Protein Engineering vol.15 no.11 pp.913–921, 2002

Modification of biologically active peptides: production of a novel lipohexapeptide after engineering of *Bacillus subtilis* surfactin synthetase

Hanka Symmank[1], Peter Franke[2], Wolfram Saenger[1] and Frank Bernhard[1,3,4]

[1]Institut für Kristallographie, Takustrasse 6 and [2]Institut für Biochemie, Thielallee 63, Freie Universität Berlin, D-14195 Berlin, Germany

[3]Present address: Institut für Biophysikalische Chemie, Universität Frankfurt/Main, Marie Curie Strasse 9, D-60439 Frankfurt/Main, Germany

[4]To whom correspondence should be addressed.
E-mail: fbern@bpc.uni-frankfurt.de

The *Bacillus subtilis* strain ATCC 21332 produces the lipoheptapeptide surfactin, a highly potent biosurfactant synthesized by a large multimodular peptide synthetase. We report the genetic engineering of the surfactin biosynthesis resulting in the production of a novel lipohexapeptide with altered antimicrobial activities. A combination of *in vitro* and *in vivo* recombination approaches was used to construct a modified peptide synthetase by eliminating a large internal region of the enzyme containing a complete amino acid incorporating module. The remaining modules adjacent to the deletion were recombined at different highly conserved sequence motifs characteristic of amino acid incorporating modules of peptide synthetases. The primary goal of this work was to identify permissive fusion sites suitable for the engineering of peptide synthetase genes by genetic recombination. Analysis of the rearranged enzymes after purification from *B.subtilis* and from the heterologous host *Escherichia coli* revealed that the selection of the recombination site is of crucial importance for a successful engineering. Only the recombination at a specific HHII × DGVS sequence motif resulted in an active peptide synthetase. The expected lipohexapeptide was produced *in vivo* and first evidence of a reduced toxicity against erythrocytes and an enhanced lysis of *Bacillus licheniformis* cells was shown.

Keywords: antibiotics/peptide synthetase/protein design

Introduction

Small microbial peptides produced by a non-ribosomal pathway are of great interest owing to their immense potential of biological acitvities and their growing economic value. Some prominent examples are the immunomodulatory agent cyclosporin A, the penicillins and lipopeptides such as surfactin. The peptides usually consists of less than 20 residues, they are cyclic, branched or linear and they often contain modifications and a high variety of non-proteinogenic amino acids as constituents. The peptide bioynthesis is catalysed by large multifunctional non-ribosomal peptide synthetases (NRPSs) via the multiple thiotemplate mechanism (Kleinkauf and von Döhren, 1996; Marahiel *et al.*, 1997; von Döhren *et al.*, 1997; Konz and Marahiel 1999; Mootz and Marahiel, 1999). The NRPSs show a modular architecture and a standard elongation module with a typical mass of ~110 kDa is responsible for the incorporation of each amino acid residue into the growing peptide product. A complete module consists of the three domains C–A–T (Mootz and Marahiel, 1997). The adenylation domain (A-domain) recognizes and activates specific amino acids by adenylation. It is followed by a thioester forming domain (T-domain) homologous to acyl carrier proteins of fatty acid synthetases and polyketide synthetases (PKSs), which covalently binds the activated amino acid at a specific serine residue via a thioester linkage to a 4′-phosphopantetheine cofactor (Stein *et al.*, 1996). The peptide bond between two activated amino acids is finally formed by the condensation domain (C-domain) (de Crécy-Lagard *et al.*, 1995; Belshaw *et al.*, 1999; Mootz and Marahiel, 1999). Additionally, further enzymatic domains responsible for modifications of the amino acid residue, e.g. methylation or epimerization, can be inserted. A large number of sequences of NRPS genes are available and sequence alignments of the single modules revealed several highly conserved motifs characteristic for the specific domains.

NRPSs and the related PKSs have been analysed recently by extensive engineering approaches referred to as combinatorial biosynthesis in order to generate the production of new antimicrobial substances (Cane *et al.*, 1998; Rodriguez and McDaniel, 2001; Staunton and Wilkinson, 2001). A promising approach to altering the substrate specificity of peptide synthetases has been the directed mutagenesis of putative substrate binding pockets in selected A-domains (Stachelhaus *et al.*, 1999; von Döhren *et al.*, 1999). The amino acid sequence of non-ribosomally produced peptides was further altered by the addition, deletion or exchange of enzymatic domains or even complete amino acid incorporating modules as implicated by the modular structure of NRPSs (Stachelhaus *et al.*, 1995; Schneider *et al.*, 1998; Doekel and Marahiel, 2000; Mootz *et al.*, 2000, Schauwecker *et al.*, 2000). Sequence comparisons, mutagenesis studies and limited proteolysis indicated that one module consists of ~1000 amino acid residues. However, only limited structural information of distinct NRPS domains is available so far (Conti *et al.*, 1997; Weber *et al.*, 2000). In addition, interactions between the different domains are also mostly unknown. The manipulation of complex proteins by insertion, exchange or deletion of large internal regions could have a major impact on the overall structural conformation and folding pathway. Choosing the correct site for recombination of peptide synthetase genes should therefore be important for the activity of the resulting enzyme. Approaches have been initiated to analyse recombinant NRPSs generated by using fusion sites located in variable surface accessible putative linker regions (Stachelhaus *et al.*, 1995; Schneider *et al.*, 1998; Doekel and Marahiel, 2000; Mootz *et al.*, 2000). However, linkers may have a crucial role in controling communication and intermodular protein–protein contacts between different modules or domains of multi-modular proteins and they are important to direct correlated movements of the various domains (Gokhale and Khosla, 2000). Linkers between the C–A-domains of NRPSs have been shown to be important and the C–A-domain couple may form a catalytic unit with a specific interface

Table I. Bacterial strains and plasmids

Strain/plasmid	Relevant characteristics	Reference
DH5a	*E.coli*, recA1, endA1, lacZDM15	New England Biolabs
M15	*E.coli*, F$^-$ lac ara gal mtl	Qiagen
ATCC 21332	*B.subtilis* wild-type, surfactin producer	ATCC[a]
$R13_{TD}$	Derivative of *B.subtilis* ATCC 21332, srfA–A replaced by srf_{TD}-M1/2–3	This work
$R13_{CDM}$	Derivative of *B.subtilis* ATCC 21332, srfA–A replaced by srf_{CDM}-M1–2/3, surfactin ADL_2 producer	This work
ATCC 14580	*B.licheniformis* indicator strain	ATCC[a]
pQE30	Ap^r, expression vector	Qiagen
pREP4	Km^r, lacI	Qiagen
pMMN13	Suicide vector for *B.subtilis*	Nakano *et al.* (1989)
$pIPsrfA\text{-}DLeu_{TD}$	Insertion plasmid for marker exchange	This work
$pIPsrfA\text{-}DLeu_{CDM}$	Insertion plasmid for marker exchange	This work
$psrf_{TD}$-M1/2–3	pQE30, expression of SrfA–A with a deletion of amino acids 1004–2040	Symmank *et al.* (1999)
$psrf_{CDM}$-M1–2/3	pQE30, expression of SrfA–A with a deletion of amino acids 1195–2233	Symmank *et al.* (1999)
$psrf_{ADH}$-M1/2–3	pQE30, expression of SrfA–A with a deletion of amino acids 859–1897	Symmank *et al.* (1999)
psrfA–A	pQE30, expression of wild-type SrfA–A	Symmank *et al.* (1999)

[a]American Type Culture Collection.

