase complexes. Consequently, the next paragraph gives (i) a brief introduction in the mechanism of polyketide synthesis, and (ii) a summary of two *in vivo* studies, in which the

well known erythromycin PKS assembly line was used as model for first steps towards a true combinatorial biosynthesis.

2.9 The first steps towards combinatorial biosynthesis in polyketide synthases.

Polyketide synthases (PKSs) are large multi-enzyme complexes, which catalyze the synthesis of pharmacologically important secondary metabolites, including the antibiotic erythromycin and the immunosuppressant FK506. Like NRPSs, type I PKS complexes can be subdivided into individual modules and catalytic domains, dedicated to recognition, activation and loading of specific acyl units. The nature and order of the single catalytic units reflects the stepwise assembly of the ketide chain, and the overall mechanism resembles that of the synthesis of fatty acids.

The biosynthesis of polyketides is initiated by the loading module, and terminated – like in NRPSs – by a thioesterase domain (TE), which releases the product. Each elongation module contains at least three catalytic core domains: acyl transferase (AT), acyl carrier protein (ACP) and ketosynthase (KS), which essentially have the same function as A, PCP and C domain in NRP biosynthesis, respectively. The AT domain recognizes an appropriate (cognate) acyl-CoA extender unit and transfers it to the prosthetic group of the associated ACP. The KS domain catalyzes a decarboxylative Claisen condensation between the growing polyketide chain and the activated, ACP-bound acyl unit, yielding the corresponding β-ketoacyl-ACP. In addition to the mentioned core domains, elongation modules contain optional ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains, required for the processing of the β-keto group. Consequently, depending on the domain organization, a given elongation module may lead to the formation of hydroxylated, unsaturated and saturated C-C bonds.

The DEBS multi-enzyme complex, responsible for the synthesis of the erythromycin aglycone precursor 6-deoxyerythroenolide-B (6-dEB), represents the best characterized example of a type I PKS complex. It consists of three enzymes - DEBS1, DEBS2 and DEBS3 -, constituted of three, two and two modules, respectively (Fig. 7). The encoding DEBS genes are organized in a polycistronic operon, comprising about 50 kb in the chromosome of the producer strain *Saccaropolyspora erythrea*.

One propionyl-CoA and six methyl-malonyl-CoA units are required for the synthesis of the 6-deoxyerythroenolide B (6-dEB). The loading module, consisting of AT and ACP domains,

binds the starter propionyl-CoA, while the extender modules catalyze the chain elongation, by stepwise addition of methylmalonyl-CoA units. Product release and cyclization are catalyzed by the C-terminal TE domain.

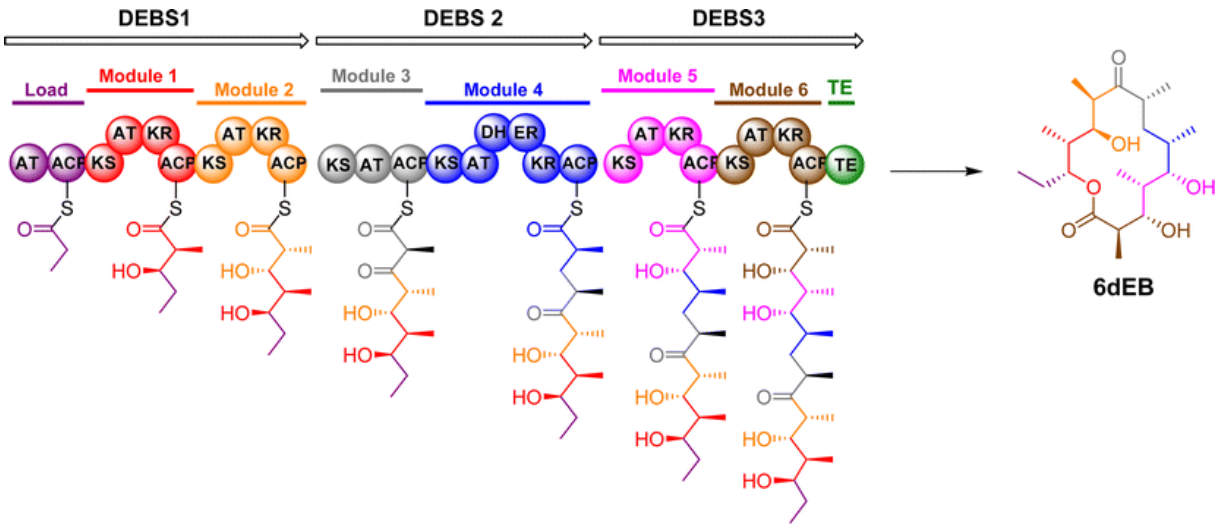


Fig. 7 Biosynthesis of the erythromycin precursor 6-deoxyerythroenolide-B (6-dEB)
The biosynthetic assembly line consists of three synthases: DEBS1, DEBS2 and DEBS 3.[Menzella et al., 2005]

A significant step towards the reprogramming of 6-dEB biosynthetic pathway was provided by the genetic manipulation of the erythromycin producer *Saccaropolispora erythraea* [McDaniel et al., 1999]. A library of more than 50 macrolides was generated by genetic manipulation of the DEBS pathway. Single and multiple domain deletions and substitutions were introduced, generating novel biosynthetic assembly lines, which were subsequently analyzed for their ability to synthesize the expected 6-dEB derivatives. Hereby, different AT and optional β-carbon processing domains were substituted against corresponding counterparts derived from the rapamycin biosynthetic pathway. Finally, all possible permutations of functional single mutants were generated. As mentioned above, this "combinatorial" approach gave rise to more than 50 novel macrolides. At the same time, the study also revealed two major limitations of this approach: i) each novel hybrid assembly line, and consequently each product to be generated, requires a separate engineering step, and ii) the genetic manipulation of the 6-dEB biosynthetic pathway in *S. erythraea* is technically difficult and time-consuming.

A more powerful strategy, which actually represents a step towards true biocombinatorial synthesis, is the so-called "multiple plasmid approach". The general principle is illustrated in Fig. 8, for a hypothetic operon, consisting of the three genes a, b and c. x variants of a, y variants of b and z variants of c are cloned in three independent plasmids, and subsequently combined in (x) x (y) x (z) different permutations. Hence, the advantage of this approach is

that it requires only (x) + (y) + (z) manipulations, in order to generate a library, constituted of (x) x (y) x (z) different compounds. Consequently, the "multiple plasmid approach" allows for the construction of relatively large libraries from a comparable small number of modified genes. Taking advantage of the principle, Xue et al. constructed four variants of DEBS1, two variants of DEBS2, and eight variants of DEBS3, The encoding genes were cloned into three compatible *Streptomyces* vectors, and used (in all possible combination) for the co-transformation of the heterologous host *S. lividans*. The resulting, 64 triple transformants were subsequently investigated for the PK production. This analysis revealed that a library of 43 polyketides was obtained, including 15 completely novel polyketides. [Xue et al., 1999].

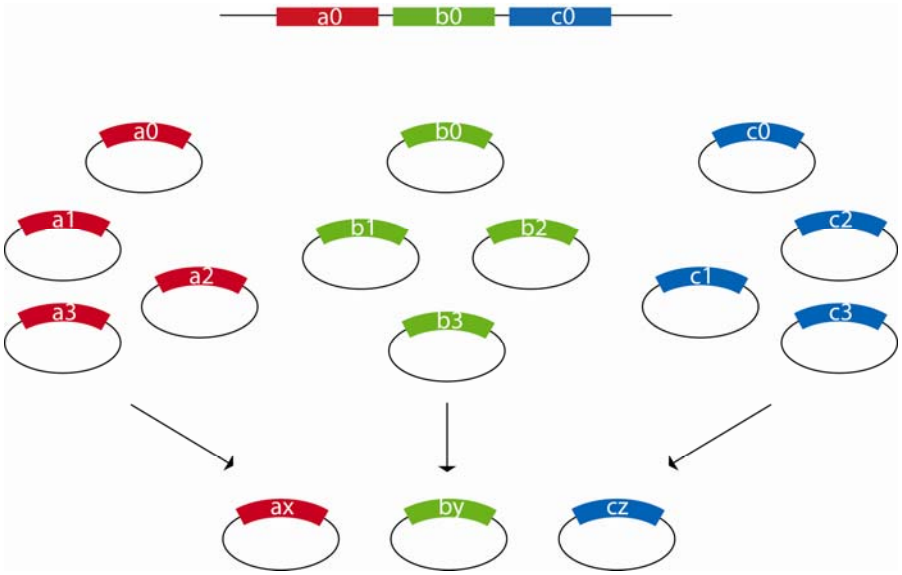


Fig. 8 Schematic presentation of the multiple plasmid approach.
The wild type genes and variants thereof are cloned into individual expression vectors, which can be maintained in the heterologous host. The plasmids are combined in all possible permutations, and the resulting strains are subsequently tested for the capability to synthesize the corresponding products.

The described multitple plasmid approach represents a significant step towards combinatorial biosynthesis. However, the approach is still limited by the fact that every engineered host strain is only able to give rise to one biosynthetic assembly line, and consequently only one sinlge product. A more challenging goal, though, would be the construction of a library of (hybrid) PKS or NRPS assembly lines, in which all enzymes can interact with each other, randomly arranging into different biosynthetic assembly lines, and producing an array of different products at the same time. Achieving this goal, however, requires an in-depth knowledge about the factors, regulating protein-protein interactions within biosynthetic complexes. For PKSs, several *in vitro* and *in vivo* studies led to the identification of intra- and intermolecular regions, referred to as linker or docking domains, which are responsible for the

defined protein-protein interaction between PKSs [Gokhale et al., 1999]. Only recently, functional analogous of PKS docking domains could be also identified in NRPS complexes.

2.10 Protein-protein communication in NRPS.

The majority of NRPS complexes consist of several enzymes, whose coordinated interplay leads to the synthesis of the final peptide product. Hence, the biosynthesis of a defined NRP relies on the NRPS's capability to establish correct interactions between partner enzymes within the multi-enzyme complex, and to avoid false interactions between non-partner enzymes, which would lead to the synthesis of wrong peptide products. For example, formation of the full-length lipoheptapeptide surfactin, requires the coordinated interaction between the last module of SrfA-A and the first module of SrfA-B, as well as the last module of SrfA-B and the first (only) module of SrfA-C, respectively (see Fig. 12). In contrast, the direct interaction between SrfA-A and SrfA-C would lead to the formation of a shortened lipotetrapeptide. Recently, the discovery of small communication-mediating (COM) domains, located at the N-terminus and C-terminus of acceptor and donor enzymes, respectively, shed light on the mechanism that ensures the selective interaction between NRPSs.

The existence of C-terminal donor COM domains (COM^D) was first proven for the initiation module TycA, derived from the tyrocidine biosynthetic complex. Based on *in vitro* studies it was known that the C-terminal half of TycA (domain organization: A-PCP-E) is indispensable for the establishment of a heterodimeric complex with the partner elongation module TycB1. By investigation of C-terminal deletion mutants of TycA, Hahn *et al.* found that the removal of the six most C-terminal amino acids already led to complete disruption of protein-protein communication. In contrast, the activity of the TycA E domain was not affected even after deletion of the 23 most C-terminal amino acids. This observation suggested that the loss of protein communication between TycA and TycB1, rather than an inactive E domain was the reason for the loss of DKP formation [Hahn and Stachelhaus, 2004].

Determination of TycA's partner COM domain, located at the N-terminus of the acceptor module TycB1, was experimentally a bit more demanding since simple generation of N-terminal deletion mutants of TycB1 was not a practicable approach. Consequently, chimeric derivatives of the acceptor modules TycB1 and TycC1, as well as the donor modules TycA and TycB3 were constructed by substituting their putative COM domains. Studying all possible combinations of acceptor and donor modules, Hahn et al. observed that efficient

product formation could be only determined for systems, harbouring matching pairs of COM domains. In contrast, the product formation completely failed when donor and acceptor module were equipped with non-matching pairs of COM domains. Interestingly, this observation also concerned partner NRPSs, when only one COM domain had been exchanged. For example the donor COM domain of TycA (COM^{D}_{TycA}) was replaced by the donor COM domain of the third module of TycB (COM^{D}_{TycB3}), leading to the hybrid protein TycA::COM^{D}_{TycB3}. This hybrid enzyme was uncapable to productively interact *in vitro* with the first module of TycB (organization of the system. TycA::COM^{D}_{TycB3}/TycB1). However, the activity of these systems could be restored, when the corresponding partner NRPS was also provided with the unnatural, but matching COM domain (i.e. TycA::COM^{D}_{TycB3}/COM^{A}_{TycC1}::TycB1).

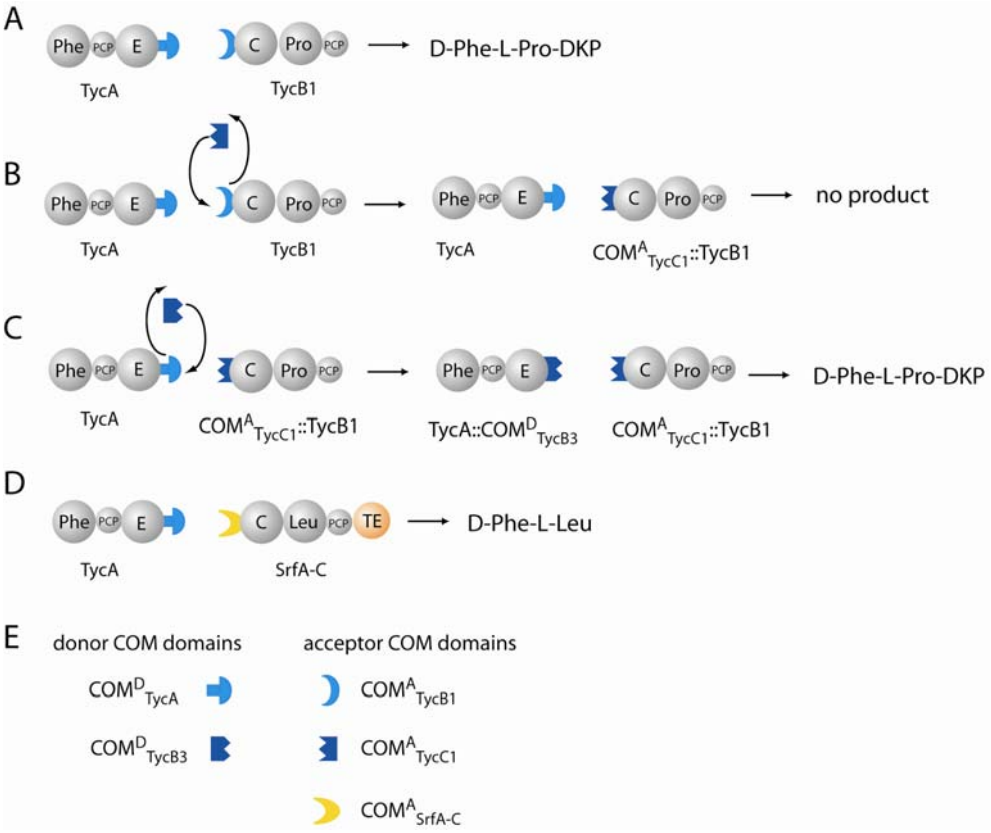


Fig. 9 *In vitro* investigation of COM domains swapping and crosstalk between different NRPS biosynthetic systems. Dimodular NRPS systems were used to investigate the *in vitro* swapping of COM domains. Productive interactions were indicated by detection of the predicted dipeptides. **(A)** The interaction between the natural partners TycA and TycB1, derived from the tyrocidine assembly line, leads to the formation of D-Phe –L-Pro DKP. **(B)** Substitution of the acceptor COM domain COM^{A}_{TycB1} with COM^{A}_{TycC1} created the hybrid enzyme COM^{A}_{TycC1}::TycB1, which is not able to interact with TycA, due to the incompatibility of acceptor and donor COM domains. **(C)** The hybrid enzyme TycA::COM^{D}_{TycB3}, created by swapping of the donor COM domain COM^{A}_{TycB3}, could productively communicate with the hybrid enzyme COM^{A}_{TycC1}::TycB1, leading to D-Phe-LPro-DKP formation. **(D)** An example of crosstalk between two different NRPS systems: given the compatibility of their donor and acceptor COM domains, TycA from the tyrocidine assembly line can interact with SrfA-C from the surfactin assembly line to give rise to the dipeptide D-Phe-L-Leu. **(E)** List of symbols used to represent acceptor and donor COM domains.

All these experiments (i) proved the presence of N-terminal COM-domain, (ii) verified the generality of COM domains at the junction between different partner NRPSs, and (iii) demonstrated that matching pairs of COM domains are essential for the establishment of productive interactions between NRPSs. Most remarkably, the study of Hahn et al. also showed that matching pairs of COM domains can be used to enforce protein-protein communication between non-partner enzymes (i.e. TycA/COM^{A}_{TycB1}::TycC1 and TycB3::COM^{D}_{TycA}/TycB1), and even between NRPSs derived from different NRP biosynthetic systems (see Fig. 9D).

These data allowed for the establishment of the following model for the description and understanding of protein-protein-communication within multi-enzymatic complexes.

Within a hypothetical tri-enzyme, tri-modular NRPS system A-B-C, the flux of reaction intermediates – and hence, the specific formation of the product a-b-c – is controlled by the COM domains pairs COM^{D}_{A}/COM^{A}_{B} and COM^{D}_{B}/COM^{A}_{C}. Given the COM domain's selectivity, the initiation module A can only interact with the elongation module B, yielding the formation of the enzyme-bound intermediate a-b-S-Ppant. The dipeptidyl moiety is subsequently transferred from the elongation module B to the termination module C, whose interaction is facilitated by the COM domain pair COM^{D}_{B}/COM^{A}_{C}. As a consequence, the tri-peptide product a-b-c-S-Ppant is formed, and subsequently released from the biosynthetic template. However, if the tri-enzyme system used the same COM domain pairs in order to mediate the communication between A and B, as well as B and C, then the initiation module A should be likewise able to establish productive interactions with the elongation enzyme B (complex: A-B-C) and the termination module C (complex: A-C) (see Fig. 10A). Consequently, this degenerated system should catalyze the formation of a mixture of the dipeptide a-c, as well as the tri-peptide a-b-c, rather than the specific formation of just one defined product. This, however, would meet the requirements for true biocombinatorial NRP biosynthesis.

The validity of this general concept of an universal-communication-system (UCS) for the biocombinatorial synthesis of NRP products could be recently proven, using the initiation module TycA, the elongation module COM^{A}_{TycB1}::BacB2(CAT)::COM^{D}_{TycA} and the termination module SrfA-C, leading to the simultaneous *in vitro* production of the dipeptide D-Phe-L-Leu and the tripeptide D-Phe-L-Orn-L-Leu [Hahn and Stachelhaus, 2006]. The simple UCS can be interpreted as the first step towards a true biocombinatorical synthesis of NRPs.

An example hereof is shown in Fig 10B, in which a pool of NRPSs is represented, carrying compatible acceptor and donor enzymes. Random combination of those compatible enzymes could generate an array of novel biosynthetic NRPS pathways.

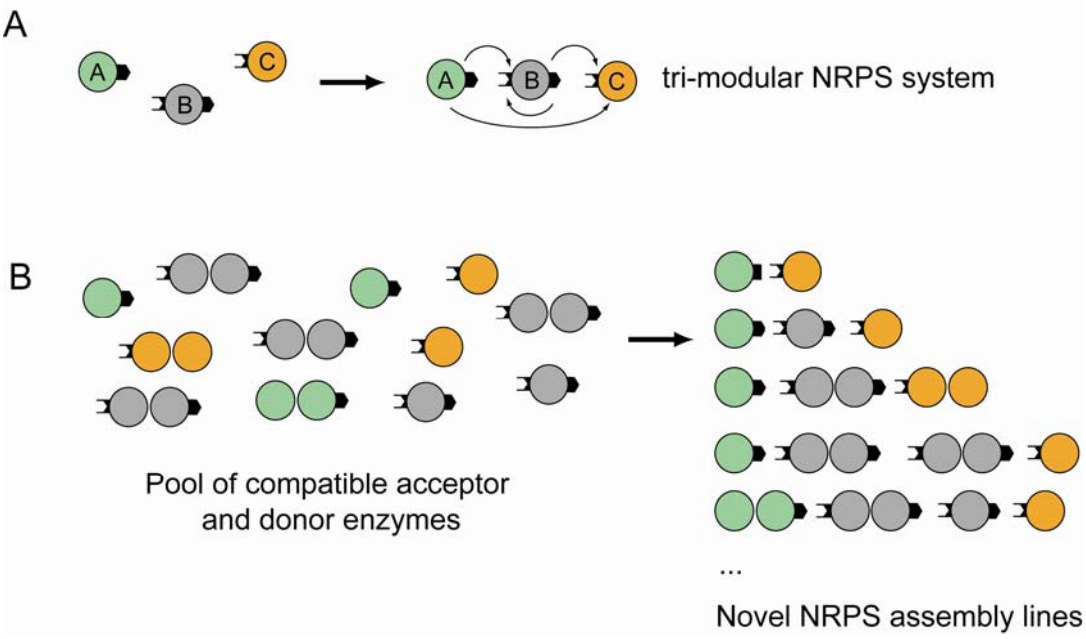


Fig. 10 Combinatorial potential of the universal communication system
(A) Schematic representation of a tri-modular NRPS system in which initiation, elongation and termination modules communicate through the same COM domain pair. In such a system, protein A should be able to communicate with both enzymes, B and C, resulting in the simultaneous formation of an array of different peptide products. In **(B)** the same COM domain pair could mediate the interaction between a pool of initiation (in green), elongation (in grey) and termination modules (in orange). The random combination of those NRPS modules would generate novel NRPS assembly lines in which, due to the abrogation of the selectivity barrier imposed by different pairs of COM domains, all acceptor and donor enzymes can productively interact.

2.11 Heterologous expression of NRPS genes.

Historically, investigations of the gramicidin S and tyrocidine biosynthetic complexes of the Gram-positive bacterium *B. brevis* have significantly contributed to our current understanding of NRP biosynthesis. For example, characterization of the biosynthetic gene clusters provided first insights into the modular organization of the multi-enzymatic complexes. A more careful in-depth analysis of individual catalytic domains and modules, however, was hampered by the genetic inaccessibility of the producing organisms. Consequently, corresponding gene fragments had to be sub-cloned into suitable expression vectors, in order to allow for their heterologous expression. Host organism of choice was the well-characterized Gram-negative bacterium *E. coli*, which in fact was successfully used for the production of functional NRPS fragments. The recombinant proteins were purified (usually by taking advantage of affinity tags), biochemical characterized, and in some cases even crystallized, in order to understand

catalytic mechanisms involved in NRP biosynthesis (on the atomic level) [Mootz and Marahiel, 1997; Stachelhaus and Marahiel, 1995b; Stachelhaus et al., 1996].

This concept for studying NRP biosynthesis worked out fine, as long as it concerned only simple mono- or didomainal constructs, or single modules. In contrast, heterolohous production of active di- and tri-modular enzymes in *E. coli*, was often hindered by several factors: i) insolubility of the synthesized protein, ii) incorrect folding, and iii) protein degradation. Even more challenging was (and still is) the expression of NRPS genes derived from filamentous fungi and actinomyces, given their high GC content, and different codon usage when compared to *E. coli*. In these cases, problems are already encountered at the stage of gene expression, and therefore alternative hosts like *Strepomyces lividans* and *Neurospora crassa* were usually preferred, which allowed for biochemical characterization of the corresponding NRPSs [Schauwecker et al., 2000; Smith et al., 1990].

Apart form the problems just mentioned, another problem during the heterologous production of NRPSs is represented by the absence of a suitable phosphopantetheine transferase (PPTase), the enzyme committed to the post-translational modification of NRPS PCP domains.

2.12 Post translational modification of NRPS

The conversion of NRPSs from the inactive apo-into the active holo-form is accomplished by the covalent loading of a 4´-Ppant moiety of CoA onto the side chain of a conserved PCP serine residue. This post-translational modification is catalyzed by specific Ppant transferases (PPTases). Members of this enzyme family include Sfp of *B. subtilis* and Gsp of *B. brevis*, which modify PCPs of the surfactin and gramicidin S biosynthetic pathways, respectively. *In vitro* modification experiments demonstrated that Sfp is a rather promiscuous enzyme, accepting not only PCPs of the surfactin biosynthetic assembly line as substrates, but also heterologous PCPs, as well as ArCPs and ACPs involved in PK and fatty acid biosynthesis [Lambalot et al., 1996; Quadri et al., 1998]. Consistent with these results are *in vivo* investigations, showing that Sfp can complement the phenotype of an AcpS-deficient *E. coli* strain [Mootz et al., 2002b]. This versatility renders Sfp a fundamental tool for two important *in vitro* applications: i) the modification of the mentioned carrier protein domains into the active holo-form, and ii) the loading of chemically synthesized aminoacyl- and peptidyl-CoA substrates onto apo-PCPs [Belshaw et al., 1999; Sieber et al., 2003].

Interestingly, the mentioned substrate promiscuity appears to be a peculiarity of certain PPTases, involved in the production of secondary metabolites. In fact, biochemical

characterization of the two native *E. coli* PPTase, AcpS and EntD, revealed a high specificity of these enzymes for their homologous substrates: *EntD* of the enterobactin biosynthetic pathway (secondary metabolism), and AcpS of the FAS (primary metabolism), are both specific for their homologous peptidyl and ACP substrates, respectively. Consequently, the heterologous expression of NRPS genes in *E. coli* leads to the production of inactive apo-enzymes [Gocht and Marahiel, 1994; Stachelhaus and Marahiel, 1995a], which have to be modified *in vitro* before undergoing biochemical characterization. However, recently it was shown that *E. coli* is able to carry out the apo- to holo-enzyme conversion, once provided with corresponding genes, encoding for either Sfp or Gsp [Gruenewald et al., 2004; Mootz et al., 2002b]. Notably, co-expression of *gsp* and the DNA coding region of the truncated A-PCP initiation module of the gramicidin S biosynthetic pathway led to predominant (80%) *in vivo* production of functional holo-enzyme [Ku et al., 1997].

2.13 Purification of NRPSs from the natural producer strain

As already mentioned, NRPs are produced by a variety of different microorganisms, including Gram-positive and Gram-negative bacteria, as well as filamentous fungi. Given the differences in codon usage and the incompatibility of the promoters, finding a surrogate host for NRPS gene expression is particularly difficult. A possible alternative, which was widely exploited in the early years of NRPS research, is provided by the utilization of the native producer strain, ensuring for the biosynthesis of soluble, correctly folded and post-translationally modified holo-enzymes. However, protein purification from the native producer suffers from at least two major disadvantages: i) the impossibility to overproduce the protein by using strong, inducible promoters, and ii) the necessity of performing a multi-step purification, given the lack of appropriate affinity tags. For example, the surfactin biosynthetic complex was partially purified from crude extracts of the native producers *B. subtilis* ATCC 21332 and OKB105, using five purification steps, including ammonium sulphate precipitation, gel filtration, ultracentrifugation in sucrose gradient and two ion-exchange chromatography runs [Menkhaus et al., 1993]. All this makes purification of NRPSs from the natural producer rather time-consuming, and – given the usually low expression levels – enforces large-scale production, in order to obtain the enzyme quantities required for biochemical characterization.

2.13.1 Immobilized metal ion affinity chromaptography (IMAC)

A possible solution to overcome the limitations would be to combine the obvious advantages of protein production in the natural host organism, with fast and effective purification schemes as e.g. provided by single-step immobilized-metal affinity chromatography (IMAC). The suitability of this approach for the purification of large NRPSs was already shown by the successful purification of recombinant di- and tri-modular enzymes, which were heterologously produced *in E. coli* [Duerfahrt et al., 2004; Symmank, 2002]. The corresponding adaptation for the purification of NRPSs from the natural producer strain, however, would require the fusion of the coding sequence of a hexahistidine affinity tag onto the 5'- or 3'-end of the corresponding NRPS gene, making this methodology only applicable for genetically accessible microorganisms.

2.13.2 Immunoaffinity chromatography, using polyol-responsive monoclonal antibodies (PR-mAB)

Another powerful alternative to traditional purification procedures is represented by immuno-affinity chromatography. Given the tight and specific binding between antigen and antibody, this method commonly yields highly pure protein preparations. However, the method has only limited applicability for the purification and biochemical characterization of NRPSs, since it requires large quantities of pure NRPS protein in the first place, in order to raise the specific antibodies. Even more importantly, the conditions required to elute the antigen, or rather to disrupt antigen/antibody complex, are extremely harsh, leading to the denaturation of the desired NRPSs (see Fig. 11). Examples are the utilization of extreme pH values (pH 3 or 10), denaturing agents (8 M urea or 6 M guanidinium hydrochloride) or chaotropic salts (3 M KSCN) that disrupt protein structure.

Recently a new technology was developed in R. Burgess' laboratory, which is based on the utilization of so-called polyol-responsive monoclonal antibodies (PR-mAb). These antibodies also very tightly bind to their antigen, however, the corresponding complex can be easily destroyed under mild, non-denaturating conditions, using small polyhydroxylated compounds (polyols). Elution under those gentle conditions permitted the purification of multi-enzyme complexes [Thompson et al., 2003] associated proteins. This technology was successfully exploited for the purification of fully-active multi-subunit RNA polymerase holo-enzymes from different microorganisms. Most notably, the methods also allowed for a co-purification of associated sigma factors (e.g. σ70), which are known to be only weakly bound to the RNA polymerase core enzymes [Bergendahl et al., 2003].

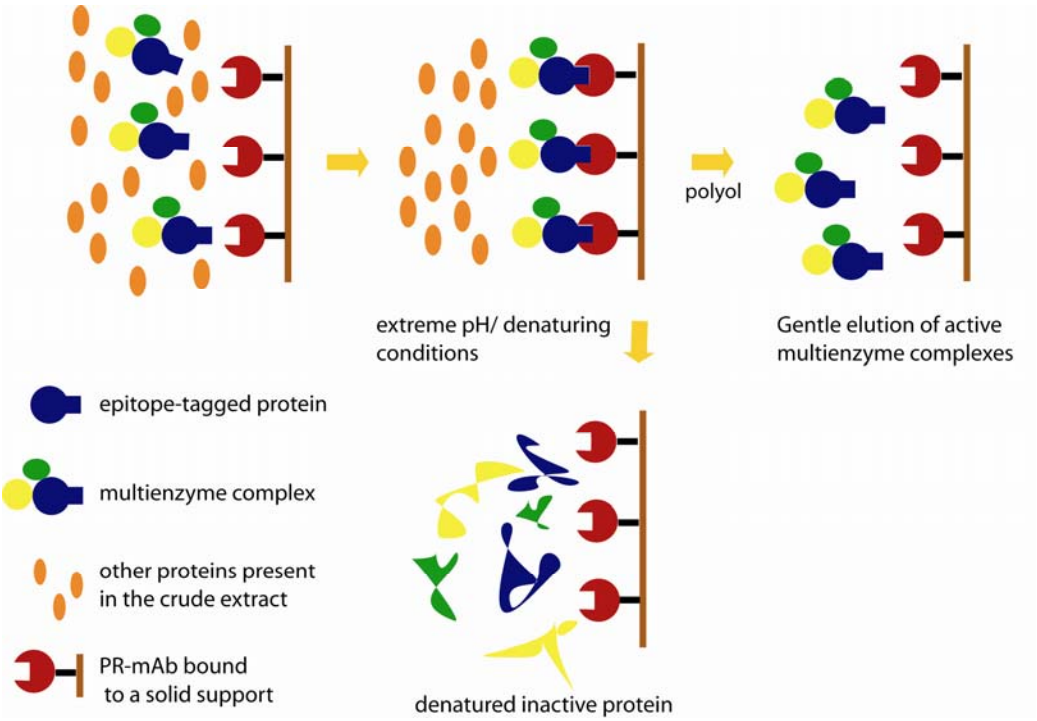


Fig. 11 Immunoaffinity chromatography using polyol responsive monoclonal antibodies (PR-mAb). Immunoaffinity chromatography is not widely used because of the denaturing conditions required to destroy the binding between antibody and antigen. The use of mPR-Abfor immunoaffnity chromatography allows for the elution of the antigen under gentle conditions: in presence of a low molecular weight polyhydroxylated compound (polyol) and non chaotropic salts.

2.14 Goal

The presented work was intended to serve two different purposes: (i) the investigation of protein-protein communication within the context of a natural NRP biosynthetic assembly line, as well as (ii) the establishment of a gentle method for the purification multi-modular NRPSs. As a model system for both studies, the well-characterized surfactin biosynthetic complex of the Gram-positive soil *bacterium B. subtilis* was chosen.

Recent *in vitro* studies revealed the importance of short, so-called communication-mediating (COM) domains for the control of the selective interaction between NRPSs. Within the scope of this work, this decisive role of COM domains for the establishment and/or prevention of productive interactions between NRPSs should be verified for the first time *in vivo*, within the context of a natural NRP assembly line. To this end, COM domains swapping experiment should be carried out, and the consequences of the swaps evaluated by means of HPLC and MS analysis. The experiments were intended (i) to verify the functionality of COM domains within the context of heterologous, biosynthetic systems (and host organisms), (ii) to achieve the targeted reprogramming of a NRPS assembly line, (iii) to achieve the intentional skipping of certain biosynthetic enzymes, and (iv) to abrogate the proposed selectivity-barrier provided by COM domains, in order to allow for the true biocombinatorial synthesis of NRPs.

In the second part of this work a quick and reliable method for purification of fully-active, multi-modular NRPSs from a natural NRP producer strains should be developed. In contrast to the multi-step procedures traditionally used, the applicability of single-step chromatography procedures like i) immobilized metal ion affinity chromatography (IMAC), and ii) immunoaffinity chromatography, using polyol responsive monoclonal antibodies, should be tested. PR-mABs were originally developed for the purification of functional multi-enzyme complexes (namely: RNA polymerase core enzyme). Consequently, the latter method should be also evaluated for the possibility to co-purify partner NRPSs and other associated proteins.

3 Material

3.1 Microorganisms

Tab. 3.1: Microorganisms

Strain	Relevant genotype/description
B. subtilis ATCC 21332	Wild type strain, surfactin producer [Cooper et al., 1981]
B. subtilis OKB105	Wild type strain, surfactin producer OKB105 Surfactin-positive transformant of *B. subtilis* JH642, *pheA1 sfp* [Nakano et al., 1988]
E. coli DH5α	*F´/endA1, hsd*R17, *(r^-_K, m^+_K), glnV*44, *thi-1, recA1, gyrA, (Nal^r)(relA1, ΔlacIZYA-orgF) U*169, *deoR (Φ80dlacΔ(lacZ/*M15*)* [Woodcock et al., 1989]
E. coli TOP10	*F´- mcrA Δ (mrr-hsdRMS-mcrBC) Φ*80*lacZΔM15 ΔlacZY74 recA1 araD*139 *Δ(ara-leu)*7697 *galU galK rpsL (Str^R) endA1 nupG* Invitrogen, Karlsruhe, Germany

3.2 Plasmids

3.2.1 pQE60

The plasmid pQE60 (3,4 kb) (Qiagen, Hilden, Germany) belongs to the family of the pDS-plasmid [Bujard et al., 1987] and is a pDS56/RBSII derivative [Stüber, 1990]. It contains the ColE1 replication origin and the β-lactamase gene *bla* from pBR322 [Sutcliffe, 1979] along with an optimized promoter/operator element, consisting of the *E. coli* T5 phage promoter P_{N25} [Gentz and Bujard, 1985] and two recognition sequences for the *lac*-operator. The synthetic ribosomal binding site RBSII ensures an improved mRNA recognition and binding. The plasmid contains a specific multiple cloning site (MCS), containing the restriction sites for the endonucleases *Nco*I, *Bam*HI and *Bgl*II. The ATG start codon, integrated in the *Nco*I recognition site is located in an optimal distance from the synthetic RBSII The coding

sequence for a hexahistidine tag is located downstream of the MCS, followed by stop codons in all three reading frames and the transcriptional terminator T_0 of the λ-phages [Schwarz E, 1987].

3.2.2 pDG646 /pDG782

The plasmids pDG646 / pDG792 are representatives of a vector collection harbouring different antibiotic resistance cassettes, which could be excised and subcloned for their utilisation in *B. subtilis* [Guerout-Fleury et al., 1995]

The plasmid pDG646, a derivative of pSB119 [Bouvier et al., 1991] contains the *mls* resistance cassette (1,6 kb) obtained from plasmid pE194 of *Staphylococcus aureus* [Horinouchi and Weisblum, 1982b] conferring resistance against the antibiotic mixture erythromycin and lincomycin. For the excision of the resistance marker, the upstream located endonuclease recognition sites of *Hind*III, *Sph*I, *Pst*I, *Sal*I, *Acc*I, *Xba*I and *Bam*HI as well as the downstream located recognition sites for the endonucleases *Nsi*I, *Cla*I, *Bam*HI, *Sma*I, *Asp*718, *Eco*RI and *Hind*III could be used.

The plasmid pDG783 is a pMTL22 derivative [Chambers et al., 1988] harbouring the *kan* cassette (1,5 kb) derived from the plasmid pAT21 of *Streptococcus faecalis* [Trieu-Cuot and Courvalin, 1983], which confers resistance against kanamycin. The resistance marker excision could be performed by the utilization of the downstream located recognition sequences of the endonucleases *Eco*RI, *Hind*III, *Sph*I, *Pst*I as well as the upstream located recognition sequences of the endonucleases *Hind*III, *Xba*I, *Bam*HI, *Sma*I, *Asp*718, *Sac*I and *Eco*RI.

3.2.3 pCm::Tc

The plasmid pCm::Tc [Steinmetz and Richter, 1994], which was used for the co-transformation of *B. subtilis*, is a derivative of pLC177. This plasmid contains the replication origin ColE1 for *E. coli* derived from pBR322 [Sutcliffe, 1979] as well as the *B. subtilis*-ori derived from pE194 [Horinouchi and Weisblum, 1982a]. Therefore, the plasmid can be replicated in both microorganisms.

In pCm::Tc, the original chloramphenicol resistance cassette derived from the plasmid pC194 of *Staphylococcus aureus* is interrupted with a tetracycline resistance cassette, which confers tetracycline resistance in *B. subtilis*.

3.2.4 pTOPO-XL

pTOPO-XL PCR vector (3,5 kb), purchased from Invitrogen™, belongs to the family of pUC vectors and carries the replication origin ColE1. The plasmid, provided in linear form, carries at the 3'-terminus of both strands a thymine overhang and is used for ligation of PCR products carrying an adenine overhang. The ligation reaction is catalyzed by the topoisomerase I, which is covalently bound to the 5'-terminus of both TOPO-XL DNA strands by a lysine residue. The energy set free after the breakage of the phospho-ester bond is used by the enzyme for the ligation reaction. A kanamycin and a zeomycin resistance gene are available for selection of the transformants.

3.2.5 pKE27

pKE27 [Eppelmann et al., 2001] is a pDR66 derivative [Ireton, 1993], containing an origin of replication for maintenance in *E.coli*, an amp cassette conferring resistance to ampicillin *in E.coli* and a cm selectable marker for *B. subtilis,* conferring chloramphenicol resistance. Furthermore this vector carries a copy of the *comS* gene, flanked by the two homolgous region *amyE* front and *amyE* back, which allow insertion of the vector by homolgous recombination in the non essential gene *amyE.*

3.3 Media

For cultivation of *E. coli* and *B. subtilis,* LB medium [Sambrook et al., 1989] and Difco Sporulation Medium (DSM) [Nakano et al., 1988] were used respectively. For preparation of solid media 1,5% agar was added before sterilization by autoclaving.

LB-Medium:

Bactotrypton	5 g/l
Yeast extract	10 g/l
NaCl	5 g/l

DSM:

Nutrient broth	8 g/l
$MgSO_4$ x 7 H_2O	120 mg/l
KCl	1 g/l
NaOH	600 mg/l

After autoclaving 1 ml/l of the following solutions were added:

$CaNO_3$	1 M
$MnCl_2$	100 mM
$FeSO_4$	1 mM

Antibiotics

The antibiotics utilized were sterilized by filtration and added to the medium in the following concentration:

E. coli: ampicillin 50 µg/ml, kanamycin 25 µg/ml

B. subtilis: chloramphenicol 5 µg/ml, kanamycin 10 µg/ml, erytromycin 1 µg/ml, lincomycin 25 µg/ml, tetracyclin 20 µg/ml.

3.4 Chemicals and laboratory products

Chemicals which are not listed in the table below were purchased from the companies Fluka (Darmstadt, Germany) or Sigma (Deisehofen, Germany).

