Chemistry & Biology 13, 899–908, August 2006 DOI 10.1016/j.chembiol.2006.06.015

In Vivo Biocombinatorial Synthesis of Lipopeptides by COM Domain-Mediated Reprogramming of the Surfactin Biosynthetic Complex

Claudia Chiocchini,[1] Uwe Linne,[1] and Torsten Stachelhaus[1,2,*]
[1] Philipps-University Marburg
Faculty of Chemistry/Biochemistry
Hans-Meerwein-Strasse
D-35032 Marburg
Germany

Summary

The intermolecular communication within NRPS complexes relies on the coordinated interplay of donor and acceptor communication-mediating (COM) domains. In this study, the potential of COM domains was exploited in vivo by establishing a system for the true biocombinatorial synthesis of lipopeptides via directed reprogramming of a natural NRP biosynthetic assembly line (i.e., surfactin). By means of COM domain swapping, these experiments verified the decisive role of COM domains for the control of protein-protein interactions between NRPSs, demonstrated the functionality of COM domain pairs even in the context of a heterologous host and NRPS system, and allowed for the intended skipping of a biosynthetic enzyme within a multienzymatic biosynthetic complex. Ultimately, abrogation of the selectivity barrier provided by COM domains afforded the successful simultaneous, biocombinatorial synthesis of distinct lipopeptide products.

Introduction

Nonribosomal peptides (NRPs) and polyketides (PKs) represent two diverse groups of pharmacologically important natural products, which are synthesized by large multimodular assembly lines, termed nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), respectively [1]. Prominent examples include antibiotics like vancomycin and erythromycin, antitumor drugs like bleomycin and epothilone, and immunosuppressants like cyclosporine and FK506. According to the molecular logic employed by NRPSs and PKSs, the biosynthesis of a defined product relies on the selectivity of individual modules and their coordinated interplay. In multienzyme complexes, which actually stand for the vast majority of known assembly lines, biosynthesis also requires the proper interaction between partner enzymes and concomitantly prevention of undesired interactions between nonpartner enzymes.

In the latter context, correct channeling of reaction intermediates along an assembly line is provided, at least for the most part, by short terminal structures, referred to as PKS docking domains [2, 3] and NRPS communication-mediating (COM) domains [4, 5]. In various studies, it has been shown that matching pairs of donor and acceptor domains promote the correct positioning of enzymes within multienzyme complexes and the selective translocation of intermediates between adjacent synth(et)ases. The NMR structure of a cognate pair of donor and acceptor docking domains revealed that in the case of the erythromycin biosynthetic system, the complex contains two distinct, noninteracting four-α-helix bundles [6]. Interestingly, based on secondary structure predictions, NRPS COM domains are likewise believed to possess α-helical structures that may provide the interfaces for the selective differentiation between partner and nonpartner NRPSs [5].

Despite obvious similarities like terminal localization and function, there are many striking differences between NRPS COM domains and PKS docking domains: first, the NMR structure suggests that C- and N-terminal PKS docking domains comprise 80–100 and 20–30 amino acid residues, respectively [6]. In contrast, based on domain swapping experiments, donor and acceptor NRPS COM domains appear to be significantly shorter, comprising only 15–30 amino acid residues, respectively [4, 5]. Second, it has been shown that type I PKSs are composed of two identical subunits [7, 8]. This organization was also reflected in the NMR structure of the PKS docking domain complex [6], suggesting that dimerization represents an important aspect of protein-protein communication between PKSs. Interestingly, sound evidence has been provided that NRPSs are monomeric and do not form dimers [9]. This suggests that the solved structure of the PKS docking domain complex does not apply for NRPSs and that a complex of donor and acceptor COM domains must be different from the two four-α-helix bundles, determined for PKSs. Third, in vitro studies have demonstrated that nonmatching pairs of docking domains between natural partner PKSs are causing a decrease in the rate of product formation [10, 11]. In contrast, comparable experiments for NRPS systems revealed that product formation was not just impaired, but rather completely ceased, when donor and acceptor enzymes were equipped with noncompatible COM domains [4, 5]. All these examples clearly indicate that NRPS COM domains and PKS docking domains represent more than just two variations of the same theme.

As for NRPSs, in vitro studies have shown that a donor COM domain COM^{D}_{X}, situated at the C terminus of an aminoacyl- or peptidyl-donating NRPS "X," and an acceptor COM domain COM^{A}_{Y}, located at the N terminus of the accepting partner enzyme "Y," form a compatible (cognate) pair that is crucial for establishing the productive interaction between both enzymes [4, 5]. In contrast, within a hypothetical assembly line "X-Y-Z," the COM domains COM^{D}_{X} and COM^{A}_{Z} of the nonpartner NRPSs "X" and "Z" are considered incompatible (noncognate), preventing their futile contact. Accordingly, the establishment of a defined assembly line and synthesis of a distinct NRP product is ensured by the grouping of exclusively cognate pairs of COM domains.

*Correspondence: torsten.stachelhaus@aureogen.com
[2] Present address: AureoGen Biosciences, Inc., 4717 Campus Drive, Kalamazoo, MI 49009.

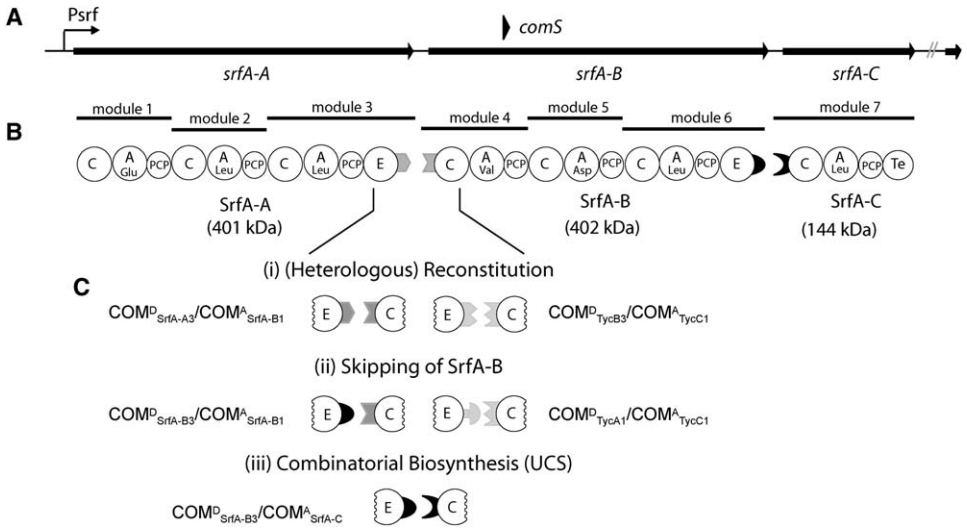


Figure 1. The Surfactin Biosynthetic System

The enzymatic assembly line of surfactin consists of three NRPSs, encoded by the polycistronic genes *srfA-ABC* (A). The synthetases SrfA-A, SrfA-B, and SrfA-C are composed of three, three, and one module(s), respectively (B). Two cognate pairs of COM domains, COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B} and COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}, facilitate the selective interaction between partner enzymes. Within this study, the first COM domain pair was replaced by means of domain swapping on the genetic level against different cognate, miscognate, and noncognate sets of COM domains (C).