(Belshaw *et al.*, 1999). The proposed editing function of the C-domains might further contribute to the failure of rearrangements in that regions or to the very low activities of the resulting hybrid enzymes (Stachelhaus *et al.*, 1995; Schneider *et al.*, 1998). Furthermore, linkers have been shown to be specific and essential for the inter-modular product transfer upon engineering of PKSs (Gokhale *et al.*, 1999; Ranganathan *et al.*, 1999). We therefore attempted to identify permissive sites suitable for the rearrangement of NRPSs within conserved motifs. As a model system, we chose the surfactin synthetase complex consisting of the two three-modular enzymes SrfA–A and SrfA–B and the monomodular enzyme SrfA–C. The internal L-leucine activating module of SrfA–A was deleted by genetic engineering and the two terminal modules were recombined at different sequence motifs located in the A-, T- and C-domains, resulting in various SrfA–A derivatives lacking the second L-leucine-incorporating module. This strategy ensured that no changes were introduced in variable linker regions which might be involved in directing domain interactions.

We demonstrate here that the conserved sequence motif HHII×DGVS located in the C-domains of peptide synthetase modules is suitable as a recombination site for the genetic engineering of NRPSs. In our conserved motif fusion approach, only a recombinant surfactin synthetase subunit fused at this region showed high activities in all enzymatic assays and resulted in the production of the expected lipohexapeptide in *B.subtilis*. We further present evidence for an altered profile of antimicrobial activity of the newly synthesized surfactin derivative.

Materials and methods

Strains, plasmids, media and DNA techniques

Strains and plasmids used in this study are listed in Table I. The *B.subtilis* strain ATCC 21332 was used as a source of surfactin synthetase genes and for the engineering of surfactin synthetase and strain DH5a was used for cloning procedures and propagation of plasmids. Bacterial cells were cultivated in Luria broth (LB) or in Landy medium (Landy *et al.*, 1948) supplemented with 0.1% yeast extract and 2 mg/l phenylalanine (Vollenbroich *et al.*, 1993), at temperatures of 28 or 37°C. If appropriate, ampicillin was added to a final concentration of 100 mg/ml. NRPSs were overproduced in *Escherichia coli* strain M15 (× pREP4) with an N-terminal poly$(His)_6$-tag using the vector pQE30. The construction of the plasmids $psrf_{CDM}$-M1–2/3, $psrf_{TD}$-M1/2–3 and $psrf_{ADH}$-M1/2–3 for the overproduction of recombinant SrfA–A proteins was described previously (Symmank *et al.*, 1999). For marker exchange mutagenesis of *B.subtilis*, DNA fragments of the constructed *srfA–A* derivatives including the fusion sites were cloned into the suicide vector pMMN13 (Nakano and Zuber, 1989). DNA techniques such as purification and recombination of DNA followed standard protocols (Sambrook *et al.*, 1989). PCR was performed with Vent-polymerase (New England Biolabs). Transformation of *B.subtilis ATCC 21332* was essentially done as described (Cutting and Vander Horn, 1990).

Isolation and purification of proteins

Overproduction and purification of proteins from *E.coli* were carried our as already described (Symmank *et al.*, 1999). *B.subtilis* cells were suspended in lysis buffer containing a final concentration of 0.1% lysozyme and lysed upon repeated freeze–thaw cycles at –20°C. The cells were then centrifuged at 30 000 *g* for 2 h or subsequently disrupted with a French press at a maximum pressure of 700 psi. The three-modular enzymes SrfA–A and SrfA–B were purified by loading the bacterial extract on an Ultrogel AcA 34 column (100×4 cm i.d.) and separated at a flow-rate of 0.5 ml/min in 50 mM Tris–HCl, pH 8.0, 100 mM NaCl and 5 mM DTT. Bimodular enzymes were purified with a Sephacryl S-200 HR column (60×3.5 cm i.d.) at a flow-rate of 1 ml/min in the same buffer. The peptide synthetases eluted in the void volume. Proteins were routinely analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) (Laemmli, 1970) and quantified with the Bradford assay (Bradford, 1976). If not used directly for enzymatic assays, the proteins were stored in 5% (v/v) glycerol at –70°C.

ATP–PPi exchange assay

The amino acid adenylation activity was determined with the ATP–PPi exchange assay (Symmank *et al.*, 1999). The assay was performed in a reaction volume of 200 µl containing 2 mM substrate amino acid, 2.5 mM $MgCl_2$, 0.5 mM ATP, 0.1 mM $Na_4P_2O_7$, 50 mM MES–HEPES, pH 6.5 and 0.11 µCi (~240 000 c.p.m.) ^{32}P-labelled PP_i. The reaction was started by addition of 3–50 pmol enzyme and incubated at 37°C. The reaction was stopped by addition of 500 µl of cold stop solution

(1% acid-washed pulverized charcoal (Fluka, ultrapure) in 0.1 M $Na_4P_2O_7$ and 14% $HClO_4$) and incubation for 10 min on ice. The charcoal was filtered through a GF92 glass filter (Schleicher & Schuell), washed with water and the amount of bound [^{32}P]ATP was determined with a liquid scintillation counter.

In vitro peptide formation

The *in vitro* peptide formation by recombinant surfactin synthetase subunits was analysed by incubating 50 pmol of purified enzyme with 20 mM amino acid substrates, 0.5 μM ATP, 10 mM $MgCl_2$, 500 mM MES–HEPES pH 6.5 and 2.7 μCi ^{14}C-labelled amino acid (200 mCi/mmol). To initiate the reaction, each sample contained 160 μM 3-hydroxytetradecanoyl-coenzyme A. The reaction was performed in a total volume of 100 μl and incubated for 60 min at 37°C. After phenol extraction, the reaction mixture was separated by thin-layer chromatography (TLC) on a silica gel 60 plate (Merck) with chloroform–methanol–H_2O (65:25:4, v/v/v) as the mobile phase. After separation, the TLC plates were dried and exposed to X-ray film for ~3 weeks.