Tab.3.2: Chemicals and laboratory products

Manufacturer (Location)	Product
AppliChem (Darmstadt, Germany)	Yeast extract, Bactotryptone, antibiotics,
Beton Dickinson GmbH (Heidelberg, Germany)	Columbia Agar with 5% sheep blood
Difco (Detroit, USA)	Yeast extract, Bactotryptone, Nutrient Broth
Eurogentech (Seraign Belgium)	Agarose, cuvettes for electroporation
GE Healthcare Life Sciences (Freiburg, Germany)	Nyonmembran Hybond N^+, ECL Random Prime Labelling and Detection System, RNAse A, Lysozym, antibiotics, Proteinase K

Manufacturer (Location)	Product
Hartmann-Analytik (Braunschweig, Germany)	[^{3}H] and [^{14}C]-labelled aminoacids
Invitrogen,	NuPAGE® Bis-Tris Gels (acrilamide concentration 4% to 12%)
Kodak (Rochester, USA)	Autoradiograhy films X-Omat AR and Biomax MR, Delevoper D19
MWG Biotech (Ebersberg, Germany)	Oligonucleotides
New England Biolabs (Cologne, Germany)	Restriction endonucleases, concentrated T4-DNA-Ligase, Calf Intestine Alkaline Phosphatase (CIP), Polynucleotide Kinase (PNK), Protein- and DNA-markers
Qiagen (Hilden, Germany)	QIAquick-spin PCR purification kit
Roche (Mannheim, Germany)	Expand long template PCR system
Sartorius (Goettingen, Germany)	Nitrocellulose membrane

3.5 Equipment

Device	Manufacturer and Type description
Blot System	Strategene, Posiblot Pressure Blotter und Pressure control system
Centrifuges	Sorvall RC 26, Sorvall RC 5B, Spinco L2 65 B Hereaus, Eppendorf 5415 D
Electroporation system	BioRad Gene Pulser II
Cell Disruption system	Polytec, French Pressure Cell Press SLM Amicon
FPLC column	GE Healthcare Life Sciences, Chelating-Superose-Column HR 10/2 HiLoad^{TM}
FPLC	GE Healthcare Life Sciences, Gradient-Programmer GP250, Pump P-500, Uvivord optic- and control Unit UV-1, dual-channel flatbed potentiometric recorder REC102, Injection Valve V-7, 3-way Magnetic Valve PSV-100, Fraction Collector FRAC 100

Device	Manufacturer and Type description
HPLC-MS System	Agilent HPLC system 1100 MSD series with DAD and MSD detector, vacuum degasser, quaternary pump and HP Chemstation
Nano-HPLC system	Ultimate, Dionex, Idstein, Germany
HPLC column	Macherey & Nagel CC250/3 Nucleosil 120-3 C8 column
Hybridization oven	Bachofer, Mini 38
Mass spectrometer	Applied Biosystems, Darmstadt, Germany, QStar Pulsar
Mass spectrometer	Thermo Finnigan, Bremen Finnigan LTQ-FT
Shaker	New Brunswick Scientific, series 25 *Incubator Shaker*
Photometer	Amersham Biosensce Europe, *Ultrospec 3000 UV/Visible spectrophotometer*.
Bi-distilled water supply	Seral Seralpur Pro 90 C
Speed-Vac	Uniequip, Univap 150 H
Scintillation counter	Packard Instrument, TRI-CARB Liquid scintillation Analyzer 2100TR
Software	Lasergene DNA star 5.0 from GATC, Adobe Photoshop 7.0, Adobe Illustrator 10.0, Microsoft Office 2000
Thermo-Cycler	Mastercycler Personal Eppendorf
UV-Incubator	Stratagene, Stratalinker UV Crosslinker Model 1800
Vacuum pump	Vacuubrand, Membran Vacuumpumpe MZ2C

4 Methods

4.1 Molecular Biology

4.1.1 Construction of plasmids

DNA fragments were amplified from chromosomal DNA of *Bacillus subtilis* ATCC 21332 using the KOD Hot Start DNA Polymerase (Novagen, Merck Biosciences, Bad Soden Germany) and the "Expand long template PCR system" (Roche, Mannheim, Germany). Restriction sites (underlined) were introduced with oligonucleotides purchased from MWG-Biotech (MWG-Biotech, Ebersberg, Germany). DNA sequencing confirmed the identity of all plasmid constructed. Standard procedures were applied for all DNA manipulations [Sambrook et al., 1989].

pCC13

The 1,4 kb-comprising 5'-end of *srfA-B* was PCR amplified using the oligonucleotides srf_12F (5'-TAT AGA TCT TAG AGG TGG CAT ATG AGC AAA AAA TC-3') and srf_14R (5'-TAT TCT AGA CAC TTG GTG AAC AGC CAT TCC-3'), terminally modified using the endonucleases *Bgl*II, *Xba*I and ligated into the vector pQE60 previously cut in the same manner, to give pCC11. The mls resistence cassette obtained from DG646 via digest with *Bam*HI and *Bgl*II was inserted into pCC11 previously cut with *Bgl*II to give pCC12, which contains the *mls* cassette in the same orientation as well as s*rf*A-B. The 1,5 kb-comprising 3'-end of s*rfA-A* was PCR amplified with the oligonucleotides srf_11F (4'-ATA CCA TGG GAG CCG GAC ATC GAA GCG-3') and srf_13R (5'-ATA GGA TCC GAA AAT TTC CAT TAA TTT ATC CAG CTC-3'), terminally modified with the endonucleases *Nco*I and *Bam*HI and cloned into pCC12, cut likewise to give the disruption vector pCC13 (pQE60-*3'-srfA-A-mls-5'-srfA-B*).

pCC14

The 1,5 kb SrfA-A3'-His fragment was amplified using the oligonucleotides srf_11F (4'-ATA CCA TGG GAG CCG GAC ATC GAA GCG-3') His-BamHI_R (5'-ATA GGA TCC TCA ATG ATG ATG ATG ATG ATG-3'), terminally modified using the endonucleases *BamHI* and *NcoI* and cloned in the plasmid pCC11 previously cut likewise. After ligation the plasmid CC17 was obtained, which carries a hexahistidine tag coding region fused in frame to the *3'-srfA-A3* fragment and is followed by the 1,4 kb *5'-srfA-B* homolgous region.

pQE61-tycA/(C1) tycB1

The vector ptycA/tycB1 [Gruenewald et al., 2004] was amplified by inverse PCR using oligoneucleotides SG076 (5'-ATT GCC GGC ACC TCG GAT ATA TC-3') and SG077 (5'-ATT ATT GTT GAC CCC GAT GCA AGA GG-3') and the 9,9 kb PCR product was obtained (fragment 1) Primers SG078 and SG079 were phosphorylated and annealed. The resulting 57 bp fragment was terminally modified with the restriction enzymes *Ngo*I and *HincII* and ligated with the 9,9 kb PCR product (fragment 1) previously cut likewise and dephosphorilated to give the vector ptycA/tycB1 (A). This vector was finally digested using the restriction enzyme *Ban*II and subcloned in the plasmid ptycA/tycB1 cut in the same way to obtain the vector pQE61-tycA/(C1) tycB1.

pCC43

The fusion of the epitope-coding gene fragment to the 3' end of SrfA-A was accomplished via fusion PCR. Therefore, the 1,6 kb-comprising fragment *3'-srfA-A* was amplified using the oligonucleotides EPI_F (5'-**GAA CTG CTG AAT GCA GGC CTG** GGC GGC TCA TAA TAG AGG TGG CAT ATG AGC-3') and srf_14R (5'-TAT TCT AGA CAC TTG GTG AAC AGC CAT TCC-3'), whereas the oligonucleotides srf_11F (5'-ATA CCA TGG GAG CCG GAC ATC GAA GCG-3') and EPI_R (sequence, 5'-**CAG GCC TGC ATT CAG CAG TTC** TGC CAG TGA GAA AAT TTC CAT TAA TTT ATC CAG-3') were used for the amplification of the 1,4 kb-comprising 5' end of *srfA-B*. Based on the 21 bp comprising, overlapping regions (in bold) within the oligonucleotides EPI_R and EPI_F, both fragments were fused during the PCR reaction and the 3 kb-comprising DNA fragment *3'-srfA-A-epi-5'-srfA-B* could be amplified utilizing the oligonucleotides srf_11F and srf_14R. The PCR product was terminally modified with the endonucleases *Nco*I and *Xba*I and cloned into pQE60 previously cut in the same manner, to give the plasmid pCC42. The 1,5 kb-comprising *kan* cassette, obtained from the vector pDG782 via digestion with the endonuclease *Cla*I, was cloned into pCC42, previously cut in the same manner. Hereby, the 1,4 kb-comprising fragment *5'-srfA-B* homologous region is interrupted by the *kan* cassette in the resulting vector pCC43 (pQE60-*3'-srfA-A::epi-5'-srfA-B'-kan-5'-srfA-B*).

pCC52 and pCC63

In order to introduce a *Bam*HI and a *Sph*I recognition sites downstream of *3'-srfA-A*, the plasmid pCC42 was amplified by inverse PCR using the oligonucleotides Srf_22F (5'-ATA

ACT AGT TAG AGG TGG CAT ATG AGC AAA AAA TC-3') and Srf_23R (5'-ATA GGA TCC TTA TGA GCC GCC CAG GCC TGC-3'). The resulting PCR product was phosphorylated using the polynucleotide kinase (PKN) and religated in order to give the plasmid pCC50. With the oligonucleotides Srf_24F (5'-ATA GGA TCC AAT GAT TGC GGC ATC CCG C-3') and Srf_25R (5'-ATA ACT AGT ATT GTC ATA CCT CCC CTA ATC-3') the 400 bp-comprising DNA fragment of the surfactin promoter was amplified, terminally modified with the endonuclease *Bam*HI and *Sph*I and cloned into pCC50 cut likewise, resulting in the plasmid pCC51 (pQE-*3'-srfA-A-srf prom-5'-srfA-B*). In order to allow for a direct selection of homologous integration within the chromosome, the selectable marker *kan* and *mls* were integrated into pCC51 between the homologous regions *3'-srfA-A* and the 400 bp region containing the surfactin promoter, respectively. The *kan* resistance cassette was excised from the vector pDG782 using the endonucleases *Bam*HI and *Bgl*II and cloned into pCC51 previously cut with *Bam*HI (pQE-*3'-srfA-A-kan-srf prom-5'-srfA-B*), to give pCC52. Analogously, the mls resistance cassette obtained from DG646 by restriction digest with the endonucleases *Bam*HI and *Bgl*II was inserted into the *Bam*HI linearized plasmid pCC51 to give the disruption vector pCC63.

pCC78a

pCC78a was obtained by inverse PCR of the plasmid pCC43 using the oligonucleotides 5'-pQE/SrfA-B_inv (5'-ATA TGG TAC CAT GCG ATG CTT GAT CCG CAT TC -3') and 3'-pQE/SrfA-B_inv (5'-ATA TCC TAG GGG TCA GTT CCG TAC CAT CTT G-3'), designed to introduce the restriction sites *Avr*II and *Acc65*I between the 3'-*srfA-A-5'srfA-B* homologous regions.

pCC78

A 3 kb fragment containing the homologous regions 3'-*srfA-A-5'srfA-B* was PCR amplified from *B. subtilis* ATCC 21332 with the oligonucleotides srf_11F (5'-TAT TCT AGA CAC TTG GTG AAC AGC CAT TCC-3') and srf_14R (5'-TAT TCT AGA CAC TTG GTG AAC AGC CAT TCC-3'), terminally modified using the endonucleases *NcoI* and *XbaI* and cloned in pQE60 previously cut in the same manner to give the integration plasmid pCC78.

pCC83a

The DNA sequence encoding the cognate COM domain pair $COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}B1}$ (150 bp) was PCR amplified with oligonucleotides 5'-SrfA-B_COM (5'-ATA TGC TAG

CGA TCT TGG TGA CGA CGA TTT G-3') and 3'-SrfA-B_COM (5'-ATA TTG TAC AGC ATT CCC TCC TGC ATT GGT G-3'), cut by restriction digest with the endonucleases *NheI*, *Bsr*GI and cloned into pCC78a previously cut with *Avr*II and *Acc65*I. The endonucleases *NheI* (recognition sequence GCTAGC) and *AvrII* (recognition sequence CCTAGG) as well as the endonucleases *Bsr*GI (recognition sequence TGTACA) and *Acc65*I (recognition sequence GGTACC) generate compatible sticky ends. After ligation the plasmid pCC83 was obtained, in which the COM domain pair $COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}B1}$ is bordered by the conserved core motifs TPSD and QEGMLY.

pCC 85a

pCC78a was cut using the endonucleases *AvrII* and *Acc65I* and used for cloning of a 198 bp fragment encoding the cognate COM domain pair $COM^{D}_{SrfA\text{-}B3}$-$COM^{A}_{SrfA\text{-}C}$ and obtained by PCR amplification using the oligonucleotides 5'-srfBC_COM (5'-ATA TGC TAG CGA CTT CAG CGC CGA AGA CC-3') and 3'-srfBC_COM (5'-ATA TTG TAC AGC ATC CCT TCC TGC ATC GGC-3'). The following ligation gave rise to the plasmid pCC85.

pCC 92a

A 246 bp DNA fragment encoding the cognate COM domain pair COM^{D}_{TycB3}-COM^{A}_{TycC1} was PCR amplified using the oligonucleotides 5'-TycBC_COM (5'-ATA TGC TAG CGA CCT GGG GGA TGA AGA G-3') and 3'-TycBC_COM (5'-ATA TTG TAC AGC ATA CCC TCT TGC AAT GGG G-3'), terminally modified with the endonucleases *NheI* and *BsrGI* and cloned in pCC78a, previously cut with *Avr*II and *Acc65*I to give the plasmid pCC92.

pCC97a and pCC98a

The DNA sequences encoding the miscognate COM domain pair $COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}C}$ and $COM^{D}_{SrfA\text{-}B3}$-$COM^{A}_{A\text{-}B1}$ were obtained by gene splicing by overlapping extension (SOE) [Horton et al., 1989]. A 461 bp fragment obtained by PCR amplification using oligonucleotides 5'-srfA3_SOE (srfC) (5'-AGA AGA TAT CAG ACA TGT GCC G -3') and 3'-srfA3_SOE (srfC) (5'-**CCC TTG CGT TTT AGA AAA TTT CC**A TTA ATT TAT CCA G -3'), and a 493 bp PCR product, amplified using the oligonucleotides 5'-srfC_SOE (srfA3) (5'- **GGA AAT TTT CTA AAA CGC AAG GG**A ATT ACA GAA GGC-3') and 3'-srfC_SOE (srfA3) (5'-AAT GTG GTG GTA GCT CCA CAC C-3') were separately purified. Basing on the 24 bp comprising overlapping regions (marked in bold) the purified

PCR product were combined and PCR amplified using the oligonucleotides 5'-SrfA-B_COM (5'-ATA TGC TAG CGA TCT TGG TGA CGA CGA TTT G-3') and 3'-srfBC_COM (5'-ATA TTG TAC AGC ATC CCT TCC TGC ATC GGC-3'). The resulting 183 bp PCR product $COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}C}$ was terminally modified using the endonucleases *Nhe*I and *BsrGI* and cloned into pCC78 previously cut with the endonucleases *Avr*II and *Acc65*I to give the plasmid pCC97. Using the same strategy a 459 bp fragment was PCR amplified using. 5'-srfB3_SOE (srfB1) and 3'-srfB3_SOE (srfB1) (5'-**CCA CCT CTA TTA TTT TAA ATT CTC** CTC AAG CAT GTC -3') and a 470 bp fragment, amplified using oligonucleotides 5'-srfB1_SOE (srfB3) (5'- **GAG AAT TTA AAA TAA TAG AGG TGG** CAT ATG AGC AAA AAA TCG -3') and 3'srfB1_SOE(srfB3) (5'-TCC ATC ATA ATA TGA TGA TTG CTC C-3`). The two resulting PCR products, which shared an overlapping region of 23 bp (marked in bold). were subsequently purified, combined and used as template for PCR amplification using the oligonucleotides 5'-srfBC_COM and 3'-SrfA-B_COM. The resulting 165 bp product $COM^{D}_{SrfA\text{-}B3}$-$COM^{A}_{A\text{-}B1}$ was terminally modified using the endonucleases *NheI* and *BsrGI* and cloned into pCC78 previously cut with the endonucleases *Avr*II and *Acc65*I to give the plasmid pCC 98

pCC 106

A 259 bp PCR product, encoding the non cognate COM domain pair COM^{D}_{TycA}-COM^{A}_{TycC1} was PCR amplified from the previously described pQE61-tycA/(C1)tycB1 using the oligonucleotides 5'-tycAB_COM und 3'-tycAB_COM and after digestion with the endonucleases *NheI* and *BsrGI* was cloned into pCC78 previously cut with *Avr*II and *Acc65*I, to generate the plasmid pCC106

pCC83, pCC85 pCC92, pCC96, pC97, pCC98, pCC106

Plasmids pCC83, pCC85 pCC92, pCC96, pC97, pCC98 were generated using the same subcloning strategy. After *Pst*I digestion of pCC83a, pCC85a pCC92a, pCC96a, pC97a, pCC98a, pCC106a, the resulting 1,5 kb *Pst*I fragments were separately subcloned into pCC78, previously digested with *Pst*I and dephosphorolyated, to give rise to the plasmids pCC83, pCC85 pCC92, pCC96, pC97, pCC98 respectively

4.1.2 *B. subtilis* transformation

Natural competence is developed by *B. subtilis* at the transition of exponential to the stationary growth phase. DNA uptake requires (i) binding of DNA to the cell surface, (ii) formation of double strand breaks, (iii) assumption of a single DNA strand and degradation of the strand outside the cell. Transformation was performed following the method of Klein et al. with minor modifications [Klein et al., 1992]. *B. subtilis* cells were grown O/N in HS medium at 37°C and 250 rpm. The next day 20 ml LS medium were inoculated with 2 ml O/N culture and incubated at 30°C and 110 rpm. After the OD_{600} of the cell culture reached a value of 0,55, 1 ml cell suspension was transformed with about 10 ng of linearized DNA and incubated two additional hours at 37°C at 250 rpm. Finally the cells were plated on DSM plates supplemented with the appropriate antibiotics.

10x SBase	$(NH_4)_2SO_4$	20 g/l
	$K_2HPO_4x3H_2O$	140 g/l
	KH_2PO_4	60 g/l
	Sodium citrate x 2 H_2O	10 g/l

1 ml 1 M $MgSO_4$- was added after autoclaving

HS Medium	10 x S-Base	100 ml/l
	50% (w/v) Glucose	10 ml/l
	10% (w/v) Yeast extract	10 ml/l
	2% (w/v) Caseinhyrdrolysate	10 ml/l
	8% (w/v) Arginine / 0,4% Histidine	100 ml/l
	0,5% (w/v) Tryptophan	10 ml/l
	0,3% (w/v) Phenylalanine	15 ml/l

Sterilised by filtration

LS-Medium	10x S-Base	100 ml/l
	50% (w/v) Glucose	10 ml/l
	10% (w/v) Yeast extract	10 ml/l
	2% (w/v) Caseinhyrdrolysate	5 ml/l
	0,5% (w/v) Tryptophan	1 ml/l

0,3% (w/v) Phenylalanine	1,5 ml/l
Spermine	1 ml/l
1 M MgCl2	2,5 ml/l

Prepared immediately before utilization and sterilised by filtration.

4.1.2.1 Congression

A widely used method to manipulate *Bacillus* chromosome is the transformation of the strain of interest with an integration vector containing a selectable marker, for example an antibiotic resistance cassette, followed by selection of the transformants on the appropriate solid medium.

By this method particular (target) genes can be interrupted or deleted giving rise to a disruption strain. More challenging is the integration of DNA fragments lacking a selectable marker. In this case congression is often employed, in order to facilitate the selection of transformants. According to this approach a disruption strain, carrying a selectable marker "x", is transformed with an integration plasmid, carrying no selectable marker along with a helper vector, containing a marker "y". The transformants are subsequently selected on the appropriate solid medium. The use of a high amount of DNA (>1 μg per transformation) increases the capability of *Bacillus* to take up both DNAs. Finally the transformants are screened for the loss of marker "x", which is associated with the homologous recombination of the insertion plasmid in *Bacillus* chromosome.

4.1.3 Southern blot analysis.

Southern blot analysis [Southern, 1975] was used for characterization of specific chromosomal DNA regions. Chromosomal DNA was prepared using the method described by Cox et al. [Cox, 1968] Qiagen columns TIP20 were used for DNA purification (Qiagen Hilden, Germany). For labelling of the DNA probe and detection of the specific DNA fragments the ECL^{TM} random prime labelling and detection kit was used according to the manufacturers protocol (GE Healthcare Life Sciences, Freiburg, Germany).

4.2 Biochemical methods

SDS-PAGE [Laemmli, 1970], Coomassie stain of the SDS gels [Bennett and Scott, 1971]and other standard methods applied for protein characterization are not described in this work.

4.2.1 Preparation of *B. subtilis* cell extracts

Production of the surfactin NRPSs in *B. subtilis* was performed by growth in SpIII medium. 3,2 l of prewarmed SpIII medium were inoculated with an O/N culture of the corresponding *B. subtilis* strain in a 1:100 dilution. After 7 h of growth at 37°C and 250 rpm, the cells were harvested by centrifugation. Depending on whether the following protein purification was performed by IMAC or by immunoaffinity chromatography the cell pellet was resuspended i) in 25 ml buffer A (see 4.2.4) or ii) in 25 ml TE buffer pH 8, containing 15% sucrose and 1% PMSF respectively. The resuspended cells were either immediately disrupted or frozen at -20°C.

Subsequently, 0,1% lysozyme was added to the cell suspension and incubated for 45 min at 37°C, before cell disruption was performed by using a pre-cooled French Pressure Cell Press SLM Amicon, (Politech, Waldbronn, Germany). Two cycles of compression and decompression were performed with each cell extract (using a pressure of 1000 PSIG). Cell debris was separated via centrifugation (17000 rpm, 4 °C, 30 min) and the supernatant was harvested for immunoaffinity chromatography

4.2.2 Western blot

Western blot and immunological detection of SrfA-A-epi were used to verify the fusion of the epitope tag (SLAELLNAGLGGS) to the C-terminus of SrfA-A: According to the method of Towbin et al [Towbin et al., 1979] electrophoretically separated proteins were transferred on a nitrocellulose membrane under *semidry* conditions, using a Sartoblot-Apparatur (Satorius AG, Göttingen, Germany). The blot was performed O/N at 20 mA.

4.2.3 Immunodetection

In order to block the non specific sites, the nitrocellulose membrane was immersed in blocking solution and incubated at RT for 1 h with gentle shaking. Subsequently the membrane was incubated for 1 h at RT with a 1:1500 dilution of the NT73-MAb in blocking solution. After two washing steps of 15 min with TTBS (Tween-Tris-buffer salin) at RT, the

membrane was treated with a 1:1500 dilution of the secondary antibody, the HRP-conjugated Ab anti-mouse (HRP: *horseradish peroxidise*). The membrane was washed twice with TTBS and analysed with the ECL detection kit (GE Healthcare Life Sciences, Freiburg, Germany), according to the manufacturers protocol.

Blocking solution	10% (w/v) blocking reagent (Qiagen) in 1x blocking buffer (Qiagen)
TTBS, pH 7,6	20 mM Tris 137 mM NaCl 3,8 mM HCl 0,1% (w/v) Tween 20

4.2.4 Protein purification by Ni-NTA affinity chromatography.

The purification of SrfA-A-His was carried out by Ni-NTA affinity chromatography. This widespread and reliable technique is based on the utilization of a agarose matrix bound to Nitrilotriacetic acid (NTA), which binds Ni^{2+} ions and immobilize them to the matrix. NTA, which has four chelation sites for Ni^{2+} ions binds nickel more tightly than other metal-chelating compounds having only three sites for the interaction with metal ions. The remaining two free coordination sites of Ni^{2+} are available for binding with the hexahistidine tag, fused in frame to the protein which has to be purified. Through this interaction hexahistidine-tagged proteins bind to the Ni-NTA matrix. The protein purification was performed using Ni^{2+}-NTA agarose (Qiagen, Hilden, Germany) and a FPLC system (GE Healthcare Life Sciences, Freiburg, Germany). The cleared supernatant after cell lysis was batch incubated with the matrix at 4°C by gentle shaking; subsequently the matrix was transferred into a Chelating-Superose-Column HR 10/2 HiLoad™ (GE Healthcare Life Sciences, Freiburg Germany). The washing steps as well as the elution of the protein were followed by measuring the absorption at 220 nm.

hexahistidine-tagged proteins were purified by Ni^{2+}-NTA affinity chromatography. The soluble fraction of a crude cell extract was batch incubated with 500 µl Ni-NTA superflow (Qiagen, Hilden, Germany) previously equilibrated with buffer Hepes A (50 mM Hepes, 100 mM NaCl, (pH 8,0)). After 2 h incubation at 4°C under gentle shaking the column (HR 10/2, Amersham Biosciences, Freiburg) was packed and the purification was carried out on a “fast

performance liquid chromatography" (FPLC) system (GE Healthcare Life Sciences, Freiburg, Germany). Protein binding was monitored at A_{220}. When the A_{220} trace reached the baseline again, a 25 min linear gradient up to 50% Hepes B (50 mM Hepes, 100 mM NaCl, 250 mM imidazole (pH 8,0) followed by a 10 min gradient to 100% Hepes B with a flow rate of 1 ml/min was applied. Fractions of 2 ml were collected. Proteins were identified by SDS-PAGE. All proteins were dialyzed with standard assay buffer (25 mM Hepes, 50 mM NaCl (pH 7,0)) by using "HiTrap Desalting" columns (GE Healthcare Life Sciences, Freiburg, Germany).

4.2.5 Protein purification by immunoaffinity chromatography

SrfA-A-epi purification was performed by immunoaffinity chromatography using polyol-responsive monoclonal antibodies (PR-mAb). This procedure takes advantage of the very tight and specific binding between the PR-mAb and the recognised antigen; moreover it allows for the elution of the antigen under mild, not denaturing conditions via the utilization of the polyol 1,2-propandiol.

As chromatographic support a NT73 column was used, containing the PR-mAb NT73 immobilized to cyanogen bromide- activated sepharose. This PR-mAb binds to the C-terminal region of the β-subunit of the *E. coli* RNA polymerase core enzyme, which consists of the utilized epitope "SLAELLNAGLGGS" [Burgess and Thompson, 2002]. Protein purification was performed following the method of Burgess et al. with minor modifications. *B. subtilis* cell extracts of CC44 or CC57 were applied to the immunoaffinity chromatography resin by batch incubation at R/T with gentle shaking. After 1,5 h the material was transferred into a FPLC column. All buffers used were applied to the column with a flow rate of 1 ml/min. After application of the crude extracts, two washing steps were performed by using 4 CV of buffer W1 and 4 CV of buffer W2. Finally elution was performed by applying 10 CV of buffer E. Washing and elution of the proteins were followed by measuring the absorption at 280 nm.

Immunoaffinity chromatography buffers

Buffer W1	Tris/HCl	10 mM, pH 8
	EDTA	1 mM, pH 8
	NaCl	100 mM
	Sucrose	15% (w/v)
	PMSF	1% (w/v)
Buffer W2	Tris/Cl	10 mM, pH 8
	EDTA	1 mM, pH 8
	NaCl	500 mM
	Sucrose	15% (w/v)
	PMSF	1% (w/v)
Buffer E	Tris/Cl	10 mM, pH 8
	EDTA	1 mM, pH 8
	NaCl	700 mM
	Sucrose	15% (w/v)
	PMSF	1% (w/v)
	1,2-Propandiol	30%

Protein concentration was determined by measuring the absorption at 280 nm referring to the calculated extension coefficient.

4.2.6 ATP-PPi exchange

The selectivity of NRPS is determined by the adenylation (A)- domain, which catalyses the activation of a cognate amino acid substrate as aminoacyl-adenylate, through ATP hydrolysis and release of PP_i [Gevers et al., 1968; Lee et al., 1975]. The reversibility of this reaction is exploited to determine the substrate specificity and the catalytic activity of the corresponding A domain. Therefore, the enzyme is incubated with the cognate amino acid, together with ATP, Mg^{2+}-ions, and 0,15 µCi of [^{32}P]-PP_i. The hydrolysis of ATP is associated with the formation of aminoacyl-adenylate and free PP_i. The inverse reaction leads to the formation of [^{32}P]-ATP, due to the incorporation of [^{32}P] radioactively labelled PP_i. Contrary to PP_i, ATP can be bound to activated charcoal by absorption and thus can be separated from the reaction

mix. The amount of radioactive labelled ATP is proportional to ATP/PP_i exchange rate and to the amount of aminoacyl-adenylate synthesized by the A-domain. This method provides therefore an indirect measure of the catalytic activity of the investigated A domain. In the negative control ATP or amino acids are omitted.

The samples, prepared in parallel, contained the analyzed A domain, the amino acids, whose activation had to be tested and the reaction buffer in a final volume of 50 µl (Mix 1).

A second mix (Mix 2) was prepared containing ATP, PPi, $MgCl_2$, [^{32}P]-PPi and buffer in a final volume of 50 µl. The solution resulting from the combination of Mix1 and Mix2 was incubated at 37°C for 15 min and then the reaction was stopped by adding 500 µl of killing mix. The samples were vortexed and centrifuged at 13000 rpm for 2 min. The resulting pellet was washed twice with 1 ml H_2O. The pelletted activated charcoal was finally resuspended in 500 µl H_2O and transferred in scintillation vials containing 3,5 ml of "Rotiszint Eco Plus" solution (Roth, Karlsruhe, Germany). The samples were measured using the scintillation counter "TRI CARB liquid scintillation analyser 2100TR (Packard instrument, Meriden USA).

Reaction mix	Enzyme	200 nM
	Aminoacid	1 mM
	ATP	1 mM
	$Na_4P_2O_7$	50 µM
	$MgCl_2$	10 mM
	[^{32}P]- $Na_4P_2O_7$	0,15 µCi
	Assay buffer	to 100 µl
Killing mix	$Na_4P_2O_7$	100 mM
	$HClO_4$	500 mM
	Active carbon	1,2 %
Assay buffer	Tris/HCl, pH 8,0	10 mM
	EDTA	1 mM
	NaCl	100 mM

4.2.7 Covalent loading of NRPSs with radioactively labelled aminoacids

The covalent loading of NRPS-PCP domains with their substrate amino acids can be analysed by incubation of the enzyme with radioactively labelled amino acids in presence of ATP and $MgCl_2$. After addition of tricloroacetic acid (TCA) the covalent bound amino acids can be coprecipitated with the enzyme, whereas the free amino acids remain in the supernatant and can be separated from the enzyme. For the preparation of the negative control ATP was not added to the reaction mix.

Two separated solutions were prepared: the first (Mix 1) contains the enzyme to analyse, $MgCl_2$ and buffer in final volume of 50 µl; the second (Mix 2) contains ATP, the radioactively labelled amino acid and buffer in a final volume of 50 µl. Mix 1 and Mix 2 were subsequently combined and incubated at 37°C for 10 min, before the reaction was stopped by adding 800 µl of cold 10% TCA. In order to verify a quantitative precipitation of the proteins, 15 µl of 25 mg/ml BSA were added. After vortexing, the samples were incubated for 15 min on ice, before the precipitated proteins were pelletted by centrifugation (13000 rpm, 5 min). The supernatant was discarded and the pellet was washed two times with 800 µl 10% TCA (10%). The pellet was finally resuspended in 200 µl formic acid and transferred into scintillation vials containing 3,5 ml “Rotiszint Eco Plus” solution. The radioactivity contained in the samples was measured by using the scintillation counter “TRI CARB liquid scintillation analyser 2100TR (Packard instrument, Meriden USA). The percentage of loading of the NRPS-PCP domains was calculated referring to the specific activity of the radioactive labelled amino acid used in the assay.

Reaction mix	Enzyme	500 nM
	$[^{3}H]$ / $[^{14}C]$	1200 nM
	ATP	4 mM
	$MgCl_2$	10 mM
	Assay buffer	To 100 l

4.2.8 Surfactin and lipo-tetrapeptide preparation

The method described by Nakano et al. [Nakano et al., 1988] was used for preparation of surfactin as well as the tetrapeptide derivative. 3 ml SpIII medium was inoculated with cells of the corresponding *B. subtilis* strain and incubated O/N at 37°C and 250 rpm. The cells were subsequently separated via centrifugation and 1ml of the supernatant was extracted three

times with ½ volume of 1-buthanol. The organic phase was dried in a speed vac and the resulting pellet was resuspended in 50 µl methanol (70 %, v/v).

4.2.9 Surfactin analysis on blood agar plates.

The haemolytic activity of the buthanolic extracts were analysed on Columbia agar plates containing 5% sheep blood (Beton Dickinson GmbH, Heidelberg, Germany) via antibiotic disc diffusion assays. About 20 µl of the butanolic extracts were spotted onto sterile MM-Whatman paper (5 mm diameter). After 20 h incubation at 37°C, the formation of haemolytic zones could be observed.

4.2.10 Surfactin and lipo-tetrapeptide analysis by HPLC/MS

The formation of surfactin and the lipo-tetrapeptide was detected and analyzed by HPLC/MS on a 1100 HPLC (Agilent), using a C_8 reverse-phase column (CC250/3 Nucleosil120-3C_8 column Macherey & Nagel, Düren Germany). This method allows the separation of the molecules contained in the analyzed sample based on their hydrophobic properties, in particular on the strength of their interaction with the matrix. In reverse phase HPLC the solid phase is constituted silica, whose surface is chemically modified by covalently bound C_8 groups. The mobile phase consists of a polar solution, in this case a mixture of H_2O, methanol and formic acid. A linear gradient is applied to the column and the progressive increased concentration of methanol, the non polar component, results in the elution of the molecules bound to the column by hydrophobic interactions. The column was equilibrated to 70% buffer B (buffer A, 0,05% formic acid in H_2O; buffer B, 0,045% formic acid in methanol). Samples were separated by applying a linear gradient to 100% buffer B (flow rate 0,3 ml min^{-1}, within 30 minutes) and detected at 214 nm. Online ESI-MS analysis was performed with a Finnigan LTQ-FT (Thermo Electron Corp., Bremen, Germany). Samples were measured with a FT resolution of 100000 in the positive ion mode. Parameters were as follows: Ion Spray Voltage 4,5 kV, Sheath gas 50 units, Auxilliary gas 10 units, capillary temperature 330°C with a capillary voltage of 41 V and a Tube lens setting of 100 V. Molecules were detected in the range from 500 to 1200 m/z.

Electrospray ionization (ESI) is a method which produces multiply charged gaseous ionized molecules. The sample solution is sprayed from a stainless needle held at approximately 4000 V and the highly charged droplets migrate through an electrical field to the counter electrode.

By applying dry gas, heat or both to the charged droplets, the solvent evaporate from the surface of the droplets, generating highly charged analyte molecules. The repulsion forces between identical charges destabilize the droplet surface and as a result the droplet brakes up into smaller droplets. This process leads to the formation of free ions, which are directed into an orifice through electrostatic lenses leading to the mass spectrometer.

4.2.11 Identification of NRPSs by tryptic digestion and ESI/MS-MS analysis

Tryptic digestion followed by ESI/MS-MS-Analysis was used to characterize the proteins purified by immunoaffinity chromatography. The detection limit of this procedure is approximately 1 pmol of protein in the gel. Because this limit of detection is similar to the limit of detection of Coomassie staining, any gel band that is reasonably stained by Coomassie blue contains a protein amount sufficient for sequence analysis by ESI-MS.

After SDS-PAGE the protein band to analyse was cut out of the gel, transferred into a reaction tube and incubated with 200 µl distaining buffer (200 mM NH_4HCO_3 in 50% CH_3CN) at 56°C for 30 min. The distaining solution was then discarded and the gel slices were dried in a speed vac at 30°C for about 1 h. After addition of 1 pg/µl trypsin solution the samples were incubated O/N at 37°C. The solution was discarded and the digested proteins were dissolved in10 µl elution buffer (10% (v/v) acetonitril 1% TFA). The samples were applied to an Ultimate nano HPLC-system (Dionex, Idstein, Germany) equipped with a Switchos for automated sample concentration and a C18 PepMap column (150 mm length, 75 uM inner diameter). The eluted peptides were measured by online-nano-ESI-MS(MS) with a QStar Pulsar i mass spectrometer (Applied Biosystems, Darmstadt, Germany). Subsequent identification of the proteins was performed with MASCOT.

MASCOT is a search engine which uses mass spectrometry data to identify proteins from primary sequence databases. After MS/MS spectra from individual peptides were measured the experimental mass values are compared with calculated peptide mass or fragment ion mass values, obtained by applying cleavage rules to the entries in a primary sequence database. By using appropriate scoring algorithms, the closest match or matches can be identified. If the investigated protein is present in the sequence database the aim is to pull out that precise entry. If the sequence database does not contain the investigated protein the aim is to identify those entries which exhibit the closest homology. Sequence homology is defined by the final score, measured as the -10*Log (P), where P is the probability that the observed match is a random event.

5 Results (I)

5.1 *In vivo* investigation of protein-protein communication in the surfactin biosynthetic complex

Up to now, the role of communication mediating (COM) domains for the control of protein-protein communication in multi-enzymatic NRPS complexes has only been investigated *in vitro*. Hence; in the first part of this work, the surfactin biosynthetic complex, consisting of the three NRPSs SrfA-A, SrfA-B and SrfA-C, was used as model system i) to investigate the *in vivo* effects of COM domain swaps on protein-protein communication and ii) to alter the mutual recognition and communication between the proteins within the multi-enzyme complex, in order to create the basis for biocombinatorial synthesis.

Based on *in vitro* studies it has been established that the C- and N-terminal COM domains of two partner NRPSs – referred to as donor (COM^D) and acceptor COM domain (COM^A), respectively – form a so-called cognate COM domain pair [Hahn and Stachelhaus, 2004]. According to this definition, the surfactin assembly line should contain two cognate COM domain pair $COM^D_{SrfA\text{-}A3}/COM^A_{SrfA\text{-}B1}$ and $COM^D_{SrfA\text{-}B3}/COM^A_{SrfA\text{-}C,}$ facilitaing the communication between SrfA-A and SrfA-B, as well as SrfA-B and SrfA-C, respectively. In the presented study, the first cognate COM domain pair $COM^D_{SrfA\text{-}A3}/COM^A_{SrfA\text{-}B1}$ was exchanged against different cognate, mis-cognate and non-cognate COM domain pairs derived from the surfactin and tyrocidine biosynthetic assembly lines, in order to generate several *B. subtilis* ATCC 21332 derivatives of the general genotype *srfA-A::*COM^D_x*-*COM^A_y*::srfA-B* (Fig. 12).

For this purpose, a well-established two-step marker exchange homologous recombination approach was adopted. This strategy requires i) the interruption of the *srfA* biosynthetic operon at the transition between *srf-A-A* and *srfA-B* via integration of a selectable marker and ii) subsequent reconstitution of a functional operon and loss of the selectable marker. In our case, the reconstitution plasmids used for the latter step carried the COM domain pairs to be exchanged.

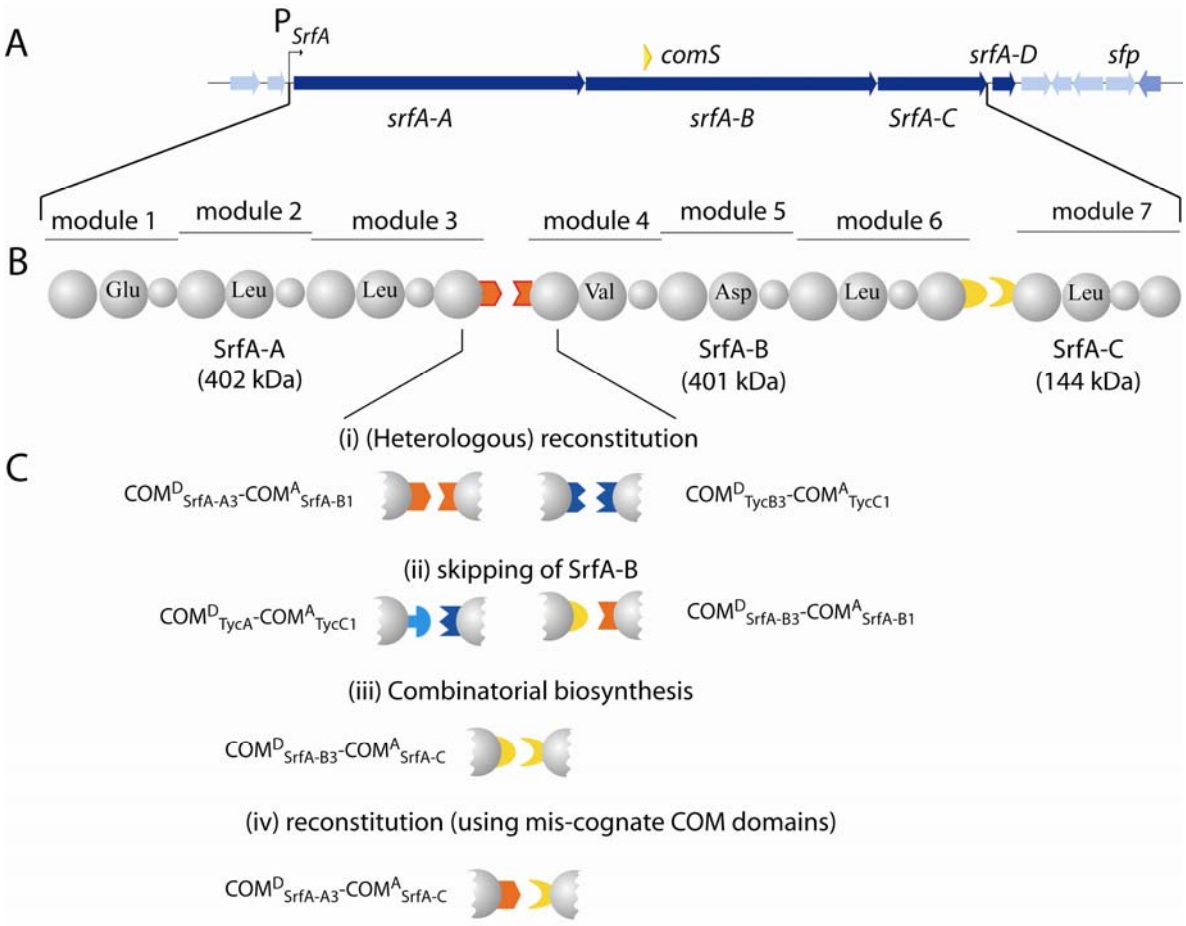


Fig. 12 The surfactin biosynthetic assembly line
The cyclic lipoheptapeptide surfactin is synthesized by three large NRPSs: SrfA-A (402 kDa), SrfA-B (401 kDa) and SrfA-C (144 kDa), encoded by the genes *srfA-A*, *srfA-B* and *srfA-C*, which are constituted of three, three and one module, respectively. Selective interaction between partner enzymes is mediated by two compatible sets of COM domains (shown in red and yellow). The first COM domain pair was replaced by means of domain swapping on the genetic level against different cognate, mis-cognate and non-cognate sets of COM domains (C).