Without the selectivity provided by different pairs of cognate COM domains, the enzymes of a given NRP biosynthetic system could arrange randomly, which eventually would lead to the synthesis of a vast array of peptide products. A corresponding abrogation of this selectivity barrier can be achieved by equipping all donor and acceptor enzymes of a multienzyme biosynthetic system with the same (cognate) donor and acceptor COM domains, respectively. A corresponding "universal communication system" (UCS) was recently established in vitro, yielding the expected and intended combinatorial biosynthesis of NRPs [5].

The presented study aimed on harnessing the biocombinatorial potential of COM domains for the first time in vivo and within the context of natural multienzyme NRP biosynthetic complex. To this end, different pairs of COM domains were exploited, in order to achieve a specific reprogramming of a biosynthetic assembly line and to establish a system for true combinatorial NRP biosynthesis. As a model system, the surfactin biosynthetic complex of the well-studied gram-positive bacterium *Bacillus subtilis* was chosen. As shown in Figure 1, the biosynthetic assembly line consists of three NRPSs (SrfA-A, SrfA-B, and SrfA-C). Thus, taken into consideration the described model for the control of protein-protein interactions, the surfactin biosynthetic complex was proposed to contain two cognate pairs of COM domains, COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B} and COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}, facilitating the selective interaction of the three synthetases and, hence, the specific formation of the lipoheptapeptide antibiotic.

We report herein the substitution of the first COM domain pair COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}, facilitating the selective interaction between SrfA-A and SrfA-B, against various cognate, miscognate, and noncognate sets of COM domains. The consequences of the corresponding in vivo COM domain swaps were investigated by means of HPLC, HPLC/MS, and high-resolution MS analysis.

Results

General Considerations for the Generation of In Vivo COM Domain Swaps

The reprogramming of the surfactin biosynthetic complex was accomplished via a well-established two-step genetic marker exchange method [12–14]. The approach uses consecutive gene disruption and reconstitution, which are both monitored by a selectable marker (Figure 2). However, for the model system under investigation, exploitation of this procedure required some special considerations.

First, the integration of a selectable marker (*mls*) at the transition between *srfA-A* and *srfA-B* not only causes the envisioned disruption of the *srfA* biosynthetic operon but also disables the coexpression of the small competence regulator gene *comS*, situated in a different reading frame within the coding region of *srfA-B* (see Figure 1) [15]. The corresponding gene product (ComS) has been shown to be essential for the establishment of the host organism's natural competence [16]. In order to maintain the genetic competence required for the second transformation step, a second copy of *comS* was integrated into a different genetic locus (*amyE*) within the chromosome of *B. subtilis* ATCC 21332. The resulting *B. subtilis* strain AM1 was then transformed with the knockout plasmid pCC63 to give the MLS resistant *srfA* disruption strain *B. subtilis* CC64. The genetic integrity of both constructs, AM1 and CC64, was verified by Southern blot and PCR analysis (data not shown).

Both strains, AM1 and CC64, were also tested for surfactin production. To this end, butanolic extracts of the cultured broths of the two *Bacillus* strains were analyzed by RP-HPLC and FT-ICR-MS, enabling even the

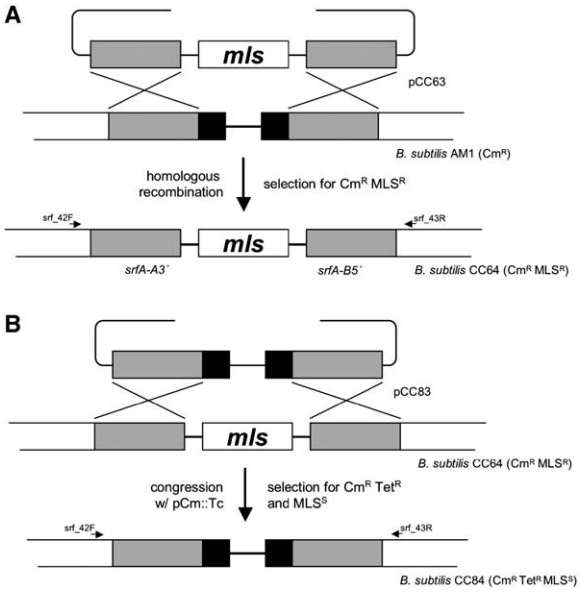


Figure 2. COM Domain Swapping

Transformation of *B. subtilis* AM1 with pCC63 led to homologous recombination and disruption of the *srfA* operon (A). The resulting strain CC64 was then cotransformed with the helper plasmid pCm::Tc and the reconstitution plasmid (e.g., pCC83) to generate the desired COM domain swap (B).

identification of trace amounts of lipopeptide products. For AM1, these tests revealed that the integration of a second copy of the *comS* gene had—as anticipated—no influence on lipoheptapeptide production (Figure 3A). As shown, surfactin represents a mixture of cyclic lipopeptides, predominantly consisting of the peptide core LGlu-LLeu-DLeu-LVal-LAsp-DLeu-LLeu, and a β-hydroxyl fatty acid with chain lengths of 12 to16 carbon atoms (Figure 3A, chemical structure). Given their slight differences in retention time, RP-HPLC analysis of natural surfactin produces a signature finger pattern, with the main peaks being derivatives with a β-hydroxyl fatty acid, constituted of 13 (Figure 3A, structure II: n = 2) and 14 (n = 3) carbon atoms, respectively. In contrast, in the case of disruption mutant CC64, the analyses unequivocally verified the expected surfactin-deficient phenotype.

The second, potential problem concerned the reconstitution of the *srfA* operon. In fact, transformation of the *srfA* disruption construct *B. subtilis* CC64 with a given COM domain swapping plasmids (i.e., pCC83) not only led to the desired reconstitution of the biosynthetic system but also the loss of the *mls* resistance cassette (Figure 2). Regrettably, there is no possibility to directly select for the loss of this genetic marker. In order to still allow for a positive selection, congression experiments were carried out with a given reconstitution plasmid along with the self-replicable helper plasmid pCm::Tc, mediating tetracyline resistance [17]. Following cotransformation, cells were first selected for the presence of the helper plasmid and only then screened for the desired loss of the *mls* resistance cassette. The genetic integrity of corresponding constructs was analyzed by PCR and verified by DNA sequencing of an about 600 bp region at the transition between *srfA-A* and *srfA-B*. After verification of the genotype (data not shown), the strains were tested for the production of lipopeptide products as described above.