Isolation of lipopeptides

Lipopeptides were isolated according to Ohno *et al*. (Ohno *et al.*, 1992), with slight modifications. Landy medium was inoculated with a fresh bacterial preculture 1:100 and incubated at 30°C and 120 r.p.m. in a shaker for 3 days. After incubation, the culture was adjusted to pH 2.0 with concentrated HCl and the lipopeptides were precipitated for 1 h at 4°C (Ohno *et al.*, 1992). The precipitate was pelleted for 20 min at 4°C at 8000 *g* and extracted with methanol for 2 h at room temperature on a shaker at 250 r.p.m. The extract was centrifuged for 10 min at 9000 *g* at room temperature and the supernatant was concentrated by rotary evaporation. The residue obtained was dissolved in a suitable volume of methanol, discolored with charcoal and used for further analysis.

Lipopeptides were analysed by TLC on silica gel 60 plates (Merck) with chloroform–methanol–H_2O (65:25:4, v/v/v) as the mobile phase. After separation, the plates were air dried and developed by spraying with water and heating slightly.

Analysis of peptides by mass spectrometry

Mass spectra of methanol extracts of the isolated lipopeptides were recorded using a Bruker Reflex MALDI-TOF instrument with delayed extraction containing a 337 nm nitrogen laser for desorption and ionization. Sample aliquots were mixed with a saturated solution of α-cyano-4-hydroxycinnamic acid in 30% aqueous acetonitrile containing 0.1% (v/v) trifluoroacetic acid and air dried. The acceleration and reflector voltages were 20 and 23 kV, respectively. Post-source decay (PSD) mass spectra were used for confirming the novel lipopeptides. Data were evaluated with Bruker FAST software.

Synthesis of 3-hydroxytetradecanoyl-coenzyme A

The synthesis of 3-hydroxytetradecanoyl-coenzyme A (β-HA-CoA) was performed in a two-step reaction. First, from DL-β-hydroxymyristic acid (β-HA), *N*-hydroxysuccinimide (NHS) and dicyclohexylcarbodiimide (DCCI), β-HA-NHS was synthesized. Second, β-HA-CoA was synthesized from β-HA-NHS and coenzyme A-SH (Blecher, 1981).

Results

In vitro analysis of product formation by recombinant surfactin synthetases

In a systematic screening for permissive fusion sites within highly conserved regions of NRPSs, recombinations at the sequence motifs FF(E/D)LGG(H/D)SL present in T-domains, HHII×DGVS present in C-domains and at the hinge region found in A-domains resulted in the amino acid activating hybrid bimodular enzymes Srf_{TD}-M1/2–3, Srf_{CDM}-M1–2/3 and Srf_{ADH}-M1/2–3 (Table II). To find out if the constructs are suitable for the synthesis of the desired lipohexapeptide *in vivo*, we first analysed the *in vitro* formation of peptide products by the recombinant bimodular synthetases. The purified enzymes were incubated with reaction mixtures containing ^{14}C-labelled L-glutamic acid and the reaction was started by adding β-hydroxytetradecanoyl-coenzme A. The samples were analysed by TLC and we could observe two putative product profiles. No clear product formation was detectable with the proteins Srf_{TD}-M1/2–3 and Srf_{ADH}-M1/2–3. The thioester formation of the enzyme Srf_{ADH}-M1/2–3 with L-glutamic acid was previously found to be drastically reduced (Table II) and this could contribute to the failure to observe a product formation. However, the result was unexpected for the enzyme Srf_{TD}-M1/2–3 as it showed a high rate of adenylation and thioacylation of its cognate amino acid substrates (Table II) (Symmank *et al.*, 1999). The hybrid T-domain in the enzyme Srf_{TD}-M1/2–3 might therefore be unable to communicate efficiently with the other enzymatic domains. The second profile included the product pattern of the recombinant enzyme Srf_{CDM}-M1–2/3 and that of the wild-type enzyme SrfA–A. A similar complex pattern of 8–9 separated bands were detected with both enzymes (data not shown). This gave the first evidence that only the two latter analysed enzymes were able to produce some products in our assay. However, owing to the lack of labelled references, we could not further identify the observed products separated by TLC.

Marker exchange mutagenesis of B.subtilis ATCC 21332

We now started to construct a *B.subtilis* mutant in order to analyse the usefulness of the selected fusion sites *in vivo*. Using a combined homologous recombination–marker exchange approach, we intended to delete the second L-leucine incorporating module of the surfactin synthetase A–A by exchanging the chromosomal *srfA–A* gene with the recombinant genes expressing the two most promising hybrid enzymes Srf_{CDM}-M1–2/3 and Srf_{TD}-M1/2–3 (Figure 1). If the constructed recombinant proteins retain all enzymatic activities, the resulting mutants should be able to produce the lipohexapeptide surfactin ADL_2, a surfactin A derivative carrying a deletion of the second L-leucine residue (Figure 1A). Approximately 1.8 kb (*srf_{CDM}-M1–2/3*) and 2.2 kb (*srf_{TD}-M1/2–3*) DNA fragments containing the corresponding fusion sites were isolated from the plasmids $psrf_{CDM}$-M1–2/3 and $psrf_{TD}$-M1/2–3 and cloned into the suicide vector pMMN13, resulting in the plasmids pIPsrfA-$DLeu_{TD}$ and pIPsrfA-$DLeu_{CDM}$. The suicide plasmids were then transformed into *B.subtilis* ATCC 21332 and plasmid integration was selected by growth on chloramphenicol. The subsequent excision of the integrated plasmids by a second homologous recombination was identified by the loss of chloramphenicol resistance. The correct recombinations in the resulting mutants *B.subtilis* $R13_{TD}$ and $R13_{CDM}$ (Figure 1B) were verified by PCR and restriction analysis.

Phenotype of the mutants B.subtilis $R13_{TD}$ and $R13_{CDM}$

The colony type of the wild-type strain *B.subtilis* ATCC 21332 and the mutant strains $R13_{TD}$ and $R13_{CDM}$ showed striking differences. In contrast to the rough and wrinkled wild-type colonies, the morphology of the mutant $R13_{TD}$ was smooth. The

Table II. Enzymatic activities of the constructed SrfA–A derivatives

Enzyme	Fusion site[a]	A[b]	T[c]	$P_{in\ vitro}$[d]	$P_{in\ vivo}$[e]
Srf_{TD}-M1/2–3	^{1000}FFEL–GGHSLA^{2046}	E:79/L:29	E:106/L:85	−	−
Srf_{CDM}-M1–2/3	^{1186}HLISDGVSL–G^{2234}	E:194/L:35	E:133/L:55	+	+
Srf_{ADH}-M1/2–3	^{856}IEYL–GREDDQ^{1903}	E:69/L:21	E:14/L:60	−	n.d.