Genotypic analysis of all mutants generated by swapping experiments was performed as follows: First, the *3'-srfA-A–5'-srfA-B* transition region was analyzed by PCR using the oligonucleotides srf_42F and srf_43R, located outside of the 3 kb *3'-srfA-A–5'-srfA-B* homologous region, in order to confirm the reconstitution of the *srfA* operon by double cross-over integration. Secondly, a 600 bp region of the generated 3 kb PCR product, containing the coding region of the substituted COM domain pair, was analyzed by DNA sequencing.
Subsequently, the effects of a specific COM domain swap on protein-protein communication were investigated by characterization of the butanolic extracts by means of RP-HPLC, and fourier transform-ion cyclotron resonance (FT-ICR) mass analysis. The latter technique was used, in order to enable the detection even of trace amounts of synthesized lipopeptides.

5.2 Maintenance of *B. subtilis* genetic competence

The described two-step marker-exchange homologous recombination approach required, as mentioned above, the preparation of a disruption strain, in which *srfA-A* and *srfA-B* are separated by a selectable marker. However, this genetic manipulation disrupts the *srfA* operon, preventing not only the transcription of *srfA-B* and *srfA-C*, but also of the small competence regulator gene *comS*, located in a different reading frame within the first module of *srfA-B* (see Fig. 12). Since the protein ComS has been shown to be necessary for the development of natural competence in *B. subtilis*, the corresponding genetic manipulation would automatically produce a genetically inaccessible *B. subtilis* mutant. To circumvent this problem, the plasmid pKE27 [Eppelmann et al., 2001], was used to insert a second copy of *comS* into the non-essential *amyE* gene of *B. subtilis* ATCC 21332. The pKE27 integration plasmid contains a chloramphenicol resistance marker (*cat*), as well as a copy of *comS* under transcriptional control of the IPTG-inducible P_{spac} promoter, which both are flanked by the homologous regions *amyE-front* and *amyE-back* required for homologous recombination (see Fig. 13). The resulting *B. subtilis* mutant AM1 was constructed and genetically characterized by Sascha Doekel and Alexandra Mees (unpublished results), and used as parental strain for all genetic manipulations of the surfactin biosynthetic operon carried out in this study.

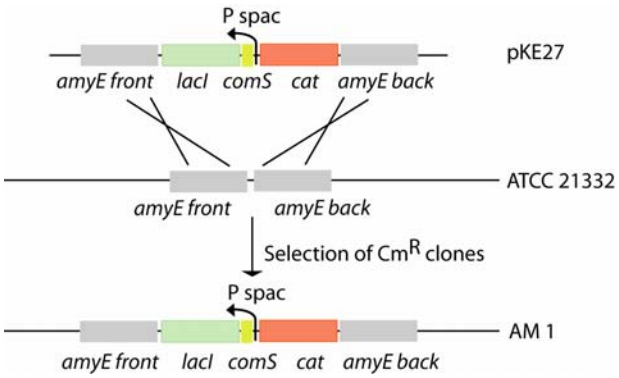


Fig. 13 Construction of the mutant *B. subtilis* AM1.
Integration of the plasmid pKE27 in the *amyE* region of the *B. subtilis* ATCC 21332 chromosome by homologous recombination generated the mutant strain AM1, containing the *cat* resistance gene and the second copy of *comS*.

5.3 Construction of the disruption strain CC64

In order to integrate a selectable resistance marker at the transition region between *srfA-A* and *srfA-B*, *B. subtilis* AM1 was transformed with the linearized disruption vector pCC13, carrying the 3'-*srfA-A* and 5'-*srfA-B* homologous regions for the integration, separated by the resistance gene *mls*. By selection for Cm^R/MLS^R phenotype, the *srfA* disruption strain *B. subtilis* CC64 could be obtained (Fig. 14A), which was genetically characterized by PCR (Fig. 14B), as well as Southern blotting (Fig. 14C) analysis. PCR amplification using *B. subtilis* CC64 chromosomal DNA along with the oligonucleotides srf_42F and srf_43R

(located outside of the used homologous regions), gave a 4,6 kb-PCR product, this way, verifying the successful disruption of the *srfA* operon. A control reaction, utilizing chromosomal DNA of ATCC 21332 as template, resulted in the expected amplification of an about 3 kb-fragment (Fig. 14B).

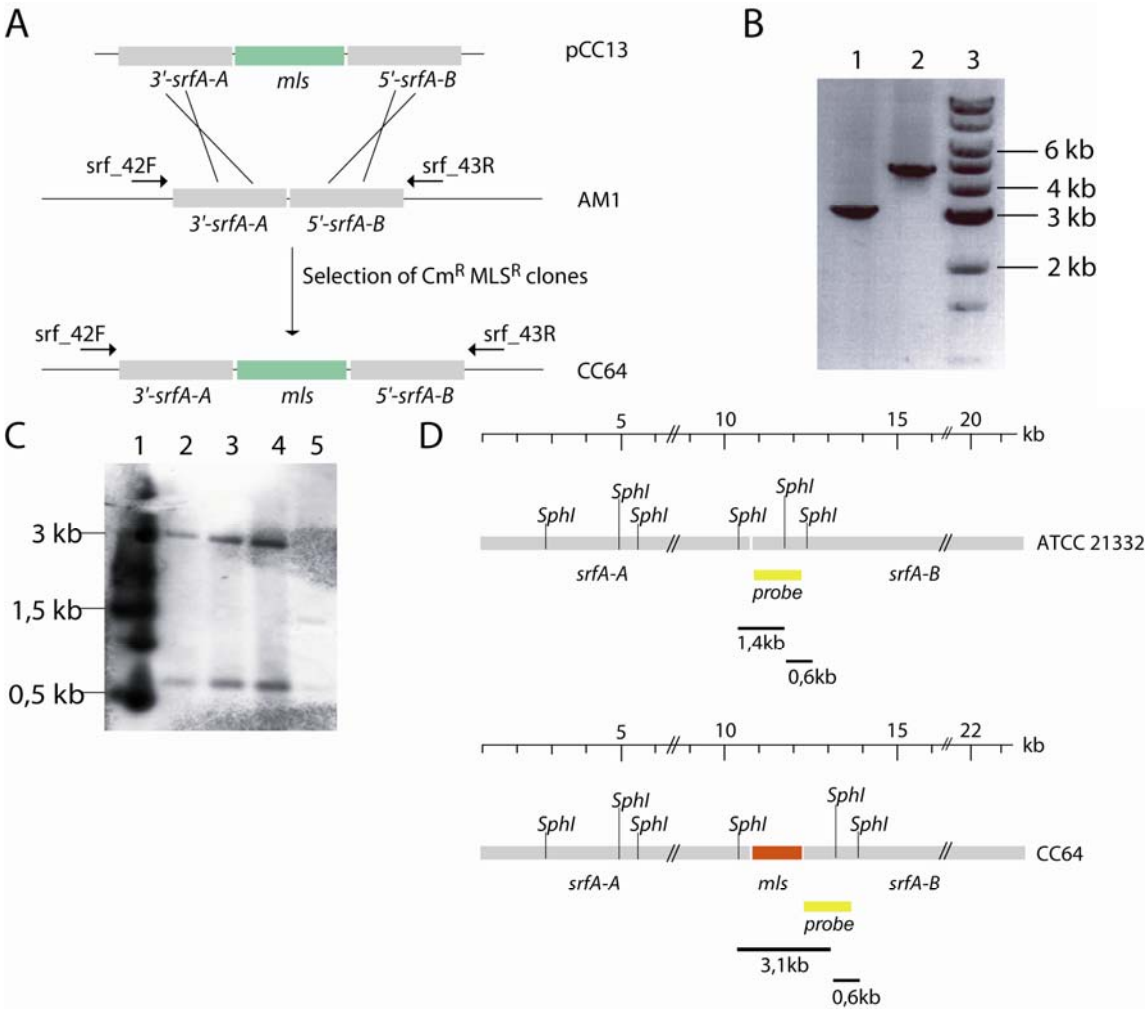


Fig. 14 Construction of the disruption strain CC64
(A) Disruption of the *srfA* biosynthetic operon in the chromosome of *B. subtilis* AM1. **(B)** PCR analysis of the resulting strain *B. subtilis* CC64, and the parental strain AM1, using the oligonucleotides srf_42F and srf_43R. Lane 1: 3 kb PCR product obtained from the amplification of ATCC21332 DNA. Lane 2: 4,6 kb PCR product obtained from the amplification of CC64 DNA. Lane 3: 1 kb DNA marker. **(C)** Southern blotting. *Sph*I digests of chromosomal DNA of CC64 (three different clones), and ATCC 21332 were loaded in lanes 2, 3, 4 and 5, respectively. In lane 1 the 1 kb DNA marker was loaded. In **(D)** the chromosomal organization of the *srfA-A-srfA-B* transition region of *B. subtilis* ATCC 21332 and CC64 is shown. The location of *Sph*I restriction sites, as well as the size of the fragments detected by hybridization with the 1,4 kb probe are shown.

Chromosomal DNA of *B. subtilis* CC64 was further analyzed by southern blotting (Fig. 14C), using the 1,4 kb *srfA-B5'* homologous region as probe, which was amplified from chromosomal DNA of *B. subtilis* AM1, using oligonucleotides srf_12F and srf_14R. As shown in Fig. 14C, the analysis of three different clones of CC64 led to the detection of the

expected 3,1 kb and 0,6 kb *Sph*I fragments, whereas for the parental strain, the hybridization with the 1,4 kb probe allowed the detection of the expected 1,4 kb and 0,6 kb *Sph*I fragments. After verification of the genotype both strains, AM1 and CC64, were analyzed for production of the lipoheptapeptide surfactin.

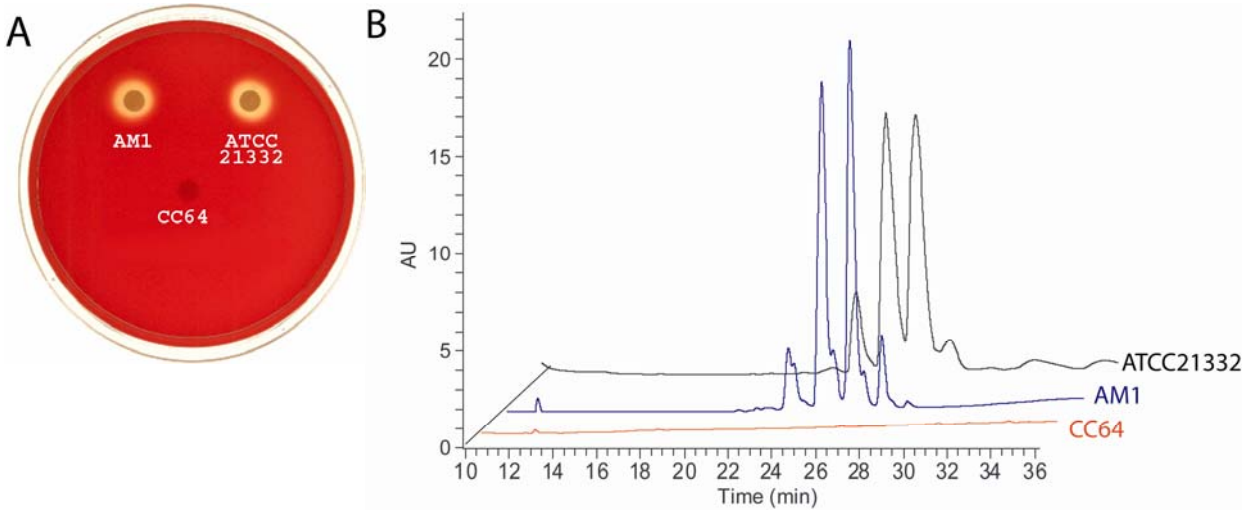
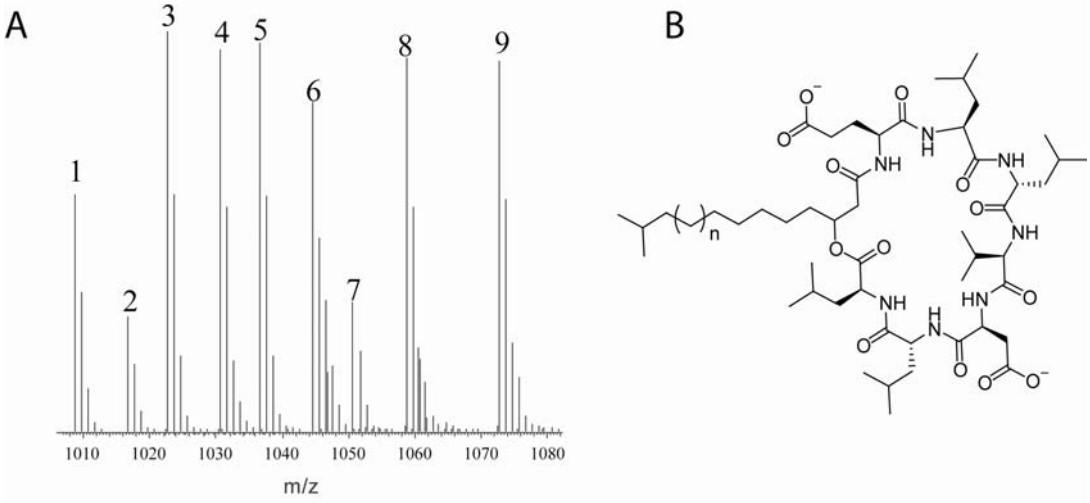


Fig. 15 Analysis of the butanolic extracts obtained from *B. subtilis* AM1, ATCC 21332 and CC64.
Butanolic extracts of the cultured broths of *B. subtilis* mutant CC64, parental strain AM1 and wild-type strain ATCC 21332 were analyzed for hemolytic activity **(A)**, and by RP-HPLC **(B)**.



C

Peak No.	Molecular Formular	Mass calculated	Mass observed
1	$C_{51}H_{90}N_7O_{13}$	1008.660	1008.661
2	$C_{50}H_{87}N_7O_{13}Na$	1016.626	1016.625
3	$C_{52}H_{92}N_7O_{13}$	1022.675	1022.676
4	$C_{51}H_{89}N_7O_{13}Na$	1030.642	1030.642
5	$C_{53}H_{94}N_7O_{13}$	1036.691	1036.692
6	$C_{52}H_{91}N_7O_{13}Na$	1044.657	1044.659
7	$C_{54}H_{96}N_7O_{13}$	1050.707	1050.710
8	$C_{53}H_{93}N_7O_{13}Na$	1058.673	1058.675
9	$C_{54}H_{95}N_7O_{13}Na$	1072.689	1072.691

Fig. 16 FT-ICR MS analysis of surfactin produced by *B. subtilis* AM1.
(A) Mass spectrum of the butanolic extracts derived from *B. subtilis* AM1 **(B)** The lipoheptapeptide surfactin FA-LGlu-LLeu-DLeu-LVal-L-Asp-DLeu-LLeu-OH is synthesized as a mixture of molecules, differing in the length of the fatty acid chain (n = 1; 2; 3 etc). **(C)** Characterization of the mass peaks shown in (A).

Butanolic extracts of the cultured broths of AM1, CC64 and ATCC 21332 were prepared as described in Methods (4.2.8), and analyzed (i) for hemolytic activity (Fig. 15A), (ii) by RP-HPLC (Fig. 15B), and (iii) by high resolution mass spectrometry (FT-ICR-MS) (see Fig. 16). As anticipated, the mutant AM1 was able to synthesize wild-type level of the lipoheptapeptide surfactin, demonstrating that the insertion of a second copy of the *comS* gene into the *amyE* site did not affect the biosynthesis of surfactin (Fig. 15A, B). In contrast, also as expected, no traces of surfactin could be detected for the disruption mutant CC64.

5.4 COM domain swapping

5.4.1 Elaboration of a cloning strategy for COM domain swapping

In analogy to the COM domains of the tyrocidine biosynthetic complex, which have been investigated *in vitro*, $COM^{D}_{SrfA\text{-}A3}$ and $COM^{A}_{SrfA\text{-}B1}$ are bordered by highly conserved amino acid sequences: a TPSD motif that is located at the transition between E and $COM^{D}_{SrfA\text{-}A3}$ domain, about 20 amino acids upstream of the C-terminus of SrfA-A, as well as a QEGMLY motif, which is located at the transition between $COM^{A}_{SrfA\text{-}B1}$ and C domain, about 15 amino acid residues downstream of the N-terminus of SrfA-B. Since both core motifs had been already successfully used in *in vitro* COM domain swap experiments [Hahn and Stachelhaus, 2004], the same fusion sites were also used for the integration of all COM^{D}-COM^{A} pairs investigated in this study (compare Fig. 12). This goal was achieved by constructing the integration vector pCC78a, which carries the required 5' and 3' flanking regions for the marker exchange homologous recombination, and features the engineered restrictions sites *Avr*II and *Acc*65I, located precisely on the coding bases of peptides "PS" and "LYH" within the mentioned core motifs (see Fig. 17). In order to maintain the primary sequence of these core motifs and, at the same time, facilitate a targeted cloning, the compatible restrictions sites *Nhe*I and *Bsr*GI were engineered onto the termini of the encoding gene fragments of the desired COM domain pairs to be cloned.

In all cases described below, the same procedure for construction of the COM domain swaps was utilized. First, the coding fragments of the corresponding COM domain pairs were amplified by PCR. Taking advantage of the simultaneously engineered restriction sites *Nhe*I and *Bsr*GI, the gene fragments were cloned into the integration vector pCC78a digested previously with *Avr*II and *Acc*65I. The resulting plasmids were subsequently used for co-transformation of the *srfA* disruption strain *B. subtilis* CC64. (Fig. 18A).

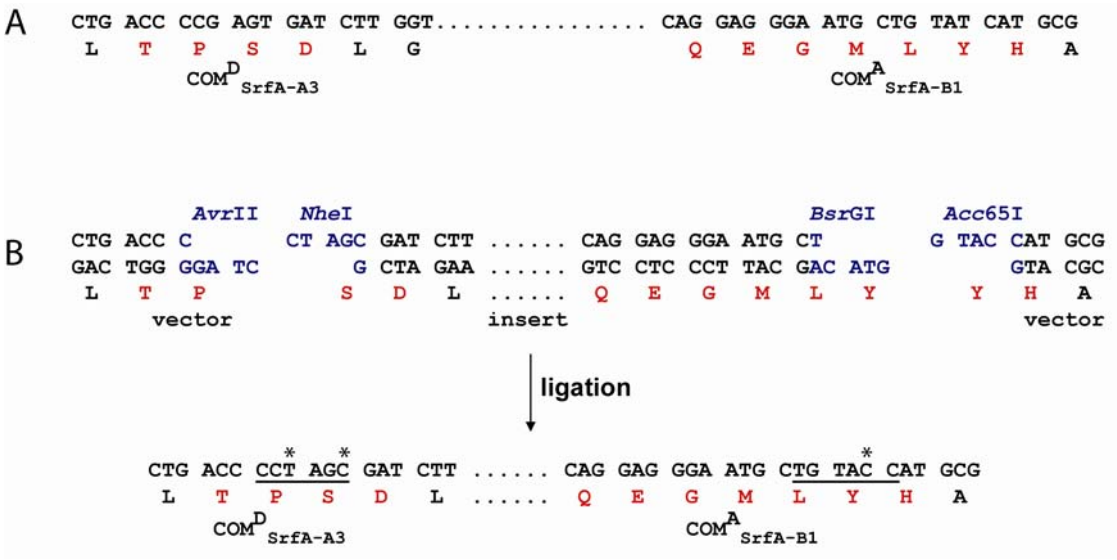


Fig. 17 Cloning of COM domains under maintenance of the primary sequence.
In order to maintain the primary sequence of the conserved sequence motifs "TPSD" and "QEGMLYH" of $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$ **(A)**, the compatible restrictions sites *Nhe*I/*Avr*II and *Bsr*GI/*Acc*65I were engineered (in blue) and used for the directed cloning of the coding regions of different COM domain pairs **(B)**. The bases coding for the peptides PS and LYH, used as fusion sites are underlined. The three point mutations introduced by the cloning strategy are marked with asterisks.

5.4.2 Swapping of $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$: a proof of principle

The feasibility of the described cloning strategy was tested by constructing a reconstitution plasmid, carrying the natural COM domain pair $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$.The resulting plamsid pCC83a was subsequently used to transform CC64, and to reconstitute the surfactin biosynthetic operon (Fig. 18).

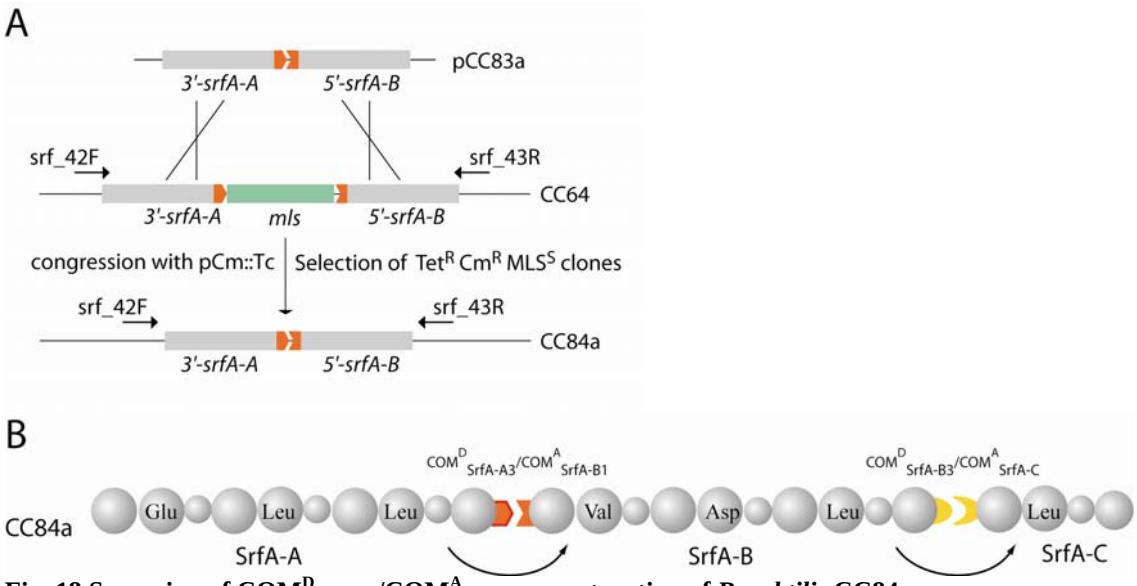


Fig. 18 Swapping of $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$: construction of *B. subtilis* CC84a
(A) Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC83a, gave the *B. subtilis* strain CC84a. **(B)** Organization of the surfactin biosynthetic complex of CC84a, generated by integration of the native, cognate COM domain pair $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$. The black arrows indicate protein-protein communication.

This experiment was important in order to verify that the three silent point mutations, which had to be introduced due the cloning strategy, had no effect on the productivity of the surfactin biosynthetic system.

Since pCC83 did not carry any selectable marker, a congression (co-transformation) experiment had to be carried out, in order to facilitate a positive selection of the successful transformants (see methods 4.1.2.1). To this end, the integration plasmid pCC83a, and the helper vector pCm::Tc, mediating tetracyline resistance, were concurrently used to transform the *srfA*-deletion strain *B. subtilis* CC64. Subsequently, five-hundred of the Tet^R and Cm^R colonies obtained were screened for MLS sensitivity. Four clones, termed CC84a (clones no. 1-4), actually showed the desired phenotype Cm^R, Tet^R and MLS^S, and were further analyzed by PCR. To this end, the *3'-srfA-A-5'-srfA-B* transition region was amplified using the primer pairs srf_42F/srf_43R. As shown in Fig. 19A, this analysis revealed the expected 3 kb-fragment of an intact *srfA* operon both for the wild-type ATCC 21332, and all four clones of CC84a. In contrast, the PCR using chromosomal DNA derived from the disruption strain CC64, gave rise to the amplification of the expected 4,6 kb-fragment.

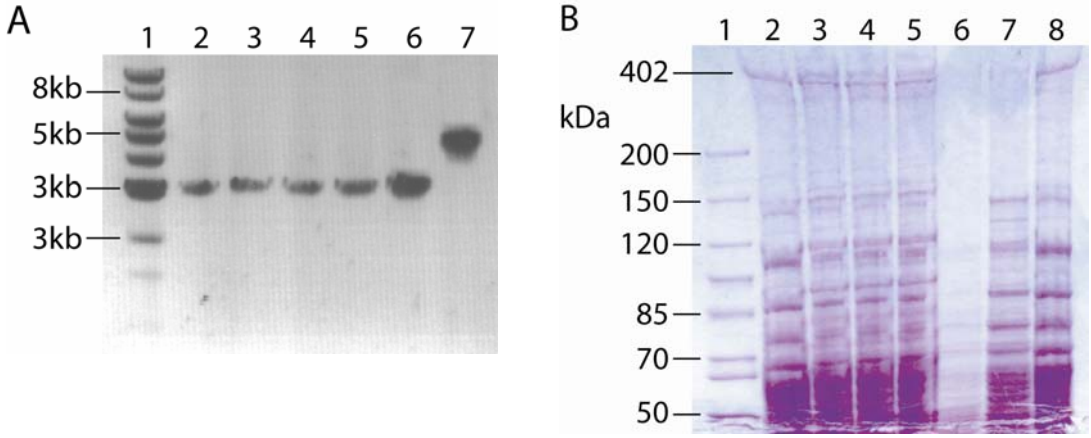


Fig. 19 Genetic and biochemical characterization of CC84a clones 1-4.
(A) PCR analysis of the genetic organization at the transition between srfA-A and srfA-B, using the oligonucleotides srf_42F and srf_43R. Lane 1: 1 kb DNA marker. Lanes 2 to 5: PCR products obtained with chromosomal DNA from CC84a clones 1 to 4, respectively. Lanes 6 and 7: PCR products obtained with DNA from *B. subtilis* ATCC 21332 and CC64, respectively. **(B)** SDS-PAGE analysis using a NuPAGE® Bis-Tris gel (acrylamide concentration 4% to 12%). Lane 1: protein marker. Lanes 2 to 5: Crude cell extracts of the four CC84a mutants no. 1-4 (lane 2 to 5) were compared with crude cell extracts obtained from the surfactin producers *B. subtilis* ATCC 21332 and OKB105 (lanes 7 and 8 respectively), as well as the crude extract of the mutant KE30 (lane 6), which is incapable to produce any of the surfactin synthetases, SrfA-A, SrfA-B and SrfA-C.

The identity of the four clones of CC84a was further verified by amplification of the 3 kb *5'-srfA-B-3'-srfA-A* transition region and subsequent DNA sequencing of a 600 bp fragment containing the integrated COM domain pair. This analysis revealed the presence of a T-to-C

single point mutation in the plasmid pCC83a (as well as in the parental vector pCC78a) and at position 448 in the coding region of the *5'-srfA-B* gene. The point mutation causes a W150R substitution within the core motif C3 (MHHxISDG(WV)S) of the first condensation domain of SrfA-B. Since the integrity of core C3 has been shown to be crucial for the activity of the condensation domain, the determined point mutation was likely to affect the mutant's ability to synthesize functional SrfA-B protein, and eventually to produce the lipoheptapeptide surfactin. Indeed, subsequent investigation of the phenotype of *B. subtilis* CC84a revealed that none of the four clones obtained was able to produce surfactin (data not shown).

In order to obtain an integration vector and plasmid without point mutation, the 3 kb *3'-srfA-A-5'-srfA-B* transition region was amplified from the chromosome of *B. subtilis* ATCC 21332 (using oligonucleotides srf_11F and srf_14R), and cloned into the plasmid pCC77 to give pCC78. After verification of the sequence, the coding fragment of $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}B1}$ was excised from pCC83a using *Pst*I - hereby excluding the undesired point mutation - and sub-cloned into pCC78 to give the plasmid pCC83 (Fig. 20).

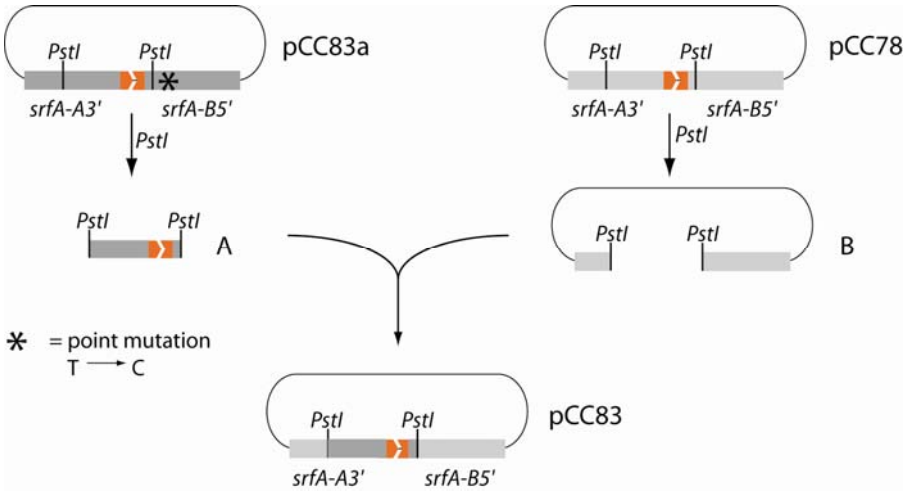


Fig. 20 Subcloning of $COM^{A}_{SrfA\text{-}A3}$-$COM^{D}_{SrfA\text{-}B1}$ in pCC78
Since the plasmid pCC83a carried an undesired point mutation, which compromised the surfactin biosynthesis, a new vector called pCC83 was generated by subcloning of the $COM^{A}_{SrfA\text{-}A3}$-$COM^{D}_{SrfA\text{-}B1}$ coding fragment **(A)** in the vector pCC78 **(B)**. The latter plasmid carried the *3'-srfA-A-5'-srfA-B* homologous region, directly amplified from the parental strain *B. subtilis* ATCC 21332. The sequence identity of all constructs was verified by DNA sequencing.

The new integration plasmid pCC83 was next used in a congression experiment to transform the disruption strain CC64, in order to achieve the reconstitution of the *srfA* operon. As described before, transformants were first selected for Tet^{R} and Cm^{R}. About 60 of the obtained clones were then analyzed for MLS sensitivity. This way, one clone named CC84 was obtained, indicating a good rate of about 2% of desired clones for the corresponding congression experiment.

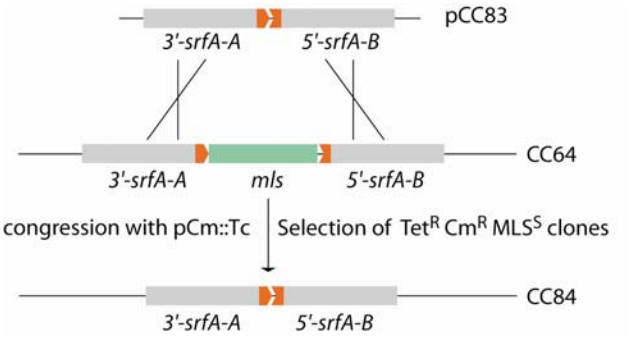


Fig. 21 Swapping of COM$^D_{SrfA-A3}$/COM$^A_{SrfA-B1}$: construction of *B. subtilis* CC84
Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC83, gave the *B. subtilis* strain CC84.

This rate was also achieved in the following experiments, regardless of the type of COM domains integrated in the *B. subtilis* chromosome. Verification of the CC84 genotype was performed as described previously (see 5.4.2, page 58). This time, the analysis of the butanolic extracts of the cultured broth of *B. subtilis* CC84 clearly revealed the expected biosynthesis of surfactin in the reconstitution strain (Fig. 22). This result supported the hypothesis that the point mutation detected in pCC78 was really the reason for the incapability of the strains CC84a 1-4 to produce surfactin, verifying the importance of the residue W150 for the functionality of NRPS C domains.

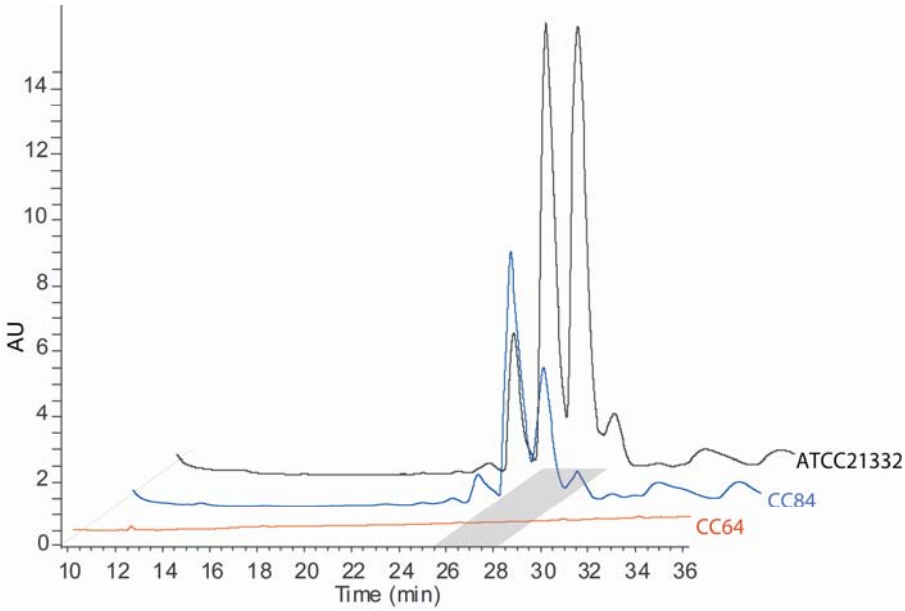


Fig. 22 HPLC analysis of the mutant *B. subtilis* CC84.
Butanolic extracts of the cultured broths of *B. subtilis* reconstitution strain CC84, disruption mutant CC64 and wild-type strain ATCC 21332 were analyzed by RP-HPLC. The lipoheptapeptide surfactin (gray area) was identified.

Quantification of the amount of surfactin produced by the reconstitution strain CC84 and the wild-type ATCC 21332 was accomplished by determination of the absorption area at 220 nm of the corresponding peaks in the HPLC chromatogram, and subsequent normalization to the cell-density of the corresponding overnight cultures. This way, it was possible to show that CC84 produced 85% of the surfactin amount produced by the wild type, indicating that the

genetic manipulation neither caused a drop in product formation nor a change in product distribution. These results clearly indicated that the used cloning strategy for the integration of the natural COM domain pair $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}B1}$ in the chromosome of *B. subtilis* was suitable to reconstitute surfactin biosynthesis, and could therefore be adopted for further COM domain swapping experiments. However, the coding fragments of all COM domain pairs were initially cloned in the vector pCC78a, which carried the described point mutation, and consequently had to be sub-cloned in the "mutation free" vector pCC78 (Fig. 20).

5.4.3 Activity of heterologous COM domains: $COM^{D}_{TycB}/COM^{A}_{TycC}$.

In order to address the question whether the native $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}B1}$ could be replaced by heterologous, cognate COM domain pairs, the functionality and performance of $COM^{D}_{TycB3}/COM^{A}_{TycC1}$, derived from the tyrocidine biosynthetic complex of *B. brevis* ATCC 8185, was tested. The coding fragment of the corresponding COM domain pair was PCR amplified from the chromosome of *B. brevis* ATCC 8185, cloned into pCC78a and subsequently sub-cloned into pCC78, using the same strategy previously executed for the native COM domain pair $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}B1}$. The resulting plasmid, termed pCC92, was used in a congression experiment for the transformation of the disruption strain CC64.

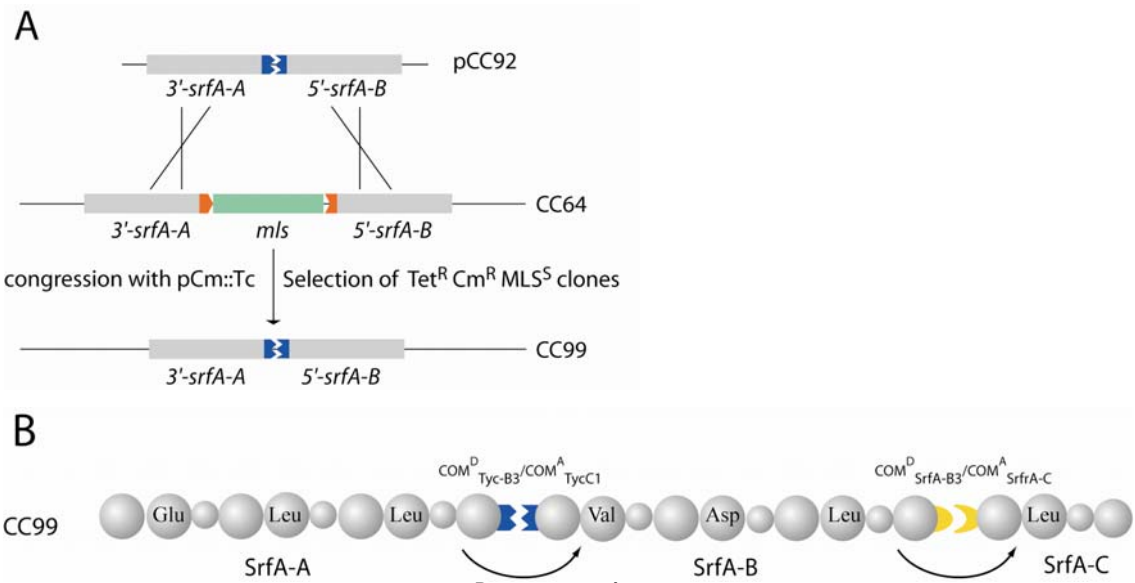


Fig. 23 Swapping of COM domains $COM^{D}_{TycB3}/COM^{A}_{TycC1}$: construction of *B. subtilis* CC99
(A) Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC92, gave the *B. subtilis* strain CC99. **(B)** Organization of the surfactin biosynthetic complex of CC99, generated by integration of the native, cognate COM domain pair $COM^{D}_{TycB3}/COM^{A}_{TycC1}$ The black arrows indicate protein-protein communication.

The following selection on Cm^{R}, Tet^{R} and MLS^{S} colonies led to the identification of the *B. subtilis* mutant CC99. PCR and DNA sequencing analyses of the *3'-srfA-A-5'-srfA-B*

transition region confirmed the correct genotype of CC99 (*srfA-A::COM$^{D}_{TycB3}$-COM$^{A}_{TycC1}$::srfA-B*) (data not shown). Remarkably, subsequent analysis of the butanolic extract revealed that CC99 was able to synthesize also nearly wild-type levels of surfactin (69%), indicating that the substitution of the natural COM domain pair COM$^{D}_{SrfA-A3}$/COM$^{A}_{SrfA-B1}$ pair by the cognate, but heterolgous COM$^{D}_{TycB3}$/COM$^{A}_{TycC1}$ did not affected the activity of the surfactin biosynthetic machinery (Fig. 24).

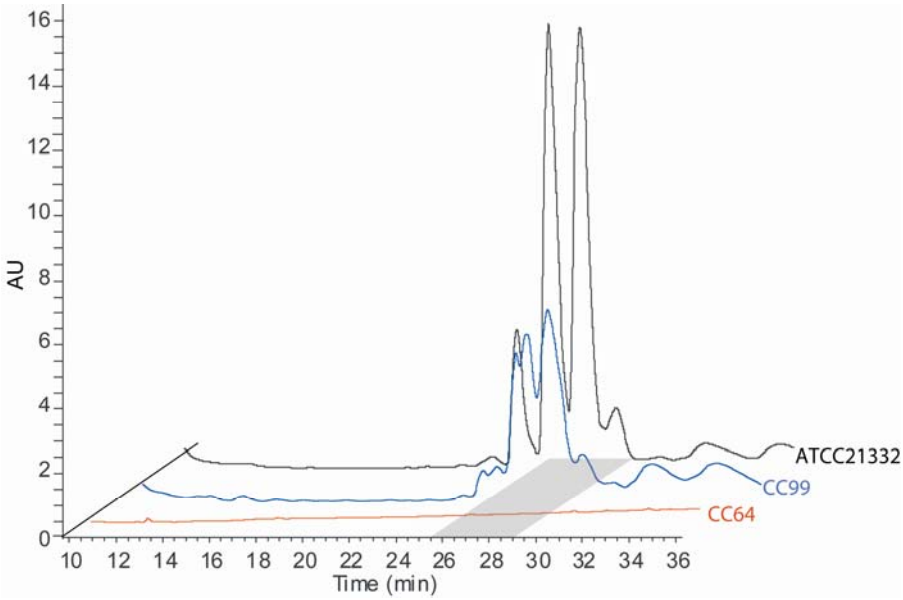


Fig. 24 HPLC analysis of the mutant *B. subtilis* CC99
Butanolic extracts of the cultured broths of *B. subtilis* reconstitution strain CC99, disruption mutant CC64 and wild-type strain ATCC 21332 were analyzed by RP-HPLC. The lipoheptapeptide surfactin (gray area) was identified

5.4.4 Skipping of SrfA-B within the surfactin biosynthetic complex: COM$^{D}_{SrfA-B3}$/COM$^{A}_{SrfA-B1}$ vs. COM$^{D}_{TycA}$/COM$^{A}_{TycC}$.