Reconstitution of the Wild-Type *srfA* Operon; $COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}B}$

In a first experiment, the *srfA* operon was reconstituted with the coding fragment of the native COM domain pair $COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}B}$, hereby restoring the wild-type situation observed in *B. subtilis* strains ATCC 21332 and AM1. Despite of being a proof-of-concept, the resulting construct CC84 was important to verify that the utilized cloning strategy (see Experimental Procedures) had no effect on the productivity of the surfactin biosynthetic system. In fact, the coding fragment of a given COM domain pair was amplified by PCR and cloned into the integration vector pCC78, carrying the 5′ and 3′ homologous regions for the marker exchange recombination (see Figure 2). This cloning step was performed in a way that ensured the maintenance of the primary

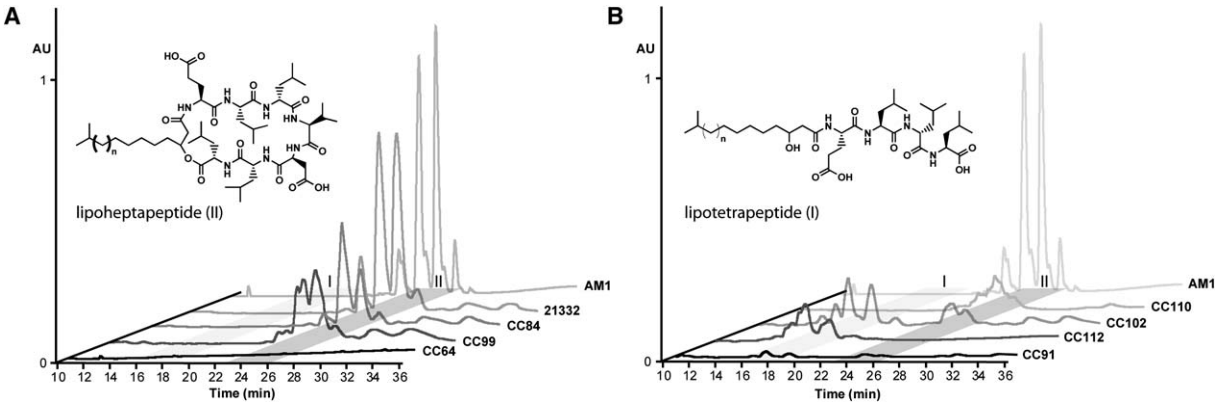


Figure 3. RP-HPLC Analysis of Butanolic Extracts Derived from *B. subtilis* Wild-Type and Mutant Strains

(A) ATCC 21332 ($COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}B}$), AM1 (second copy of *comS*; $COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}B}$), CC64 (disruption mutant), CC84 ($COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}B}$), and CC99 ($COM^{D}_{TycB}/COM^{A}_{TycC}$).

(B) AM1 ($COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}B}$), CC112 ($COM^{D}_{TycA}/COM^{A}_{TycC}$), CC102 ($COM^{D}_{SrfA\text{-}B}/COM^{A}_{SrfA\text{-}B}$), CC91 ($COM^{D}_{SrfA\text{-}B}/COM^{A}_{SrfA\text{-}C}$), and CC110 ($COM^{D}_{SrfA\text{-}A}/COM^{A}_{SrfA\text{-}C}$). Both lipotetrapeptides (I, light gray area) and lipoheptapeptides (II, dark gray area) were identified and characterized by coupled HPLC/ESI-MS and FT-ICR MS analysis (Figure 4).

Table 1. Comparison of Determined and Estimated Production Levels

Bacillus subtilis Strain	COM Domain Pair	Lipoheptapeptide: Area of Absorbance at 214 nm (in AU)[a]	Lipoheptapeptide: Production Level (in % of wt)[b]	Lipotetrapeptide: Area of Absorbance at 214 nm (in AU)[a]	Lipotetrapeptide: Estimated Production Level (in % of wt)[b]	Comments/Description
ATCC 21332	$COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$	1460 ± 80	100 (wt)	n.d.	n.d.	wild-type
AM1	$COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$	1600 ± 100	110	n.d.	n.d.	second copy of *comS*
CC64	none	n.d.	n.d.	n.d.	n.d.	*srfA* disruption
CC84	$COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$	1110 ± 100	76	n.d.	n.d.	*srfA* reconstitution
CC99	$COM^{D}_{TycB}/COM^{A}_{TycC}$	1010 ± 80	69	n.d.	n.d.	heterologous COM
CC112	$COM^{D}_{TycA}/COM^{A}_{TycC}$	n.d.	n.d.	1070 ± 50	(*73*)[c]	skipping
CC102	$COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$	260 ± 20	18	550 ± 40	(*38*)[c]	skipping vs. UCS
CC91	$COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$	40 ± 10	3	70 ± 10	(*5*)[c]	UCS
CC110	$COM^{D}_{SrfA-A}/COM^{A}_{SrfA-C}$	380 ± 30	26	n.d.	n.d.	non- versus miscognate

n.d., not detected.

[a] The areas of absorbance at 214 nm were determined for at least three individual experiments by using the implemented ChemStation software provided with the HPLC instrument. To compare the productivity per cell and take into account possible differences in growing behavior, the determined values were normalized for the OD_{600nm} of the cultures, used for the butanolic extraction. AU, arbitrary units.

[b] Production levels of mutant strains were determined by comparing the areas of absorbance at 214 nm of wild-type and mutant. The production level calculated for the wild-type was set to 100%.

[c] The comparison assumes identical molar extinction coefficients at 214 nm for both lipotetra- and lipoheptapeptide. Since this is certainly not the case, the given values for the lipotetrapeptide represent only estimations. The analytic procedure was enforced by the lack of an authentic standard and is likely to underestimate the amount of lipotetrapeptide produced.

sequences of the conserved core motifs "TPSD" and "QEGMLYH," which were used as fusion points. This, in turn, required the introduction of three silent point mutations. As shown in Figure 3A and Table 1, the analysis of *B. subtilis* CC84 revealed that the reconstitution construct was able to produce nearly wild-type levels of the lipoheptapeptide surfactin, indicating that the genetic manipulations neither resulted in a drop of product titer nor a change in product distribution.

Activity of Heterologous COM Domains; $COM^{D}_{TycB}/COM^{A}_{TycC}$

In order to address the question whether a heterologous COM domain pair could replace the native COM domain pair $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$, the functionality and performance of $COM^{D}_{TycB}/COM^{A}_{TycC}$, derived from the tyrocidine biosynthetic complex of *B. brevis* ATCC 8185, was tested. In vitro studies showed that COM^{D}_{TycB} and COM^{A}_{TycC} are forming a cognate COM domain pair, facilitating the productive interaction between the natural partner NRPSs TycB and TycC [4]. Furthermore, the COM domain pair could be fused to different donor and acceptor enzymes, still facilitating their productive interaction. On the other hand, the donor COM domain COM^{D}_{TycB} is not compatible with the heterologous acceptor COM domain COM^{A}_{SrfA-C} of SrfA-C. Taking all this into consideration, the introduction of the coding fragment of $COM^{D}_{TycB}/COM^{A}_{TycC}$ into the chromosome of the *srfA* disruption strain *B. subtilis* CC64 was expected to yield a restoration of surfactin production.