[a]Amino acid sequence of the relevant part of the recombinant enzymes. The numbers indicate the positions relative to the sequence of SrfA–A. The dash marks the site of fusion.
[b]A: adenylation of the amino acids glutamic acid (E) and L-leucine (L) as a percentage relative to the wild-type enzyme SrfA–A (Symmank *et al.*, 1999).
[c]T: Thioester formation with the amino acids glutamic acid (E) and L-leucine (L) as a percentage relative to the wild-type enzyme SrfA–A (Symmank *et al.*, 1999).
[d]$P_{in\ vitro}$: *in vitro* product formation with the cognate substrates after intiation of the reaction with β-hydroxytetradecanoyl acid.
[e]$P_{in\ vivo}$: *in vivo* lipopetide production after recombination into *B.subtilis*.

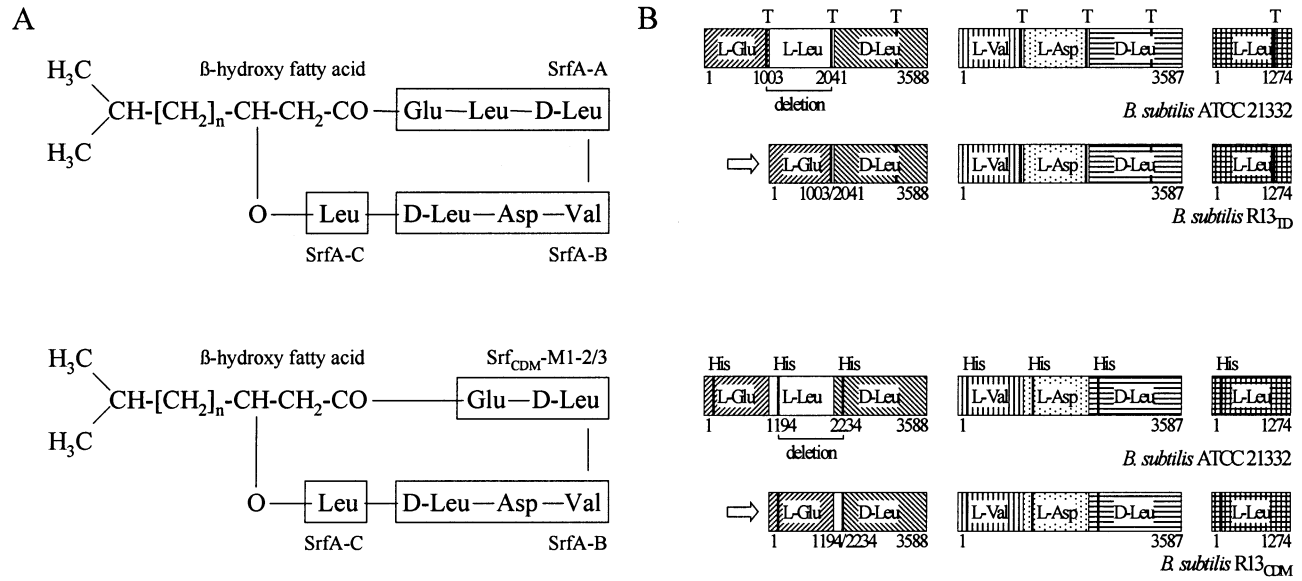

Fig. 1. Engineering of surfactin A synthesis by chromosomal mutations of *B.subtilis* ATCC 21332. (**A**) Primary structures of wild-type surfactin A and the constructed derivative surfactin ADL_2. The surfactin synthetase subunits responsible for the incorporation of the amino acid residues are indicated. (**B**) Modular organization of the surfactin synthetase subunits of the constructed mutants $R13_{TD}$ and $R13_{CDM}$ compared with that of the wild-type strain ATCC 21332. The deleted amino acid residues are indicated below the modules. The positions of the conserved sequence motifs used for recombination are indicated above the modules. His, HHII×DGVS motif; T, FF(E/D)LGG(H/D)SL motif.

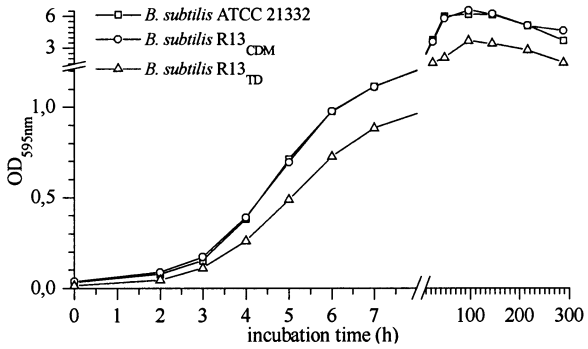

Fig. 2. Growth curve of *B.subtilis* ATCC 21332, $R13_{CDM}$ and $R13_{TD}$ in Landy medium at 28°C. Data are means of at least three determinations.

colony surface of the mutant $R13_{CDM}$ showed an intermediate appearance. The mutant $R13_{TD}$ showed furthermore an increased autolysis, which was complete after ~10 days of incubation on LB agar at room temperature. This phenotype was not observed either with the wild-type strain or with mutant $R13_{CDM}$. In addition, we observed a retarded growth of the mutant $R13_{TD}$ in LB or Landy medium (Figure 2). While the growth of the wild-type strain and mutant $R13_{CDM}$ was similar, the final cell density of mutant $R13_{TD}$ was reduced to ~60% if compared with the two former strains.

We further compared the protein profile of the three strains and could clearly detect the production of high molecular weight surfactin synthetase subunits in the wild-type strain *B.subtilis* ATCC 21332 and in the mutant $R13_{CDM}$ (Figure 3). Besides the 401 kDa SrfA–B and 140 kDa SrfA–C proteins, a band at the expected size of 280 kDa for the recombinant protein Srf_{CDM}-M1–2/3 is visible in extracts of the mutant $R13_{CDM}$. However, we could not detect any production of surfactin synthetase subunits in the mutant $R13_{TD}$ even after analysing the protein profile in different growth phases and after incubation at different temperatures. Interestingly, the strain was deficient not only in the production of the recombinant enzyme Srf_{TD}-M1/2–3 but also in the production of the enzymes SrfA–B and SrfA–C.