Given the encouraging outcome of the previous reconstitution experiments, further COM domain swaps were carried out in order i) to enforce the *in vivo* communication between non-partner NRPSs, and – at the same time – ii) to disrupt the *in vivo* interaction between natural partner NRPSs. The establishment of a corresponding system, facilitating the direct communication between SrfA-A and SrfA-C, as well as the controlled *in vivo* skipping of SrfA-B, should be detectable by the production of a shortened lipotetrapeptide (FA-LGlu-LLeu-DLeu-LLeu-OH). In order to achieve this goal and challenge this theory, the two non-cognate COM domain pairs COM$^{D}_{TycA}$/COM$^{A}_{TycC1}$ and COM$^{D}_{SrfA-B3}$/COM$^{A}_{SrfA-B1}$ were chosen.

From *in vitro* studies [Hahn and Stachelhaus, 2004] it was known that TycA and SrfA-C are productively interacting although they are belonging to different NRPS assembly lines. The corresponding crosstalk between both enzymes is possible due to the compatibility of their

respective COM domains, COM^{D}_{TycA} and $COM^{A}_{SrfA-C,}$ which form a so-called mis-cognate pair. In contrast, no effective communication has been observed between COM^{D}_{TycA} and COM^{A}_{TycC1}, which represent a non-cognate pair. Based on both observations, the integration of the non-cognate pair $COM^{D}_{TycA}/COM^{A}_{TycC1}$ at the transition region between *srfA-A* and *srfA-B* should i) allow for the direct interaction between SrfA-A and SrfA-C, and ii) prevent the communication between natural partner NRPSs SrfA-A and SrfA-B (see Fig. 25).

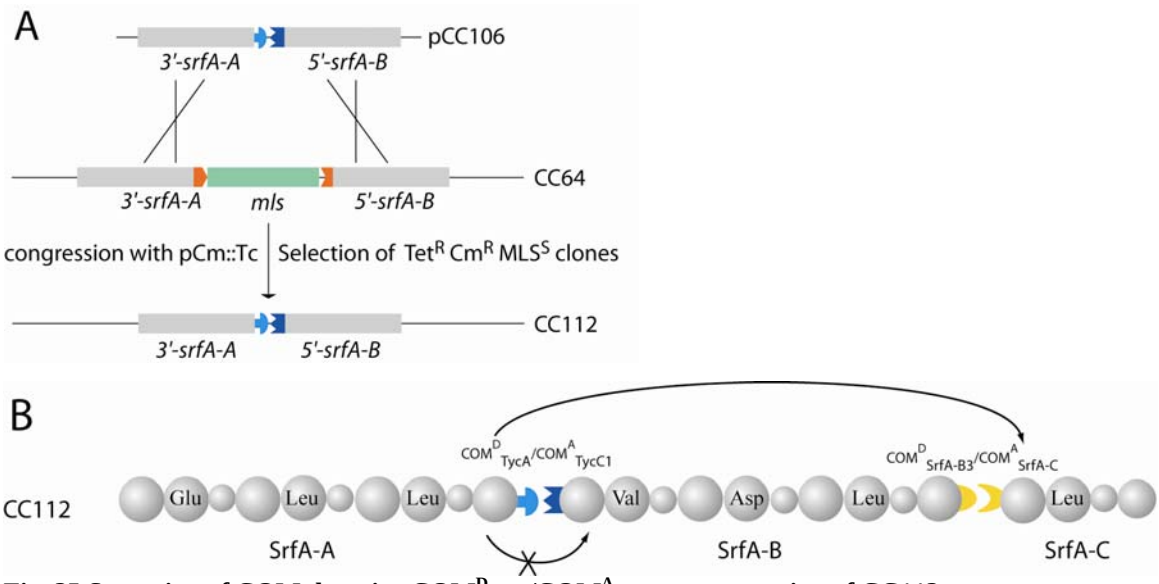


Fig. 25 Swapping of COM domains $COM^{D}_{TycA}/COM^{A}_{TycC1}$: construction of CC112
(A) Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC106, generated the *B. subtilis* strain CC112. **(B)** Organization of the surfactin biosynthetic complex of CC112, obtained by integration of the non-cognate COM domain pair $COM^{D}_{TycA}/COM^{A}_{TycC1}$ The black arrows indicate protein-protein communication.

Provided that the proposed model would also apply to the surfactin biosynthetic system, the COM domain pairs $COM^{D}_{SrfA-B3}/COM^{A}_{SrfA-B1}$, and $COM^{D}_{SrfA-A3}/COM^{A}_{SrfA-C}$ should represent non-cognate COM domain pairs, preventing the futile interactions between two molecules of SrfA-B, as well as between SrfA-A and SrfA-C within the native surfactin biosynthetic complex. Consequently, the introduction of $COM^{D}_{SrfA-B3}/COM^{A}_{SrfA-B1}$ at the transition between SrfA-A and SrfA-B was also expected to cause the intended skipping of SrfA-B, and biosynthesis of the shortened lipotetrapeptide product.

Congression experiments were performed with the plasmids pCC106 and pCC98, respectively, resulting in the construction of the *B. subtilis* mutants CC112 (*srfA-A::* COM^{D}_{TycA}-COM^{A}_{TycC1}*::srfA-B*) (Fig. 25) and CC102 (*srfA-A::* $COM^{D}_{SrfA-B3}$-$COM^{A}_{SrfA-B1}$*::srfA-B*) (Fig. 26), respectively. PCR analyses, as well as DNA sequencing data confirmed the genetic integrity of both mutants (data not shown).

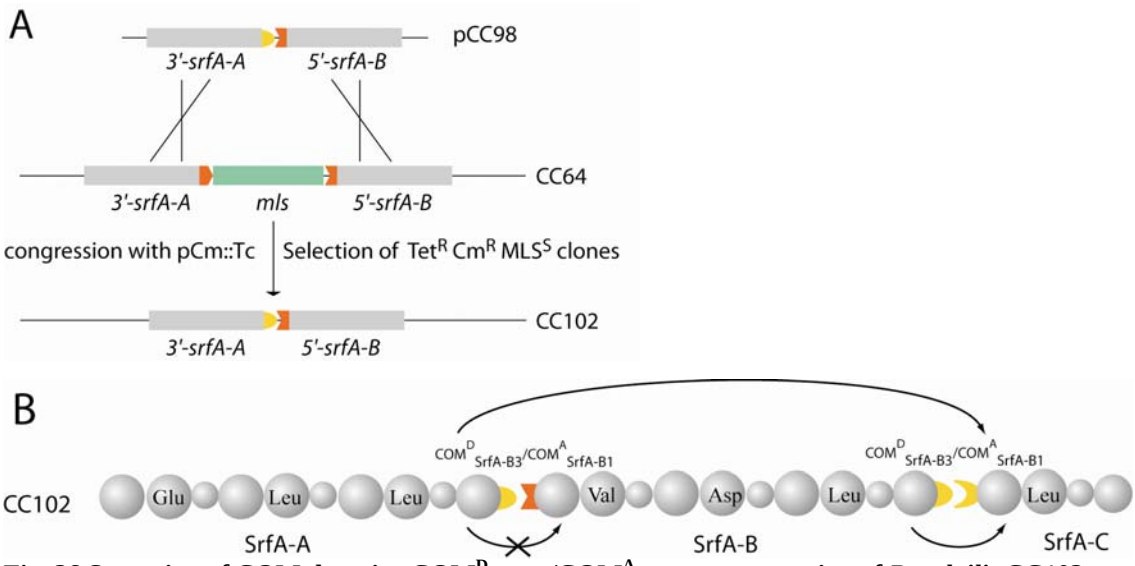


Fig. 26 Swapping of COM domains $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}B1}$: construction of *B. subtilis* CC102
(A) Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC98, gave the *B. subtilis* strain CC102. **(B)** Organization of the surfactin biosynthetic complex of CC102, generated by integration of the non-cognate COM domain pair $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}B1}$. The black arrows indicate protein-protein communication.

After verification of their genotypes, both mutants were investigated for the production of lipopeptides as described above. In case of the mutant CC112, this analysis revealed the expected production of the linear lipotetrapeptide product FA-LGlu-LLeu-DLeu-LLeu-OH. As in the case of the parental lipoheptapeptide surfactin, the fatty acid moiety of the lipotetrapeptide likewise exhibits variations in the chain length (Fig. 27). Notably, the product titer was in same order of magnitude (70%), as observed for the full-length lipoheptapeptide in the wild-type producer. In this context it should be noted that due to the lack of an authentic standard for the linear lipotetrapeptide, quantification of the products was carried out, using the area of absorption at 220 nm. Since the lipotetrapeptide contains a lower number of absorbing peptide bonds, equal areas of absorption should correspond to relatively higher amounts of lipotetrapeptide.

Interestingly, as shown in Fig. 27 and Fig. 28, the same HPLC(-MS) analysis of CC112 revealed no indication for the formation of the full-length lipoheptapeptide antibiotic surfactin; not even in trace amounts. This observation, along with the successful formation of the shortened lipotetrapeptide, is clear evidence for intended skipping of SrfA-B, and the enforced productive interaction between the natural non-partner enzymes SrfA-A and SrfA-C.

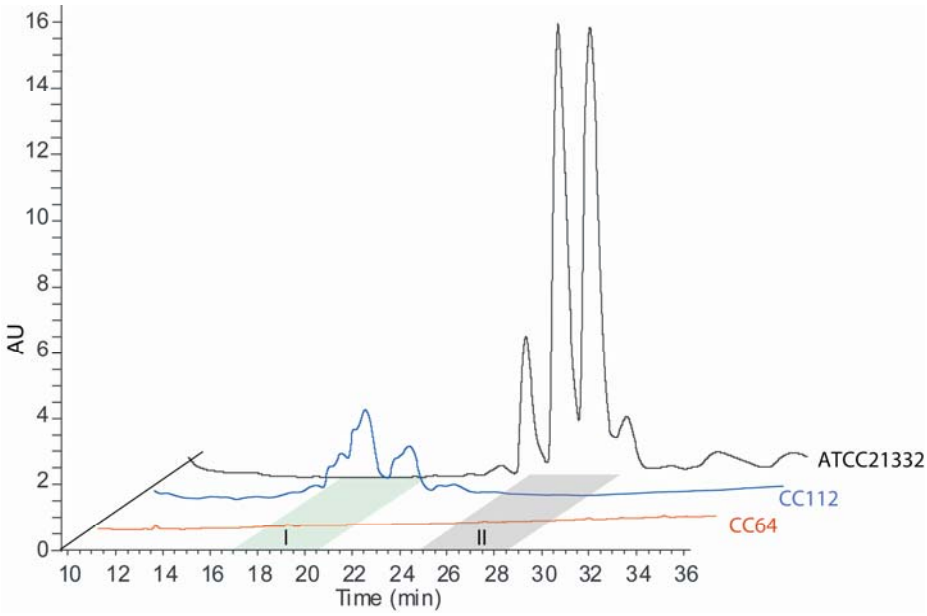
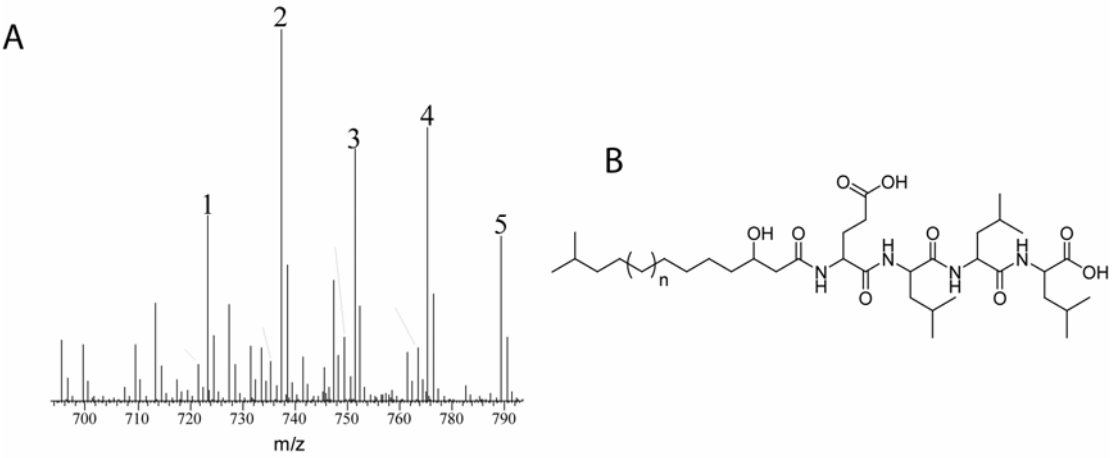


Fig. 27 HPLC analysis of the mutant *B. subtilis* CC112
Butanolic extracts of the cultured broths of *B. subtilis* reconstitution strain CC112, disruption mutant CC64 and wild-type strain ATCC 21332 were analyzed by RP-HPLC.The lipoheptapeptide surfactin (gray area) and the lipotetrapeptide FA-LGlu-LLeu-DLeu-LLeu-OH (green area) were both identified.



C

Peak No.	Molecular Formula	Mass calculated	Mass observed
1	$C_{35}H_{64}N_4O_9K$	723.431	723.430
2	$C_{36}H_{66}N_4O_9K$	737.447	737.446
3	$C_{37}H_{68}N_4O_9K$	751.462	751.462
4	$C_{38}H_{70}N_4O_9K$	765.478	765.477
5	$C_{39}H_{72}N_4O_9K$	779.494	779.493

Fig. 28 FT-ICR MS analysis of the lipotetrapeptide produced by CC112 (and CC102)
(A) Mass spectrum of the butanolic extracts derived from *B. subtilis* CC112 and CC102 **(B)** The lipotetrapeptide FA-LGlu-LLeu-DLeu-LLeu-OH is synthesized as a mixture of molecules, differing in the length of the fatty acid chain (n = 1; 2; 3 etc). **(C)** Characterization of the mass peaks shown in (A).

The second mutant constructed, *B. subtilis* CC102, also showed the production of expected lipotetrapeptide, even though at slightly lower production level (37%). Surprisingly, however, the analysis also revealed the simultaneous formation of the full-length lipoheptapeptide surfactin, although at a reduced product titer (18% of wild type level, Fig. 29 and Tab. 1).

This indicates that communication between SrfA-A and SrfA-B was not completely abrogated, and that $COM^{D}_{SrfA\text{-}B3}$ and $COM^{A}_{SrfA\text{-}B1}$ are forming a mis-cognate, rather than the postulated non-cognate COM domain pair. In the context of the NRPS hybrid system under investigation, this mis-cognate COM domain pair mediates the productive interaction between both SrfA-A, and SrfA-B, and eventually formation of the lipoheptapeptide surfactin.

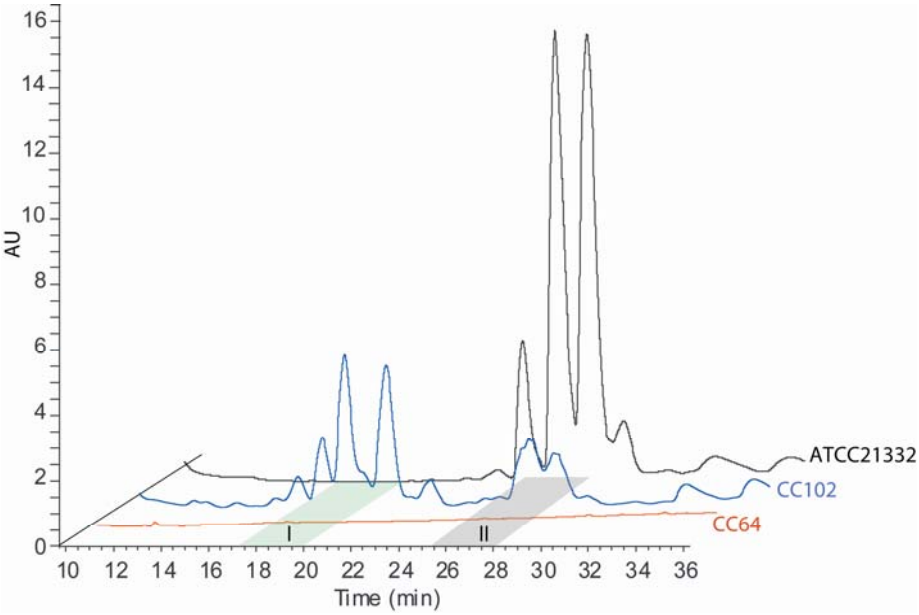


Fig. 29 HPLC analysis of the mutant *B. subtilis* CC102
Butanolic extracts of the cultured broths of *B. subtilis* reconstitution strain CC102, disruption mutant CC64 and wild-type strain ATCC 21332 were analyzed by RP-HPLC. The lipoheptapeptide surfactin (gray area) and the lipotetrapeptide FA-LGlu-LLeu-DLeu-LLeu-OH (green area) were both identified

5.4.5 Combinatorial *in vivo* biosynthesis; $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C}$

The ultimate aim of this study was the exploitation of the biocombinatorial potential of COM domains by engineering of a biosynthetic assembly line in which all donor and acceptor enzyme are equipped with the same (cognate) pair of COM domains. Due to the resulting abrogation of the selectivity-barrier provided by the COM domains, all NRPSs of the biosynthetic complex should be inter-communicable. Recently, such a so-called universal communication system has been established *in vitro*, resulting in the formation of random NRPS complexes, and giving rise to the simultaneous synthesis of an array of different peptide products [Hahn and Stachelhaus, 2006].

In order to test the feasibility of this approach *in vivo*, the biosynthetic assembly line of surfactin was manipulated by integration of the cognate COM domain pair $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C}$ at the transition between SrfA-A and SrfA-B. As a result, the hybrid enzyme SrfA-A:: $COM^{D}_{SrfA\text{-}B3}$ was expected to be able to interact with both its natural partner SrfA-B (now equipped with the compatible donor $COM^{D}_{SrfA\text{-}C}$), and the natural non-partner enzyme

SrfA-C. As shown in Fig. 30B, both possibilities should lead to the simultaneous formation of the full-length lipoheptapeptide surfactin (organization of the biosynthetic complex: SrfA-A/SrfA-B/SrfA-C), and the shortened lipotetrapeptide FA-LGlu-LLeu-DLeu-LLeu-OH (SrfA-A/SrfA-C). Technically also feasible would be an interaction between two molecules of SrfA-B, which should lead to the formation of an elongated lipodecapeptide product (SrfA-A/SrfA-B/SrfA-B/SrfA-C). To realize the corresponding COM domain swap, the plasmid pCC85, carrying the COM domain pair $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C,}$ was constructed and used in a congression experiment for the transformation of the disruption mutant *B. subtilis* CC64. The genotype (*srfA-A*::$COM^{D}_{SrfA\text{-}B3}$-$COM^{A}_{SrfA\text{-}C}$::*srfA-B*) (Fig. 30A) of the resulting strain CC91 was verified by PCR and sequence analysis as described above (see 5.1) (data not shown).

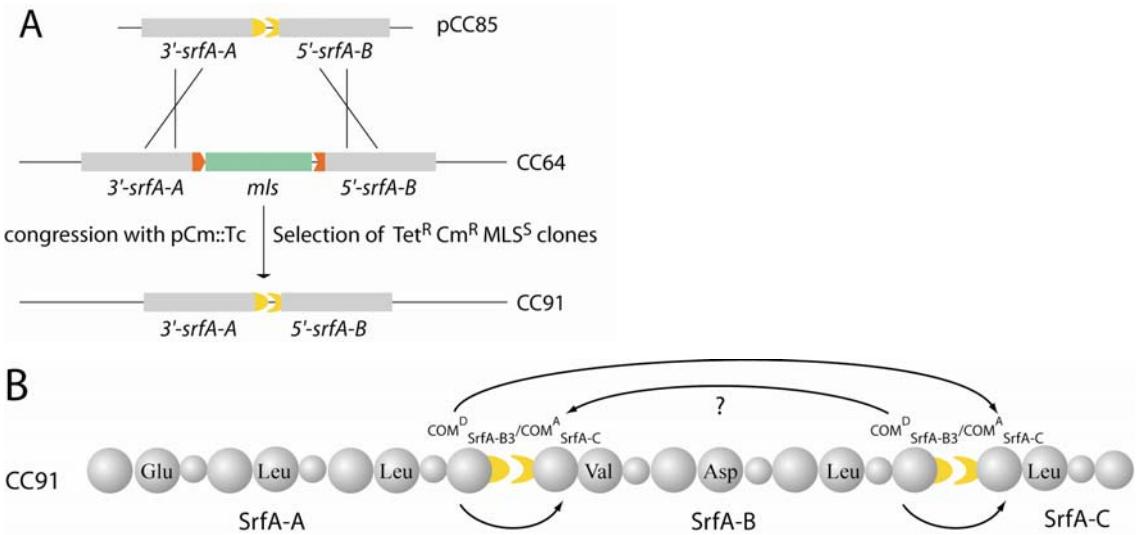


Fig. 30 Swapping of homologous COM domains $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C}$ construction of *B. subtilis* CC91
(A) Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC85, gave the *B. subtilis* strain CC91 **(B)** Organization of the surfactin biosynthetic complex of CC91, constructed by integration of the native, cognate COM domain pair $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C}$. The black arrows indicate protein-protein communication

After verification of the genetic integrity of CC91, the mutant was investigated for the production of lipopeptide products. The corresponding HPLC analysis clearly revealed the expected formation of both lipotetra- and lipoheptapeptide, while no lipodecapeptide could be observed (Fig. 31). Intriguingly, however, the production level of CC91 was relatively low, yielding only about 3% of the amount of surfactin produced by the wild-type and most other mutants (Tab. 1). Interestingly, the ratio between the amount of lipotetrapeptide and surfactin produced by CC91 was about 2:1, suggesting that the communication between the natural non-partner NRPSs SrfA-A and SrfA-C was favoured.

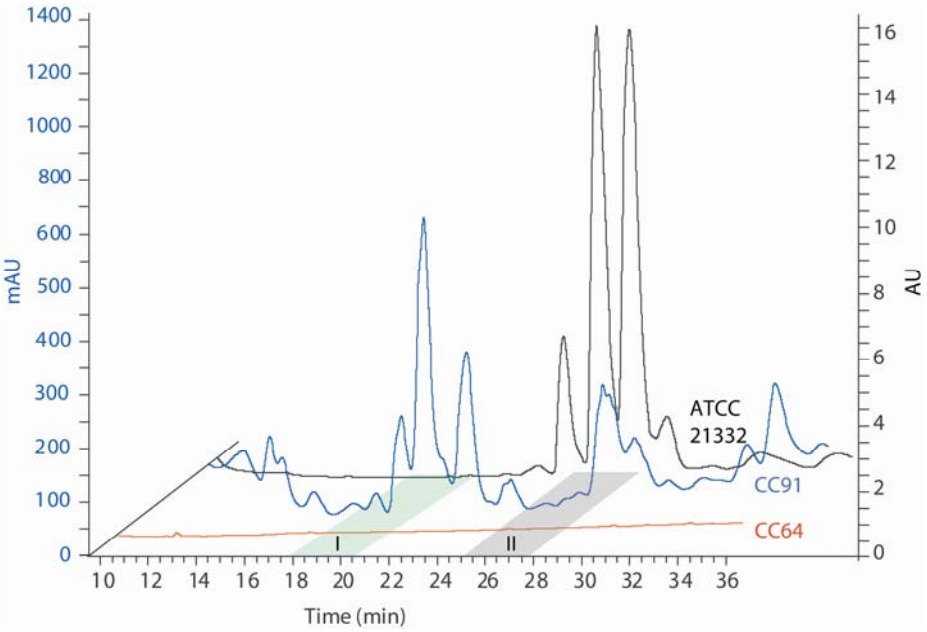


Fig. 31 HPLC analysis of the mutant *B. subtilis* CC91
Butanolic extracts of the cultured broths of *B. subtilis* reconstitution strain CC91, disruption mutant CC64 and wild-type strain ATCC 21332 were analyzed by RP-HPLC.The lipoheptapeptide surfactin (gray area) and the lipotetrapeptide FA-LGlu-LLeu-DLeu-LLeu-OH (green area) were both identified

5.4.6 Investigation of the proposed non-cognate COM domain pair $COM^{D}_{SrfA-A3}/COM^{A}_{SrfA-C}$

According to the current understanding of COM domains, donor and acceptor COM domain of natural non-partner enzymes should be incompatible (non-cognate), in order to prevent the futile interaction between both NRPSs. This hypothesis actually holds true for the *in vitro* investigation of the COM domains derived from tyrocidine biosynthetic complex [Hahn and Stachelhaus, 2004]. However, careful analysis of $COM^{D}_{SrfA-B3}/COM^{A}_{SrfA-B1}$ revealed that this presumed non-cognate COM domain was very well able to mediate a productive interaction between the natural partner NRPSs SrfA-A and SrfA-B (see this work 5.4.4). In order to evaluate whether this observation was due to an experimental artefact, or a peculiarity of COM domains derived from the surfactin biosynthetic complex, the second presumed non-cognate COM domain pair $COM^{D}_{SrfA-A3}/COM^{A}_{SrfA-C}$ was likewise investigated. According to the initial assumption, the corresponding COM domain swap should prevent the productive interaction between SrfA-A and SrfA-B and, consequently, also the biosynthesis of surfactin (Fig. 32B). In order to verify this hypothesis, the plasmid pCC106 was constructed, carrying the coding-fragment of $COM^{D}_{SrfA-A3}/COM^{A}_{SrfA-C}$, flanked by the usual homologous regions (*3'-srfA-A* and *5'-srfA-B*), required for recombination in the *B. subtilis* chromosome (Fig. 32A).

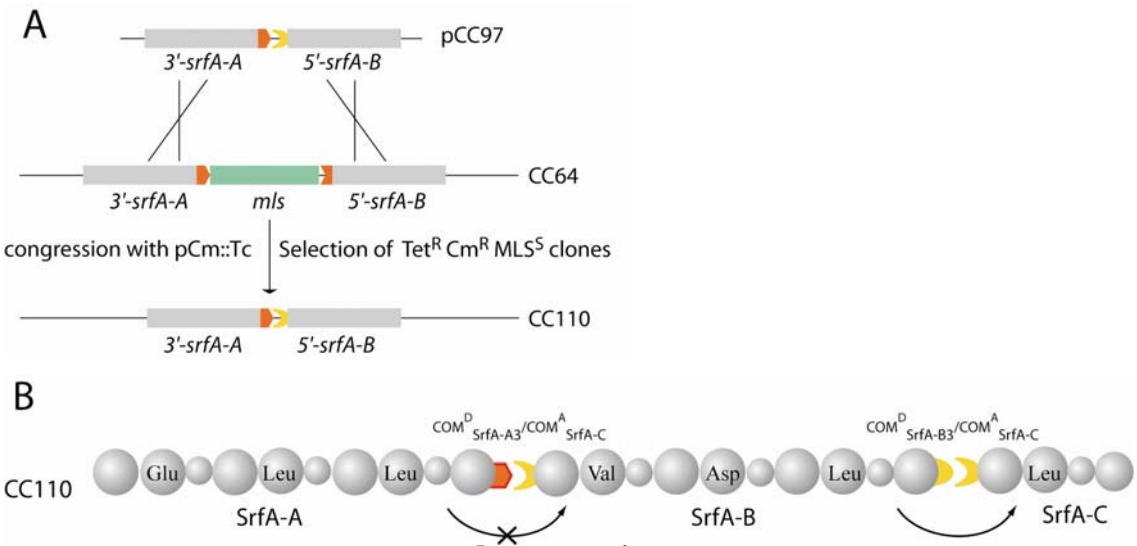


Fig. 32 Swapping of COM domains $COM^{D}_{SrfAA3}/COM^{A}_{SrfA-C}$: construction of CC110
(A) Reconstitution of the *srfA* operon in the chromosome of *B. subtilis* CC64, using plasmid pCC97, gave the *B. subtilis* strain CC110. **(B)** Organization of the surfactin biosynthetic complex of CC110, generated by integration of the native, cognate COM domain pair $COM^{D}_{SrfA-A3}/COM^{A}_{SrfA-C}$. The black arrows indicate protein-protein communication.

The plasmid pCC106 was used for the transformation of the disruption strain CC64, and the congression experiment resulted in the isolation of a Tet^{R} MLS^{S} clone. Genetic characterization was carried out as described above (see 5.4.2), verifying the integrity of *B. subtilis* strain CC110 (data not shown). Hence, the isolated strain could be next investigated for the formation of lipopeptide products.

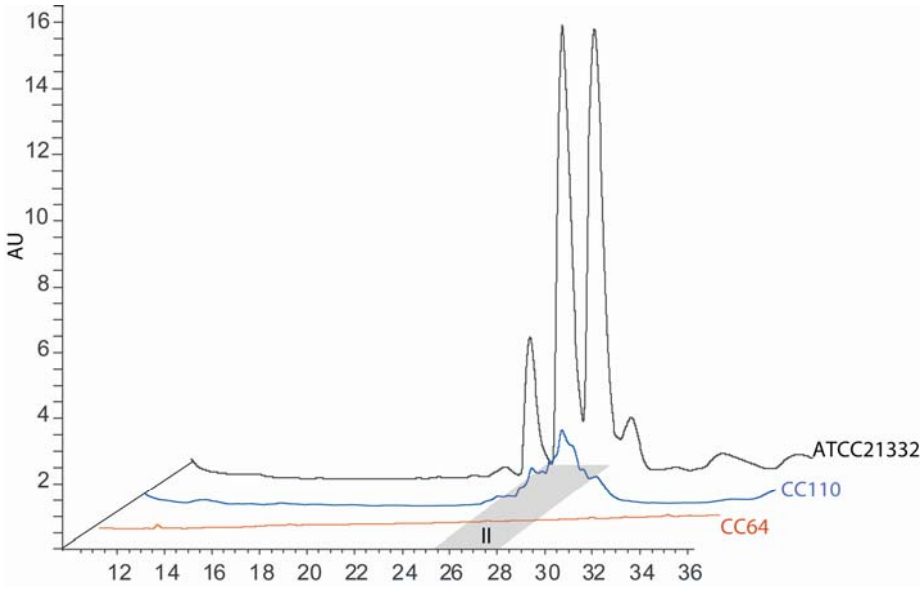


Fig. 33 HPLC analysis of the mutant *B. subtilis* CC110
Butanolic extracts of the cultured broths of *B. subtilis* reconstitution strain CC110, disruption mutant CC64 and wild-type strain ATCC 21332 were analyzed by RP-HPLC. The lipoheptapeptide surfactin (gray area) was identified

Astonishing, RP-HPLC analysis clearly showed that CC110 was still able to synthesize surfactin (26%) (see Fig. 33 and Tab. 1), indicating that the COM domain pair $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}C}$ is also able to provide a productive communication between SrfA-A and SrfA-B. Consequently, $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}C}$ has to be considered a mis-cognate, rather than a non-cognate COM domain pair.

3.5 Summary

The presented study aimed on harnessing the potential of COM domains for the directed reprogramming of the surfactin biosynthetic complex, and on establishing of an *in vivo* system for true biocombinatorial synthesis of lipopeptides. To this end, the first pair of COM domains of the surfactin biosynthetic complex, facilitating the selective interaction between SrfA-A and SrfA-B, was substituted against various cognate, mis-cognate and non-cognate sets of COM domains. These experiments demonstrated the functionality of COM domains even in the context of a heterologous host and NRPS system, and allowed for the intended skipping of biosynthetic enzymes within a multi-enzymatic NRPS complex. Abrogation of the selectivity-barrier provided by COM domains afforded the simultaneous, biocombinatorial synthesis of distinct lipopeptide products. Importantly, most of these manipulations were connected with only minor reductions in product titer (average: 51% of wild type). The following table summarizes the observed consequences of the corresponding COM domain swaps on the nature and titer of the synthesized lipopeptide product.

Strain	description (COM domains used)	lipoheptapeptide (in %area at UV_{214} nm)	% lipotetrapeptide (in %area at UV_{214} nm)
ATCC 21332	$COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}B1}$	100 ± 7	n.d.
CC64	disruption mutant	n.d.	n.d.
AM1	$COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}B1}$	110 ± 9	n.d.
CC84	$COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}B1}$	76 ± 8	n.d.
CC99	COM^{D}_{TycB3}-COM^{A}_{TycC1}	69 ± 8	n.d.
CC112	COM^{D}_{TycA}-COM^{A}_{TycC1}	n.d.	73 ± 6
CC102	$COM^{D}_{SrfA\text{-}B3}$-$COM^{A}_{SrfA\text{-}B1}$	18 ± 3	38 ± 3
CC91	$COM^{D}_{SrfA\text{-}B3}$-$COM^{A}_{SrfA\text{-}C}$	3 ± 1	5 ± 1
CC110	$COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}C}$	26 ± 4	n.d.

Tab. 1 Quantification of the lipopeptide products
The table shows the relative amounts of surfactin and the truncated lipotetrapeptide synthesized by the *B. subtilis* strains constructed in this study. The amount of surfactin produced by the wild type strain *B. subtilis* ATCC 21332 is defined as 100%. Quantification of the products was carried out by determination of the area of absorption at 220 nm. Due to the lack of an authentic standard, the measured integral for the lipotetrapeptide was also compared with the integral of full-length lipoheptapeptide surfactin. Given the differences of there corresponding extinction coefficients at 220 nm, i.e. due to the different number of peptide bonds, this procedure might underestimate the actually amount of lipotetrapeptide produced.

6 Results (II)

6.1 Method development for the gentle purification of NRPSs and entire NRPS multi-enzyme complexes

The genetic characterization of NRPSs produced by bacteria of the genus *Bacillus* provided important information about the organization and function of NRPS in general. Therefore, the corresponding NRPS genes or gene fragments were cloned into suitable expression vectors and overexpressed in the heterologous host *E. coli*. The resulting proteins were subsequently purified and biochemically characterized *in vitro*. While a successful biochemical characterization of single domains or modules could be accomplished, the heterologous overproduction of entire multi-modular NRPSs is often correlated with problems like incorrect folding of the proteins, formation of inclusion bodies, as well as protein degradation [Pfeifer et al., 2001; Symmank et al., 1999]. Furthermore *E. coli* is not equipped with an appropriate Ppant-transferase like Sfp, necessary for the post-translational modification of NRPS from the inactive apo-to the active holo-form. This deficiency of *E. coli* can be compensated either by carrying out *in vitro* modification reaction via the incubation of the purified NRPS with a designated Ppant transferase and the required substrates (Mg^{2+} and CoA), or by the *in vivo* in-parallel heterologous production of the NRPS and the Ppant transferase [Gruenewald et al., 2004; Walsh et al., 1997]. Expression of NRPS genes in the natural producer strain circumvents the mentioned problems, yielding correctly folded and fully active holo-enzymes. An additional advantage of the utilization of the natural producer strain is provided by the possibility to co-purify associated – potentially even uncharacterized or unknown – proteins, which take part in the biosynthesis of a given peptide product. For example, a 40 kDa-enzyme, probably involved in the lipo-initiation of the surfactin biosynthesis, was isolated from a crude extract of *B. subtilis* ATCC 21332 [Menkhaus et al., 1993]. Surprisingly this enzyme could not be detected in following experiments and the acyl transferase function has still to be assigned to the correct protein. Possible drawbacks of natural producer organisms with regard to protein purification, though, are the usually low production levels of NRPSs, and the need for time-consuming multi-step purification schemes.

In the present work, I aimed to develop a method for the gentle, single-step purification of holo-NRPSs and NRPS complexes. To this end, the well-characterized but still not

completely deciphered surfactin multi-enzyme complex of *B. subitlis* was selected as a model system.

6.2 Purification by Ni-NTA affinity chromatography

In a first set of experiments, Ni-NTA affinity chromatography should be tested as a tool for the single-step purification of the tri-modular surfactin synthetase SrfA-A. For this purpose, the coding sequence of a hexahistidine-affinity tag had to be fused to the *srfA-A* gene within the chromosome of *B. subtilis* OKB105, using a two-step marker exchange homologous recombination approach. However, this method, consisting of successive gene disruption and reconstitution, required some additional considerations, since integration of a resistance marker within the *3'-srfA-A-5'-srfA-B* transition region, not only leads to the desired disruption of the surfactin biosynthetic operon, but also disables the co-expression of the small competence regulator gene *comS*, situated in a different reading frame in the coding region of the first PCP domain of SrfA-B. Since the corresponding gene product, ComS, has been shown to be essential for establishment of the host organism's natural competence, the disruption would ultimately result in a genetically inaccessible deletion strain, which could not be used for further genetic manipulations. In order to circumvent this problem, a second copy of *comS* had to be integrated into the chromosome, before the actual marker exchange experiment could be started.

6.2.1 Integration of a second copy of *comS* in the *amyE*-site of the surfactin producer *B. subtilis* OKB105

For the integration of a second copy of *comS*, the genetic locus of the alpha-amylase gene *amyE* was chosen. Under the laboratory conditions used, *amyE* is not essential and hence could be disrupted without impact on the host organism's growth and surfactin production. The *comS* integration was performed, using the *amyE*-integration plasmid pKE27 (pDR66-*comS*). This plasmid harbors *comS* under the transcriptional control of the IPTG-inducible P_{spac}-promotor, as well as a chloramphenicol-conferring resistance marker for direct selection [Eppelmann et al., 2001], and was used – after linearization with the restriction endonucleases *Apa*I and *Sac*II – for the transformation of wild type strain *B. subtilis* OKB105 (Fig. 34A). After selection on DSM agar that was supplemented with 5 µg/ml chloramphenicol, three Cm^R colonies were obtained. The genetic verification of the transformants was carried out via

PCR analysis, using chromosomal DNA along with oligonucleotides RF-48F and RF-48R (see appendix), which are located outside of the used homologous regions (Fig. 34A). For all three strains, the expected PCR product of 5 kb was obtained (data not shown), indicating the stable, double-crossover integration of *comS*.

Final confirmation of the genotype of the obtained strains *B. subtilis* CC15a-c was obtained by Southern blot analysis. As controls, chromosomal DNAs of the parental strain *B. subtilis* OKB105, and *B. subtilis* KE30 [Eppelmann et al., 2001] were analyzed, containing an intact *amyE* gene or *comS* insertion within the *amyE* site, respectively. A 500 bp-fragment of the 5'-*amyE* front fragment was used as a probe, which was PCR amplified from chromosomal DNA of *B. subtilis* OKB105 with oligonucleotides 5'Amy-front and 3'AmyE-probe. As shown in Fig. 34, hybridization with the fluorescein-dUTP-labeled 500 bp-probe allowed for the expected detection of a 15 kb-*Sph*I fragment and 6,5 kb-*Hind*III fragment for the parental strain, as well as 12 kb (*Sph*I) and 3,7 kb (*Hind*III)-fragments for KE30, as well as the three clones of *B. subtilis* CC15, this way verifying the genotype of the obtained OKB105-derivatives.

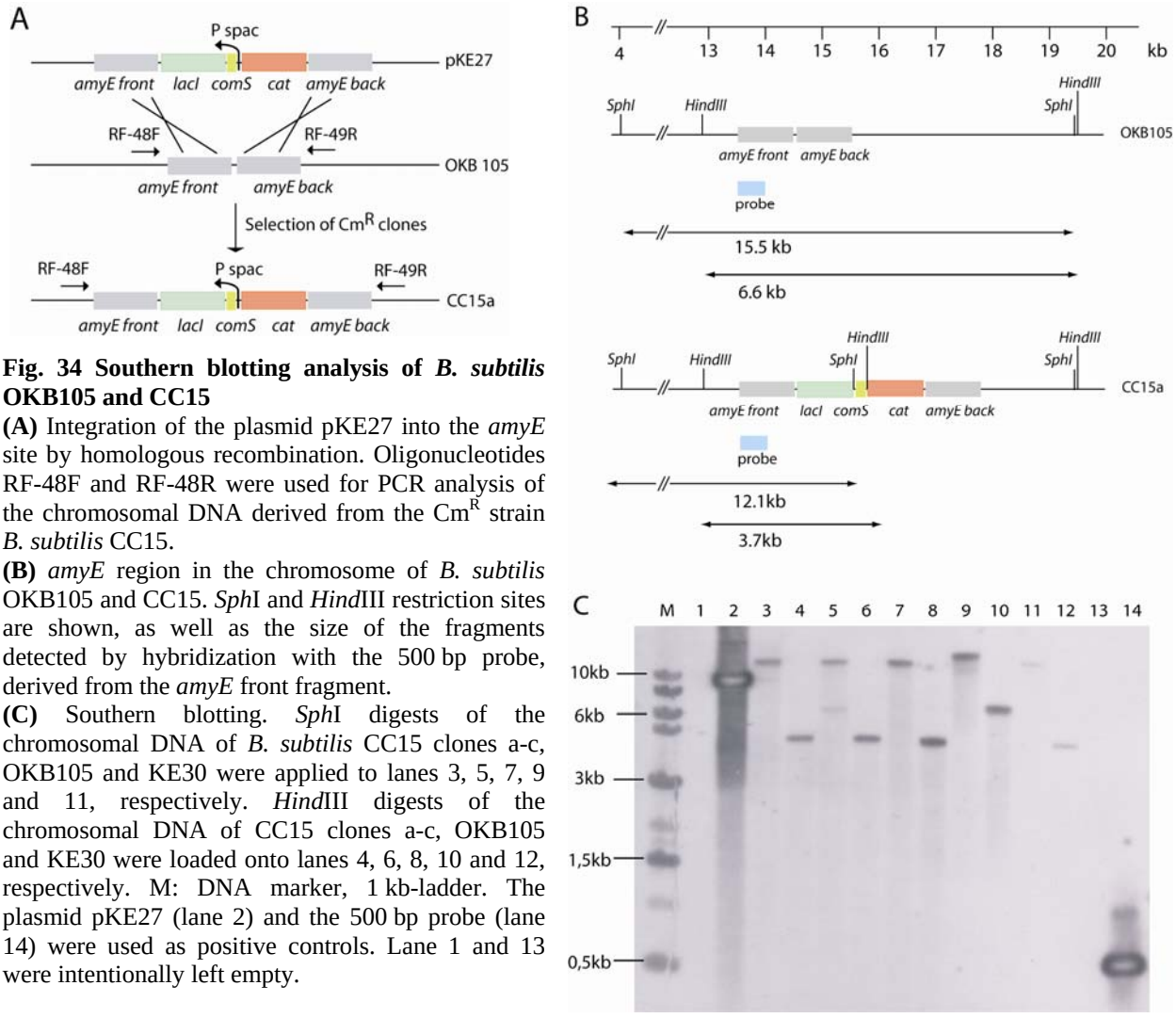


Fig. 34 Southern blotting analysis of *B. subtilis* OKB105 and CC15

(A) Integration of the plasmid pKE27 into the *amyE* site by homologous recombination. Oligonucleotides RF-48F and RF-48R were used for PCR analysis of the chromosomal DNA derived from the Cm^R strain *B. subtilis* CC15.