After verification of the correct genotype (data not shown), the resulting strain *B. subtilis* CC99 was analyzed for surfactin production. As verified by HPLC, integration of the heterologous COM domain pair $COM^{D}_{TycB}/COM^{A}_{TycC}$ only caused a minor decrease in productivity (Figure 3A and Table 1). The product titer of 69% (compared to the wild-type) indicated that the cognate COM domains do very well maintaining their activity and mutual affinity, even in the context of a heterologous host and NRPS system.

Skipping of Enzymes within an NRPS Complex; $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$ versus $COM^{D}_{TycA}/COM^{A}_{TycC}$

COM domains can be exploited to enforce the interaction between natural nonpartner enzymes and to prevent the contact between natural partner enzymes [4, 5]. With regard to the targeted engineering of NRP biosynthetic assembly line, this concept was challenged by attempting the controlled in vivo skipping of the second NRPS, SrfA-B, and—at the same time—establishing a productive interaction between the trimodular initiation enzyme SrfA-A and the monomodular termination enzyme SrfA-C (see Figure 1). This was expected to cause the rational formation of a shortened lipotetrapeptide product. Two different strategies were pursued in order to achieve the intended skipping of SrfA-B.

Based on in vitro studies, COM^{D}_{TycA} and COM^{A}_{TycC} have been shown to form a noncognate COM domain pair [4, 5]. Thus, no productive interaction between donor and acceptor enzymes, carrying these two COM domains, is observed. COM^{D}_{TycA}, however, has been shown to establish a miscognate interaction with COM^{A}_{SrfA-C}, allowing for a crosstalk between enzymes carrying these two COM domains, even if they are derived from two different biosynthetic systems [4]. In vivo, when replacing the native COM domain pair $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$, the noncognate pair $COM^{D}_{TycA}/COM^{A}_{TycC}$ should consequently facilitate a controlled skipping of the trimodular elongation enzyme SrfA-B and a selective interaction between SrfA-A and SrfA-C. The latter should be detectable by the formation of the shortened lipotetrapeptide product FA-LGlu-LLeu-DLeu-LLeu-OH.

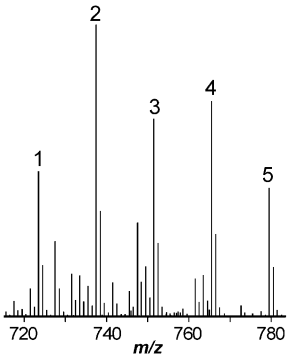
Provided that the established working model would also apply to surfactin biosynthetic system, the COM domain pairs $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$ and $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$ should represent cognate pairs of COM domains, facilitating the selective interaction between SrfA-A and SrfA-B, as well as SrfA-B and SrfA-C, respectively. On the other hand, the COM domain pairs $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-C}$ and $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$ had to be considered as noncognate. Hence, the latter

pair $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$ was likewise expected to facilitate a controlled skipping of the elongation module SrfA-B, and a direct interaction between SrfA-A and SrfA-C, when substituting the native $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$ at the transition between SrfA-A and SrfA-B.

The corresponding COM domain swaps were generated, and the integrity of the resulting *B. subtilis* strains CC112 ($COM^{D}_{TycA}/COM^{A}_{TycC}$) and CC102 ($COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$) verified as described above (data not shown). Subsequently, the lipopeptide production was analyzed. For *B. subtilis* CC112, these tests revealed the expected formation of the lipotetrapeptide FA-LGlu-LLeu-DLeu-LLeu-OH (Figure 3B and Figure 4). As in the case of the lipoheptapeptide, HPLC analysis revealed a typical finger pattern, representative for lipopeptides with slight variations in the length of the β-hydroxyl fatty acid (see Figure 3B, chemical structure of lipotetrapeptide). The identity of lipotetrapeptide was verified by high-resolution FT-ICR analysis (Figure 4).

The accurate quantification of the novel product was barred, due to the lack of an authentic standard. In order to overcome this limitation and still permit at least an estimation of production levels, the areas of absorbance of the corresponding UV signals at 214 nm were used for the comparison of lipohepta- and lipotetrapeptide production (Table 1). It should be noted that all production levels provided refer to the comparison of areas of absorbance at 214 nm, rather than real quantities of the corresponding lipopeptide products. It should be stressed that this procedure only permits an estimation of production levels, which might underestimate the actual amount of the lipotetrapeptide, given its fewer number of absorbing peptide bonds.

According to this analysis, the production level of lipotetrapeptide in *B. subtilis* CC112 ($COM^{D}_{TycA}/COM^{A}_{TycC}$) was estimated to be in the same order of magnitude (73%) (Table 1) as observed for the full-length lipoheptapeptide within the wild-type producer strain. The same examination for *B. subtilis* CC102 ($COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$) likewise revealed the formation of the expected lipotetrapeptide product at essentially the same production level (38%) (Figure 3B and Table 1). Surprisingly, however, the latter analysis also revealed the simultaneous formation of the full-length lipoheptapeptide surfactin, although at a reduced product titer (18%). This indicates that the communication between SrfA-A and SrfA-B was not completely abrogated, and that COM^{D}_{SrfA-B} and COM^{A}_{SrfA-B} are forming a miscognate, rather than the expected noncognate COM domain pair. In the context of the hybrid NRPS system under investigation (see Figure 1), this miscognate COM domain pair allows for the productive interaction between SrfA-A and SrfA-B, as well as SrfA-A and SrfA-C, and eventually formation of both lipopeptide products.



Peak	Molecular Formula	Mol. Mass (calc.)	Mol. Mass (obs.)
1	n=1; $C_{35}H_{64}N_4O_9K$	723.431	723.430
2	n=2; $C_{36}H_{66}N_4O_9K$	737.447	737.446
3	n=3; $C_{37}H_{68}N_4O_9K$	751.462	751.462
4	n=4; $C_{38}H_{70}N_4O_9K$	765.478	765.477
5	n=5; $C_{39}H_{72}N_4O_9K$	779.493	779.493

Figure 4. FT-ICR MS Analysis

High-resolution FT-ICR MS unequivocally revealed the identity of the lipotetrapeptides FA-LGlu-LLeu-DLeu-LLeu-OH. For the sake of clarity, only the $[M+K^+]$ peaks were annotated.

Combinatorial In Vivo Biosynthesis; $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$

The present data demonstrate the inherent potential of COM domains for the directed reprogramming of the NRP assembly lines. The ultimate goal of such manipulations, however, would be a biocombinatorial synthesis of NRPs, which should be achievable by suspending the selectivity barrier, provided by COM domains. Such abrogation, as has been recently described in vitro on the basis of a so-called universal communication system (UCS) [5]. In the context of the surfactin biosynthetic complex, a similar UCS can be established, e.g., by integration of the cognate COM domain pair $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$ at the transition between the SrfA-A and SrfA-B. As a result, all donor and acceptor enzymes of the surfactin assembly line would be equipped with the same (cognate) pair of COM domains (see Figure 1) and hence should be intercommunicable. In other words, SrfA-A should equally be able to interact with its natural partner SrfA-B (now equipped with the compatible acceptor COM domain COM^{A}_{SrfA-C}) and the natural nonpartner enzyme SrfA-C. 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).