The surfactin synthetases were isolated from the wild-type strain *B.subtilis* ATCC 21332 and from the mutant $R13_{CDM}$ and purified by ammonium sulphate precipitation and gel filtration. Each preparation was ~90% pure as judged by SDS–PAGE and contained the three corresponding subunits. The activation of cognate amino acids was analysed in the ATP–PPi exchange assay (Table III). Both synthetase complexes

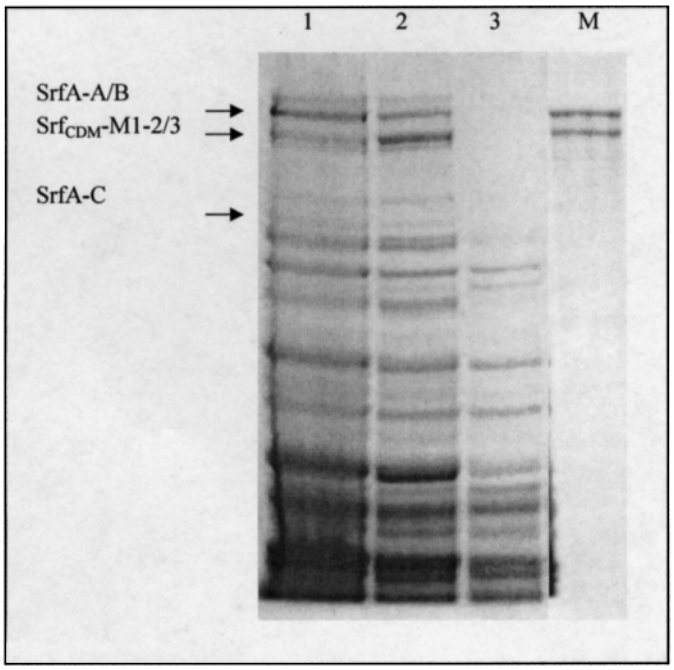

Fig. 3. Protein profile of the *B.subtilis* wild-type strain ATCC 21332 (lane 1) and the mutants R13$_{CDM}$ (lane 2) and R13$_{TD}$ (lane 3). The cells were grown at 28°C in Landy medium for 2 days. Samples of the extracts were analysed by SDS–PAGE on a 7.5% gel. A mixture of the purified proteins SrfA–A (402 kDa) and Srf$_{ADH}$-M1/2–3 (280 kDa) was used as marker (M).

Table III. Amino acid adenylation with purified surfactin synthetase complexes

Amino acid	SrfA synthetase	SrfADL$_2$ synthetase
L-Glutamic acid	1080	3100
L-Leucine	11000	4000
L-Valine	5500	2800
L-Aspartic acid	2800	870

Values were determine with the ATP–PP$_i$ exchange assay and are given in cpm/h.µg enzyme. Means of at least two determinations in the linear range of the reaction.

gave the highest activities with L-leucine, but the activity of the mutant surfactin synthetase was decreased to ~36% compared with the wild-type enzyme. This reduction was expected as one complete L-leucine activating module was deleted in the subunit Srf$_{CDM}$-M1–2/3. In addition, the activation of L-glutamic acid by the mutant surfactin synthetase was considerably higher compared with the wild-type enzyme. This result is in good agreement with our previous *in vitro* experiments, where the L-glutamic acid adenylating activity of the subunit Srf$_{CDM}$-M1–2/3 was about twice that of the wild-type subunit SrfA–A (Table II) (Symmank *et al.*, 1999). The reduced adenylation of L-valine and of L-aspartic acid by the SrfADL$_2$ synthetase indicates a decreased activity of the sububit SrfA–B. This might be due to negative effects of the engineered open reading frame (ORF) on the expression of further downstream located ORFs. Alternatively, the stability of SrfA–B might be decreased in the R13$_{CDM}$ extract.

Peptide production of the mutant B.subtilis R13$_{CDM}$

We compared the lipopeptide biosynthesis of the mutant R13$_{CDM}$ with that of the wild-type strain ATCC 21332. In both strains, the lipopeptide production started in the late exponential

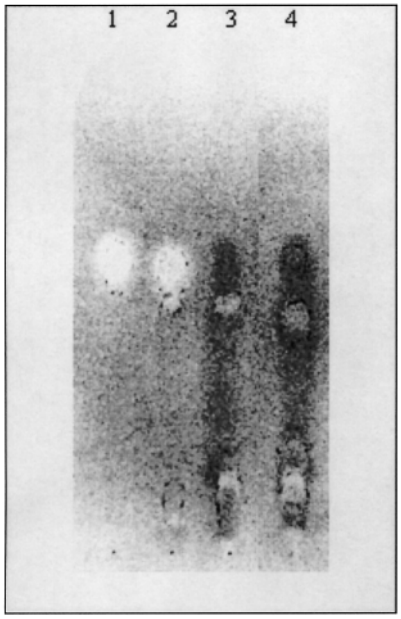

Fig. 4. TLC analysis of the lipopeptide production of the strains *B.subtilis* ATCC 21332 and R13$_{CDM}$. The HCl precipitate of 100 ml of Landy medium was extracted with 10 ml of methanol. The following amounts were used for analysis: lane 1, 25 µg of a surfactin A standard; lane 2, *B.subtilis* ATCC 21332 (5 µl); lane 3, R13$_{CDM}$ (50 µl); lane 4, R13$_{CDM}$ (100 µl).

growth phase and continued after further incubation in the stationary phase (data not shown). Lipopeptides were extracted with methanol from the culture supernatant after 3 days of growth in Landy medium and were analysed by TLC. As expected, the wild-type strain *B.subtilis* ATCC 21332 produced the lipoheptapeptide surfactin A which was visible as a white spot after development of the TLC plate (Figure 4). The R_f value of 0.51 agrees with the literature (Ullrich *et al.*, 1991; Menkhaus *et al.*, 1993). With the extract of the supernatant of the mutant R12$_{CDM}$, a predominant spot with an R_f value of 0.39 was obtained, presumably representing the lipohexapeptide surfactin ADL$_2$ lacking one L-leucine residue. Considering the amounts of the samples analysed and assuming a similar sensitivity for the detection of the two lipopeptides in the TLC assay, the production rate of surfactin ADL$_2$ in the strain R13$_{CDM}$ was estimated to account for only ~5% of the production rate of surfactin A in the wild-type strain *B.subtilis* ATCC 21332. This would correspond to ~25 mg of surfactin ADL$_2$ per litre of Landy medium.

The complete non-separated extracts of the two strains were further analysed by MALDI mass spectrometric analysis (Figure 5). Both the wild-type surfactin A and the putative surfactin ADL$_2$ showed a typical clustering of signals, caused by variations in the chain length of the β-hydroxy fatty acid moiety in the range 13–16 carbon atoms. Owing to the fermentation conditions, the clusters were dominated by Na^+ and K^+ adducts. The peaks of the wild-type cluster in the range *m/z* 1031–1089 were all in good agreement with previously published spectra of surfactin A (Leenders *et al.*, 1999). With the putative surfactin ADL$_2$, a signal cluster in the area between 918 and 976 Da was obtained. The *m/z* difference of 113 between equivalent peaks of the two clusters corresponds exactly to the deleted L-leucine residue in surfactin ADL$_2$. In order to confirm the identification of the novel surfactin ADL$_2$, a post-source decay (PSD) mass spectrum of the parent ion at