(B) *amyE* region in the chromosome of *B. subtilis* OKB105 and CC15. *Sph*I and *Hind*III restriction sites are shown, as well as the size of the fragments detected by hybridization with the 500 bp probe, derived from the *amyE* front fragment.

(C) Southern blotting. *Sph*I digests of the chromosomal DNA of *B. subtilis* CC15 clones a-c, OKB105 and KE30 were applied to lanes 3, 5, 7, 9 and 11, respectively. *Hind*III digests of the chromosomal DNA of CC15 clones a-c, OKB105 and KE30 were loaded onto lanes 4, 6, 8, 10 and 12, respectively. M: DNA marker, 1 kb-ladder. The plasmid pKE27 (lane 2) and the 500 bp probe (lane 14) were used as positive controls. Lane 1 and 13 were intentionally left empty.

Analysis of the surfactin production of *B. subtilis* CC15 revealed that – as already shown for the corresponding *B. subtilis* ATCC21332-derivative AM1 (see 5.3, page 56) – insertion of the *comS* gene in the *amyE* site did not affect the surfactin biosynthesis (data not shown).

6.2.2 Integration of a resistance marker in the transition site of *srfA-A-srfA-B*

In order to generate a disruption mutant, in which the transition region between *srfA-A* and *srfA-B* was interrupted by a selectable marker, *B. subtilis* CC15a was transformed with the *Xho*I/*Xba*I-linearized plasmid pCC13. As shown in Fig. 35A, this disruption plasmid contains *3'-srfA-A* and *5'-srfA-B* homologous regions that are separated by a MLS-resistance conferring cassette. Transformants were selected on DSM agar for Cm^R and MLS^R. The genotypes of the obtained clones were analyzed by PCR using oligonucleotides srf_42F and srf_43R (see appendix). As a control, chromosomal DNA of the parental strain *B. subtilis* CC15 was tested. As shown in Fig. 35B, the corresponding PCRs gave rise to the expected 3 kb and 4,6 kb fragment for CC15 and CC16, this way verifying the successful double cross-over integration and disruption of the *srfA* operon.

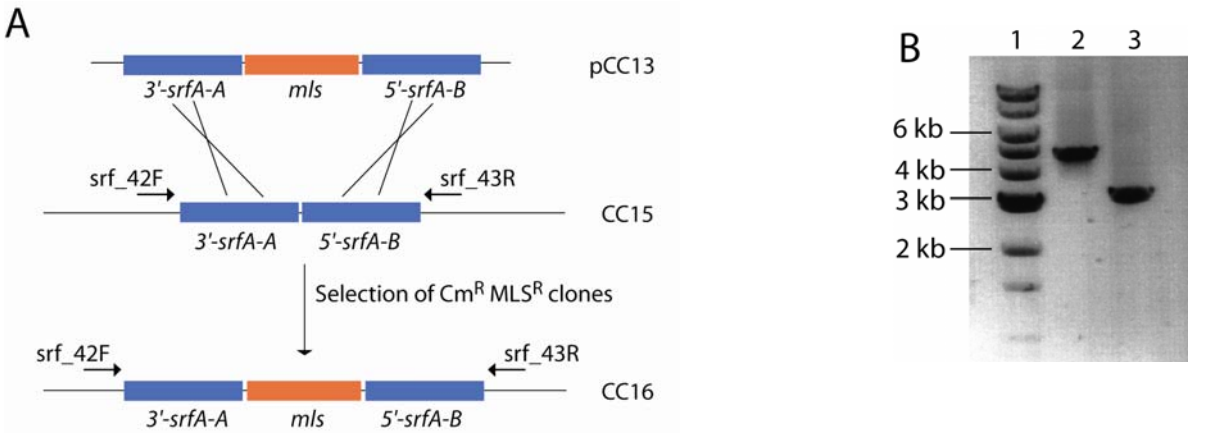


Fig. 35 Construction and characterization of the disruption strain *B. subtilis* CC16
(A) Integration of pCC13 via homologous recombination into the chromosome of *B. subtilis* CC15. **(B)** PCR analysis of the obtained strain *B. subtilis* CC16 and the parental strain CC15 using the oligonucleotides srf_42F and srf_43R Lane 1: 1 kb DNA marker. Lanes 2 and 3: PCR, using chromosomal DNA of *B. subtilis* obtained CC16 and CC15, respectively.

In order to determine, whether the generated disruption strain *B. subtilis* CC16 was still capable of surfactin production, a butanol extraction was carried out. However, subsequent HPLC/MS analysis unequivocally revealed the expected surfactin-deficient phenotype.

6.2.3 Fusion of a Hexahistidine-tag to the C-terminus of SrfA-A

In order to achieve the reconstitution of the *srfA* operon and simultaneous fusion of the coding region of a hexahistidine-tag onto the 3'-end of *srfA-A*, the reconstitution plasmid pCC14 was constructed. The plasmid contains the 1,5 kb *3'-srfA-A* homologous regions fused in frame to the His-tag coding sequence, as well as the 1,4 kb *5'-srfA-B* fragment. After linearization with the endonucleases *Xho*I and *Xba*I, plasmid pCC14 was used along with the helper plasmid pCm::Tc for the co-transformation of *B. subtilis* CC16 (Fig. 36A). The helper plasmid is self-replicable both in *E. coli* und *B. subtilis*, and confers resistance towards the antibiotic tetracycline [Steinmetz and Richter, 1994]. Hence, transformants derived from the congression experiment were first selected on DSM agar, supplemented with 5 µg/ml chloramphenicol (Note: wild-type and mutants carry a *cat* cassette at *amyE* site; see Fig. 34) and 20 µg/ml tetracycline. The resulting 300 Cm^R/Tet^R clones were subsequently screened for MLS sensitivity. Three clones with the desired genotype of Cm^R, Tet^R, MLS^S were finally analyzed by PCR, using the oligonucleotides Srf_42F and Srf_43R. As shown in Fig. 36B, PCR analysis of *B. subtilis* CC17 revealed the expected 3 kb-fragment was amplified, proving the successful reconstitution of the *srfA* operon.

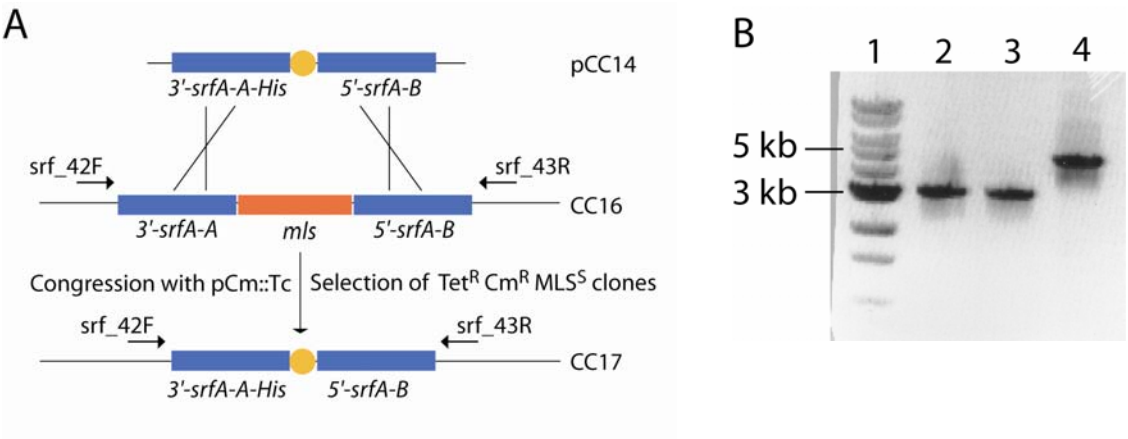


Fig. 36 Construction and characterization of *B. subtilis* CC17.
(A) Double-crossover homologous recombination and reconstitution of the *srfA* operon, using plasmid pCC14. **(B)** Chromosomal DNAs of the mutant CC17 clone 1, *B. subtilis* OKB 105 and CC16 were used as templates for PCRs, using oligonucleotides Srf_42F and Srf_43R. Lane 1: 1 kb DNA marker; lanes 2, 3 and 4 PCR products obtained with chromosomal DNA of CC17, OKB105, and CC16, respectively.

In order to establish that the appending of a hexahistidine-tag onto SrfA-A had no negative effect on surfactin production, *B. subtilis* CC17 was also investigated for the capability of producing surfactin. As controls, the wild-type producer strain OKB 105, and the disruption mutant CC17 were analyzed.

HPLC analysis revealed, that strain CC17 produced almost wild-type levels of surfactin, indicating that the introduced His-tag had no significant effect on the overall productivity of

the surfactin biosynthetic complex. After normalization (OD_{600} nm of the culture used for the butanolic extraction), the production level of the mutant strain amounted to 78%.

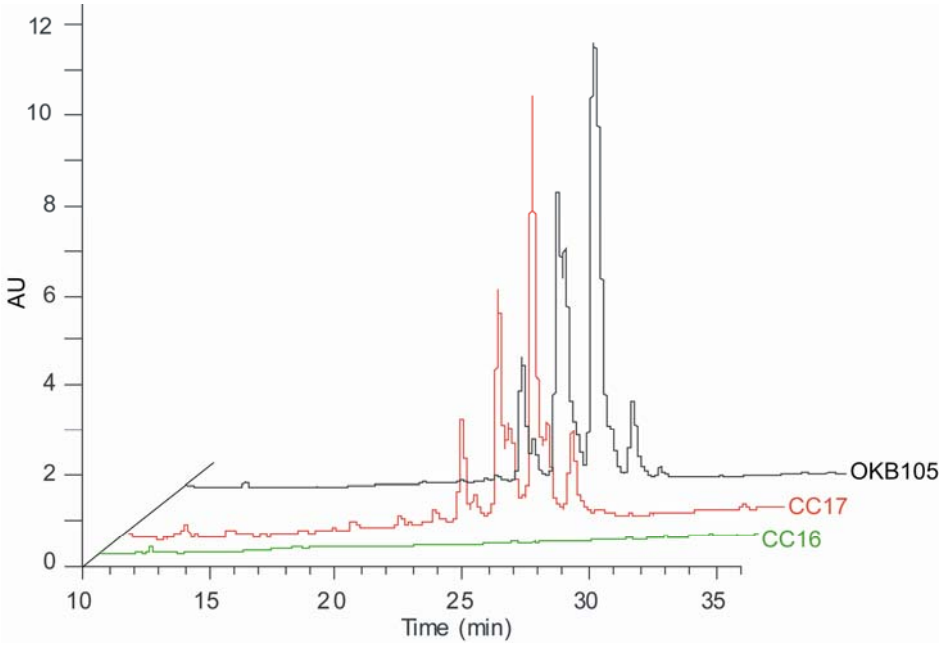


Fig. 37 HPLC analysis of butanolic extracts derived from *B.subtilis* OKB 105, CC17 and CC16. As shown, mutant CC17 (in red) is able to produce surfactin at a level comparable to the wild type strain OKB105 (in black) . In contrast, no surfactin was detected in the butanolic extract of CC16, whose *srfA* operon was interrupted by introduction of a *mls* resistance gene.

As shown in Fig. 37 the HPLC chromatogram of CC17 and wild type OKB105 differ only slightly in the dimensions of the peaks: main component in both cases was a surfactin derivative with a molecular mass m/z = 1022,7, carrying a fatty acid chain with 13 C atoms.

In order to verify whether the His-tag was successfully fused to the C-terminus of SrfA-A, *B. subtilis* CC17 was grown in 3 ml SpIII medium and the crude cell extract analyzed – after SDS-PAGE separation – by Western blotting. Immuno-detection with HRP-conjugated anti His-tag Abs revealed that a specific signal at a size of ~400 kDa could be only obtained for the mutant strain, carrying a His-tag fused to SrfA-A, but not for the wild-type (Fig. 38).

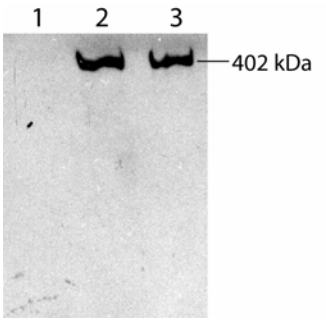


Fig. 38 Western blotting and immunodetection of SrfA-A-His. Crude extracts of the wild type OKB105 and CC17 were analyzed by Western blotting, using HRP-conjugated anti- His tag Ab. 10 µl of OKB 105 crude extract were applied to lane 1; 8 µl and 12 µl of CC17 were loaded on lanes 2 and 3 respectively.

6.2.4 SrfA-A-His purification.

The *B. subtilis* strain CC17 was grown in 2 l SpIII medium. The cleared supernatant after cell lysis was applied to a Ni-NTA affinity column in order to purify the SrfA-A-His protein. Application of a linear gradient of increasing concentrations of imidazole (7,5 to 125 mM) led to the elution of the bound SrfA-A-His protein. As revealed by the SDS-PAGE analysis (Fig. 39A), highest amounts of protein could be observed after 10-14 min in the elution fractions no. 4 - no. 6 (Fig. 39A). The elution fractions no. 4 to 9, containing the purified SrfA-A-His protein, were combined and dialysed against assay buffer, in order to remove imidazole, which could interfere with the protein's activity and stability. The dialysis led to a significant dilution/lost of protein, so that a reliable determination of the actual SrfA-A concentration was not possible. Consequently, for the subsequent biochemical characterization, as much protein as possible was used in each assay. In order to assess the activity of the SrfA-A adenylation domains, amino acid-dependent ATP-PP_i exchange reactions were carried out (see 4.2.6). Since SrfA-A consists of one glutamate- (module 1) and two leucine-activating modules (modules 2 and 3) a higher exchange activity in the presence of leucine was expected. As shown in Fig. 39B, leucine indeed revealed the highest exchange activity, while glutamate gave an about 7-fold lower, which was, however, clearly higher than the negative control. This outcome was not unexpected, since earlier studies frequently revealed a rather inefficient *in vitro* activation of acid amino acid [Schneider, 1996] [Ishihara and Shimura, 1974]. In order to test for the activity of the PCP domains, covalent loading assays was performed in the presence of radiolabeled substrate amino acid, ATP and Mg^{2+}. The corresponding thiolation reactions can only take place, when NRPS PCP domains had been converted beforehand in the active, phospopantetheinylated holo-form.

The amount of radioactively labelled amino acid incorporated was taken as indirect measurement of the PCP domain activity. A specific covalent loading of [^{3}H]-leucine was detected (450 pM), whereas only very weak incorporation of the cognate substrate [^{14}C] glutamate could be measured (0,9 pM) (Fig. 39 C). Negative controls were performed in absence of ATP. Still, the outcome of the loading assays clearly indicated that SrfA-A-His had been produced as functional holo-enzyme by its native host *B. subtilis*, this way rendering unnecessary an *in vitro* apo-to-holo-conversion of the NRPS protein after purification.

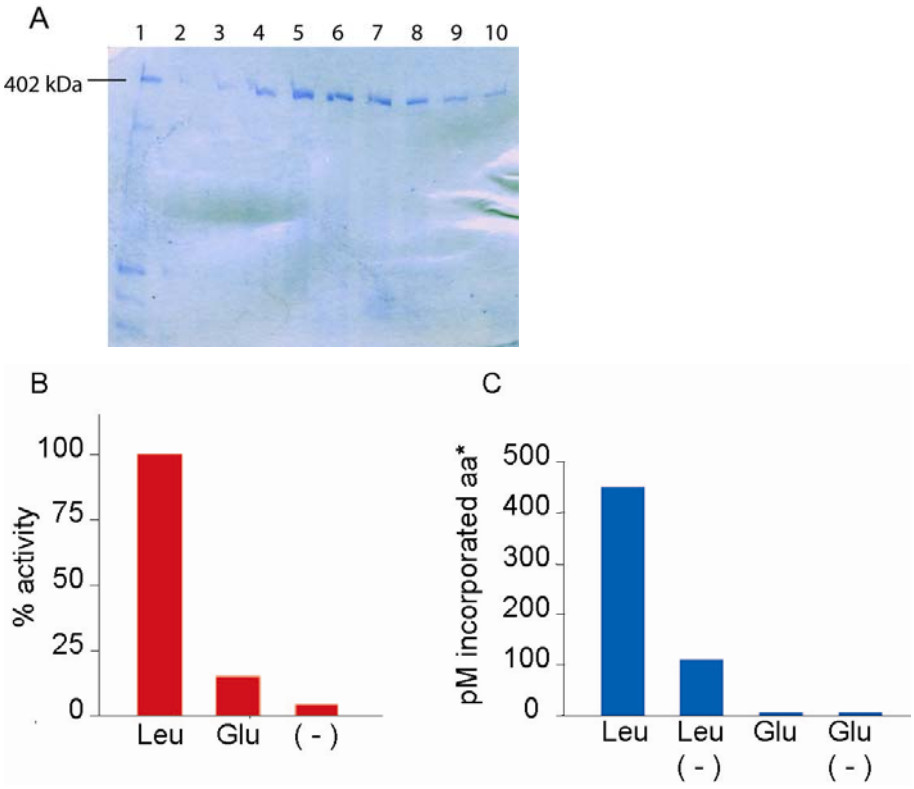


Fig. 39 Purifiction and biochemical characterization of SrfA-A-His
(A) Coomassie-stained SDS-polyacrylamide gel. (4% stacking gel, 5% running gel). Lane 1: Crude cell extract of the protein SrfA-A (402 kDa) used as a control for the size of the purified protein. Lane 2 to 10: SrfA-A-His elution fractions no. 1 to no. 9. **(B**) ATP-PPi exchange reactions. The highest activity detected was set to 100%. The negative control was carried out in absence of substrate aa. **(C)** Covalent loading assay. Negative controls were performed with radioactively labelled amino acids, but in the absence of ATP.

6.3 Immunoaffinity chromatography using polyol-responsive monoclonal antibodies.

By the utilization of Ni-NTA affinity chromatography, the tri-modular 402 kDa-protein SrfA-A-His could be purified from crude extracts of *B. subtilis* CC17 in its active, posttranslational modified holo-form, as well as and in amounts sufficient for subsequent biochemical characterizations. However, Ni-NTA chromatography is considered to be a relatively harsh method for protein purification, making it particularly harmful for the purification of large, fragile enzymes. Furthermore, the high salt concentrations used (300-500 mM sodium chloride) actually do suppress protein-protein interactions, preventing a possible co-purification of associated proteins or entire multi-enzymatic complexes.

Recently, Burgess and co-workers developed a powerful and gentle method for the efficient purification of functional multi-enzyme complexes [Burgess and Thompson, 2002]. Using polyol-responsive monoclonal antibodies (PR-mAbs) for immunoaffinity chromatography,

they were able to purify entire RNA-polymerase complexes, including the – as is known - weakly associated sigma factors [Bergendahl et al., 2003].

Given the very strong and specific interaction between antigen and antibody, immunoaffinity chromatography has been recognized and used for many years as a powerful and suitable method for protein purification.

However, the binding between antigen and antibody is usually that strong, that it can only be destroyed under denaturating conditions, using either extreme pH or strong detergents. Burgess and co-workers, however, found that some antibodies are able to release the bound antigen under gently, non-denaturating conditions, in the presence of low-molecular-weight polyhydroxylated compounds. Thus, these so-called polyol-responsive monoclonal antibodies (PR-mAbs) represent ideal tools for the purification of multi-enzymatic complexes (seeFig. 11). One polyol-responsive antibody characterized by Burgess and co-workers was the PR-mAb NT73, specifically recognizing the epitope SLAELLNAGLGGS located at the C-terminal region of the β´-subunit of the *E. coli* RNA polymerase core enzyme. This tag has been successfully fused to different proteins, in order to mediate their immunoaffinity purification, using PR-mAb NT73. In order to exploit the potential of PR-mAbs for the gentle purification of SrfA-A (and possibly the entire surfactin biosynthesis complex), the epitope tag SLAELLNAGLGGS had to be fused likewise to the C-terminus of the trimodular NRPS SrfA-A.

6.3.1 Fusion of the epitope tag at the C-terminus of SrfA-A

In order to append the epitope tag SLAELLNAGLGGS to the C-terminus of SrfA-A, essentially the same strategy already used for the fusion of the Hexahistidine tag was adopted. Consequently, the first step, consisting in the interruption of the *srfA* operon, had been already accomplished by preparation of the disruption mutant CC16 (see 6.2.2). For the second step, the reconstitution plasmid pCC42 was constructed, harboring the 1,5 kb homologous region *3'-srfA-A*, fused in frame to the coding region of the epitope tag SLAELLNAGLGGS, followed by the 1,4 kb *5'-srfA-B* fragment. The subsequent transformation of the disruption mutant CC16 was carried out by congression, using the *Xho*I/*Xba*I-linearized plasmid pCC42 along with the self-replicable helper vector pCm::Tc [Steinmetz and Richter, 1994]. As described previously, the transformants were first selected on DSM agar, supplemented with 20 µg/ml tetracycline and 5 µg/ml chloramphenicol, and subsequently screened for MLS sensitivity.

In the previous experiment for the construction of *srfA-A::his*, one out of 300 transformants (0,3%) revealed the expected phenotype, and given the similarities, a comparable quote was expected for the construction of *srfA-A::epi*.

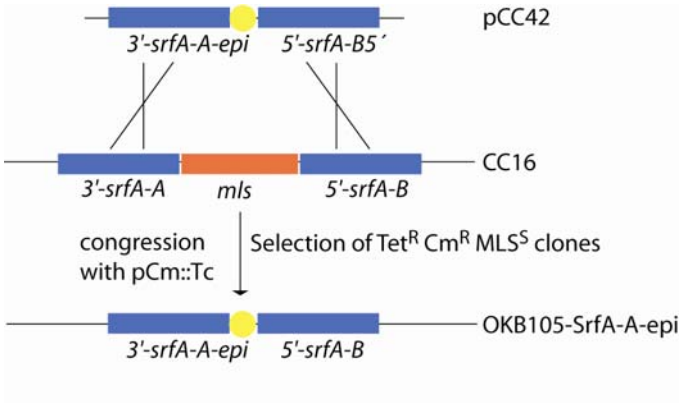


Fig. 40 Construction of OKB105-SrfA-A-epi by congression.
Double-crossover homologous recombination and reconstitution of the *srfA* operon, using the plasmid pCC42**.**

Surprisingly, even after having screened more than 9,000 Tet^R Cm^R transformants, none of them revealed a MLS-sensitive phenotype. In order to rule out the possibility of a feasible toxicity of the epitope tag, an alternative strategy was elaborated for the fusion of the epitope tag to SrfA-A.

For this approach, the plasmid pCC43 was constructed, differing from pCC42 by the insertion of a *kan* resistance marker into the 5'-region of fragment *5'-srfA-B*. This organization actually prevents the reconstitution of the *srfA* operon, but leads to the desired fusion of the epitope tag onto the chromosomal copy of *srfA-A*, and – most importantly – allows for positive selection of the recombinant clones after homologous recombination (Fig. 41A). After transformation of *B. subitilis* CC16 with the *Xho*I/*Xba*I-linearized plasmid pCC43, the transformants were again first selected on DSM agar, containing 10 µg/ml kanamycin and 5 µg/ml chloramphenicol, and Cm^R/Km^R clones then were screened for MLS sensitivity. Six out of 7 transformants tested (termed CC44) showed the desired phenotype (Cm^R Km^R MLS^S), whose genetic integrity was verified by PCR analysis, using the oligonucleotides Srf_42F, Srf_43-R (data not shown). Subsequently, six clones of the obtained strain CC44 clone 1-6 were analyzed for the production of the protein SrfA-A-epitope (SrfA-A-epi).

As shown in Fig. 41B, all analyzed clones were apparently able to produce SrfA-A-epi.

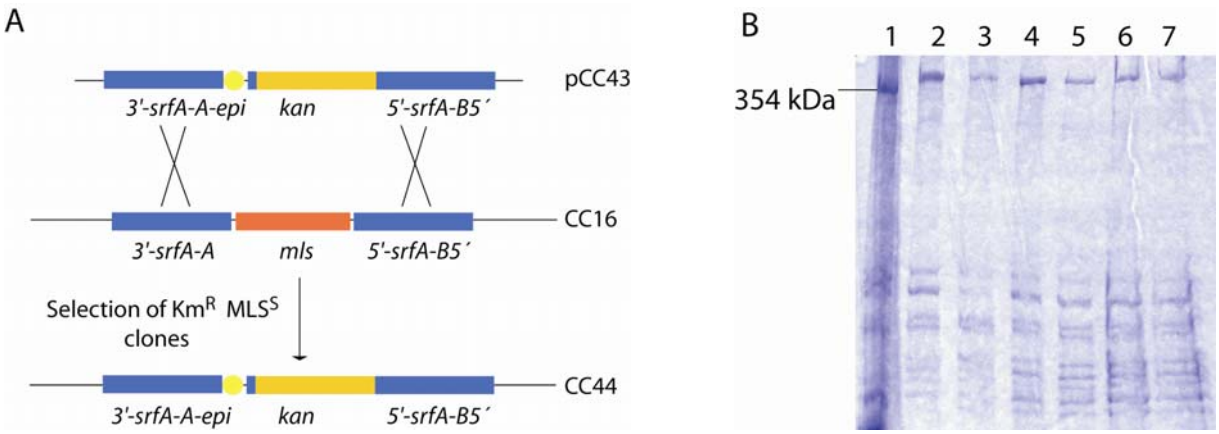


Fig. 41 Construction of the *B. subtilis* strain CC44
(A) Integration of pCC43 by homologous recombination into the chromosome of *B. subtilis* CC16, followed by selection of Cm^R, Km^R and MLS^S transformants, led to the isolation of the CC44.
(B) Coomassie-stained SDS-polyacrylamide gel. Lane: Size marker (354 kDa-protein. Lanes 2 to 7: crude cell extracts after test expression of six clones of *B. subtilis* CC44.

6.3.2 SrfA-A-epi purification by immunoaffinity chromatography

In order to test for the successful fusion of the epitope tag onto SrfA-A, *B. subtilis* CC44 (clone 1) was grown in 2 l SpIII medium. Following cell lysis, the cleared cell crude extract was applied to a NT73-sepharose column, and the bound protein was eluted using increasing concentrations of polyol. Subsequently, the different fractions obtained were analyzed by SDS-PAGE. However, since, only a low amount of SrfA-A-epi could be detected in the elution fractions after Coomassie Blue-staining of the acrylamide gel (data not shown), a Western blot analysis was carried out, utilizing PR-mAb NT73 as primary and a HRP-conjugated anti-mouse Ab as secondary antibody.

Within the western blot analysis, specific signals for a protein of the expected size of ~400 kDa (see Fig. 42) could be observed in the crude cell extract, in the flow-through and in the elution fractions. These results verify the fusion of the epitope-tag to SrfA-A, as well as the tight binding of the fusion protein to the PR-mAB NT73, since no SrfA-A-epi protein was observed within the wash fractions. The analysis also revealed a slight degradation of the SrfA-A-epi protein (Fig. 42 lanes 1 and 2).

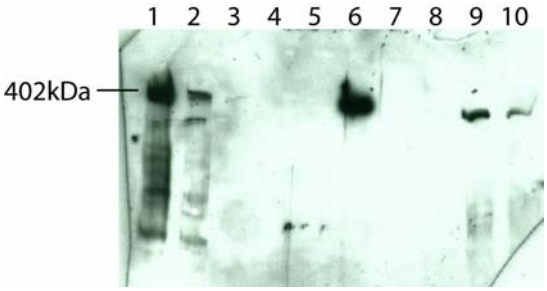


Fig. 42 Western-blot analysis of the immuno-affinity purification of SrfA-A-epi. Lane 1: supernatant obtained from ultracentrifugation of the crude extract. Lane 2: flow-through after sample application. Lane 3, 4, 5, wash fractions. Lane 6: crude extract before French press Lane 7 to 9: elution fractions no. 1 to no. 4.

In order to improve the yield of purified protein, the soluble fraction of an analogously prepared crude cell extract was batch-incubated under gentle shaking with 4 ml of the NT73-sepharose immunoaffinity resin at RT for 1 h. Subsequently, the resin was packed into a FPLC-column, and protein purification carried out as described (see 4.2.5). As shown in Fig. 43, this batch-procedure yielded significantly higher quantities of SrfA-A-epi in the elution fractions, which were sufficient to carry out biochemical characterization of SrfA-A-epi. Hereby, the activity and selectivity of the adenylation domains were characterized via amino acid-dependent ATP-PP_i exchange reaction (4.2.6). In agreement with the previous assay conducted with SrfA-A-His (6.2.4), SrfA-A-epi was found to selectively activate leucine (activity set to 100%), while only a weak activation for the second cognate substrate, glutamate, (1%) could be measured (see Fig. 43).

The following covalent loading assay was carried out to determine the activity of the SrfA-A-epi PCP domains. As described previously, the amount of labelled amino acid incorporated represents a direct measure of the activity of the corresponding PCP domain. A specific thioester-binding was observed for [^{3}H]-leucine (143 pM), whereas only marginal incorporation of [^{14}C]-glutamate (1 pM) could be detected. Controls were conducted separately with each of the radioactively labelled substrates in absence of ATP.

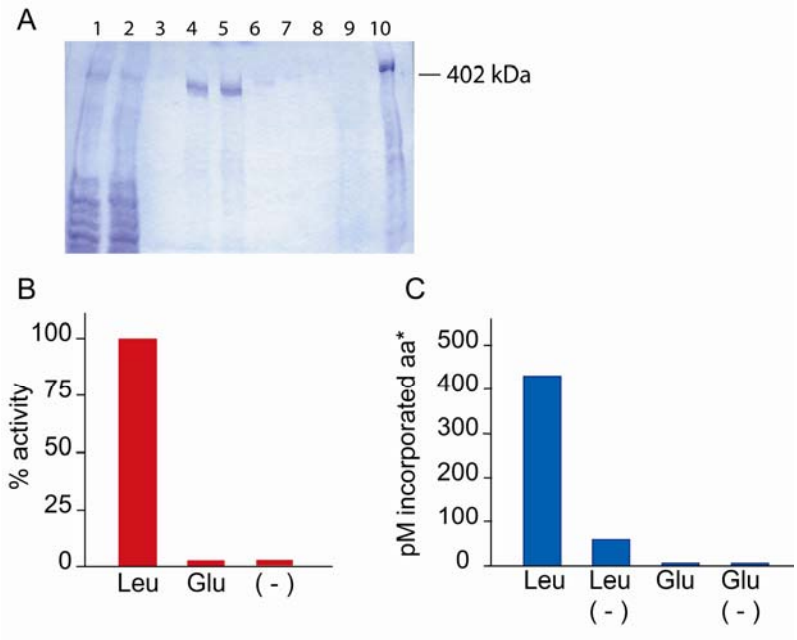


Fig. 43 Purifiction and biochemical characterization of SrfA-A-epi
(A) Coomassie-stained SDS-polyacrylamide gel of the protein purification of SrfA-A-epi. Lane 1: Soluble fraction of the crude extract. Lane 2: flow-through. Lane 3 to 9: elution fractions 2; 3; 4; 5, 6; 7; 8. Lane 10: SrfA-A-epi crude extract. **(B)** ATP-PP_i exchange assay of purified SrfA-A-epi (elution fraction 4). The highest activity, measured for leucine, was set to 100%.**(C)** Covalent loading assay of purified SrfA-A-epi. The assay revealed the formation of a specific thioester bond in case of [^{3}H]-leucine.

6.3.3 Reconstitution of the surfactin biosynthetic complex

Although the reason for the failure of attempt of fusing the epitope tag onto SrfA-A still remains unclear (see 6.3.1), the previous experiment unequivocally demonstrated that it was not due to a possible toxicity of the epitope. The experiment also demonstrated that SrfA-A-epi could be purified in active form by immunoaffinity chromatography utilizing the PR-mAb NT73. Actual goal of this whole study was, however, to attempt purification of the entire surfactin biosynthetic complex by immunoaffinity chromatography. In order to achieve this goal, fusion of the epitope-tag encoding gene fragment to the 3'-terminus of *srfA-A* had to be accomplished under reconstitution of surfactin biosynthesis.

Since this goal could not be achieved using pCC42 (see 6.3.1), a slightly different approach – using the novel reconstitution plasmid pCC53 – was adopted. This plasmid represents a derivative of pCC42, carrying the fusion of the homologous region *3'-srfA-A* to the epitope coding region directly followed by the *kan* resistance marker to allow for positive selection of plasmid integration. In continuation, the 500 bp-comprising *srfA*-promotor region was cloned directly in front of the 3' homologous region *5'-srfA-B* to restart expression of the remaining surfactin biosynthetic genes. *B. subtilis* CC16 was transformed with *Fsp*I/*Xho*I-linearized plasmid DNA of pCC53 and transformants were selected for MLS^S and Km^R (Fig. 44). The genotype verification of the five obtained mutants of *B. subitlis* strain CC57 was carried out as described before via PCR using the oligonucleotides Srf_42F and Srf_44R (see appendix).

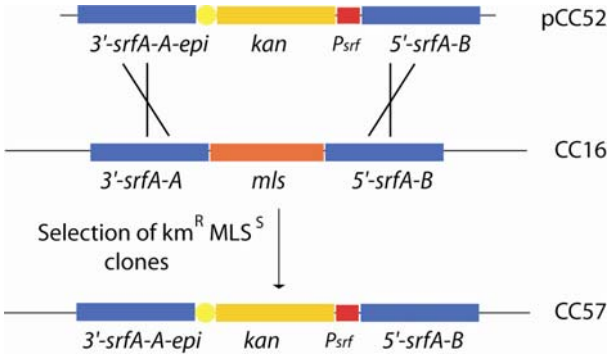


Fig. 44 Construction of the *B. subtilis* mutant CC57. The integration by homologous recombination of the linearized plasmid pCC52 into the chromosome of the disruption mutant CC16 generated the mutant CC57.

In order to verify the reconstitution of the surfactin biosynthesis, the wild-type strain *B. subitlis* OKB105, the parental strain CC16, the five clones of CC57 were grown in SpIII-medium. Subsequently, the cultured broths were butanol-extracted and analyzed by HPLC-MS.

As shown in Fig. 45, the HPLC chromatograms of mutant CC57 and the parental strain OKB105 did not reveal any relevant differences, both showing the characteristic HPLC profile of surfactin.

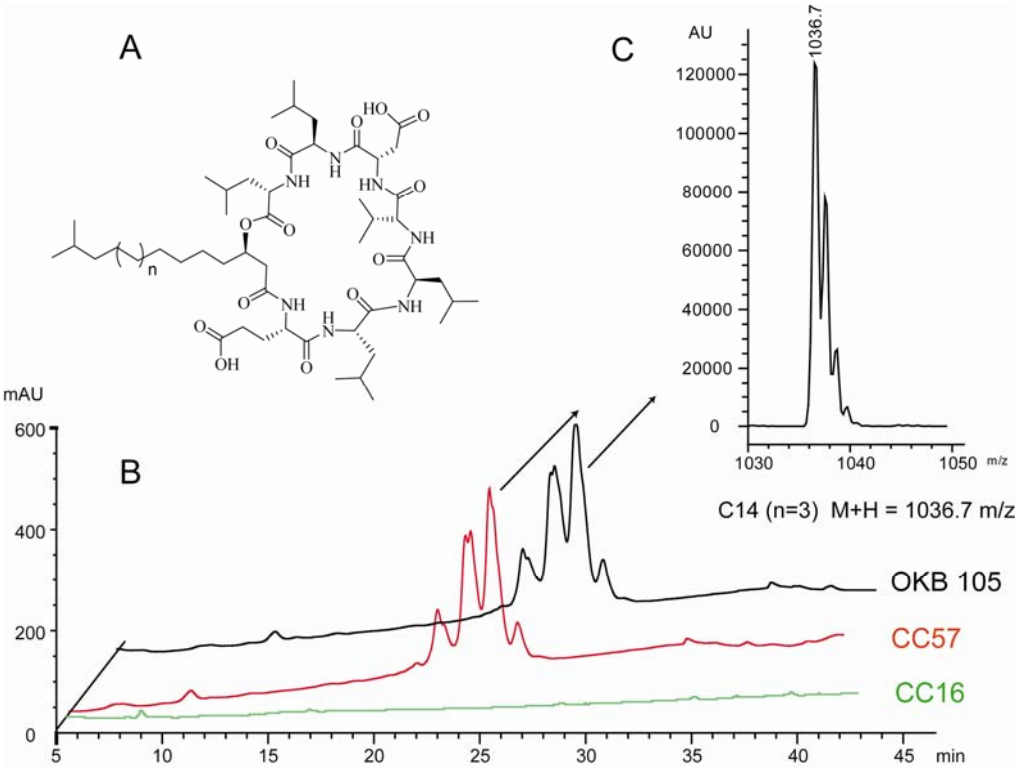


Fig. 45 HPLC-MS analysis of mutant CC 57
(A) Structure of surfactin. **(B)** HPLC diagram showing the mutant CC57 (in red) in comparison with the wild type OKB 105 (in dark) and CC64 (in green). **(C)** Surfactin detection by MS analysis of the peak eluted after 25,8 min.

Quantification of surfactin was carried out as described in section 5.4.2, page 61. The amount of surfactin produced by the wild-type was set to 100%. As shown in Fig. 46, the mutant CC57 produced almost wild-type levels of the lipopeptide surfactin (about 97%), indicating that the epitope tag did not had a negative influence on surfactin production, as previously suspected. Hence CC57 was selected for further investigations.

After having shown the successful reconstitution of surfactin biosynthesis, *B. subtilis* CC57 was next exploited for purification of SrfA-A-epi.

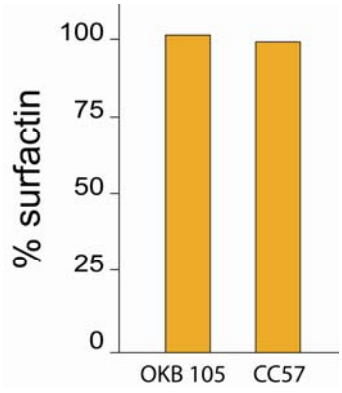


Fig. 46 Quantification of the amount of surfactin produced by OKB105 and CC57 mutants
Basing on the UV signal obtained after HPLC assay the amount of surfactin produced by the wild type *B. subtilis* OKB105 and by the mutant CC57 was quantified. As shown in the grafic CC57 produces an amount of surfactin comparable to the wild type.

In order to allow for a possible co-purification of associated proteins, like NRPSs, the acyl-transferase enzyme and/or the external type II thioesterase (TE II), the purification protocol was slightly modified compared to earlier attempts (see 4.2.5). For example, to avoid the early elution of weakly associated proteins, the ionic-strength of the chromatography buffer W2 was lowered (300 mM NaCl) and immuno-affinity column was washed only with two – instead of four – column volumes (CV) of washing buffer. The highest amount of SrfA-A-epi was detected in elution fractions no. 3 and no. 4. Interestingly, both fractions also contained three additional proteins of molecular weights between 300 and 60 kDa (Fig. 47A).