The corresponding COM domain swap was created on the genetic level, and the integrity of the resulting *B. subtilis* strain CC91 verified (data not shown). HPLC analysis of the biosynthetic products synthesized by the reprogrammed assembly line clearly revealed the expected formation of lipotetra- and lipoheptapeptide, while no lipodecapeptide could be observed (Figure 3B). Intriguingly, however, the production level of *B. subtilis* CC91 ($COM^{D}_{SrfA-B}/COM^{D}_{SrfA-C}$) was relatively low, yielding about 1/20th of the amount of surfactin produced by the wild-type under the same conditions (Table 1). Also interesting was the observation that the specificity of the reprogrammed NRPS complex was shifted in favor of the shortened lipotetrapeptide (ratio lipotetrapeptide-to-lipoheptapeptide: approximately 2:1), although

this calls for the interaction between the natural nonpartner enzymes SrfA-A and SrfA-C.

Discussion

Ever since elucidation of the modular organization of NRP and PK biosynthetic complexes, the reprogramming of their enzymatic assembly lines and eventually biocombinatorial synthesis of novel natural product derivatives represented the ultimate goals of research in this area [18, 19]. For NRPs, corresponding efforts included (1) the exchange of A-PCP minimal modules [13, 14], (2) the exchange or deletion of C-A-PCP elongation modules [20, 21], (3) the change-of-selectivity mutagenesis of A domains [12], and (4) the translocation of the terminal, product-releasing Te domains [22]. Although all these approaches led to the formation of the anticipated NRP products, the genetic manipulations were usually involved with significant drops in product titer due to ambiguous problems in the interplay between native and newly introduced catalytic domains and/or the processing of alternative substrate amino acids.

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 COM domain pair of the surfactin biosynthetic complex ($COM^{D}_{SrfA\text{-}A}$/$COM^{A}_{SrfA\text{-}B}$), facilitating the selective interaction between SrfA-A and SrfA-B, was substituted against various cognate, miscognate, and noncognate sets of COM domains.

In a first set of experiments, the disrupted surfactin biosynthetic gene cluster of *B. subtilis* CC64 (see Figure 2) (genotype: *srfA-A::mls::srfA-B*) was reconstituted with the gene fragments encoding for the native COM domain pair $COM^{D}_{SrfA\text{-}A}$/$COM^{A}_{SrfA\text{-}B}$ and the heterologous COM^{D}_{TycB}/COM^{A}_{TycC}, derived from the tyrocidine biosynthetic system, respectively. Both COM domain pairs, featuring compatible (cognate) COM domains, facilitated the productive interaction between SrfA-A and SrfA-B and consequently led to the expected restoration of surfactin biosynthesis. Just as important, the generated (hybrid) NRPS systems revealed essentially the same productivity as the wild-type. This represents a major improvement toward previous experiments, where already the simple reconstitution of the wild-type system was connected with significant reductions in product titer [13, 14, 20].

Similarly to the first experiments, the heterologous, noncognate pair COM^{D}_{TycA}/COM^{A}_{TycC} was exploited to (1) bypass the trimodular SrfA-B, (2) facilitate—due to the crosstalk between COM^{D}_{TycA} and $COM^{A}_{SrfA\text{-}C}$—a direct interaction between the natural nonpartner NRPSs SrfA-A and SrfA-C, and (3) cause the directed synthesis of the novel lipotetrapeptide product FA-LGlu-LLeu-DLeu-LLeu-OH. Utilization of the cognate COM domain pair $COM^{D}_{SrfA\text{-}B}$/$COM^{A}_{SrfA\text{-}C}$, on the other hand, led to the establishment of a so-called universal communication system (UCS) [5], in which all donor and acceptor NRPSs were equipped with the same (compatible) COM^{D} and COM^{A} domains, respectively. Due to the abrogated selectivity of protein-protein communication, the resulting universal communication system (UCS) was capable of biocombinatorial peptide synthesis. In fact, the NRPS system was able to form different assembly lines (SrfA-A/SrfA-B/SrfA-C and SrfA-A/SrfA-C) for the synthesis of the lipoheptapeptide and the lipotetrapeptide, at the same time and in roughly comparable amounts.

Given the compatibility of all donor and acceptor COM domains, (theoretically) also feasible should be the communication between two molecules of SrfA-B. 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 an elongated lipodecapeptide. However, analysis of the butanolic extracts, obtained from the supernatant and the pellet fraction of *B. subtilis* crude extracts revealed no evidence for such product. The absence of lipodecapeptide formation could be due to (1) the toxicity of the product, (2) the incapability of the SrfA-C Te domain to process the alternative lipodecapeptidyl-PCP substrate, (3) the hydrolysis of the unnatural reaction intermediates from the protein template as catalyzed by the cleaning enzyme Srf-Te [23], or (4) the inherent instability of the decapeptide product. The most likely explanation, however, is the limited availability of SrfA-B. In fact, the polycistronic organization of the surfactin biosynthetic genes ensures for the production of equimolar quantities of SrfA-A, SrfA-B, and SrfA-C, whereas the formation of the biosynthetic assembly line of the lipodecapeptide would require an excess of SrfA-B.

The most intriguing result of this study, though, was the determination of both lipopeptide products (rather than just the lipotetrapeptide) when using the presumed noncognate COM domain pair $COM^{D}_{SrfA\text{-}B}$/$COM^{A}_{SrfA\text{-}B}$. From previous studies on the tyrocidine biosynthetic complex, it was known that COM^{D} and COM^{A} domains of nonpartner enzymes are incompatible, supposedly to prevent the futile contact between, i.e., the first (TycA) and third (TycC) enzyme of the NRPS assembly line [4, 5]. Against this background, it was expected that the putative incompatible COM domain pair $COM^{D}_{SrfA\text{-}B}$/$COM^{A}_{SrfA\text{-}B}$ would provide a skipping of SrfA-B and a direct interaction between the natural nonpartner enzymes SrfA-A and SrfA-C, when introduced at the transition between SrfA-A and SrfA-B. The observed biocombinatorial synthesis of both lipopeptides, however, pointed toward a still functional interaction between the latter two enzymes.

How can this be explained? Previous in vitro studies indicated that the homology between COM domains might provide a clue for the possible crosstalk between natural nonpartner NRPSs. In this context, an interaction between donor enzymes carrying COM^{D}_{TycA} and acceptor enzymes bearing the (mis)cognate domains COM^{A}_{TycB}, COM^{A}_{GrsB}, or $COM^{A}_{SrfA\text{-}C}$ was experimentally proven [4, 5]. No interaction, in contrast, was observed between enzymes, carrying COM^{D}_{TycA} and the noncognate COM^{A}_{TycC}. The average sequence homology among the mentioned (miscognate) acceptor COM^{A} domains of TycB, GrsB, and SrfA-C amounts to 79%, compared to only 50% identity observed between the (noncognate) acceptor COM^{A} domains of TycB and TycC. Interestingly, the sequence identity between the two COM^{A} domains of the surfactin biosynthetic complex, $COM^{A}_{SrfA\text{-}B}$ and $COM^{A}_{SrfA\text{-}C}$, amounts to astonishing 63%, which may still

be high enough to render COM domains from either pair, $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$ and $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$, mutually interchangeable.