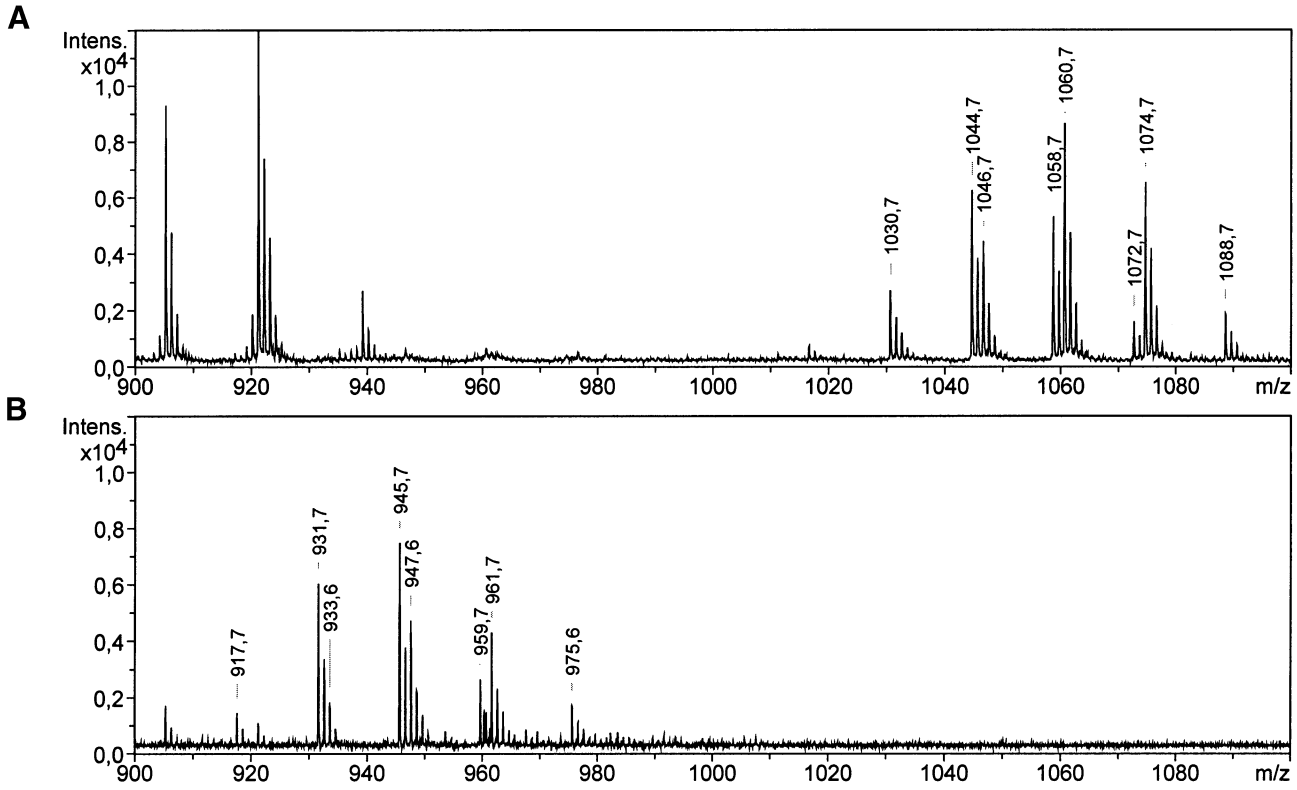

Fig. 5. MALDI mass spectra of the lipopeptides produced from *B.subtilis* ATCC 21332 (**A**) and *B.subtilis* $R13_{CDM}$ (**B**) within the range *m/z* 900–1100. The typical signal clustering caused by variations in the chain length of the fatty acid residue is present for both surfactin types. The values (1030.7, 1044.7, 1058.7, 1072.7) and (1046.7, 1060.7, 1074.7, 1088.7) correspond to the calculated Na^+ and K^+ adducts of surfactin A with a fatty acid residue ranging from 13 to 16 carbon atoms. The values (917.7, 931.7, 945.7, 959.7) and (933.7, 947.7, 961.7, 975.7) correspond to the calculated Na^+ and K^+ adducts of surfactin ADL_2 also with fatty acid bodies ranging from 13 up to 16 carbon atoms.

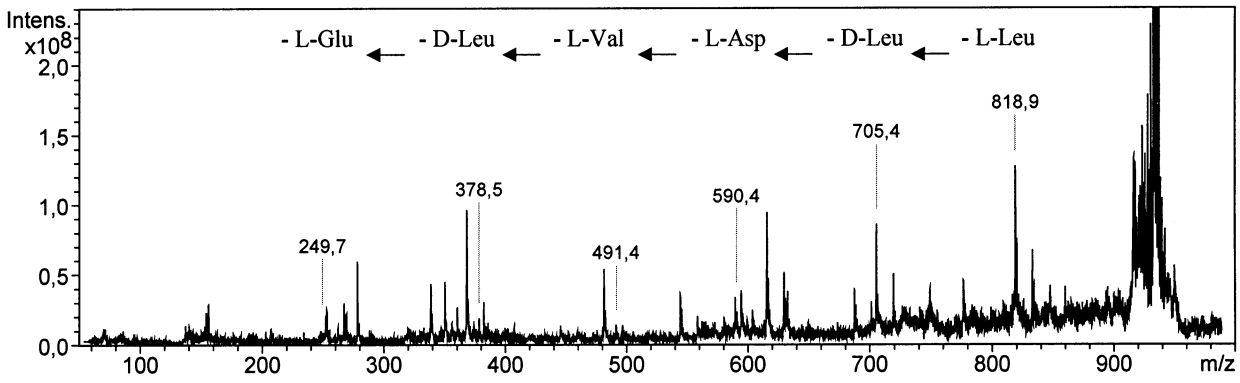

Fig. 6. MALDI-TOF post-source decay mass spectrum of surfactin ADL_2 from the methanol extract of $R13_{CDM}$ cells. The parent ion at *m/z* 931.7 corresponds to the Na^+ adduct containing a fatty acid residue with 14 carbon atoms. Fragments representing the calculated masses derived from the stepwise cleavage of the amino acids L-leucine (818.9), D-leucine (705.4), L-aspartic acid (590.4), L-valine (491.4), D-leucine (378.5) and L-glutamic acid (249.4) from surfactin ADL_2 are marked.

m/z 931.7 corresponding to the Na^+ adduct with a C_{14} fatty acid side-chain body was recorded (Figure 6). We could detect the expected fragment ion pattern resulting from the successive cleavage of the six amino acids from surfactin ADL_2, providing good evidence that surfactin ADL_2 was produced by the strain $R13_{CDM}$.

Bioactivity of the mutant $R13_{CDM}$

The lipopeptide surfactin A causes hemolysis and has an inhibiting activity against a broad range of microorganisms. Commercial blood agar plates were inoculated with the strains *B.subtilis* ATCC 21332 and $R13_{CDM}$ and incubated for 24 h at 37°C. The strain *B.subtilis* ATCC 21332 produced a clear hemolytic zone surrounding the bacterial colonies (Figure 7A). In contrast, no hemolysis was visible with the mutant $R13_{CDM}$. Similar results were obtained by using methanol extracts of culture supernatants from the two strains grown in Landy medium.