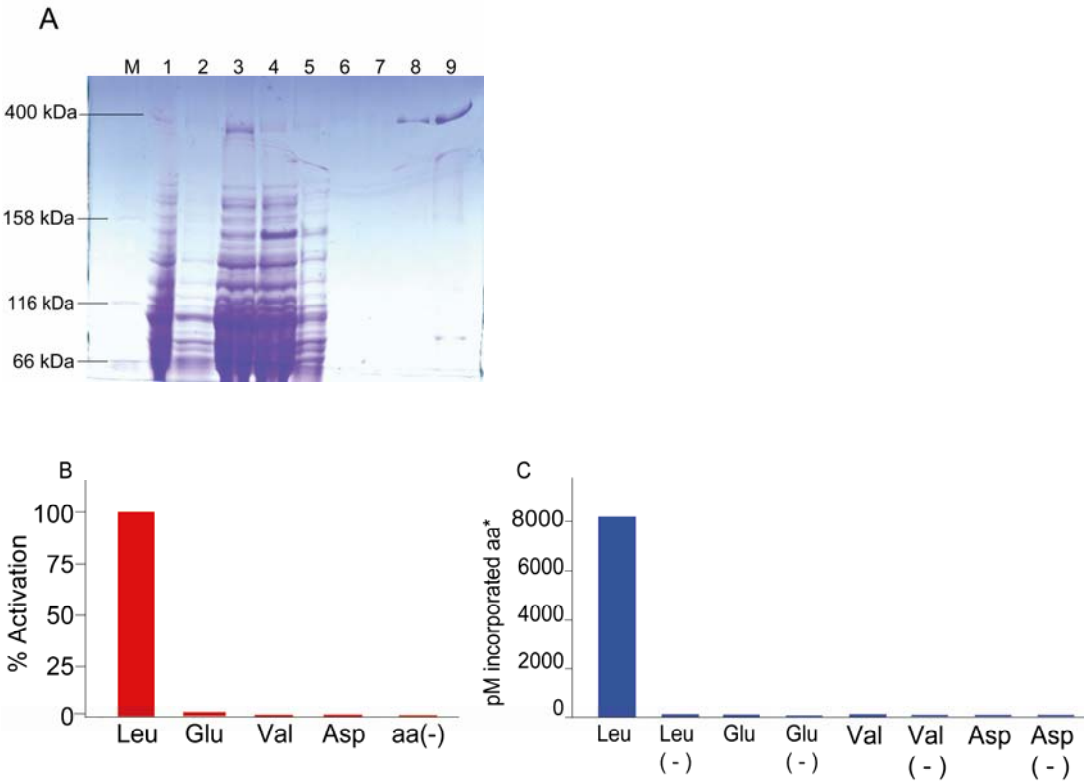


Fig. 47 Immunoaffinity chromatography and biochemical characterization of SrfA-A-epi purified from CC57. (A). Coomassie stained SDS-polyacrylamide NuPAGE® Bis-Tris gel (acrylamide concentration 4% to 12%), showing SrfA-A-epi purification from the crude cell extract of the mutant CC57. Lane M: protein marker; lane 1: crude extract; lanes 2 and 3: pellet and supernatant after cell lysis, respectively. Lane 4: flow through after batch incubation. Lanes 5, 6 and 7: wash fractions 1, 3 and 6 respectively; lanes 8 and 9: elution fractions number 3 and 4, respectively**. (B)** ATP-PP_i exchange assay. All cognate amino acids of the surfactin biosynthetic machinery were used as substrates **(C)** Covalent loading assay. Specific covalent binding could be observed for leucine, but not for the other amino acid substrates**.**

Elution fraction no. 3 (lane 8 of the acrylamide gel in Fig. 47A) was used for biochemical characterization of the purified proteins using the ATP-PP_i exchange and the covalent loading assay. In contrast to the substrate amino acid leucine, which is activated by all three surfactin synthetases, the amino acids aspartate and valine are specific, cognate substrates of the NRPS SrfA-B. Since the activation of those two amino acids could provide hints about the possible co-purification of SrfA-B, these latter two amino acids were used – in addition to SrfA-A-epi substrates leucine and glutamate – in amino acid-dependent ATP-PP_i exchange assays.

However, as shown in Fig. 47B, only a specific activation of leucine could be detected. This result was confirmed by the following covalent loading assay, performed to determine the activity of the PCP domain. [^{3}H]-leucine could be specifically bound from the Ppant cofactor of PCP. However, no significant loading was measured for [^{14}C]-glutamate, [^{14}C]-aspartate and [^{14}C]-valine (Fig. 47C)

Despite the fact that the previous biochemical characterization of SrfA-A-epi did not reveal any direct evidence for a feasible co-purification of associated biosynthetic enzymes, the elution fractions with the highest amount of SrfA-A-epi protein were more thoroughly investigated. This analysis was particularly encouraged by the three smaller proteins in elution fractions no. 3 and no. 4, which might be involved in surfactin biosynthesis and were not observed in previous purifications (see elution fractions no. 3 and no. 4). To this purpose, elution fractions no. 3 and no. 4 were dialyzed overnight against TE buffer, containing 15% sucrose, and subsequently 5-fold concentrated, using an Amicon cell. Different amounts of the concentrated samples were then analyzed on a gradient NuPAGE® Bis-Tris gel (acrylamide concentration: 4% to 12%) and on a 10% acrylamide gel. In the event of a successful purification of the surfactin multi-enzyme complex, the three NRPSs SrfA-A, SrfA-B and SrfA-C should be detectable in the gradient gel, whereas the smaller proteins i.e. Sfp (28 kDa), and the hypothetic acyl transferase (about 40 kDa), should be separated and found in the 10% acrylamide gel. As shown in Fig. 48, two high molecular weight proteins of about 400 kDa represented the most prominent proteins detectable on the gradient gel. Given their sizes, the corresponding double-band could be attributed to the expected NRPSs SrfA-A and SrfA-B. Coomassie-staining also revealed the presence of several minor proteins, with molecular weights ranging between 400 and 40 kDa.

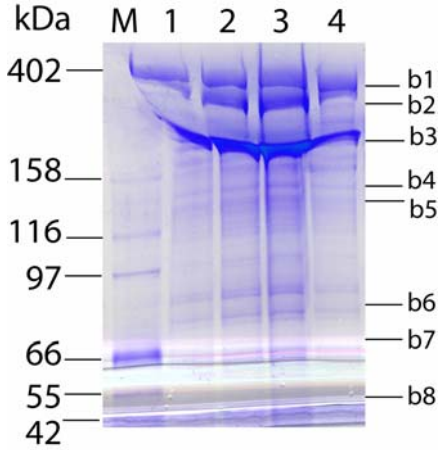


Fig. 48 SDS-PAGE of SrfA-A-epi purified from the mutant CC57. Elution fraction 3 and 4 obtained from SrfA-A-epi purification (Fig.13) were concentrated 5-fold and loaded on a NuPAGE® Bis-Tris Gels (acrylamide concentration 4% to 12%). M: protein marker. Lanes 1, and 2 contain 10 µl, 20 µl of concentrated elution fraction no. 3, respectively; lanes 4 and 5 contain 10 µl, 20 µl of concentrated elution fraction no. 4, respectively. The protein bands b1 to b8 were cut off from the gel and characterized by nano HPLC-ESI-MS/MS.

In contrast, the 10%-SDS-polyacrylamide gel only showed two proteins of about 30 and 60 kDa, respectively, in addition to the dominating double-band of SrfA-A/SrfA-B (data not shown). These two protein bands, as well as eight protein bands of the gradient gel were cut-off from both gels, digested with trypsin, and analyzed by nano-HPLC-ESI-MS/MS.

The primary sequences of the generated tryptic fragments were compared with the sequences available in public accessible databases using Mascot (Matrix science Ltd, London, UK). Interestingly, only peptides belonging to SrfA-A and SrfA-B could be detected with a score of 1.500 and 400, respectively. Although the program did not allow for a precise quantification of the relative amounts of SrfA-A and SrfA-B proteins present in the sample, the data suggested that SrfA-B might be present in low amounts when compared to SrfA-A. Presumably the amount of co-purified SrfA-B was not sufficient for reaching the minimal concentration required to carry out biochemical characterization. In fact, previous studies indicated that ~1 pM enzyme represent the minimal amount of protein required, to pick-up a detectable ATP-PP_i exchange activity [Neumüller, 2001]. The data obtained by nano-HPLC-MS/MS and primary sequence analysis were reproducible and confirmed the presence of peptides derived from the fragmentation of SrfA-A and SrfA-B. Unfortunately no additional proteins belonging to the surfactin biosynthetic complex could be detected.

7 Discussion

NRPSs represent the template and catalytic machinery for the synthesis of many pharmacologically and industrially important peptides. Their modular organization renders them suitable for genetic manipulations, aimed at generating novel “unnatural” natural products. In the past, numerous attempts for the reprogramming of NRP biosyntheses have been performed. Examples are the substitution of A-PCP minimal modules [Stachelhaus et al., 1995], the translocation of the C-terminal Te domain [de Ferra et al., 1997], the fusion and deletion of entire elongation modules [Mootz et al., 2002a; Mootz et al., 2000], and the modification of the selectivity of the A domain by directed mutagenesis [Eppelmann et al., 2002; Stachelhaus et al., 1999]. Although most of these strategies led to the biosynthesis of the predicted NRPs, the manipulations were often connected with considerable reductions in yield of synthesized products, mostly due to unproductive communication between catalytic domains within the modified multi-enzymatic system.

The recent discovery of so-called communication-mediating (COM) domains, and their determined key role in protein-protein communication opened new possibilities for the modification and reprogramming of NRPS assembly lines. These short terminal peptides, located at the C- and N-termini of donor and acceptor NRPSs respectively, were identified by means of *in vitro* studies on protein-protein interaction within the tyrocidine biosynthetic system [Hahn and Stachelhaus, 2004]. These experiments, conducted on simple di and tri-modular (hybrid) systems, demonstrated that the compatibility of a donor (COM^{D}) and an acceptor (COM^{A}) COM domain is a fundamental requirement for the productive interaction between two NRPSs. They even allow the enforcement of protein-protein communication between natural non-partner NRPSs, and the crosstalk between different systems. Incompatible COM domain pairs, in contrast, are responsible for the prevention of the futile interaction between non-partner enzymes. They can, however, also be exploited to prevent communication between natural partner NRPSs.

In this work, the potential of COM domains was exploited in order to redesign the surfactin assembly line and to create the basis for true biocombinatorial synthesis.

In the surfactin biosynthetic system of *B. subtilis*, the cognate COM domain pairs $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$ and $COM^{D}_{SrfA\text{-}B3}$/$COM^{A}_{SrfA\text{-}C}$ facilitate the selective interactions between SrfA-A and SrfA-B, as well as SrfA-B and SrfA-C, respectively. Within the scope of the presented *in vivo* study, the natural cognate COM domain pair $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$ was exchanged against alternative cognate, non-cognate and mis-cognate COM domain pairs derived from the surfactin and from the heterologous tyrocidine biosynthetic system. The

consequences of these COM domain swaps were investigated by analyzing the cultured broth of the corresponding *B. subtilis* mutants with regard to lipopeptide production.

Another aim of this work was to set up a fast and robust method for the purification of large NRPSs and NRPS biosynthetic complexes in active holo-form. In this context, the well-characterized surfactin biosynthetic system was chosen again as a model system. By using the natural producer strain *B. subtilis*, problems like incorrect folding, insolubility or degradation of the proteins, which are often encountered during the production of large di- and tri-modular NRPSs in heterologous hosts, could be circumvented. Moreover, in contrast to *E. coli*, the surfactin producer strain *B. subtilis* is equipped with the Ppant transferase Sfp, required for the post-translational modification of NRPSs in their active holo-form, offering an additional advantage over the utilization of the heterologous host *E. coli*. In order to allow for a single-step purification, the coding sequences of two different affinity-tags (a hexahistidine tag, as well as a 13aa-epitope tag) were fused to the 3'-end of the *srfA-A* gene within the chromosome of *B. subtilis*. Subsequently, affinity purification of the tagged, tri-modular NRPS SrfA-A – and potentially co-purification of associated proteins of the surfactin biosynthetic complex – was attempted using either i) IMAC or ii) immunoaffinity chromatography using PR-mAb.

7.1 Elaboration of a strategy for COM domain swapping

The first evidence for the existence of COM domains was obtained by the *in vitro* investigation of protein-protein communication in the tyrocidine biosynthetic system, consisting of three NRPSs: TycA, TycB and TycC. Both donor enzymes of this NRPS complex, TycA and TycB, harbour an E domain at their C-terminal end. Hence, the communication between donor and acceptor enzymes takes place at the interface between E and C domain. Serial deletions at the C-terminal end of TycA were made and analyzed for their effect on protein-protein communication between TycA and TycB. These experiments revealed that already the deletion of six amino acid residues was sufficient to destroy the communication between TycA and TycB [Hahn and Stachelhaus, 2004]. In order to characterize the amino acid sequences presumably involved in the interactions of other partner enzymes, the primary sequences of the E-C transition regions of NRPSs derived from different biosynthetic systems were compared.

Fig. 49 shows a sequence alignment of 15 E-C transition regions from eight different NRPS systems. Interestingly, it was found that both COM^D and COM^A domains possess only low overall conservation.

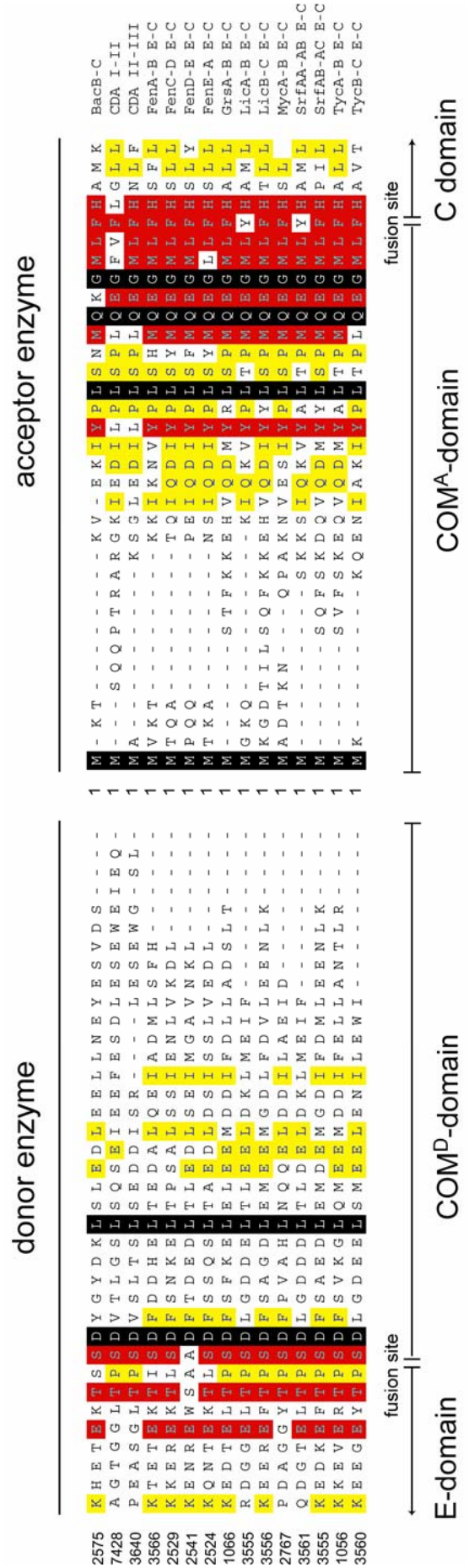


Fig. 49 Sequence alignment of inter-molecular E-C transition regions
Alignment of 15 E-C transition regions, derived from eight different biosynthetic systems (bacitracin, CDA, fengicin, gramicidin S, lichenisin, mycosubtilisin, surfactin and tyrocidine). COM^D and COM^A consist of variable amino acid sequences, bordered by conserved core motifs, which were chosen as fusion sites for the COM domain swapping experiments described.

However, their variable amino acid sequences are bordered by two highly conserved core motifs, TPSD and L(T/S)P(M/L)QEG, localized at the transition between COM^D and E domain, as well as COM^A and C domain, respectively (Fig. 49).

These core motifs were used as fusion sites for the COM domain swapping experiments conducted in this study (Fig. 50).

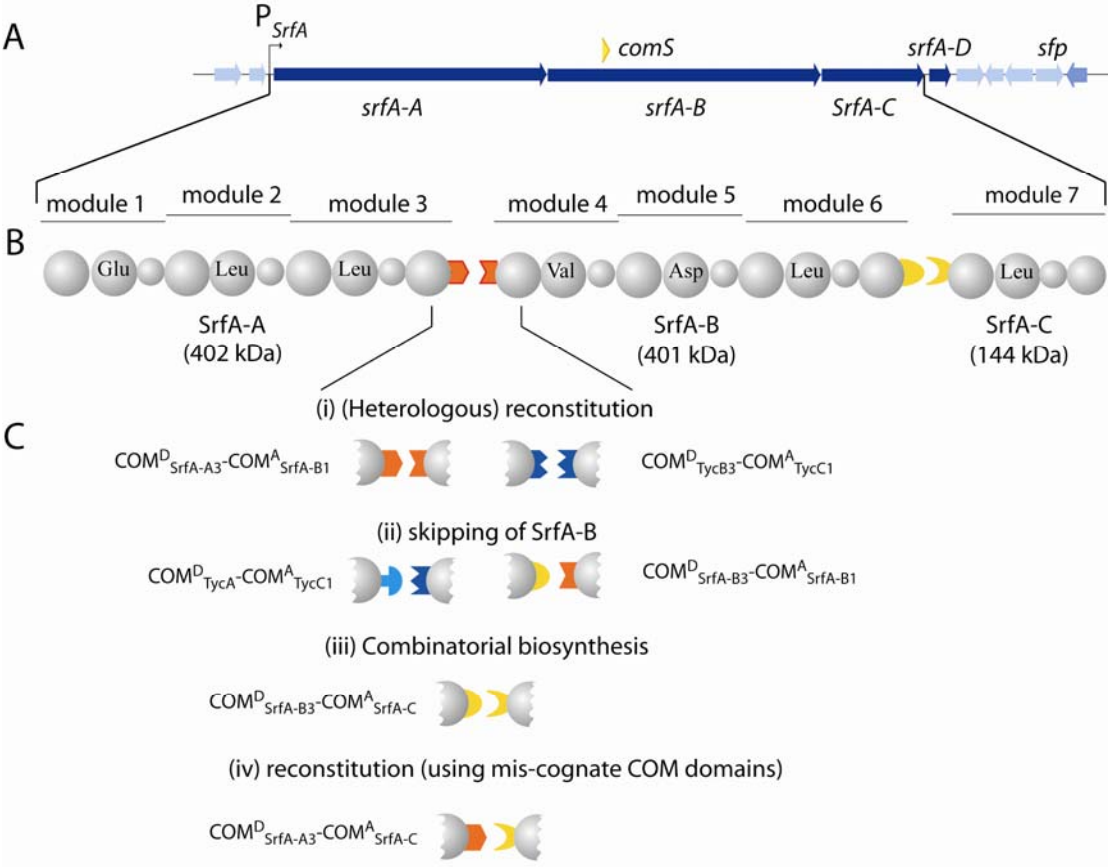


Fig. 50 *In vivo* COM domain swapping within the surfactin biosynthetic complex
(A) The surfactin biosynthetic assembly line consists of the three synthetases SrfA-B, SrfA-B and SrfA-C. The COM domain pairs, facilitating the selective protein-protein interactions within the system, are shown in red and yellow, respectively. **(B)** Cognate, non-cognate and mis-cognate COM domain pairs, derived from the tyrocidine and surfactin biosynthetic systems, were used for the substitution of the native cognate COM domain pair $COM^D_{SrfA-A3}/COM^A_{SrfA-B1}$.

The envisioned COM domain swaps should be accomplished under conservation of the primary sequence of the the N- and C-terminally flanking core motifs. This goal was achieved by constructing the integration vector pCC78a, which carried the required 5' and 3'-homologous regions for the marker exchange homologous recombination. The vector also contained the engineered restrictions sites *Avr*II and *Acc*65I, located precisely on the coding bases of peptides "PS" and "LYH" within the mentioned core motifs (see Fig. 51). In order to maintain the primary sequence and, concurrently, facilitate a targeted cloning, the compatible restrictions sites *Nhe*I and *Bsr*GI were engineered onto the termini of the encoding gene

fragments of the desired COM domain pairs. As shown in Fig. 51, ligation of the coding sequence of the cognate COM domain pair $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$ into the integration vector pCC78a caused the introduction of three silent point mutations: two in the region of the upstream core motif "TPSD", and one in the region of the downstream core motif "QEGMLYH".

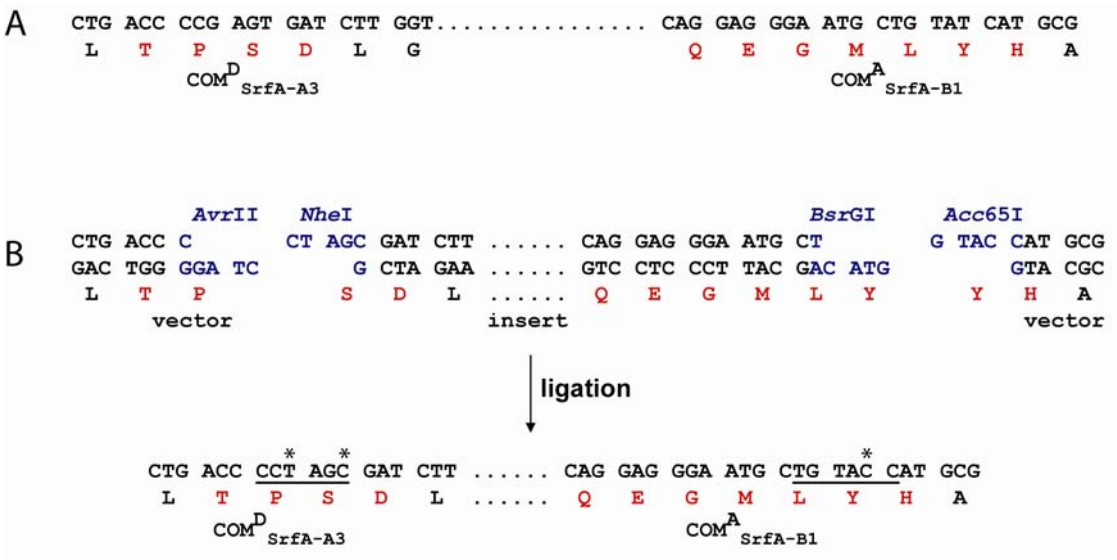


Fig. 51 Cloning strategy for COM domain swapping
(A) Conserved core motifs flanking the COM domain pair $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$. The conserved amino acid residues are marked in red. **(B)** The recognition sequences of the compatible restriction sites *Nhe*I/*Avr*II and *Bsr*GI/*Acc*65I were engineered onto the termini of vector and insert (in blue). The (re-)cloning of the insert occurs under maintenance of the original primary sequence. The fusion sites are underlined, and the three silent point mutations are marked by asterisks.

In a proof-of-concept, the validity of the described cloning strategy was investigated by the reconstitution of surfactin biosynthetic gene cluster in the chromosome of *B. subtilis* CC84 (genotype: *srfA-A::*$COM^{D}_{SrfA\text{-}A3}$-$COM^{A}_{SrfA\text{-}B1}$*::srfA-B*). The subsequent analysis revealed that this strain was in fact able to produce almost wild type levels of the lipoheptapeptide. This experiment not only validated the cloning strategy, which was subsequently used for swapping of all other COM domains pairs, but also showed that the generated *B. subtilis* mutant CC84 was able to produce surfactin without a drop of productivity. This represented a substantial improvement as opposed to previous experiments, in which already the simple reconstitution of the wild type system was always accompanied by significant reduction in product titer [Mootz et al., 2002a; Schneider et al., 1998; Stachelhaus et al., 1995].

7.2 Reconstitution of surfactin biosynthesis using heterologous COM domains

In vitro investigations provided the first evidence that COM domains maintain their functionality even when they were separated from their natural partner domains and fused to internal domains [Hahn and Stachelhaus, 2006]. For example, the donor COM domain COM^{D}_{TycA}, which is located at the C-terminus of TycA E domain in the native system, had been directly fused to the upstream located PCP domain. Notably, the obtained hybrid protein TycA::$COM^{D}\Delta$E was perfectly able to communicate with its partner enzyme TycB1 within product formation assays. A similar proof could be obtained for the TycB1 acceptor COM domain COM^{A}_{TycB1}, which was fused to the internal C domain of BacB2. The resulting hybrid NRPS was likewise able to productively interact with TycA (the natural partner of TycB1) [Hahn and Stachelhaus, 2006]. This so-called portability of COM domains is a very important feature, since it allows mobilizing COM domain from their natural context and fusing them to different domains within a biosynthetic template, thus broadening their application for NRP biosynthesis manipulation.

In the presented work, the portability of COM domains was further investigated by evaluating their *in vivo* functionality within the context of a heterologous biosynthetic complex. To this end, the COM domain pair $COM^{D}_{SrfA-A3}$/$COM^{A}_{SrfA-B1}$ of the surfactin biosynthetic operon was exchanged against the heterologous, cognate COM domain pair COM^{D}_{TycB3}/COM^{A}_{TycC1}, derived from the tyrocidine biosynthetic complex of *B. brevis* ATCC 8185. The functionality of this latter COM domain pair was verified by checking the surfactin production for the generated *B. subtilis* mutant strain CC99. The very robust product titer of 69% compared to the wild type indicated that the two hybrid proteins SrfA-A::COM^{D}_{TycB3} and COM^{A}_{TycC1}::SrfA-B could communicate. This *in vivo* experiment also unequivocally showed that COM domains conserve their functionality even in a heterologous NRPS system and in a heterologous host. This represents an interesting and important property for future applications of COM domains in combinatorial biosynthesis, in particular for the generation of hybrid biosynthetic NRPS assembly lines.

7.3 COM domains as a tool for combinatorial biosynthesis

One of the main objectives of this work was to exploit the biocombinatorial potential of COM domain, by creating *in vivo* a so-called universal communication system (UCS), in which all donor and acceptor COM domains of a given NRPS system are compatible, potentially

leading to the formation of random NRPS complexes, and the simultaneous synthesis of different peptide products.

The first UCS was realized *in vitro* and consisted of the initiation module TycA::$COM^{D}\Delta E$ (carrying a deletion of the N-terminal E domain), the elongation module COM^{A}_{TycB1}::BacB2(CAT)::COM^{D}_{TycA}, and the termination module SrfA-C [Hahn and Stachelhaus, 2006]. The resulting tri-modular system, equipped with compatible donor and acceptor COM domains (COM^{D}_{TycB1} and COM^{D}_{TycA} can both productively interact with COM^{A}_{SrfA-C}), allowed for the concurrent *in vitro* synthesis of the dipeptide D-Phe-L-Leu and the tripeptide D-Phe-L-Orn-L-Leu. Theoretically, based on the compatibility of its own COM^{A} and COM^{D} domains, the elongation module COM^{A}_{TycB1}::BacB2(CAT)::COM^{D}_{TycA} should also be able to interact with itself, giving rise to the tetrapeptide D-Phe-L-Orn-L-Orn-L-Leu. However, this latter product was not observed, presumably to due the equal (stoichiometric) amounts of initiation, elongation and termination enzyme used in this experiment.

In the presented work, an UCS should be realized *in vivo*, within the context of a natural biosynthetic assembly line. This goal was achieved by exchanging the COM domain pair $COM^{D}_{SrfA-A3}$/$COM^{A}_{SrfA-B1}$ of the surfactin biosynthetic assembly line against $COM^{D}_{SrfA-B3}$/$COM^{A}_{SrfA-C.}$ This manipulation eliminated the selectivity barrier provided by the COM domains, and allowed for the establishment of two different biosynthetic assembly lines: the first, consisting of SrfA-A/SrfA-B/SrfA-C and capable of producing surfactin; and the second, formed by SrfA-A/SrfA-C, and capable of synthesizing the truncated lipo-tetrapeptide FA-L-Glu-L-Leu-D-Leu-L-Leu. The analysis of butanolic extract derived from the corresponding, reprogrammed *B. subtilis* strain CC91 revealed the concurrent synthesis of both lipopeptide products in roughly comparable amounts, demonstrating the feasibility of *in vivo* biocombinatorial synthesis via UCS.

Given the compatibility of all donor and acceptor COM domains, the communication between two molecules of SrfA-B would be (theoretically) also feasible. This would lead to the formation of a third assembly line (SrfA-A/SrfA-B/SrfA-B/SrfA-C), eventually resulting in the synthesis of a lipodecapeptide. However, analysis of the butanolic extract obtained from both, the supernatant and the pellet fraction of *B. subtilis* CC91 crude extracts revealed no traces of lipodecapeptide. The absence of lipodecapeptide formation might be due to different factors: i) the eventual toxicity of the product, ii) the incapability of the SrfA-C Te domain to process the alternative lipodecapeptidyl-PCP substrate, iii) the hydrolysis of the unnatural reaction intermediates from the protein template as catalyzed by the cleaning enzyme TeII

[Schwarzer et al., 2002], iv) the inherent instability of the formed product, and v) the limited availability of SrfA-B in the constructed strain. The latter problem could be actually circumvented by increasing the intra-cellular concentration of SrfA-B in *B. subtilis* CC91. This could be achieved by integration of a second copy of $COM^{A}_{SrfA\text{-}B3}$*::srfA-B* at a different locus of the *B. subtilis* chromosome. This way, it should be possible to generate an excess of SrfA-B (compared to SrfA-A and SrfA-C), eventually forcing the interaction between two molecules of SrfA-B.

7.4 Reprogramming of the surfactin biosynthetic assembly line: skipping of SrfA-B.

Previous *in vitro* studies indicated that the exchange of the natural COM^{A} domain of TycB1 against COM^{A}_{TycC1} was sufficient to prevent TycB1's interaction with its natural partner enzyme TycA. Hence, the conclusion was drawn that COM^{D}_{TycA} and COM^{A}_{TycC1} are forming a non-cognate COM domain pair, preventing the futile contact between the non-partner NRPSs TycA and TycC within the context of the natural trimeric biosynthetic complex. On the other hand, it was demonstrated that compatible COM domain pairs may also enable a productive interaction between enzymes derived from different biosynthetic pathways. An example for this crosstalk was represented by the productive interaction between the non partner enzymes TycA and SrfA-C, which led to the *in vitro* synthesis of the di-peptide product D-Phe-L-Leu [Hahn and Stachelhaus, 2004]. Consequently, the two hetereologous COM domains COM^{D}_{TycA} and $COM^{A}_{SrfA\text{-}C}$ are forming a mis-cognate COM domain pair.

Taking both mentioned observations into consideration, the cognate COM domain pair $COM^{D}_{SrfA\text{-}A3}$/$COM^{A}_{SrfA\text{-}B1}$ was exchanged against the non-cognate, heterologous COM^{D}_{TycA}/COM^{A}_{TycC}. Given the incompatibility of the latter two COM domains, this should prevent the productive communication between SrfA-A and its natural partner enzyme SrfA-B. On the other hand, given the known crosstalk between COM^{D}_{TycA} and $COM^{A}_{SrfA\text{-}C}$, the exchange should concurrently enable the interaction between SrfA-A and its non-natural partner SrfA-C, this way giving rise to the formation of the shortened lipotetrapeptide (FA-LGlu-LLeu-D-Leu-L-Leu). Analysis of the corresponding *B. subtilis* mutant strain CC112 indeed revealed the intended skipping of SrfA-B, and hence production of the novel lipotetrapeptide, while no lipoheptapeptide surfactin – not even in trace amounts – could be observed. Although a precise quantification of the produced lipo-tetrapeptide was not possible, due to the lack of an authentic standard, the amount of the lipotetrapeptide was

estimated to be in the same order of magnitude (73%), as observed for the full-length lipopeptide surfactin produced by the wild type *B. subtilis* ATCC 21332.

The achieved skipping of SrfA-B further corroborated the potential of COM domains for the reprogramming of NRP biosynthetic assembly lines. In previous, state-of-the-art module swapping experiments, the about 3,2 kb-coding region of a NRPS module had to be exchanged, in order to accomplish the substitution of a single amino acid in the final product. In contrast, only a small, about 200 bp-fragment had to be exchanged for presented substitution of the coding region of a single COM domain pair, giving rise to a completely different NRPS biosynthetic template, due to the reorganisation of the protein-protein interactions. Furthermore, the described experiment also showed that the specific crosstalk between COM^{D}_{TycA} and COM^{A}_{SrfA-C}, which was previously only verified *in vitro*, is also feasible under much more stringent *in vivo* conditions.

Provided that the *in vitro* results obtained for the COM domains of the tyrocidine biosynthetic complex would also apply for the surfactin system, it was expected that the $COM^{D}_{SrfA-A3}$/$COM^{A}_{SrfA-B1}$ and $COM^{D}_{SrfA-B3}$/$COM^{A}_{SrfA-C1}$ represent cognate COM domain pairs, facilitating the productive interaction between partner-NRPSs. On the other hand, the COM domain pairs $COM^{D}_{SrfA-A3}$/COM^{A}_{SrfA-C} and $COM^{D}_{SrfA-B3}$/$COM^{A}_{SrfA-B1}$ had to be considered as non-cognate, preventing the futile contact between non-partner enzymes. Based on these assumptions it was expected that the intended skipping of SrfA-B within the surfactin biosynthetic assembly line could be also achieved by the substitution of $COM^{D}_{SrfA-A3}$/$COM^{A}_{SrfA-B1}$ against the proposed non-cognate COM domain pair $COM^{D}_{SrfA-B3}$/$COM^{A}_{SrfA-B1}$. Surprisingly, however, the generated *B. subtilis* mutant CC102 produced not only the intended lipotetrapeptide (38%), but also full-length lipoheptapeptide surfactin, even though in lower amounts (17% of wild type level). The latter result indicates that the communication between SrfA-A and SrfA-B is not abrogated, and that $COM^{D}_{SrfA-B3}$ and $COM^{A}_{SrfA-B1}$ are forming a mis-cognate, rather than the postulated non-cognate COM domain pair. In the context of the hybrid NRPS system under investigation, this mis-cognate COM domain pair allows for the productive interaction between SrfA-A and SrfA-B, and eventually formation of the lipoheptapeptide surfactin.

How can this be explained? Previous studies indicated that the crosstalk between different NRPSs could be predicted on the basis of their sequence homology. In particular, *in vitro* experiments showed that a donor enzyme carrying COM^{D}_{TycA} could communicate with non-partner enzymes harbouring the cognate COM^{A}_{TycB1}, or the mis-cognate COM^{A}_{GrsB1} or COM^{A}_{SrfA-C}. This experimental evidence finds an explanation in the high sequence homology

of the mentioned COM^A domains, which amounts to 79%. In contrast, the sequence homology between COM^A_{TycB1} (the cognate COM domain of TycA) and the non-cognate domain COM^A_{TycC} amounts to only 50%.

Remarkably, the sequence homology between the two COM^A domains of the surfactin complex: $COM^A_{SrfA\text{-}B1}$ and $COM^A_{SrfA\text{-}C}$ amounts to 63%, which could be sufficient to justify a productive interaction between $COM^D_{SrfA\text{-}B1}$ and $COM^A_{SrfA\text{-}B3}$ (see Fig. 52).

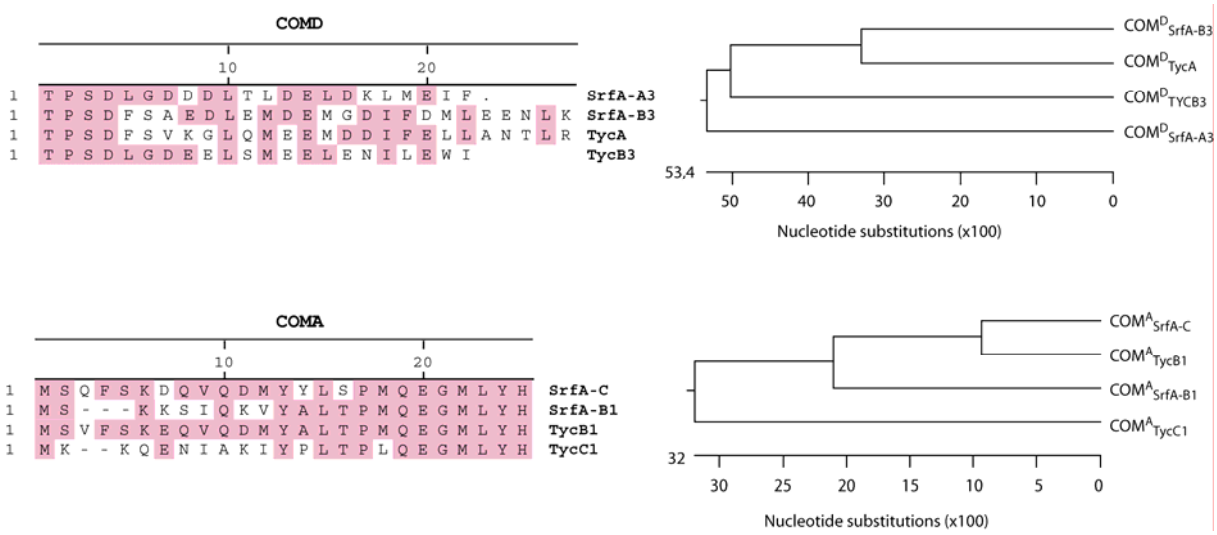


Fig. 52 Homology between COM^D and COM^A of the surfactin and tyrocidine biosynthetic systems
The proposed sequences of the donor and acceptor COM domains of the surfactin and tyrocidine biosynthetic systems were aligned using the ClustAL algorithm. On the right side, the corresponding phylogenetic trees are shown.

Additional arguments to explain the apparently productive interaction between $COM^D_{SrfA\text{-}B3}$ and $COM^A_{SrfA\text{-}B1}$ is provided by the so-called 5-residue model [Hahn and Stachelhaus, 2006].

7.5 The 5-residue model: identification of the selectivity-conferring residues of COM domains

Recently, a 5-residue-model was proposed to describe the interaction between COM^D and COM^A domains, and to predict the respective amino acid residues, which facilitate their contact. According to this model, a cognate pair of COM domains forms a leucine-zipper like motif, with COM^D and COM^A each contributing one helix (see Fig. 53). The proposed selectivity-conferring residues can be determined based on their relative location towards the highly conserved core motifs "TPSD" (COM^D) and "L(T/S)P(M/L)QEG" (COM^A), which act as structural anchor. Having done this assignment for all twelve COM domain pairs of the tyrocidine, gramicidin S, surfactin, lichenisin, fengicin, and bacitracin biosynthetic complexes, it was found that of the 120 putative selectivity-conferring residues, >96% were either polar or charged residues. Furthermore, astounding 59 of 60 proposed amino acids pairs

(>98%) would lead to the establishment of productive (non-repulsive) polar or electrostatic interactions.

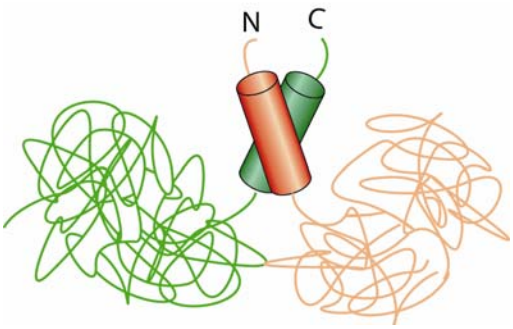


Fig. 53 Leucine-zipper motif
According to the 5-residue model, partner COM domains are believed to form a leucine-zipper-like structure, with each single donor and acceptor COM domain contribute to one helix.

The model could be actually validated by means of mutational analysis [Hahn and Stachelhaus, 2006]. Based on the 5-residue model, the productive interaction between the non-cognate COM^{D}_{TycA} and COM^{A}_{TycC1} is prevented, because of the formation of one repulsive interaction between two lysine residues of the donor and acceptor, respectively. In the cognate COM domain COM^{A}_{TycB1}, the corresponding lysine moiety is replaced by an aspartate residue that – in the context of the cognate COM domain pair $COM^{D}_{TycA}/COM^{A}_{TycB1}$ would give rise to a strong, productive, electrostatic interaction. Hence, conversion of the basic residue Lys 9-residue of COM^{A}_{TycC1} was expected to turn the non-cognate pair $COM^{D}_{TycA}/COM^{A}_{TycC1}$ into a cognate one. This hypothesis could be confirmed by means of site-directed mutagenesis. The corresponding point-mutant COM^{A}_{TycC1}(K9D) was unable to interact with its natural partner COM^{D}_{TycB3} (loss of function), but gained the ability to productively communicate with the non-cognate COM^{D}_{TycA} (gain-of-function) [Hahn and Stachelhaus, 2006].

The 5-residues model also provides a plausible explanation for the outcome of the COM domain swapping experiment between $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}B1}$ and $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}B1}$. As shown in Fig. 54, the establishment of a productive contact between two NRPSs of the surfactin biosynthetic complex requires at least four proliferous (polar or electrostatic) interactions. In addition, the cognate COM domain pair $COM^{D}_{SrfA\text{-}B3}/COM^{A}_{SrfA\text{-}C}$ also shows one, presumably repulsive contact (note: this represents the only mis-match observed so far in the 5-residue model) between a Glu residue of the COM^{D} domain, and an Asp residue of COM^{A}. In contrast, the proven non-cognate pair $COM^{D}_{TycA}/COM^{A}_{TycC1}$ does establish one repulsive, but only three productive interactions (Fig. 54). As importantly, when comparing the quality of contacts formed between $COM^{D}_{SrfA\text{-}B3}$ and the cognate partner $COM^{A}_{SrfA\text{-}C}$ or the mis-cognate $COM^{A}_{SrfA\text{-}B1}$, it is found that the same number of electrostatic and polar interactions is formed. This would explain the observed productive interaction between SrfA-

A and SrfA-B in the hybrid NRPS systems of CC102, and the resulting biocombinatorial synthesis of lipohepta- and lipotetrapeptide. Interestingly, the model also predicts a mis-cognate interaction between $COM^{D}_{SrfA\text{-}A3}$ and $COM^{A}_{SrfA\text{-}C}$, which was experimentally confirmed in this study by constructing the corresponding COM domain swap. In the corresponding hybrid NRPS system of *B. subtilis* CC110 (genotype: *srfA-A::*$COM^{D}_{SrfA\text{-}A3}$*-*$COM^{A}_{SrfA\text{-}C}$*::srfA-B*), SrfA-A was very well able to establish a mis-cognate interaction with SrfA-B, leading to the restoration of lipoheptapeptide surfactin production (product titer compared to wild-type: 26%).