Recently, we also proposed a five-residue model to describe the interaction between donor and acceptor COM domains and to predict the respective amino acid residues, facilitating their contact [5]. 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. The proposed selectivity-conferring residues—five from each COM domain—are determined based on their relative location toward the highly conserved core motifs "TPSD" (COM^{D}) and "L(T/S)P(M/L)QEG" (COM^{A}) (Figure 5A). Having done this assignment for all twelve COM domain pairs of the tyrocidine, gramicidin S, surfactin, lichenisin, fengycin, and bacitracin biosynthetic complexes, it was found that more than 96% of the 120 putative selectivity-conferring residues are either polar or charged. Furthermore, astounding 59 of 60 proposed amino acids pairs (>98%) lead to the establishment of productive (nonrepulsive) polar or electrostatic interactions [5].

It should be stressed that this five-residue model was postulated based on the in silico analysis of NRPSs. It is, though, supported by the observation that major differences between (mis)cognate and noncognate (acceptor) COM domains occur every three amino acid residues, this way, matching the known rise per repeating unit within an α-helical structure. Even more compelling, the model is in agreement with biochemical data obtained for a mutational analysis of the acceptor COM domain COM^{A}_{TycC} (Figure 5B) [5].

As for the surfactin biosynthetic complex, the five-residue model predicts that the productive interaction between two partner enzymes requires at least four proliferous (polar or electrostatic) contacts (Figure 5). The cognate COM domain pair $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$ also shows one, presumably repulsive contact (note: this is the only mismatch observed so far in this model) between a Glu residue and an Asp residue. In contrast, the proven noncognate pair $COM^{D}_{TycA}/COM^{D}_{TycC}$ does establish one repulsive but only three proliferous contacts. Just as important, when comparing the quality of contacts, it is found that the same number of electrostatic and polar interactions is formed between COM^{D}_{SrfA-B} and the cognate COM^{A}_{SrfA-C} or the miscognate COM^{A}_{SrfA-B}. This would explain the observed productive interaction between SrfA-A and SrfA-B in the hybrid NRPS systems of CC102 ($COM^{A}_{SrfA-B}/COM^{D}_{SrfA-B}$) and the resulting biocombinatorial synthesis of lipohepta- and lipotetrapeptide. Interestingly, the model also predicts a miscognate interaction between the COM^{D}_{SrfA-A} and COM^{A}_{SrfA-C} (Figure 5), which was experimentally confirmed by constructing the corresponding COM domain swap (data not shown). In the resulting hybrid NRPS system of strain CC110 (genotype: *srfA-A::COM*$^{D}_{SrfA-A}$*-COM*$^{A}_{SrfA-C}$*::srfA-B*), SrfA-A was able to establish a miscognate interaction with SrfA-B, leading to a robust restoration of surfactin production (26%) (Figure 3B and Table 1).

The predicted differences between selectivity-conferring residues of (mis)cognate and noncognate COM domain pairs appear to be rather minor. It was, however, already shown that as little as one point mutation can

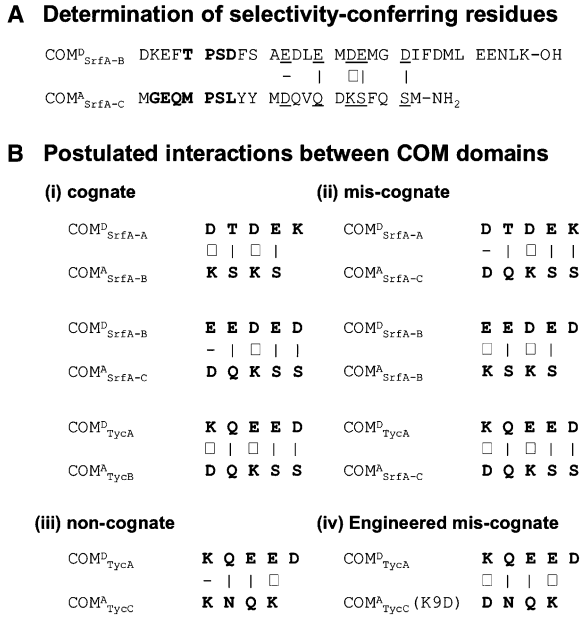


Figure 5. Five Residue Model

(A) The highly conserved core motifs "TPSD" (COM^{D}) and "L(T/S)P(M/L)QEG" (COM^{A}) act as structural anchors for the determination of putative selectivity-conferring residues, e.g., within the COM domain pair $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$.

(B) Proposed interactions within representative cognate, miscognate, and noncognate COM domain pairs. Electrostatic interactions are indicated by squares, polar interactions by vertical bars, and putative repulsive interactions by minuses.

be sufficient to prevent a COM domain from interacting with its native partner and/or to render it, instead, a compatible counterpart of a natural nonpartner enzyme [5]. Hence, the described five-residue model provides a plausible explanation for the product patterns observed in the course of our COM domain swap experiments. On the other hand, 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 nonpartner enzymes are equipped with compatible (cognate or miscognate) pairs of COM domains. In this context, it is interesting to realize that the COM domain pair $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-C}$ only causes a miscognate interaction between natural partner enzymes (i.e., SrfA-A and SrfA-B in *B. subtilis* CC110), but not nonpartner NRPSs (i.e., SrfA-A and SrfA-C in *B. subtilis* CC110 or ATCC 21332). This may be an indication for additional subsidiary structures, situated outside the COM domains, which may influence protein-protein communication.

Where could such additional subsidiary structures be located? One of the best-studied systems to narrow down this question is the dimodular TycA/TycB1 (domain organization: PheA-PCP-E-COM^{D}/COM^{A}-C-ProA-PCP). Deletion studies showed that the E domain of TycA has little or no impact on the protein-protein interaction between both enzymes [5]. Mispriming experiments and mutational studies revealed that the C domain of TycB1 selects for the stereochemistry of the donor substrate [24, 25]. Additional influence of the donor substrate on protein-protein interaction, however,

has not been observed [4, 5]. That leaves A and PCP domains as the most likely candidates, whereby PCP domains have been frequently shown to possess recognition elements (particularly helix 3 of the four-α-helix bundle), which have a major impact on interaction between acyl-S-Ppant PCPs and associated domains [26, 27]. Further studies are required to clarify this point.

Despite such additional structures that may have an influence on protein-protein communication, our study clearly confirmed the decisive role of COM domain pairs for the establishment and/or prevention of productive interactions within NRP biosynthetic complexes. These properties could be exploited for (1) the restoration of surfactin biosynthesis with a heterologous COM domain pair, (2) the directed skipping of SrfA-B within the surfactin biosynthetic assembly line, (3) the substantiation of the soundness of the five-residue model for the description of the interactions between donor and acceptor COM domains, and (4) the establishment of a hybrid system for true biocombinatorial synthesis of lipopeptides.