We further compared the growth inhibition of microorganisms by the two surfactins. Cells of the Gram positive indicator strain *Bacillus licheniformis* were embedded in LB-

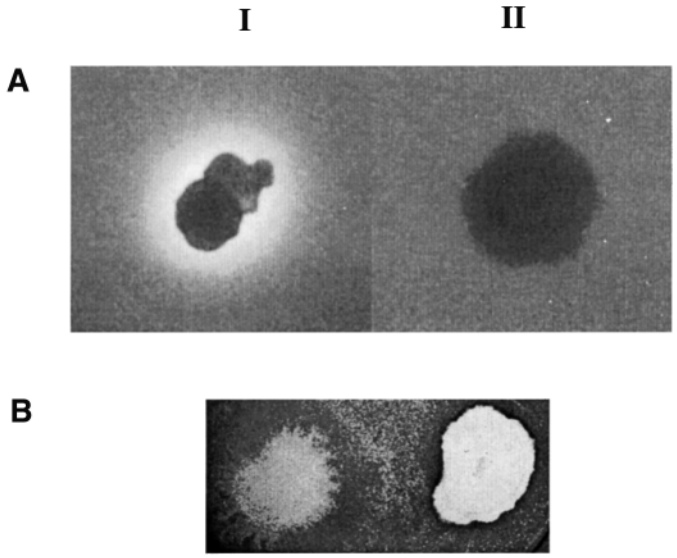

Fig. 7. Inhibitory effects of the *B.subtilis* strains ATCC 21332 and $R13_{CDM}$. (**A**) Hemolysis after incubation on blood agar plates (Merck) for 24 h at 37°C. I, ATCC 21332; II, $R13_{CDM}$. (**B**) Inhibition of *B.licheniformis*. Cells of *B.licheniformis* embedded in LB-agar were covered with a cellophane membrane containing single colonies of *B.subtilis* ATCC 21332 (I) and *B.subtilis* $R13_{CDM}$ (II). The inhibition zones were visible after 24 h of incubation at 30°C.

agar in Petri dishes and the solidified agar was overlayed with a sterile cellophane membrane containing single colonies of the two surfactin producing strains *B.subtilis* ATCC 21332 and $R13_{CDM}$ and incubated for 1 day at 30°C. A clear inhibition zone with a sharp border was produced by the mutant $R13_{CDM}$ (Figure 7B). In contrast, the wild-type strain *B.subtilis* ATCC 21332 caused a turbid inhibition zone with an uneven border, indicating some escape of the indicator cells from the inhibitory effect of surfactin A. In similar experiments, we could not find differences between the two strains in the inhibition of *E.coli* and of *Pichia pastoris* cells (data not shown).

Discussion

We could identify a permissive fusion site for the engineering of NRPSs located in a highly conserved sequence motif of condensation domains and demonstrate its use for the production of designed lipopeptides *in vivo*. Other analysed recombinations within sequence motifs of A- and T-domains failed to result in the production of the desired peptide. A flexible hinge region divides the A-domain into a large N-terminal subdomain providing a substrate binding cleft and into a small C-terminal subdomain which might form a lid over the cleft upon substrate binding (Conti *et al.*, 1997). Despite high homologies between A-domains of different origins and with different substrate specificity, recombinations at various highly conserved sequence motifs within the two subdomains were found to be non-permissive and always resulted in inactive hybrid A-domains (Elsner *et al.*, 1997) presumably because of severe disturbance of the functional integrity of the two subdomains. In contrast, recombinations at or very close to the hinge region were permissive and the hybrid A-domains retained their enzymatic activities (Elsner *et al.*, 1997; Symmank *et al.*, 1999). However, in the context of a multimodular enzyme, hybrid A-domains fused at the hinge region seem to be unable to transfer the activated amino acid adenylate to the T-domain in order to form a thioester with the 4′-phosphopantetheine cofactor (Symmank *et al.*, 1999). This result gives evidence for some specific interactions between the N-terminal subdomain of the A-domain and their cognate T-domains. A recognition mechanism is further supported by the observation that aminoacylation of isolated T-domains from the surfactin synthetase *in trans* is only possible with their cognate A-domains derived from the same module (Weinreb *et al.*, 1998). In contrast, fusions within the interdomain A–T linker has been used to construct functional dimodular NRPSs between modules derived from the tyrocidin and bacitracin operons (Doekel and Marahiel, 2000). The observed dipeptide formation *in vitro* implicated no specific interactions between A and T domains in that system.

The T-domains have a symmetric structure and consist of ~75 amino acid residues with a structural core spanning 37 amino acid residues in both directions from an invariant serine in the conserved sequence motif D/(D/N)FF×LGGHS(L/I), which serves as a 4′-phosphopantetheine binding site (Weber *et al.*, 2000). Our results localize the proposed interface between A- and T-domains within amino acid residues located N-terminal to the invariant serine residue. Interactions between the A- and T-domains obviously remained intact after recombination within this motif and resulted in a fully active thioester forming hybrid T-domain.

The lack of any detectable products synthesized from Srf_{TD}-M1/2–3 *in vitro* could point to an essential interaction between T- and C-domains, which then also might involve the N-terminal half site of the T-domain as the hybrid T-domain preserving the C-terminal part was obviously not sufficient to communicate with its cognate C-domain. This result agrees with the observation that the homologous ACP–KS couple in PKSs needs to be preserved during engineering (Gokhale *et al.*, 1999; Ranaganathan *et al*., 1999). However, the T–C linker connecting two distinct elongation modules has been successfully used as a fusion site to recombine modules from the tyrocidine synthetase and the expected peptide formation *in vitro* could be shown (Mootz *et al.*, 2000). Taken together with the above-mentioned somewhat contradictory results using the A–T linker as a fusion site, these data suggest that interdomain or intermodular communications might be different within the surfactin synthetase and the tyrocidine synthetase.

A peptide production from strain $R13_{TD}$ could not be expected as no production of surfactin synthetases was observed. This could indicate some changes in general regulation networks. Effects of mutations in the surfactin synthetase operon on efficient sporulation and competence development have been reported and regulatory genes such as *comS* have been found to be inserted in that region (Nakano *et al.*, 1991; Hamoen *et al.*, 1995). The deletions in the SrfA–A coding region introduced upon construction of the mutant $R13_{TD}$ might have affected the expression of genes essential for the regulation of sporulation and surfactin biosynthesis.