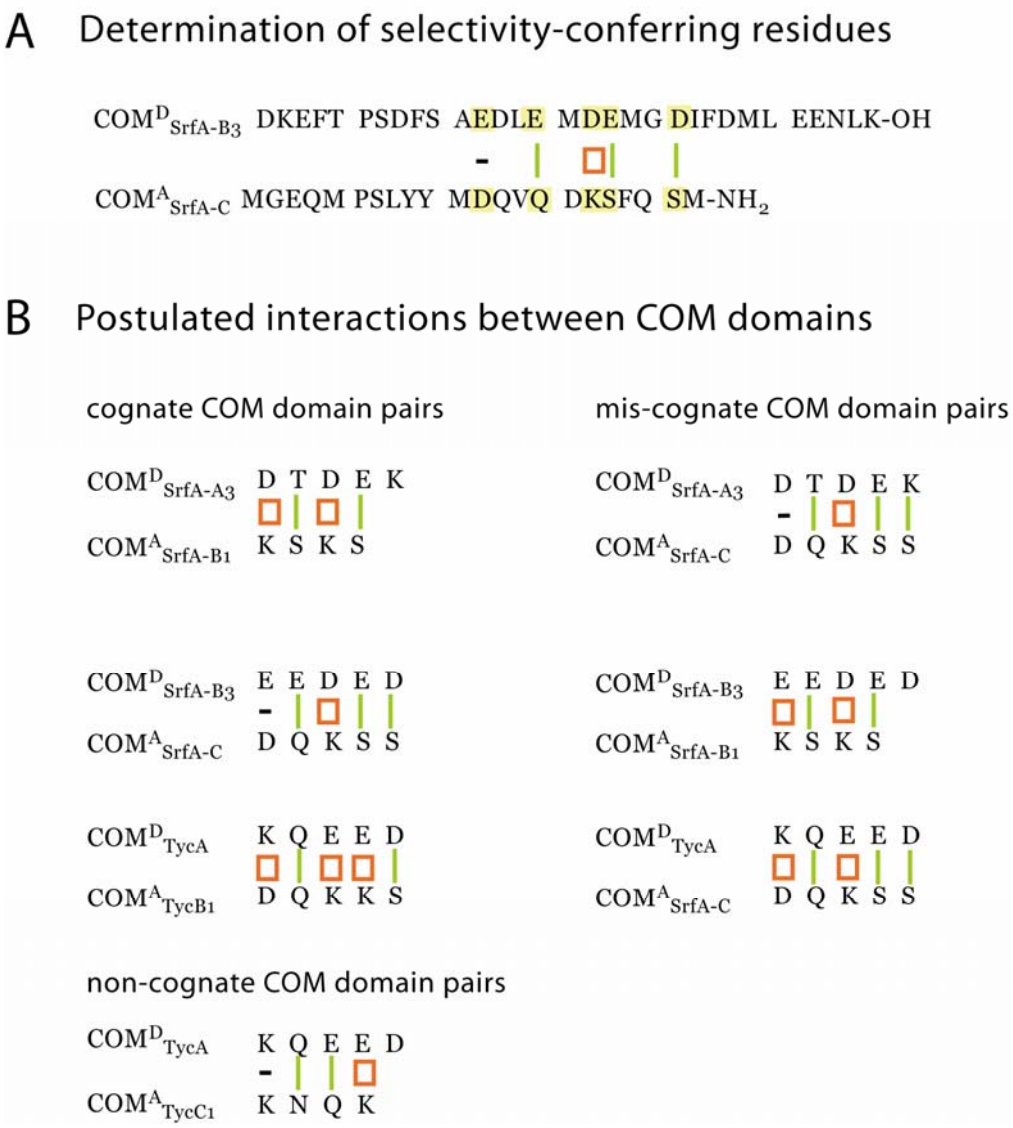


Fig. 54 The 5-residue model: Determination of the selectivity-conferring residues
(A) The putative selectivity-conferring residues within the cognate COM domain pair $COM^{D}_{SrfA\text{-}B1}$/$COM^{A}_{SrfA\text{-}C}$ were determined, using the highly conserved core motifs "TPSD" and "L(T/S)P(M/L)QEG" as structural anchors. **(B)** Proposed interactions between cognate, mis-cognate and non-cognate COM domain pairs. Electrostatic interactions are indicated by red squares, polar interactions by vertical bars, and putative repulsive interactions by minuses.

Apparently, the described model provides an explanation for the product patterns observed in the course of the COM domain swapping experiments. On the other hand, however, it leaves open the question of how the formation of a defined biosynthetic complex (and eventually synthesis of a specific peptide product) is controlled, when both partner and non-partner enzymes are equipped with compatible (cognate or mis-cognate) pairs of COM domains. In this context, it is interesting to realize that the COM domain pair $COM^{D}_{SrfA\text{-}A3}/COM^{A}_{SrfA\text{-}C}$ only causes a mis-cognate interaction between natural partner enzymes (i.e. SrfA-A and SrfA-B in *B. subtilis* CC110), but not non-partner NRPSs (i.e. SrfA-A and SrfA-C in *B. subtilis* CC110 or the wild-type). This may be an indication for additional, subsidiary structural features, situated outside the COM domains within the main body of partner-enzymes (in case of SrfA-A and SrfA-B, a region of more than 3,500 aa!), which influence their protein-protein interaction. Further studies are required to clarify this point, and potentially identify these additional factors.

7.6 COM domains and docking domains: differences and homologies.

Intra- and intermolecular linkers play a fundamental role in protein-protein communication within natural product synthesis. When it comes to the selective communication between donor and acceptor enzymes biosynthetic complexes, the correct channelling of the substrate intermediates is controlled by either COM domains in NRPSs, or so-called linker- or docking domains in PKSs .

NRPS COM domains and PKS docking domains share several similarities, for example they are both located at the C- and N-termini of the corresponding partner enzymes, and they mostly consist of charged and polar amino acid residues.

Unfortunately, a more detailed comparison of docking and COM domains on the structural level is not possible, since no structural data of COM domains are currently available. In contrast, the structures of docking domains have been determined by NMR analysis for the corresponding domains, connecting the PKSs DEBS2 and DEBS3 in the biosynthetic assembly line of erythromycin (see Fig. 55). Consistent with secondary structure predictions, the elucidation of NMR structure of the PKS docking domains of DEBS2-DEBS3 (dock 2-3) revealed a multi-helical structure. Hereby, the donor docking domain of DEBS 2 (amino acid residues 1-80) contributes three α-helices, while the N-terminal acceptor docking domain of DEBS 3 (amino acid residues 83-120) contributes only a single, but longer α helix.

Helices 1 and 2, as well as 3 and 4 are forming the dimeric domains A and B, respectively.

In this connection, domain A belongs entirely to the C-terminus of DEBS2, and consists of a four α-helix bundle, whose monomers (1, 1´, 2 and 2´) are connected by short loops. Domain B, connected to A by a long flexible loop, also contains four α-helixes, 3, 3´, 4 and 4´. Two helixes (4 and 4´) form a coiled-coil dimer, and represent the N-terminus of DEBS3, while the other two (3 and 3´) are part of the C-terminus of DEBS2. Hence, the B domain represents the actual area of contact between the donor docking domain of DEBS2, and acceptor docking domain of DEBS3.

Further analysis of this NMR structure also revelaed that protein-protein communication between PKSs is established by hydrophobic interactions at the interface between the docking domains. However, decisive for the discrimination between partner and non-partner enzymes are several hydrogen bridges (electrostatic interactions) between acidic and basic amino acid residues. In particular, repulsive ionic forces are believed to prevent the undesired interaction between non-partner enzymes.

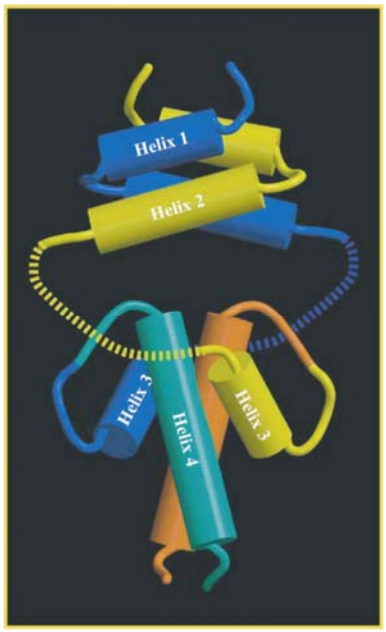


Fig. 55 Structure of DEBS dock 2-3.
Docking domains facilitating the interactions between DEBS2 and DEBS3 are organized in two dimeric α-helix bundle, in which monomers are connected by flexible loops. Three α-helixes (1-3) belong to the C-terminal donor docking domain of DEBS2, while one helix (4) belongs to the N-terminal acceptor docking domain of DEBS3. The fusion point between docking domains is in the middle of the loop connecting helix 3 and 4.

Up to now, no comparable structural information are available for NRPS COM domains. Nonetheless, secondary structure predictions indicate that COM domains may likewise form α-helical structures (Fig. 56). This forecast is also supported by the analysis of the crystal structure of the free-standing C domain VibH of the vibriobactin NRPS system [Keating et al., 2002]. This single domain should be equipped with COM^{A} and COM^{D}, mediating a selective interaction with its partner NRPSs VibB and VibF, respectively. Interestingly both, the N- and the C-terminus of VibH possess α-helical structure.

Apart from the mentioned similarities between NRPS COM domains and PKS docking domains concerning, i.e. localization, structure, and function, there are, however, also at least two striking differences. First, COM^{D} (20-30 aa) and COM^{A} (15-25 aa) domains are significantly smaller then the corresponding docking domains (80-120 aa and 30-50 aa, respectively), therefore they are likely to form two anti-parallel oriented α-helices, associated in a leucine zipper-like structure, rather than multi-helix bundles as observed in the case of PKS docking domains.

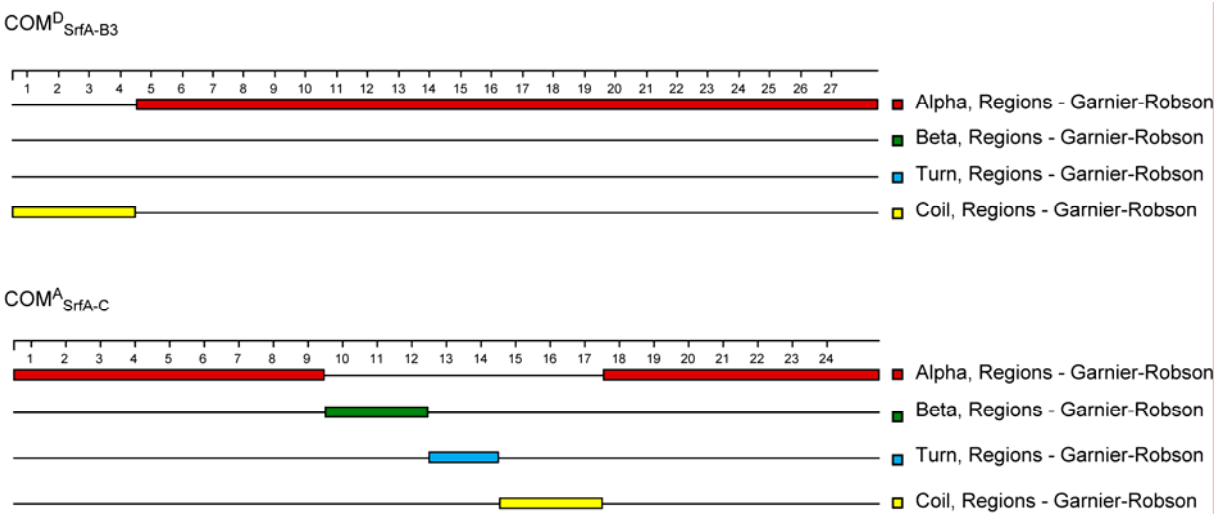


Fig. 56 Prediction of the secondary structure for $COM^{D}_{SrfA-B3}$ and COM^{A}_{SrfA-C}
Analysis of the primary structure of $COM^{D}_{SrfA-B3}$ and COM^{A}_{SrfA-C} with the computer program Protean suggests that both COM domains possess α-helical structures.

Futhermore, it was observed that the donor substrate can also influence the efficiency of communication between compatible docking domains [Wu et al., 2002]. In contrast, recent *in vitro* studies demonstrated that compatible COM domains can give rise to a productive interactions independently from the nature of the presented donor substrate [Hahn and Stachelhaus, 2004]. For example, in case of the mis-cognate COM domain pair $COM^{D}_{TycA}/COM^{A}_{SrfA-C}$ it was shown that SrfA-C was very well able to process the donor substrate D-Phe and form the di-peptide D-Phe-L-Leu, although the native donor substrate would have been D-Leu.

7.7 Outlook

In conclusion, the *in vivo* investigations conducted in this work verified the crucial role of COM domains in protein-protein communication for the first time *in vivo*, and demonstrated their enormous biocombinatorial potential as well as for the engineering and reporgrammning of NRPS assembly lines. Furthermore, the *in vivo* portability of COM domains could be

demonstrated i.e. by the successful utilization of heterologous COM domains from the tyrocidine biosynthetic system for the restoration of surfactin biosynthesis.

A single COM domain swapping experiment was exploited to destroy the communication between natural partner NRPSs (SrfA-A and SrfA-B), and – at the same time – to establish a productive interaction between non-partner enzymes (SrfA-A and SrfA-C). Within the context of this study, this strategy was used for the intended skipping of an internal enzyme (SrfA-B) within a given NRPS assembly line, and synthesis of a shortened peptide product, without the need of deleting the corresponding NRPS gene (Fig. 57A). A similar approach should be also feasible for the integration of an additional NRPS into a NRPS assembly line, leading to the targeted reprogramming of the NRPS complex, and formation of a novel NRP product. In order to achieve this goal, a gene "x", encoding the NRPS "X", could be cloned outside of the NRPS operon, and expressed *in trans*. If provided with appropriate COM domains, the resulting NRPS can be integrated into the biosynthetic template, and productively interact with the desired partner enzymes (Fig. 57B). This approach may be only limited by the C-domains donor-site selectivity, i.e. of the first module of the newly introduced NRPS ("X") and the first module of the native NRPS, accepting the unnatural reaction intermediate (enzyme "C" in Fig. 57B).

The construction of an *in vivo* UCS, in which all donor and acceptor enzymes were equipped with the same cognate COM domain pair, allowed for the concurrent biosynthesis of two NRPs (surfactin and the lipotetrapetide). This result opened interesting prospects for the engineering NRPS pathways. For example, the UCS could be enlarged by construction and incorporation of additional (heterologous) NRPSs. In this connection, insertion of an additional enzyme "X", carrying the appropriate terminal acceptor and donor COM domains would considerably increase the number of possible biosynthetic templates, and eventually the size and diversity of the corresponding NRP library (Fig. 57C).

Finally, heterologous COM domains could be used in order to differentiate a large multi-modular NRPS in a multi-enzyme complex, composed of two or more smaller NRPSs, this way simplifying purification and biochemical characterization of individual modules. For example, the approach could be applied to the 1,4 MDa-enzyme CssA, responsible for the synthesis of the immunosuppressant cyclosporine, and consisting of 11 modules. Using different compatible COM domain pairs, this mega-enzyme could be subdivided into smaller mono- or di-modular proteins, which could be assembled to reconstitute the original biosynthetic pathway.

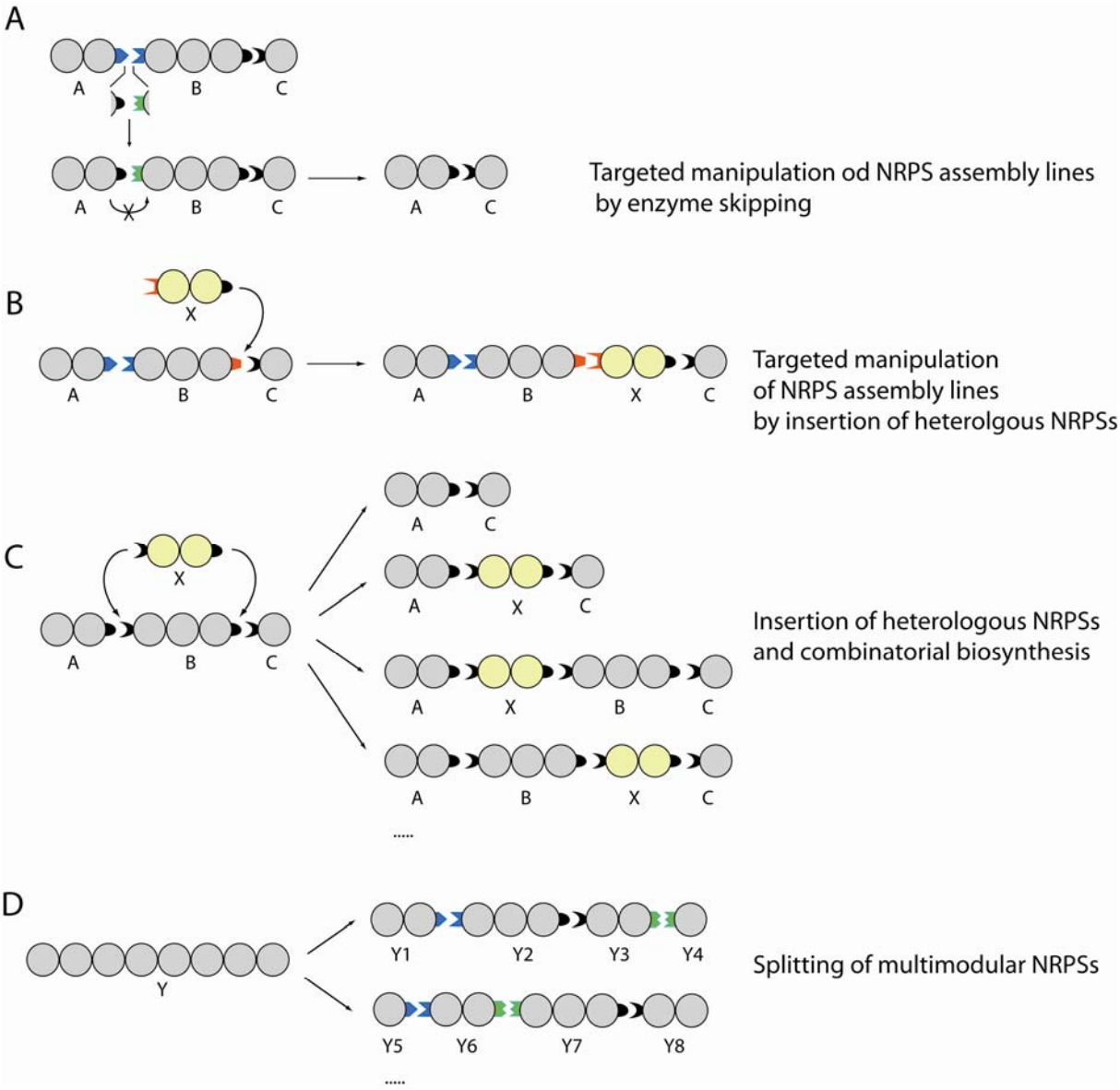


Fig. 57 Prospects for the utilization of NRPS COM domain
(A) Targeted manipulation of a NRP pathway: intended enzyme skipping. By substitution of a given COM domain pair, it is possible to interrupt the communication between the NRPSs A and B, and to enforce the direct communication between A and C. **(B)** Targeted modification of a NRPS assembly line by insertion of a heterologous NRPS "X". A given NRPS "X" can be integrated into a defined point of a biosynthetic assembly by adequate utilization of compatible and non-compatible COM domain pairs. **(C)** Insertion of a heterologous NRPS "X" for biocombinatorial synthesis. In a so-called UCS, all donor and acceptor enzymes are provided with the same set of COM domain pairs. The introduction of a heterologous NRPS "X" equipped with the same COM domains would enlarge the number of possible combinations, thus increasing the size and diversity of the corresponding NRP library. **(D)** Splitting of a multi-modular NRPS. The biochemical characterization of a mega-enzyme Y, consisting of 8 modules, could be facilitated by differentiating the huge enzyme in several mono- and di-modular NRPSs. Appropriate utilization of cognate COM domain pairs, ensures the formation of a defined biosynthetic template, and consequently formation of the desired NRP product. Two possibilities are shown in (D).

7.8 Purification of large NRPSs and multi-enzymatic NRPS complexes.

The work's second objective was the elaboration of a fast and gentle method for the purification of large NRPSs, and – potentially – multi-enzymatic NRPS complexes.

Many producer organisms of natural products are difficult to cultivate under laboratory conditions, and/or genetically inaccessible. None of these problems actually concerns the soil bacterium *B. subtilis*, making the Gram-positive one of the best-studied NRP producers known. In 1993, for example, Menkhaus et al. described the isolation of the surfactin biosynthetic complex from *B. subtilis* ATCC 21332 and *B. subtilis* OKB 105, which, however, required a laborious, and time-consuming five-step purification scheme [Menkhaus et al., 1993].

In the present work, the genetic accessibility of *B. subtilis* OKB 105 was exploited to append two different affinity tags to the C-terminus of SrfA-A, in order to ease purification of the tri-modular enzyme NRPS, and potentially even associated constituents of the surfactin biosynthetic complex, using a single step affinity chromatography. In order to achieve the envisioned genetic manipulation of the *srfA* operon, the previously described two marker-exchange homologous recombination approach was used, which had been also exploited for the COM domain swapping experiments, (see 5.1; 5.3; 5.4.2).

In a first set of experiments, the coding sequence of a hexahistidine affinity tag was fused to the 3'-end of *srfA-A*. The resulting mutant strain, *B. subtilis* CC17, was shown to be still able to produce wild type levels of surfactin, indicating that the affinity tag did not affect the protein-protein communication between SrfA-A-epi and SrfA-B. Subsequent purification of SrfA-A-His was accomplished by single-step Ni-NTA affinity chromatography, which provided pure protein in sufficient quantities for the following biochemical characterization. Amino acid-dependent ATP-PP_i exchange reactions, as well as covalent loading assays revealed that the purified SrfA-A-His protein was active, and had been converted into its phosphopantetheinylated holo-form. The covalent loading assay revealed decent loading activity for the substrate amino acid leucine, while the measured incorporation of the cognate substrate glutamate was rather weak. This latter observation, however, is in agreement with previous *in vitro* analyses of synthetases of surfactin and bacitracin biosynthetic complexes, where likewise only a weak loading activity of the acidic substrate amino acids was determined [Ishihara and Shimura, 1974; Schneider, 1996]. In conclusion, these biochemical data revealed that IMAC can be successfully used for the purification of a high molecular-weight proteins like SrfA-A-His. However, conditions of IMAC are considered rather harsh

and render this method inappropriate for the envisioned co-purification of associated proteins and members of the surfactin biosynthetic complex.

Burgess and co-workers recently developed an alternative – potentially more gentle – method for the purification of large proteins and labile multi-enzymatic complexes. The method is based on a single-step immunoaffinity chromatography, using so-called polyol-responsive monoclonal antibodies (PR-mAb). These particular antibodies are able to establish the known strong and specific interaction with the cognate antigen. However, PR-mAbs release their antigen under gentle, non-denaturating conditions in the presence of non-chaotropic salts, and low molecular weight poly-hydroxylated compounds (polyols).

One of the PR-mAb developed and characterized by the Burgess´ lab, called NT73, specifically recognizes the C-terminus (sequence: SLAELLNAGLGGS) of the β`-subunit of the *E. coli* RNA polymerase. The entire *E. coli* RNA polymerase core enzyme, composed by five subunits (α_2,β.β´,ω) plus weakly associated sigma factors, was purified in fully-active form by immunoaffinity chromatography using a NT73 column [Bergendahl et al., 2003; Thompson et al., 1992]. The successful purification of this large holo-enzyme complex (MW = 449 kDa) clearly demonstrated the potential of this method as a tool for the purification of labile multi-enzyme complexes.

In the following years, the method was further developed, and the determined epitope SLAELLNAGLGGS exploited as a general tag for the affinity purification of various enzymes. The tag was fused to the green fluorescence protein GFP, and the resulting recombinant protein successfully purified to homogeneity by loading the corresponding *E. coli* crude extract onto a NT73 immunoaffinity column [Burgess and Thompson, 2002; Thompson et al., 2003].

Based on these observations the coding sequence of the epitope tag was fused to the 3'-end of of *srfA-A*, and the encoded protein Srf-A-A-epi purified by immunoaffinity chromatography. In agreement with previous results obtained for SrfA-A-His, biochemical characterization revealed the specific activation and incorporation of the cognate substrate leucine by SrfA-A-epi, while nearly no activation of the acidic amino acid glutamate could be observed. The generated *B. subtilis* mutant CC57 was also still capable of producing wild-type levels of the lipoheptapeptide product surfactin, indicating that the epitope tag had no effect on protein-protein interaction.

The most important question to be answered concerned the possible co-purification of associated proteins. SDS-PAGE analysis of elution fractions of the NT73 immunoaffinity chromatography revealed the co-purification of several proteins, ranking in size between 30-

400 kDa. LC-MS/MS analysis of the most prominent protein bands (besides SrfA-A-epi), unequivocally verified the co-purification of the SrfA-A partner-enzyme SrfA-B. However, under the conditions tested, no other – potentially associated – proteins involved in surfactin biosynthesis (e.g. SrfA-C, Sfp, Srf-TeII, acyl transferase) could be co-purified. Nonetheless, it is worth mentioning that the di-modular complex of SrfA-A-epi/SrfA-B, which was purified by NT73 immunoaffinity chromatography, already combines a MW of over 800 kDa, almost doubling the size of the RNA polymerase complex (MW = 449 kDa) purified by Burgess & co-workers.

7.9 Method optimization for purification of large NRPSs and multi-enzymatic NRPS complexes

Within the scope of this work, two different methods (immobilized metal-ion affinity chromatography and immunoaffinity chromatography) for the single-step purification of functional multi-modular NRPSs have been successfully implemented. In the case of immunoaffinity chromatography, the purification even enabled the co-purification of the associated partner-NRPS. However, little is known about the actually binding affinities between partner-NRPSs, as well as between NRPS and associated proteins (e.g. external type II thioesterase, Ppant transferase). Previous studies only indicated that the interaction between NRPS might be rather dynamic process [Gruenewald et al., 2004], and that standard multi-step purifications of NRPSs from the natural producer strain – under appropriate conditions – can afford the co-purification of the associated type II thioesterase (Stachelhaus, personal communication). In this regard, “appropriate conditions” have to be determined empirically, leaving plenty of options for the possible optimization of methods established for the purification of multi-enzymatic complexes. Some of the parameters that could be altered include i.e. the composition of the buffer system, the temperature, the choice of the resin, and the nature and position of the affinity tag.

7.9.1 Optimization of IMAC

The composition of the buffer used for any purification procedure is in general of great importance in order to obtain a good protein preparation. The choice of the buffer system is even more problematic when it comes to the purification of large enzymes or labile enzyme complexes. An *ideal* buffer for the purification of large proteins and protein complexes should (i) ensure a specific and sufficiently strong binding of the protein to the solid support, (ii) stop

the unspecific binding or association of proteins, and – at the same time – (iii) stabilize the protein complex, preventing from disintegration.

In case of IMAC, the selective binding of the His-tagged protein to the column is usually facilitated by the use of high-salt concentrations (300-500 mM NaCl), disfavouring the unspecific binding of other proteins present in the crude cell extracts. However, the high-ionic strength of such a buffer system also destabilizes protein-protein interactions within a multi-enzymatic complex. Consequently, a possible way of improving IMAC for the purification of multi-enzymatic complexes could be the step-wise reduction of the buffer system's ionic strength. Another problem encountered during purification of large proteins by Ni-NTA chromatography is the low yield of the final protein preparation due to incomplete binding of the protein to the affinity column. In the past, the efficiency of binding could be improved primarily by using batch-incubation during the binding of the His-tagged protein to the affinity resin [Duerfahrt et al., 2003; Trauger and Walsh, 2000]. This strategy promotes the efficient binding of the His-tagged protein, especially when the tag is not fully accessible, or when the desired protein is present only at a very low concentration.

Another possibility is provided by the utilization of NTA affinity resins, which had been charged with different divalent metal ions, like Zn^{2+} or Cu^{2+}. Especially immobilized Cu^{2+} ions were shown to be more effective than Ni^{2+} ions in retaining the His-tagged protein [Casey et al., 1995]. Unfortunately, this higher affinity of Cu^{2+}-NTA resins usually also promotes the unspecific binding of untagged proteins present in the crude extract, bearing the risk of getting undesired contaminations in the elution fractions.

Another prospect for increasing the binding affinity between protein and matrix would be the utilization of longer deca- or octahistidine tags, or by the exploitation of two consecutive hexahistidine tags, separated by an amino acid spacer. The latter approach had been successfully used by Khan et al. For the purification of the green fluorescent protein GFP [Khan et al., 2006]. However, in the worst case, the histidine tag remains buried or only partially accessible for the binding to the solid support. In this case, a new position for the affinity tag, or a completely different tag has to be chosen.

7.9.2 Optimization of immunoaffinity chromatography

When it comes to the purification of multi-enzyme complexes, a potential problem arises from the requirement to maintain the protein-protein interactions within the protein complex. The strong binding between antibody and antigen is provided by hydrophobic and electrostatic interactions; the same forces that stabilize a protein's tertiary structure.

Consequently, it is not surprising that in immunoaffinity chromatography usually denaturing conditions are required in order to release the antigen from the column. However, given the special properties of PR-mAbs, very mild conditions can be utilized to the disintegration of antibody/antigen complex, allowing for the purification of active tagged protein, and in some cases even multi-enzymatic complex. Interestingly, the non-chaotropic salts and polyols, used for the decomposition of the antigen/antibody complex, are both protein-stabilizing agents. Therefore, it was not immediately clear why the combination of those two agents could break the complex. The most likely explanation is that their concerted action strengthens the interactions within the enzyme and consequently destabilizes the Ab-Ag binding [Thompson et al., 1990].

Several additional parameters appear to play an important role in immunoaffinity chromatography. The temperature, for example, can influence the interaction between the epitope-tagged protein and the PR-mAb. For instance, it has been noted that moderate temperatures (20-25°C) apparently favour strong interactions, and at the same time positively affect the stability of both the PR-mAb and the target protein.

Like in the case of IMAC, the composition of the buffers used during the purification procedure might also influence the outcome of a given experiment. A difficult task, for instance, is to find a reasonable compromise between i) stabilization of the tertiary structure of the single proteins, ii) maintenance of the interactions among the components of the multi-enzymatic complex, and iii) suppression of the unspecific binding of contaminating proteins from the crude cell extracts. In order to stabilize the large surfactin synthetases, all buffers used for immunoaffinity chromatography were supplemented with 15% sucrose. However, the addition of this stabilizing agent could easily lead to a decomposition of the multi-enzyme complex, given the known interference of sucrose with weak protein-protein interactions, as observed e.g. between NRPSs and associated enzymes. Hence, alternative agents (i.e. mild non-ionic detergents like Triton X-100) could be tested, in order to avoid the degradation of the surfactin biosynthetic complex, and at the same time stabilize the fragile mega-enzyme.

Another factor, which should be considered for the improvement of the purification scheme for multi-enzymatic protein complexes is the positioning of the affinity tag. The experiments presented in this study demonstrated that epitope tag (as well as the hexahistidine tag) could be appended to the C-terminal end of SrfA-A without affecting its interaction with the partner enzyme SrfA-B. Based on this observation, likewise no interference with the protein-protein communication is expected when the same tag is fused to C-terminus of SrfA-B or SrfA-C. Since SrfA-B represents the central enzyme of the biosynthetic assembly line, functionally

interacting with both SrfA-A and SrfA-C, the construction of a SrfA-B-epi protein could provide a better chance to achieve the purification of the entire surfactin biosynthetic complex.

7.10 Evaluation of different methods for the genetic manipulation of *B. subtilis* chromosome

Many of the *B. subtilis* mutants constructed in this study were obtained using the previously described two step marker exchange homologous recombination method. In the first step, the DNA locus of interest is hereby interrupted with a selectable marker (here: a *mls* cassette). In the second step, a so-called reconstitution plasmid, is used for the transformation of the resulting disruption mutant. The following marker exchange homologous recombination event leads to the excision of the selectable marker (*mls*), and – simultaneously – the reconstitution of the biosynthetic gene cluster. Since the *B. subtilis* mutants obtained can not be screened directly for the loss of the resistance marker *mls*, a second, so-called helper plasmid, carrying an independent selectable marker, is used in a co-transformation (or congression) experiment, in order to facilitate a positive selection (for the independent marker) (see 4.1.2.1) (Fig. 58A). Only then, successful transformants are screened for the loss of the primary marker, whereby usually about 0.1 to 1% of clones exhibit the desired phenotype.

The described congression procedure was successfully applied for all COM domain swapping experiments described in this study, as well as for the construction of the *B. subtilis* mutant CC17, in which the coding sequence of the hexahistidine-affinity tag had been fused to the 3'end of *srfA-A*. In contrast, the congression method failed for the corresponding construction of *srfA-A::epi*.

Although the actual reason for the failure of this experiment remained obscure, the results from different congression experiments suggested that the rate of double-crossover events obtained might very-well depend on the *B. subtilis* strain utilized for transformation. In fact, within the scope of this study, two different disruption mutants, CC64 and CC16, with essentially the same genotype (*srfAA-mls-srfA-B*) were constructed within the genetic background of the native surfactin-producer strains *B. subtilis* ATCC 21332 and OKB 105, respectively. Surprisingly, subsequent congression experiments revealed significantly different transformation efficiencies for these two different disruption mutants, although exactly the same reconstitution plasmids were used. Indeed, for all COM domain swapping experiments, the frequency of double-crossover events varied by about one order of

magnitude between the *B. subtilis* ATCC 21332 derivative CC64 (2%), and the OKB 105 derivative CC16 (0.3%).

Apart from the observed differences in recombination efficiency between different strains of *B. subtilis*, the lack of a positive selection represents a general bottleneck of congression experiments. This problem could be potentially solved by the utilization of alternative strategies for the genetic manipulation. A well-known method frequently utilized for the genetic engineering of *E. coli, Bacillus spp.* and other bacteria, is based on the utilization of temperature-sensitive (TS) *origin of replication* (*ori*) within the reconstitution plasmid [Link et al., 1997; Yakimov et al., 2000]. Based on a mutation, the activity of this *ori* is temperature-dependent. At a permissive temperature, the *ori* is active and the corresponding plasmid behaves like a self-replicative vector. However, at higher, non-permissive temperatures the *ori* becomes inactive, and the plasmid's genetic information can only be maintained after its integration in the bacterial chromosome by single crossover recombination. This approach – not unlike the described congression experiments – relies on successive gene disruption and gene reconstitution. However, the reconstitution step is carried out without participation of an additional helper plasmid. Instead, the reconstitution plasmid with the temperature-sensitive *ori* also carries a selectable marker, i.e. an antibiotic resistance cassette, located outside of the DNA region to be integrated into the chromosome. After transformation, cells are first transferred onto a selective solid medium and incubated at a non-permissive temperature (usually 45-48°C), in order to enforce the integration of the plasmid by single crossover recombination. By transferring the mutants onto a non-selective medium and incubating them at the permissive temperature (about 30°C), the subsequent excision of the integrated plasmid is induced. Under these permissive conditions, the temperature-sensitive *ori* is activated, and due to the fact that two functional replication orgin within the same DNA fragment (in our case within the chromosome) can not co-exist, the plasmid's excision is enforced. Depending on the position, at which the second single crossover event is taking place, this excision either leads to the desired mutant or a so-called revertant, carrying the genetic organization of the parental strain. This method was already successfully used for the manipulation of the surfactin biosynthesis operon in *B. subtilis* OKB 105. Hereby, the exchange of first glutamate-activating module of the surfactin biosynthetic complex, against the first glutamine-activating module of the lichenisin assembly line, led to the construction of the *B. subtilis* mutant 1D1, which produced the new lipoheptapeptide Gln^1-surfactin [Yakimov et al., 2000].

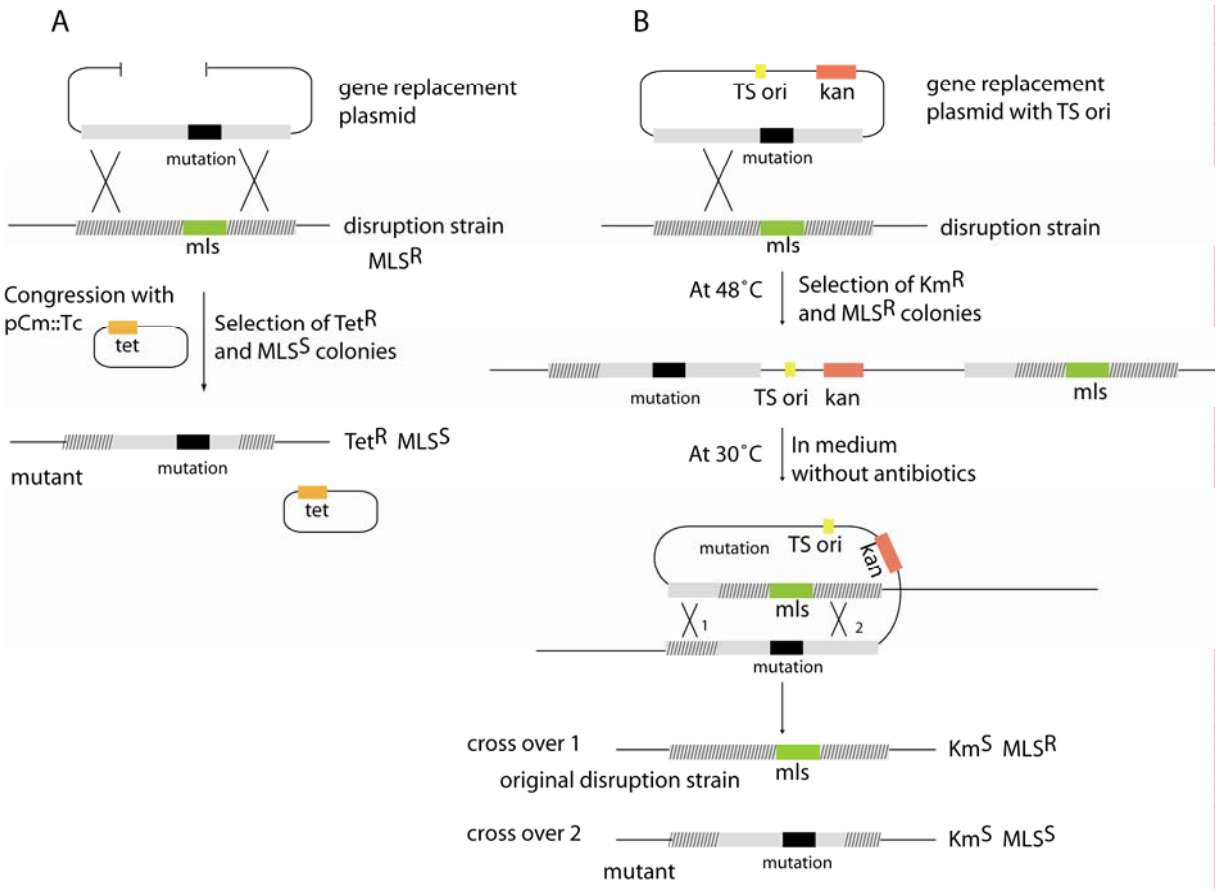


Fig. 58 Two-step marker exchange homologous recombination methods
(A) Congression: an appropriate disruption mutant is co-transformed with two plasmids: i) an integration plasmid, carrying the mutation, flanked by two homologous regions (in gray), and ii) a helper plasmid, carrying an independent selectable marker. Positive clones, carrying the desired mutation are screened on selective solid medium for the appropriate phenotype. **(B)** Gene replacement using a TS plasmid. The gene replacement plasmid, carrying the mutation to be integrated, flanked by two homologous regions (in gray), also harbors a TS origin of replication, and a selectable marker located outside of the homologous regions. Two consecutive temperature switches induce the integration and following excision of the TS plasmid. 1 and 2 indicate the two possible locations for the second crossover event, which can generate a revertant or the desired mutant.

In both methods described above, transformants are screened for the loss of a selectable marker, located within the DNA fragment to be exchanged (negative selection). A possible way to simplify the genetic manipulation is presented by the utilization of a so-called counter-selection method. According to this strategy, the DNA fragment to be modified or substituted has to be interrupted by a selectable marker, whose excision – under appropriate conditions – allows for the cell's survival (positive selection).