Significance

NRPSs catalyze the formation of structurally diverse and pharmacologically important peptide natural products. Given their modular organization, they hold an enormous potential for the generation of novel bioactive compounds. Up until now, corresponding approaches were generally limited to module swaps, which (1) require major manipulations of the biosynthetic template, (2) were mostly connected with significant drops in product titer, and (3) constitutionally lead to the synthesis of only one product per experiment. A powerful alternative is provided by the exploitation of COM domains, which have an important role during the intermolecular channeling of reaction intermediates along the biosynthetic assembly line. In the presented study, the inherent potential of COM domains was utilized for the first time in vivo for the directed reprogramming of a NRP assembly line (i.e., surfactin) and the setting up of a system for true biocombinatorial peptide synthesis. 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 a biosynthetic enzyme within the multienzyme 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 the production level (average: 51% of wild-type). Future applications of COM domains will include (1) the generation of biocombinatorial libraries of pharmacologically important NRPs, (2) the coordinated crosstalk between different NRP biosynthetic system, and (3) the segmentation of large monoenzymatic NRP assembly line into easier manageable, multienzymatic complexes. An improved understanding of the structural basis of the interaction between donor and acceptor COM domains may also allow the development of inhibitors for the prevention of protein-protein communication, e.g., in cases were the NRP product is involved in the virulence of a given pathogenic producer strain.

Experimental Procedures

Bacterial Strains and Culture Media

Strains of *Bacillus subtilis* (see the Supplemental Data available with this article online) were grown in Difco sporulation medium (DSM) [28] or modified SpII medium [13]. The media were supplemented—when applicable—with 1 μg/ml erythromycin, 25 μg/ml lincomycin, 5 μg/ml chloramphenicol, and/or 10 μg/ml tetracycline. *Escherichia coli* strains were grown in Luria Broth (LB) medium, supplemented with 50 μg/ml amplicillin.

Plasmid Construction

When not indicated otherwise, DNA fragments were amplified from chromosomal DNA of *B. subtilis* ATCC 21332 with either KOD Hot start DNA Polymerase (Novagen, Merck Biosciences, Bad Soden, Germany) or the "Expand long template PCR system" (Roche, Mannheim, Germany) in accordance to the manufacturer's protocols. The primers (Supplemental Data) were purchased from MWG-Biotech (Ebersberg, Germany) and utilized to introduce the desired terminal restriction sites for the subsequent cloning of PCR products. Standard procedures were applied for DNA manipulations [29], and DNA sequencing confirmed the identity of all plasmids constructed.

The 1.6 kb fragment ′*srfA-A3* and the 1.4 kb fragment *srfA-B5*′ were amplified with the oligonucleotides EPI_F and srf_14R as well as srf_11F and EPI_R, respectively. Based on the 21 bp overlapping region in EPI_F and EPI_R, both fragments were used simultaneously as template for the subsequent fusion PCR with the oligonucleotides srf_11F and srf_14R. The resulting PCR product was modified with *Nco*I and *Xba*I and cloned into the *E. coli* expression vector pQE60 previously cut in the same manner, to give pCC42. Inverse PCR, using the oligonucleotides srf_22F and srf_23R, was used for engineering a *Bam*HI restriction into pCC42 to give pCC50. Subsequently, this restriction site was exploited for cloning of a 1.6 kb *Bam*HI/*Bgl*II fragment, containing the *mls* resistance cassette, from pDG646 into pCC50 to give the final disruption vector pCC63 [30].

The 3 kb fragment ′*srfA-A3-srfA-B5*′ was amplified by using the oligonucleotides srf_11F and srf_14R. After digestion with *Nco*I and *Xba*I, the fragment was ligated into pQE60 to give pCC77. Subsequently, inverse PCR with the oligonucleotides 5′-pQE/srfAB_inv and 3′-pQE/srfAB_inv was used for engineering of the restriction sites *Avr*II and *Acc*65I. The resulting plasmid pCC78 was then used for cloning of all COM domain pairs investigated in this study. To this end, the corresponding gene fragments were amplified via PCR, terminally modified with *Nhe*I and *Bsr*GI, and ligated into pCC78 previously cut with *Avr*II and *Acc*65I. The 150 bp gene fragment of $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-B}$ was amplified using primers 5′-srfAB_COM and 3′-srfAB_COM to yield (after cloning) pCC83. Similarly, the gene fragment of $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-C}$ was obtained by using the oligonucleotides 5′-srfBC_COM and 3′-srfBC_COM, to construct pCC85. The 246 bp fragment of $COM^{D}_{TycB}/COM^{A}_{TycC}$ was amplified from chromosomal DNA of *B. brevis* ATCC 8185 by using the oligonucleotides 5′-TycBC_COM and 3′-TycBC_COM, to obtain pCC92. The 259 bp fragment $COM^{D}_{TycA}/COM^{A}_{TycC}$ was amplified from pQE61-TycA-(C1)TycB1 by using the oligonucleotides 5′-tycAB_COM und 3′-tycAB_COM, to give pCC106. Fusion-PCRs were carried out for the construction and amplification of the gene fragments of the miscognate COM domain pairs $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$ and $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-C}$. To this end, the 459 bp fragment of COM^{A}_{SrfA-B} was amplified by using the oligonucleotides 5′-srfB1_SOE(srfB3) and 3′-srfB1_SOE(srfB3), while the 470 bp fragment of COM^{D}_{SrfA-B} was amplified by using the oligonucleotides 5′-srfB3_SOE(srfB1) and 3′-srfB3_SOE(srfB1). After purification, both DNA fragments were combined and used as template for the amplification of the 165 bp fragment of $COM^{D}_{SrfA-B}/COM^{A}_{SrfA-B}$ by using the oligonucleotides 5′-srfBC_COM and 3′-srfAB_COM. Cloning into pCC78 yielded pCC98. Analogously, the 461 bp fragment of COM^{D}_{SrfA-A} was amplified by using 5′-srfA3_SOE(srfC) and 3′-srfA3_SOE(srfC), whereas the 493 bp fragment of COM^{A}_{SrfA-C} was amplified by using 5′-srfC_SOE(srfA3) and 3′-srfC_SOE(srfA3). After purification, both fragments were used as template for the amplification of the 183 bp fragment of $COM^{D}_{SrfA-A}/COM^{A}_{SrfA-C}$ by using the oligonucleotides 5′-srfAB_COM and 3′-srfBC_COM. Cloning into pCC78 gave pCC97.

B. subtilis Strain Construction

Transformations were carried out as described by Klein et al. [31]. The genotypes were verified via PCR and/or Southern blotting analysis. Transformation of *B. subtilis* ATCC 21332 with *Apa*I/*Sac*II-linearized pKE27 [32] resulted in the Cm^R mutant AM1, which contains a second copy of *comS* in the *amyE* site. *B. subtilis* AM1 was transformed with pCC63, previously linearized with *Xba*I and *Xho*I, to give the Cm^R and MLS^R disruption strain CC64 (genotype: *′srfA-A-mls-srfA-B′*).