An example of such a counter-selection system has been recently adopted for *B. subtilis*, and is based on the catalytic activity of the enzyme uracyl-phosphoribosyl-transferase (UPRTase), encoded by the *upp* gene of *B. subtilis* [Fabret et al., 2001; Fabret et al., 2002]. This enzyme is involved in a so-called salvage pathway, catalyzing the conversion of uracil and 5'-phosphoribosyl-α-1-pyrophosphate to UMP. UPRTase is also able to accept the substrate-analogon 5-fluorouracile, leading to the synthesis of 5-fluoro-UMP, which is further

metabolized to 5-fluoro-dUMP. This latter compound, however, represents a potent inhibitor of the essential enzyme thymidylate synthase, which is responsible for the synthesis of dTMP. and is therefore toxic for bacterial cells expressing the *upp* gene. After the substrate 5-fluoro UMP has bound the thymidylate synthase the first two steps of the normal enzymatic reaction take place, before the enzyme is irreversibly tramped and arrested in a dead-end complex.
Bacillus cells, harboring a deletion of the *upp* gene, are able to grow in a medium, which had been supplemented with 5-fluorouracile, while those, containing a functional copy, cannot survey. Genetic manipulations, elaborating this *upp* counter-selection method, would be carried out as follows: i) construction of a *B. subtilis* strain, in which the natural copy of *upp* gene has been deleted, ii) disruption of desired chromosomal locus (e.g. NRPS gene), by integration of *upp*, and iii) transformation of this disruption mutant with an appropriate plasmid, whose integration leads to the reconstitution of the parental gene, and – at the same time – loss of the *upp* gene. The cells of the last transformation step are selected plated on medium containing 5-fluorouracil, allowing for a positive selection of the those mutants, carrying the desired genotype due to the loss of the *upp* gene

7.11 Outlook

Within the course of this work, two different approaches for the affinity purification of SrfA-A – and potentially associated proteins– have been established and tested. Immunoaffinity chromatography, using the polyol-responsive antibody NT73, actually allowed for the successful purification of an approx. 800 kDa dimeric complex, consisting of SrfA-A-epi and SrfA-B. However, under the tested purification conditions, neither the third synthetase of the surfactin biosynthetic complex, SrfA-C, nor any other enzyme, which may participate in the surfactin biosynthesis, could be co-purified. As mentioned above, the efficacy of (immuno-) affinity chromatography is significantly affected by manifold factors. Hence, it is reasonable to assume that the purification conditions could be optimized, in order to eventually allow for the purification of the entire surfactin multi-enzymatic complex; a hope, which is particularly based on the described examples of the successful purification of RNA polymerase core enzyme complexes. In case of the surfactin biosynthetic system, efforts to finally achieve the purification of the entire multi-enzymatic complex are certainly worthwhile, given its prospects for possible co-purification of weakly associated proteins like the Ppant transferase Sfp, the type II thioesterase SrfA-TE, and especially the acyl transferase (and other enzymes?) involved in the lipo-initiation. Indeed, characterization of the latter enzyme(s) would finally

shed light on this poorly understood mechanism, eventually clarifying how the β-hydroxy-fatty acid moiety is transferred onto the glutamate residue, activated by the first module of SrfA-A. As already mentioned (see 2.6), a hypothetical acyl transferase was already purified by Menkhaus and co-workers [Menkhaus et al., 1993]. However, this protein has not yet been biochemically characterized, and – even more puzzling – could not be detected in subsequent experiments. Therefore, it still remains to be clarified, whether an unknown specific acyl transferase, or an already known enzyme involved in the primary metabolism is required for lipo-initiation. In this context, it was for instance recently proposed that the external type II thioesterase SrfA-TE, which had been proven to posses potent proof-reading activity for the cleaning of mis-acylated PCP domains [Schwarzer et al., 2002], could be also responsible for lipo-initiation [Steller et al., 2004].

8 Appendix

8.1 Strains and plasmids

Tab. 2 *B.subtilis* strains constructed in this work

B subtilis strain	Genotype
AM1	*(amyE::-cat- pspac-comS-lacI::amyE)*
CC16	*srfA-A-mls-srfA-B*
CC64	*srfA-A-mls-srfA-B*
CC84	*(srfA-A::COM$^{D}_{srfA-A3}$-COM$^{A}_{srfA-B1}$::srfA-B) (amyE´-cat pspac-comS-lacI-amyE)*
CC99	*(srfA-A::COM$^{D}_{TycB3}$-COM$^{A}_{TycC1}$::srfA-B) (amyE´-cat pspac-comS-lacI-amyE)*
CC112	*(srfA-A::COM$^{D}_{TycA}$-COM$^{A}_{TycC1}$::srfA-B) (amyE´-cat pspac-comS-lacI-amyE)*
CC91	*(srfA-A::COM$^{D}_{SrfA-B3}$-COM$^{A}_{SrfA-C}$::srfA-B) (amyE´-cat pspac-comS-lacI-amyE)*
CC102	*(srfA-A:: COM$^{D}_{SrfA-B3}$-COM$^{A}_{SrfA-B1}$::srfA-B) (amyE´-cat pspac-comS-lacI-amyE)*
CC110	*(srfA-A:: COM$^{D}_{SrfA-A3}$-COM$^{A}_{SrfA-C}$::srfA-B) (amyE´-cat pspac-comS-lacI-amyE)*

Tab. 3: Plasmids constructed in this work

Plasmid	Description
pKE27	pDR66-*comS*
pCC13	pQE60-*3'srfA-A-mls-5'srfA-B*
pCC14	pQE60-*3'srfA-A::His--5'srfA-B*
pCC42	pQE60-*3'-srfA-A::epi-mls-srfA-B5´*
pCC43	pQE60-*3´-srfA-A::epi-5'srfA-B-kan-5'-srfA-B*
pCC77	pQE60-*3'-srfA-A-srfA-B5´*
pCC78 (a)	pQE60-*3'-srfA-A-AvrII-Acc65I-5'-srfA-B´*
pCC83 (a)	pQE60-*3'-srfA-A::COM*$^{D}_{srfA\text{-}A3}$-*COM*$^{A}_{srfA\text{-}B1}$*::5'-srfA-B*
pCC92 (a)	pQE60-*3'-srfA-A::COM*$^{D}_{TycB3}$-*COM*$^{A}_{TycC1}$*::5'-srfA-B*
pCC106 (a)	pQE60-*3'-srfA-A::COM*$^{D}_{TycA}$-*COM*$^{A}_{TycC1}$*::5'-srfA-B*
pCC85 (a)	pQE60-*3'srfA-A::-COM*$^{D}_{srfA\text{-}B3}$-*COM*$^{A}_{srfA\text{-}C}$*::5'-srfA-B*
pCC98 (a)	pQE60-*3'-srfA-A::COM*$^{D}_{srfA\text{-}B3}$-*COM*$^{A}_{srfA\text{-}B1}$*::5'-srfA-B*
pCC97 (a)	pQE60-*3'srfA-A::COM*$^{D}_{srfA\text{-}A3}$-*COM*$^{A}_{srfA\text{-}C}$*::5'-srfA-B*

8.2 Oligonucleotides

Tab. 4: Oligonucleotides

Name	Sequence
5’-pQE/srfAB_inv	5’-ATA TGG TAC CAT GCG ATG CTT GAT CCG CAT TC -3’
3’-pQE/srfAB_inv	5’-ATA TCC TAG GGG TCA GTT CCG TAC CAT CTT G-3’
5’_srfAB_COM	5’-ATA TGC TAG CGA TCT TGG TGA CGA CGA TTT G-3’
3’_srfAB_COM	5’-ATA TTG TAC AGC ATT CCC TCC TGC ATT GGT G-3’
5’_srfBC_COM	5’-ATA TGC TAG CGA CTT CAG CGC CGA AGA CC-3’
3’_srfBC_COM	5’-ATA TTG TAC AGC ATC CCT TCC TGC ATC GGC-3’
5’_tycBC_COM	5’-ATA TGC TAG CGA CCT GGG GGA TGA AGA G-3’
3’_tycBC_COM	5’-ATA TTG TAC AGC ATA CCC TCT TGC AAT GGG G-3’
5’-srfB3_SOE(srfB1)	5’-AAC CTG CGC CGC ATT CCG-3’
3’-srfB3_SOE(srfB1)	5’-CCA CCT CTA TTA TTT TAA ATT CTC CTC AAG CAT GTC -3’
5’-srfB1_SOE(srfB3)	5’- GAG AAT TTA AAA TAA TAG AGG TGG CAT ATG AGC AAA AAA TCG -3’
3’-srfB1_SOE(srfB3)	5’-TCC ATC ATA ATA TGA TGA TTG CTC C-3’
5’-srfA3_SOE(srfC)	5’-AGA AGA TAT CAG ACA TGT GCC G -3’
3’-srfA3_SOE(srfC)	5’-CCC TTG CGT TTT AGA AAA TTT CCA TTA ATT TAT CCA G -3’
5’-srfC_SOE(srfA3)	5’- GGA AAT TTT CTA AAA CGC AAG GGA ATT ACA GAA GGC-3’
3’-srfC_SOE(srfA3)	5’-AAT GTG GTG GTA GCT CCA CAC C-3’
5’-Amy-front	5’-GCT GTT TCA TTT GGT TCT GG-3’
3’-Amy back	5’- CGT TGG TTG TAT CCG TGT C-3’
AmyE-probe	5’-AAT GAA TTC TGC GTG ACA TC-3’
EPI_F	5’-GAA CTG CTG AAT GCA GGC CTG GGC GGC TCA TAA TAG AGG TGG CAT ATG AGC-3’
EPI_R	5’-CAG GCC TGC ATT CAG CAG TTC TGC CAG TGA GAA AAT TTC CAT TAA TTT ATC CAG-3’
His-BamHI_R	5’-ATA GGA TCC TCA ATG ATG ATG ATG ATG ATG-3’
RF-48F	5’-AGC ATG TCG AAC TGG TAC TG -3’
RF-48R	5’-GCA TTT ACC TGG CTC CAA TG -3’
SG077	5’-ATT ATT GTT GAC CCC GAT GCA AGA GG-3’
SG078	5’-ATT GCC GGC ACC TCG GAT ATA TC-3’
Srf_11F	5’-ATA CCA TGG GAG CCG GAC ATC GAA GCG-3’
Srf_12F	5’-TAT AGA TCT TAG AGG TGG CAT ATG AGC AAA AAA TC-3’
Srf_13R	5’-ATA GGA TCC GAA AAT TTC CAT TAA TTT ATC CAG CTC-3’
Srf_14R	5’-TAT TCT AGA CAC TTG GTG AAC AGC CAT TCC-3’
Srf_22F	5’-ATA ACT AGT TAG AGG TGG CAT ATG AGC AAA AAA TC-3’
Srf_23R	5’-ATA GGA TCC TTA TGA GCC GCC CAG GCC TGC-3’
Srf_24F	5’-ATA GGA TCC AAT GAT TGC GGC ATC CCG C-3’
Srf_25R	5’-ATA ACT AGT ATT GTC ATA CCT CCC CTA ATC-3’
Srf_42F	5’-AAG TGG ACC GCA AAG CCT TG-3’
Srf_43R	5’-TCC GTT TCG CTT GCT CTT CA-3’

8.3 The conserved core motifs of NRPS domains

Tab. 5 Sequence motifs of the core domains

Domain	Core motif	Consensus sequence
Adenylation (A) domain	A1	L(TS)YxEL
	A2	LKAGxAAYL(VL)P(LI)D
	A3	LAYxxYTSG(ST)TGxPKG
	A4	FDxS
	A5	NxGPTE
	A6	GELIxIxGx(VL)ARGYL
	A7	Y(RK)TGDL
	A8	GRxPxQVKIRGxRIELGEIE
	A9	LPxYM(IV)P
	A10	NGK(VL)DR
Peptidyl carrier proteins (PCP) domain	T	LGG(DH)SL
Condensation (C) domain	C1	SxAQXR(LM)(WY)xL
	C2	RHExLRTxF
	C3	MHHxISDG(WV)S
	C4	YxD(FY)AVW
	C5	(IV)GxFVNT(QL)(CA)xR
	C6	(HN)QD(YD)PFE
	C7	RDxSRNPL
Te domain	TE	GxSxG

Tab. 6 Sequence motifs of the optional domain

Domain	Core motif	Consensus sequence
E domain	E1	PIQxWF
	E2	HHxISDG(WV)S
	E3	DxLLxAxG
	E4	EGHGRE
	E5	RTVGWFTxxYP(YV)PFE
	E6	PxxGxGYG
	E7	FNYLG(QR)
Heterocyclization domain	Cy1	FLP(TS)xxQxAYxxGR
	Cy2	RH(IM)L(PAL)x(ND)GxQ
	C3	(DNR)4DxxS
	Cy3	(LI)Pxx(PALx(LPF)P
	Cy4	(TS)(PA)3x(LAF)6x(IVT)LxxW
	Cy5	(GA)(DQN) FT
	Cy6	P(IV)VF(TA)SxL
	Cy7	QV(LI)Dx(QH)11xW(DYF)
N-methylation domain	M1 (SAM)	VL(DE)GxGxG
		NELxYRYxAV
		VExSxARQxGxLD
Oxydation domain	Ox1	KYxYxSAGxxY(PG)VQ
	Ox2	GxxxG(LV)xxGxYYY(HD)P
Reductase domain	R1	V(LF)LTGATGFLGAxLLxxLL
	R2	VYCLVRA
	R3	GPLxxPxLGL
	R4	NVxGT
	R5	GYxQSKWVAExxVxxAxxRGL
	R6	G(LF)LxxP

8.4 Biosynthetic systems

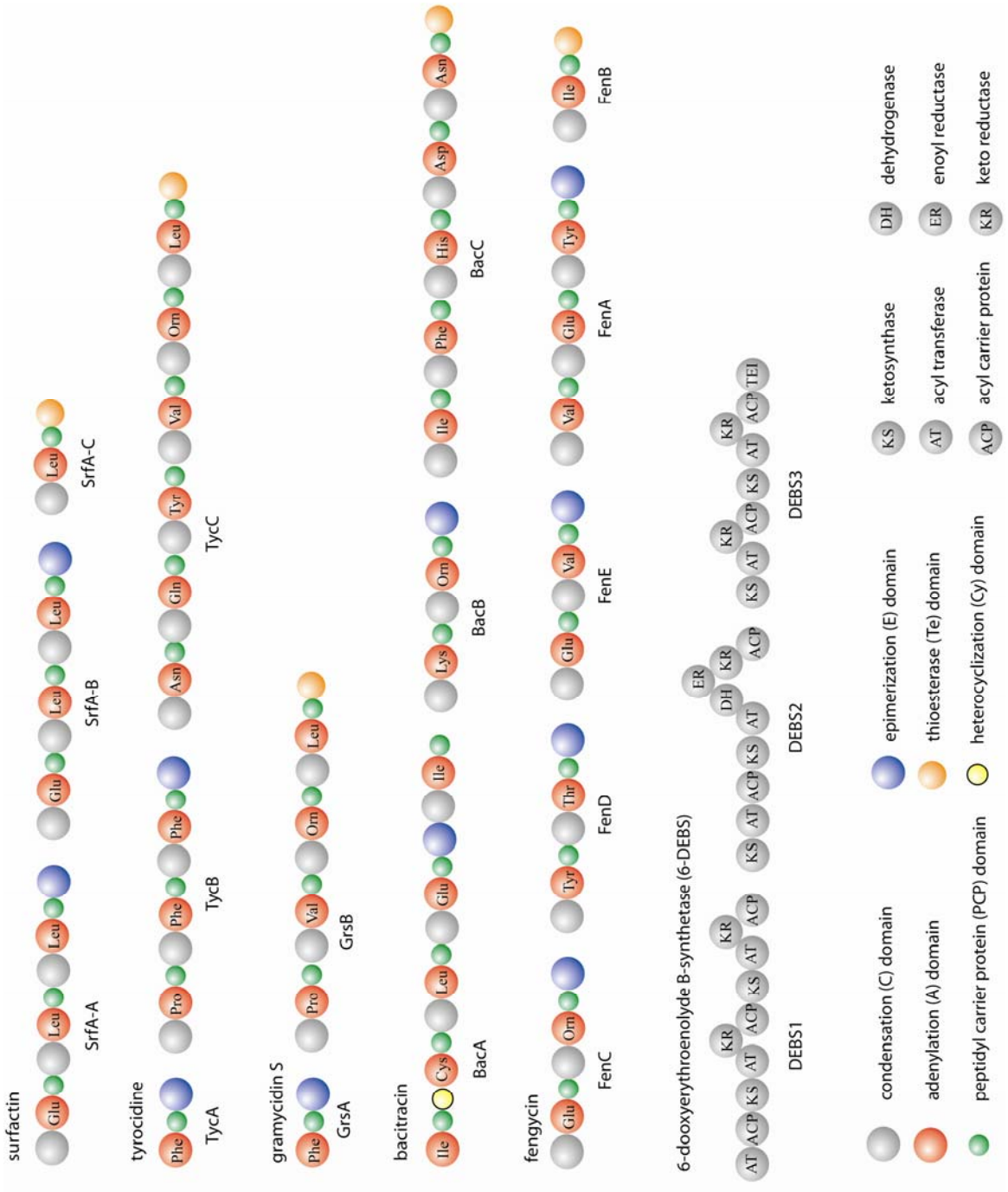


Fig. 59 Biosynthetic systems
Representation of the biosynthetic systems of surfactin, produced by *Bacillus subtilis*, tyrocidine and gramicidin S, both produced by *Bacillus brevis*, bacitracin, produced by *Bacillus licheniformis*, fengicin, produced by *Bacillus subtilis* and 6-deoxyerythroenolide B, produced by *Saccaropolyspora erithrea*.

Literature

Arima, K., Kakinuma, A. and Tamura, G. (1968) Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subitlis*: isolation, characterization and its inhibition of fibrin clot formation. *Biochem Biophys Res Commun*, **31**, 488-494.

Belshaw, P.J., Walsh, C.T. and Stachelhaus, T. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284**, 486-489.

Bennett, J. and Scott, K.J. (1971) Quantitative staining of fraction I protein in polyacrylamide gels using Coomassie brillant blue. *Anal Biochem*, **43**, 173-182.

Bergendahl, V., Thompson, N.E., Foley, K.M., Olson, B.M. and Burgess, R.R. (2003) A cross-reactive polyol-responsive monoclonal antibody useful for isolation of core RNA polymerase from many bacterial species. *Protein Expr Purif*, **31**, 155-160.

Bernheimer, A.W. and Avigad, L.S. (1970) Nature and properties of a cytolytic agent produced by Bacillus subtilis. *J Gen Microbiol*, **61**, 361-369.

Bonmatin JM, G.M., Labbé H, Ptak M. (1994) Solution three-dimensional structure of surfactin: A cyclic lipopeptide studied by ^{<FONT SIZE='-1'>1</FONT>}H-nmr, distance geometry, and molecular dynamics. *Biopolymers*, **34**, 975-986.

Bouvier, J., Pugsley, A.P. and Stragier, P. (1991) A gene for a new lipoprotein in the dapA-purC interval of the Escherichia coli chromosome. *J Bacteriol*, **173**, 5523-5531.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T. and Stubbs, M.T. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure (Camb)*, **10**, 301-310.

Bujard, H., Gentz, R., Lanzer, M., Stueber, D., Mueller, M., Ibrahimi, I., Haeuptle, M.T. and Dobberstein, B. (1987) A T5 promoter-based transcription-translation system for the analysis of proteins in vitro and in vivo. *Methods Enzymol*, **155**, 416-433.

Burgess, R.R. and Thompson, N.E. (2002) Advances in gentle immunoaffinity chromatography. *Curr Opin Biotechnol*, **13**, 304-308.

Casey, J.L., Keep, P.A., Chester, K.A., Robson, L., Hawkins, R.E. and Begent, R.H. (1995) Purification of bacterially expressed single chain Fv antibodies for clinical applications using metal chelate chromatography. *J Immunol Methods*, **179**, 105-116.

Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. (1988) The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene*, **68**, 139-149.

Clough, J., Chen, S., Gordon, E.M., Hackbarth, C., Lam, S., Trias, J., White, R.J., Candiani, G., Donadio, S., Romano, G., Ciabatti, R. and Jacobs, J.W. (2003) Combinatorial

modification of natural products: synthesis and in vitro analysis of derivatives of thiazole peptide antibiotic GE2270 A: A-ring modifications. *Bioorg Med Chem Lett*, **13**, 3409-3414.

Conti, E., Stachelhaus, T., Marahiel, M.A. and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non- ribosomal biosynthesis of gramicidin S. *Embo J*, **16**, 4174-4183.

Cooper, D.G., Macdonald, C.R., Duff, S.J. and Kosaric, N. (1981) Enhanced Production of Surfactin from Bacillus subtilis by Continuous Product Removal and Metal Cation Additions. *Appl Environ Microbiol*, **42**, 408-412.

Cox, R.A. (1968) The use of guanidinium hydrochloride in the isolation of nucleic acids. *Methods Enzymol.*, **12B**, 120-129.

de Ferra, F., Rodriguez, F., Tortora, O., Tosi, C. and Grandi, G. (1997) Engineering of peptide synthetases. Key role of the thioesterase-like domain for efficient production of recombinant peptides. *J Biol Chem*, **272**, 25304-25309.

Doekel, S. and Marahiel, M.A. (2001) Biosynthesis of natural products on modular peptide synthetases. *Metab Eng*, **3**, 64-77.

Du, L., Chen, M., Sanchez, C. and Shen, B. (2000) 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*, **189**, 171-175.

Duerfahrt, T., Doekel, S., Sonke, T., Quaedflieg, P.J. and Marahiel, M.A. (2003) Construction of hybrid peptide synthetases for the production of alpha-l-aspartyl-l-phenylalanine, a precursor for the high-intensity sweetener aspartame. *Eur J Biochem*, **270**, 4555-4563.

Duerfahrt, T., Eppelmann, K., Muller, R. and Marahiel, M.A. (2004) Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. *Chem Biol*, **11**, 261-271.

Egan, W.J., Walters, W.P. and Murcko, M.A. (2002) Guiding molecules towards drug-likeness. *Curr Opin Drug Discov Devel*, **5**, 540-549.

Ehmann, D.E., Trauger, J.W., Stachelhaus, T. and Walsh, C.T. (2000) Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol*, **7**, 765-772.

Eppelmann, K., Doekel, S. and Marahiel, M.A. (2001) Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host Bacillus subtilis. *J Biol Chem*, **276**, 34824-34831.

Eppelmann, K., Stachelhaus, T. and Marahiel, M.A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry*, **41**, 9718-9726.

Fabret, C., Ehrlich, D.S. and Noirot, P. (2001) Positive selection for the loss of a gene in Bacillus subtilis: its application to functional genomics. *poster, Conference of Bacilli, San Diego, USA*.

Fabret, C., Ehrlich, S.D. and Noirot, P. (2002) A new mutation delivery system for genome-scale approaches in Bacillus subtilis. *Mol Microbiol*, **46**, 25-36.

Finking, R. and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol*, **58**, 453-488.

Gehring, A.M., DeMoll, E., Fetherston, J.D., Mori, I., Mayhew, G.F., Blattner, F.R., Walsh, C.T. and Perry, R.D. (1998) Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by Yersinia pestis. *Chem Biol*, **5**, 573-586.

Gentz, R. and Bujard, H. (1985) Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. *J Bacteriol*, **164**, 70-77.

Gevers, W., Kleinkauf, H. and Lipmann, F. (1968) The activation of amino acids for biosynthesis of gramicidin S. *Proc Natl Acad Sci U S A*, **60**, 269-276.

Gocht, M. and Marahiel, M.A. (1994) Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. *J Bacteriol*, **176**, 2654-2662.

Gokhale, R.S., Tsuji, S.Y., Cane, D.E. and Khosla, C. (1999) Dissecting and exploiting intermodular communication in polyketide synthases. *Science*, **284**, 482-485.

Gruenewald, S., Mootz, H.D., Stehmeier, P. and Stachelhaus, T. (2004) In vivo production of artificial nonribosomal peptide products in the heterologous host Escherichia coli. *Appl Environ Microbiol*, **70**, 3282-3291.

Grünewald, J., Sieber, S.A., Mahlert, C., Linne, U. and Marahiel, M.A. (2004) Synthesis and derivatization of daptomycin: a chemoenzymatic route to acidic lipopeptide antibiotics. *J Am Chem Soc*, **126**, 17025-17031.

Guerout-Fleury, A.M., Shazand, K., Frandsen, N. and Stragier, P. (1995) Antibiotic-resistance cassettes for Bacillus subtilis. *Gene*, **167**, 335-336.

Hahn, M. and Stachelhaus, T. (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc Natl Acad Sci U S A*, **101**, 15585-15590.

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

Hann, M.M., Leach, A.R. and Harper, G. (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci*, **41**, 856-864.

Horinouchi, S. and Weisblum, B. (1982a) Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J Bacteriol*, **150**, 815-825.

Horinouchi, S. and Weisblum, B. (1982b) Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibodies. *J Bacteriol*, **150**, 804-814.

Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61-68.

Imai, Y.S., H.; Fujita, T.; Kakinuma, A. 1971, ,. (1971) Hypocholesterolemic effect of surfactin, a novel bacterial peptidelipid. *Takeda Kenkyusho Res. lab*, **30**, 728-734.

Ireton, K.R., D. Z. Siranosian, K. J. Grossman, A. D. (1993) Integartion of multiple developmental signals in *Bacillus subtilis* through the SpoA transcription factor. *Genes Dev.*, **7**, 283-294.

Ishihara, H. and Shimura, K. (1974) Biosynthesis of bacitracin III. Partial purification of a bacitracin-synthesizing enzyme system from *Bacillus licheniformis*. *Biochimica et Biophysica Acta*, **338**, 588-600.

Jain, R.K., Trias, J. and Ellman, J.A. (2003) D-Ala-D-lac binding is not required for the high activity of vancomycin dimers against vancomycin resistant enterococci. *J Am Chem Soc*, **125**, 8740-8741.

Kameda, Y., Matsui, K., Kato, H., Yamada, T. and Sagai, H. (1972) Antitumor activity of Bacillus natto. 3. Isolation and characterization of a cytolytic substance on Ehrlich ascites carcinoma cells in the culture medium of Bacillus natto KMD 1126. *Chem Pharm Bull (Tokyo)*, **20**, 1551-1557.

Keating, T.A., Marshall, C.G., Walsh, C.T. and Keating, A.E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol*, **9**, 522-526.

Khan, F., He, M. and Taussig, M.J. (2006) Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic Acid surfaces. *Anal Chem*, **78**, 3072-3079.

Klein, C., Kaletta, C., Schnell, N. and Entian, K.D. (1992) Analysis of genes involved in biosynthesis of the lantibiotic subtilin. *Appl Environ Microbiol*, **58**, 132-142.

Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A. and Walsh, C.T. (2001) Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry*, **40**, 7099-7108.

Kohli, R.M., Walsh, C.T. and Burkart, M.D. (2002) Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature*, **418**, 658-661.

Konz, D., Klens, A., Schorgendorfer, K. and Marahiel, M.A. (1997) The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol*, **4**, 927-937.

Ku, J., Mirmira, R.G., Liu, L. and Santi, D.V. (1997) Expression of a functional non-ribosomal peptide synthetase module in Escherichia coli by coexpression with a phosphopantetheinyl transferase. *Chem Biol*, **4**, 203-207.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C. and Walsh, C.T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol*, **3**, 923-936.

Lee, S.G., Littau, V. and Lipmann, F. (1975) The relation between sporulation and the induction of antibiotic synthesis and of amino acid uptake in Bacillus brevis. *J Cell Biol*, **66**, 233-242.

Link, A.J., Phillips, D. and Church, G.M. (1997) Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. *J Bacteriol*, **179**, 6228-6237.

Linne, U., Doekel, S. and Marahiel, M.A. (2001) Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. *Biochemistry*, **40**, 15824-15834.

Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **46**, 3-26.

Mahlert, C., Sieber, S.A., Grunewald, J. and Marahiel, M.A. (2005) Chemoenzymatic approach to enantiopure streptogramin B variants: characterization of stereoselective pristinamycin I cyclase from Streptomyces pristinaespiralis. *J Am Chem Soc*, **127**, 9571-9580.

Marahiel, M.A., Stachelhaus, T. and Mootz, H.D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev*, **97**, 2651-2674.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M. and Ashley, G. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci U S A*, **96**, 1846-1851.

Menkhaus, M., Ullrich, C., Kluge, B., Vater, J., Vollenbroich, D. and Kamp, R.M. (1993) Structural and functional organization of the surfactin synthetase multienzyme system. *J Biol Chem*, **268**, 7678-7684.

Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A. and Santi, D.V. (2005) Combinatorial polyketide biosynthesis by de

novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol*, **23**, 1171-1176.

Miller, D.A., Luo, L., Hillson, N., Keating, T.A. and Walsh, C.T. (2002) Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem Biol*, **9**, 333-344.

Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., Gorlach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J. and Ligon, J.M. (2000) The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. *Chem Biol*, **7**, 97-109.

Mootz, H.D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D. and Marahiel, M.A. (2002a) Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J Am Chem Soc*, **124**, 10980-10981.

Mootz, H.D. and Marahiel, M.A. (1997) The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol*, **179**, 6843-6850.

Mootz, H.D., Schorgendorfer, K. and Marahiel, M.A. (2002b) Functional characterization of 4'-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of Saccharomyces cerevisiae lys5. *FEMS Microbiol Lett*, **213**, 51-57.

Mootz, H.D., Schwarzer, D. and Marahiel, M.A. (2000) Construction of hybrid peptide synthetases by module and domain fusions. *Proc Natl Acad Sci U S A*, **97**, 5848-5853.

Mootz, H.D., Schwarzer, D. and Marahiel, M.A. (2002c) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem*, **3**, 490-504.

Nakano, M.M., Marahiel, M.A. and Zuber, P. (1988) Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in Bacillus subtilis. *J Bacteriol*, **170**, 5662-5668.

Neumüller, A. (2001) Untersuchungen zur Bacitracin-Selbst-Resistenz in *Bacillus licheniformis* ATCC 10716 und Grundlagen zur gerichteten Protein-Evolution von Adenylierungs-Domänen aus Peptidsynthetasen. *Dissertation, Fachbereich Chemie, Philipps Universität Marburg.*

Nicolaou, K.C., Cho, S.Y., Hughes, R., Winssinger, N., Smethurst, C., Labischinski, H. and Endermann, R. (2001) Solid- and solution-phase synthesis of vancomycin and vancomycin analogues with activity against vancomycin-resistant bacteria. *Chemistry*, **7**, 3798-3823.

Oprea, T.I., Davis, A.M., Teague, S.J. and Leeson, P.D. (2001) Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci*, **41**, 1308-1315.

Ortholand, J.Y., Ganesan A. (2004) Natural produtcs and combinatorial chemistry: back to the future. *Current Opinion in Chemical Biology*, **8**, 271-280.

Page M.GP, W.C. (2005) Antimicrobials. *Current Opinion in Microbiology*, **8**, 459-497.

Peypoux, F., Bonmatin, J.M. and Wallach, J. (1999) Recent trends in the biochemistry of surfactin. *Appl Microbiol Biotechnol*, **51**, 553-563.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. and Khosla, C. (2001) Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. *Science*, **291**, 1790-1792.

Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P. and Walsh, C.T. (1998) Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry*, **37**, 1585-1595.

Raja, A., LaBonte, J., Lebbos, J. and Kirkpatrick, P. (2003) Daptomycin. *Nat Rev Drug Discov*, **2**, 943-944.

Razafindralambo, H., Popineau, Y., Deleu, M., Hbid, C., Jacque, P., Thonart, P. and Paquot, M. (1998) Foaming properties of lipopeptides produced by *Bacillus subtilis*: effect of lipid and peptide structural attributes. *J Agric Food Chem*, **46**, 911-916.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Lab. Press, NY.

Schauwecker, F., Pfennig, F., Grammel, N. and Keller, U. (2000) Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. *Chem Biol*, **7**, 287-297.

Schneider, A. (1996) Nichtribosomale Peptidsynthese: Modulaustausch und Funktion der Thioesterase-ähnlichen Proteine. *Dissertation, Fachbereich Biologie, Philipps-Universität Marburg, Deutschland*.

Schneider, A. and Marahiel, M.A. (1998) 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*, **169**, 404-410.

Schneider, A., Stachelhaus, T. and Marahiel, M.A. (1998) Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Mol Gen Genet*, **257**, 308-318.

Schwarz E, S.G., Hobom G, Kossel H. (1987) Nucleotide sequence of cro, cII and part of the O gene in phage lambda DNA. *Nature*, **30**, 410-414.

Schwarzer, D., Mootz, H.D., Linne, U. and Marahiel, M.A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A*, **99**, 14083-14088.

Shaw-Reid, C.A., Kelleher, N.L., Losey, H.C., Gehring, A.M., Berg, C. and Walsh, C.T. (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem Biol*, **6**, 385-400.

Sieber, S.A., Tao, J., Walsh, C.T. and Marahiel, M.A. (2004) Peptidyl thiophenols as substrates for nonribosomal peptide cyclases. *Angew Chem Int Ed Engl*, **43**, 493-498.

Sieber, S.A., Walsh, C.T. and Marahiel, M.A. (2003) Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J Am Chem Soc*, **125**, 10862-10866.

Smith, D.J., Burnham, M.K., Edwards, J., Earl, A.J. and Turner, G. (1990) Cloning and heterologous expression of the penicillin biosynthetic gene cluster from penicillum chrysogenum. *Biotechnology (N Y)*, **8**, 39-41.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, **98**, 503-517.

Stachelhaus, T. and Marahiel, M.A. (1995a) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J Biol Chem*, **270**, 6163-6169.

Stachelhaus, T. and Marahiel, M.A. (1995b) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J Biol Chem*, **270**, 6163-6169.

Stachelhaus, T., Mootz, H.D., Bergendahl, V. and Marahiel, M.A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J Biol Chem*, **273**, 22773-22781.

Stachelhaus, T., Mootz, H.D. and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol*, **6**, 493-505.

Stachelhaus, T., Schneider, A. and Marahiel, M.A. (1995) Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science*, **269**, 69-72.

Stachelhaus, T., Schneider, A. and Marahiel, M.A. (1996) Engineered biosynthesis of peptide antibiotics. *Biochem Pharmacol*, **52**, 177-186.

Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R. and Morris, H.R. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J Biol Chem*, **271**, 15428-15435.

Stein, T., Vater, J., Kruft, V., Wittmann-Liebold, B., Franke, P., Panico, M., Mc Dowell, R. and Morris, H.R. (1994) Detection of 4'-phosphopantetheine at the thioester binding site for L-valine of gramicidinS synthetase 2. *FEBS Lett*, **340**, 39-44.

Steinmetz, M. and Richter, R. (1994) Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in Bacillus subtilis, through in vivo recombination. *Gene*, **142**, 79-83.

Steller, S., Sokoll, A., Wilde, C., Bernhard, F., Franke, P. and Vater, J. (2004) Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry*, **43**, 11331-11343.

Stüber, D., Matile, H. & Garotta, G. (1990) *System for high level production in Escherichia coli and rapid purification of recombinant proteins: application to epitope mapping, preparation of antibodies and structure-function analysis. In Immunological methods pp.121-152*. I. Levkovitz & B. Pernis.

Sutcliffe, J.G. (1979) Complete nucleotide sequence of the Escherichia coli plasmid pBR322. *Methods Enzymol.*, **43**, 77-90.

Symmank, H., Saenger, W. and Bernhard, F. (1999) Analysis of engineered multifunctional peptide synthetases. Enzymatic characterization of surfactin synthetase domains in hybrid bimodular systems. *J Biol Chem*, **274**, 21581-21588.

Symmank, H.F., P. Saenger, W. Bernahrd, F. (2002) Modification of biologically active peptides: production of a novel lipohexapeptide after engineering of *Bacillus subtilis* surfactin synthetase. *Protein Eng*, **15**, 913-921.

Thompson, N.E., Aronson, D.B. and Burgess, R.R. (1990) Purification of eukaryotic RNA polymerase II by immunoaffinity chromatography. Elution of active enzyme with protein stabilizing agents from a polyol-responsive monoclonal antibody. *J Biol Chem*, **265**, 7069-7077.

Thompson, N.E., Arthur, T.M. and Burgess, R.R. (2003) Development of an epitope tag for the gentle purification of proteins by immunoaffinity chromatography: application to epitope-tagged green fluorescent protein. *Anal Biochem*, **323**, 171-179.

Thompson, N.E., Hager, D.A. and Burgess, R.R. (1992) Isolation and characterization of a polyol-responsive monoclonal antibody useful for gentle purification of Escherichia coli RNA polymerase. *Biochemistry*, **31**, 7003-7008.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A*, **76**, 4350-4354.

Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. and Walsh, C.T. (2000) Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature*, **407**, 215-218.

Trauger, J.W., Kohli, R.M. and Walsh, C.T. (2001) Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry*, **40**, 7092-7098.

Trauger, J.W. and Walsh, C.T. (2000) Heterologous expression in Escherichia coli of the first module of the nonribosomal peptide synthetase for chloroeremomycin, a vancomycin-type glycopeptide antibiotic. *Proc Natl Acad Sci U S A*, **97**, 3112-3117.

Trieu-Cuot, P. and Courvalin, P. (1983) Nucleotide sequence of the Streptococcus faecalis plasmid gene encoding the 3'5"-aminoglycoside phosphotransferase type III. *Gene*, **23**, 331-341.

Tsukagoshi, N., Tamura, G. and Arima, K. (1970) A novel protoplast-bursting factor (surfactin) obtained from Bacillus subtilis IAM 1213. I. The effects of surfactin on bacillus megaterium KM. *Biochim Biophys Acta*, **196**, 204-210.

van Wageningen, A.M., Kirkpatrick, P.N., Williams, D.H., Harris, B.R., Kershaw, J.K., Lennard, N.J., Jones, M., Jones, S.J. and Solenberg, P.J. (1998) Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol*, **5**, 155-162.

Vollenbroich, D., Özel, M., Vater, J., Kamp, R.M. and Pauli, G. (1997) Mechanism of inactivation of enveloped viruses by the biosurfactant surfactin from *Bacillus subtilis. Biologicals*, **25**, 289-297.

Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E. and Flugel, R.S. (1997) Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr Opin Chem Biol*, **1**, 309-315.

Weber, G., Schorgendorfer, K., Schneider-Scherzer, E. and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet*, **26**, 120-125.

Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A. and Holak, T.A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure Fold Des*, **8**, 407-418.

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z. and Graham, M.W. (1989) Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res*, **17**, 3469-3478.

Wu, N., Cane, D.E. and Khosla, C. (2002) Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry*, **41**, 5056-5066.

Xu, R., Greiveldinger Guy , Marenus Linda E., Cooper Alan, and Ellman Jonathan A. (1999) Combinatorial Library Approach for the Identification of Synthetic Receptors Targeting Vancomycin-Resistant Bacteria. *J Am Chem Soc*, **121**, pp 4898 - 4899.

Xu Ruo, G.G., Marenus Linda E., Cooper Alan, and Ellman Jonathan A. (1999) Combinatorial Library Approach for the Identification of Synthetic Receptors Targeting Vancomycin-Resistant Bacteria. *J Am Chem Soc*, **121**, pp 4898 - 4899.

Xue, Q., Ashley, G., Hutchinson, C.R. and Santi, D.V. (1999) A multiplasmid approach to preparing large libraries of polyketides. *Proc Natl Acad Sci U S A*, **96**, 11740-11745.

Yakimov, M.M., Giuliano, L., Timmis, K.N. and Golyshin, P.N. (2000) Recombinant acylheptapeptide lichenysin: high level of production by Bacillus subtilis cells. *J Mol Microbiol Biotechnol*, **2**, 217-224.

Yeh, E., Kohli, R.M., Bruner, S.D. and Walsh, C.T. (2004) Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem*, **5**, 1290-1293.

Acknowledgements

I would like to thank Dr. Torsten Stachelhaus for his excellent scientific supervision during my PhD thesis. His constant support allowed me to solve out many scientific problems and accomplish this thesis. I would also like to thank him for giving me the chance to participate in interesting international meetings and present my work to an international audience.
I would like to thank Prof. Dr. Mohamed Marahiel for his constant interest in my work and for supervising my PhD thesis during the last year, in particular for giving me the possibility to become a member of the Graduiertenkolleg “Proteinfunktion auf atomarer Ebene” and for being in my thesis committee.
Thank you also to Prof. Dr. Wolfgang Buckel for accepting the external supervision of my PhD thesis and for being in my thesis committee.

I would like to thank Dr.Uwe Linne for his fundamental help with HPLC and high resolution mass spectrometry and also Frank Wiegeshoff for his help with computer problems.
I would like to thank Dr.Stephan Grünewald for providing the vector pQE61-tycA/(C1)tycB1, and Dr.Sascha Dökel and Dr.Alexandra Mees for providing the *B. subtilis* strain AM1.
I am grateful to all members of the Stachelhaus and Marahiel working groups for their collaboration and for the nice atmosphere. I would like to thank the former group members Dr. Katrin Eppelmann, Dr. Stephan Grünewald, Dr. Martin Hahn, Dr. Jürgen May, Dr. Richard Nadvornik, and Martina Pfeiff for their generous help as well as for the good time inside and outside the lab. I would also like to thank, Karen Hunger, Björn Wagner and Frank Wiegeshoff for the constructive scientific discussions and for the nice free time spent together.
I am grateful to the technical assistants Antje Schäfer, Christiane Bomm and Gabi Schimpff Weiland for preparation of the culture media and for helping to solve out technical problems.

I am indebted to Dr. Katrin Eppelmann and Dr. Torsten Stachelhaus for the critical and fast proof reading of this manuscript.

I would like to thank my boyfriend Martin von Meltzer for his great moral support, for motivating me in difficult moments and for his precious contribution to the graphical part of this manuscript.

I am grateful to Alex Peter and Sigrid Marahiel, for helping me in learning the German language, making my first year in a completely new environment enjoyable.

I would like to thank my parents and my brother, who always supported me in the course of my PhD sharing with me nice and difficult moments. I wish to dedicate this thesis to them.