Generation of the desired COM domain swaps was achieved by congression experiments. To this end, the disruption strain CC64 was transformed with 1 μg of the corresponding linearized reconstitution plasmid (pCC83, pCC92, pCC106, pCC98, pCC85, or pCC97, respectively), along with 50 ng of the helper plasmid pCm::Tc, carrying a selectable tetracycline resistance marker [17]. Transformants were selected on Cm^R and Tc^R, followed by a screening on MLS^S. These transformation gave the *B. subtilis* mutants CC84 (*srfA-A::COM*$^{D}_{SrfA-A}$-*COM*$^{A}_{SrfA-B}$*::srfA-B*), CC99 (*srfA-A::COM*$^{D}_{TycB}$-*COM*$^{A}_{TycC}$*::srfA-B*), CC112 (*srfA-A::COM*$^{D}_{TycA}$-*COM*$^{A}_{TycC}$*::srfA-B*), CC102 (*srfA-A::COM*$^{D}_{SrfA-B}$-*COM*$^{A}_{SrfA-B}$*::srfA-B*), CC91 (*srfA-A::COM*$^{D}_{SrfA-B}$-*COM*$^{A}_{SrfA-C}$*::srfA-B*), and CC110 (*srfA-A::COM*$^{D}_{SrfA-A}$-*COM*$^{A}_{SrfA-C}$*::srfA-B*), respectively.

Product Analysis

Purification of lipopeptides was performed as described previously [13]. Extracts were analyzed by RP-HPLC on a 1100 series instrument (Agilent, Waldbronn, Germany) with a CC250/3 Nucleosil 120-3C_8 column (Macherey & Nagel, Düren, Germany), equilibrated to 70% buffer B (buffer A, 0.05% [v/v] formic acid in H_2O; buffer B, 0.045% [v/v] formic acid in methanol). Samples were separated by applying a linear gradient to 95% buffer B over 30 min at a flow rate of 0.3 ml min^{-1}. Products were determined at 214 nm and quantification of the lipotetra- and lipoheptapeptides was achieved by integration of the corresponding UV signals.

Online mass spectrometric analysis was performed with a Finnigan LTQ-FT mass spectrometer (Thermo Electron, Bremen, Germany). Samples were measured in the positive ion mode (mass range: 198–1,980) with a Fourier transformation (FT) resolution of 100,000. Parameters were as follows: ion spray voltage 4.5 kV, sheath gas 50 units, auxiliary gas ten units, capillary temperature 330°C with a capillary voltage of 41 V and a tube lens setting of 100 V.

Supplemental Data

Supplemental Data include the strains, plasmids, and oligonucleotides used in this study and are available at http://www.chembiol.com/cgi/content/full/13/8/899/DC1/.

Acknowledgments

We are indebted to Katrin Stachelhaus for discussions and critical reading of the manuscript, Sascha Doekel and Alexandra Mees for providing *B. subtilis* AM1, as well as Stephan Gruenewald for plasmid pQE61-TycA-(C1)TycB1. The authors also wish to thank Mohamed Marahiel for allowing us to carry out this study at the Institute of Biochemistry. C.C. is a young research fellow of the European Commission's training network on directed evolution of functional proteins (ENDIRPRO) and member of the Deutsche Forschungsgemeinschaft "Graduiertenkolleg" protein function on atomic level. The Federal Ministry of Education and Research sponsored this work within the scope of its BioFuture program.

Received: April 9, 2006
Revised: May 30, 2006
Accepted: June 26, 2006
Published: August 25, 2006

References

1. Schwarzer, D., and Marahiel, M.A. (2001). Multimodular biocatalysts for natural product assembly. Naturwissenschaften *88*, 93–101.
2. Kumar, P., Li, Q., Cane, D.E., and Khosla, C. (2003). Intermodular communication in modular polyketide synthases: structural and mutational analysis of linker mediated protein-protein recognition. J. Am. Chem. Soc. *125*, 4097–4102.
3. Gokhale, R.S., and Khosla, C. (2000). Role of linkers in communication between protein modules. Curr. Opin. Chem. Biol. *4*, 22–27.
4. Hahn, M., and Stachelhaus, T. (2004). Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. Proc. Natl. Acad. Sci. USA *101*, 15585–15590.
5. Hahn, M., and Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. Proc. Natl. Acad. Sci. USA *103*, 275–280.
6. Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F., and Weissman, K.J. (2003). The structure of docking domains in modular polyketide synthases. Chem. Biol. *10*, 723–731.
7. Kao, C.M., Pieper, R., Cane, D.E., and Khosla, C. (1996). Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. Biochemistry *35*, 12363–12368.
8. Staunton, J., Caffrey, P., Aparicio, J.F., Roberts, G.A., Bethell, S.S., and Leadlay, P.F. (1996). Evidence for a double-helical structure for modular polyketide synthases. Nat. Struct. Biol. *3*, 188–192.
9. Sieber, S.A., Linne, U., Hillson, N.J., Roche, E., Walsh, C.T., and Marahiel, M.A. (2002). Evidence for a monomeric structure of nonribosomal Peptide synthetases. Chem. Biol. *9*, 997–1008.
10. Tsuji, S.Y., Cane, D.E., and Khosla, C. (2001). Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. Biochemistry *40*, 2326–2331.
11. 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.
12. 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.
13. 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.
14. 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.
15. D'Souza, C., Nakano, M.M., and Zuber, P. (1994). Identification of comS, a gene of the srfA operon that regulates the establishment of genetic competence in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA *91*, 9397–9401.
16. D'Souza, C., Nakano, M.M., Frisby, D.L., and Zuber, P. (1995). Translation of the open reading frame encoded by comS, a gene of the srf operon, is necessary for the development of genetic competence, but not surfactin biosynthesis, in *Bacillus subtilis*. J. Bacteriol. *177*, 4144–4148.
17. 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.
18. Walsh, C.T. (2002). Combinatorial biosynthesis of antibiotics: challenges and opportunities. ChemBioChem *3*, 125–134.
19. Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. Nat. Prod. Rep. *20*, 275–287.
20. Mootz, H.D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., and Marahiel, M.A. (2002). 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.
21. 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.
22. 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.

23. 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. USA *99*, 14083–14088.
24. Stachelhaus, T., and Walsh, C.T. (2000). Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry *39*, 5775–5787.
25. 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.
26. Lai, J.R., Fischbach, M.A., Liu, D.R., and Walsh, C.T. (2006). A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. Proc. Natl. Acad. Sci. USA *103*, 5314–5319.
27. 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.
28. 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.
29. Sambrook, J., Fritsch, E.F., and Maniatis, F. (1989). Molecular Cloning: A Laboratory Handbook (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
30. Guerout-Fleury, A.M., Shazand, K., Frandsen, N., and Stragier, P. (1995). Antibiotic-resistance cassettes for *Bacillus subtilis*. Gene *167*, 335–336.
31. 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.
32. 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.