The C-domains are characterized by the active site motif HHII×DGVS involved in the catalysis of non-ribosomal peptide bond formation. A catalytic mechanism has been proposed where the second histidine residue of that motif is essential for peptide bond formation (de Crécy-Lagard *et al.*, 1995). C-domains have been differentiated into an unspecific N-terminal donor and a specific C-terminal acceptor pocket (Belshaw *et al.*, 1999; Ehmann *et al.*, 2000; Linne and Marahiel, 2000). The fatty acid–L-glutamic acid intermediate in the hybrid Srf_{CDM}-M1–2/3 instead of a fatty acid–L-glutamic acid–

L-leucine intermediate in the wild-type enzyme SrfA–A should therefore be accepted as a substrate by the donor pocket of the C-domain. In fact, out of the analysed fusion sites only a hybrid bimodular enzyme fused at this motif produced *in vitro* a similar product pattern if compared with the wild-type enzyme SrfA–A and the expected peptide was synthesized *in vivo*. Furthermore, the HHII×DGVS motif has been successfully used as a fusion site for the module exchange between the highly homologous NRPSs surfactin A synthetase from *B.subtilis* and lichenysin A synthetase from *B. licheniformis* (Yakimov *et al.*, 2000). In this work, the entire amino acid incorporating modules 1 and 5 of the surfactin synthetase have been replaced by the corresponding modules of lichenysin synthetase, showing that this fusion site seems to be generally suitable also for module swaps within the surfactin synthetase.

The identified lipohexapeptide $SrfADL_2$ from strain $R13_{CDM}$ is one of the first rational designed peptide constructed by the directed deletion of a complete internal module of a peptide synthetase. We could further show that the variety of the incorporated β-hydroxy fatty acids into surfactin $SrfADL_2$ is identical with the pattern of wild-type surfactin A (Peypoux *et al.*, 1991; Leenders *et al.*, 1999), indicating that the activity of the associated β-hydroxy fatty acid transferring acyltransferase is not influenced by the rearrangement of the surfactin synthetase. However, the productivity of surfactin $SrfADL_2$ was only estimated at ~5% to that of wild-type surfactin A, giving a yield of ~25–50 mg of $SrfADL_2$ per litre of medium compared with 0.5–1 g/l for SrfA (Peypoux and Michel, 1992; de Ferra *et al.*, 1997). We could show that an altered expression rate of the recombinant surfactin synthetase genes is obviously not responsible for this effect as the enzymes were clearly detectable by SDS–PAGE. The reduced chain length of peptide $SrfADL_2$ could severely affect its three-dimensional structure. The wild-type surfactin A has a horse saddle-like conformation with the polar side chains of the L-glutamic acid and L-aspartic acid residues opposing the non-polar β-hydroxy fatty acid (Bonmatin *et al.*, 1992, 1994). The deletion of the second L-leucine residue in surfactin $SrfADL_2$ will most probably alter that conformation with possible effects on the stability of $SrfADL_2$ in the bacterial cell. Rapid *in vivo* degradation and less efficient export mechanisms might further contribute to the observed relatively low production of $SrfADL_2$. While a suboptimal conformation of the constructed hybrid C-domain might also reduce the specific activity of the recombinant surfactin synthetase, the observed high yields of modified lipoheptapetides after module exchange using the HHII×DGVS motif (Yakimov *et al.*, 2000) gave further evidence of some problems in stability or transport of the synthesized lipohexapeptide.

First preliminary bioassays revealed that the constructed surfactin derivative $SrfADL_2$ has clearly altered biological activities compared to the wild-type surfactin SrfA. Most interesting was the lack of any detectable hemolytic activity concomitant with an increase in growth inhibition of bacterial cells. Considering the reduced $SrfADL_2$ production in strain $R13_{CDM}$ compared to the SrfA production in strain ATCC 21332, this difference is even more striking. This result gives first evidence that $SrfADL_2$ or similar surfactin derivatives might exhibit a reduced toxicity against eukaryotic cells, which could improve their therapeutic applications.

The results obtained with the engineering of NRPSs indicate that a general optimal fusion site suitable for rearrangements might not exist. Linker regions which have been successfully used as fusion sites for the recombination of the tyrocidine or bacitracin synthetases seem not to be suitable for the recombination of surfactin synthetase. In contrast, the potential of the HHII×DGVS motif for combinatorial biosynthesis of further surfactin derivatives has already been demonstrated (Yakimov *et al.*, 2000). It will now be interesting to prove the versatility of this motif for the recombination of other NRPSs.

Acknowledgements

We are grateful to Clemens Langner and Rita Diehl for technical assistance. We thank Ferdinand Hucho for providing access to mass spectrometry facilities. This work was supported by the European Union, Grant PL 950176.

References

Belshaw,P.J., Walsh,C.T. and Stachelhaus,T. (1999) *Science*, **284**, 486–489.
Blecher,M. (1981) *Methods Enzymol*., **72**, 404–408.
Bonmatin,J.M. *et al.* (1992) *Biochimie*, **74**, 825–836.
Bonmatin,J.M., Genest,M., Labb□Œ,H. and Ptak,M. (1994) *Biopolymers*, **34**, 975–986.
Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
Cane,D.E., Walsh,C.T. and Khosla,C. (1998) *Science*, **282**, 63–68.
Conti,E., Stachelhaus,T., Marahiel,M.A. and Brick,P. (1997) *EMBO J.*, **16**, 4174–4183.
Cutting,S.M. and Vander Horn,P.B. (1990) In Harwood,C.R. and Cutting,S.M. (eds), *Molecular Biological Methodes for Bacillus*. Wiley, Chichester.
de Crécy-Lagard,V., Marliere,P. and Saurin,W. (1995) *C. R. Acad. Sci. III: Life Sci.*, **318**, 927–936.
de Ferra,F., Rodriguez,F., Tortora,O., Tosi,C. and Grandi,G. (1997) *J. Biol. Chem.*, **272**, 25304–25309.
Doekel,S. and Marahiel,M.A. (2000) *Chem. Biol*., **7**, 373–384.
Ehmann,D.E., Trauger,J.W., Stachelhaus,T. and Walsh,C.T. (2000) *Chem. Biol*., **7**, 765–772.
Elsner,A., Engert,H., Saenger,W., Hamoen,L., Venema,G. and Bernhard,F. (1997) *J. Biol. Chem.*, **272**, 4814–4819.
Gokhale,R.S. and Khosla,C. (2000) *Curr. Opin. Chem. Biol*., **4**, 22–27.
Gokhale,R.E., Tsuji,S.Y., Cane,D.E. and Khosla,C. (1999) *Science*, **284**, 482–485.
Hamoen,L.W., Eshuis,H., Jongbloed,J., Venema,G. and van Sinderen,D. (1995) *Mol. Microbiol.*, **15**, 55–63.
Kleinkauf,H. and von Döhren,H. (1996) *Eur. J. Biochem.*, **236**, 335–351.
Konz,D. and Marahiel,M.A. (1999) *Chem. Biol.*, **6**, 39–48.
Landy,M., Warren,G.H., Rosenman,S.B. and Colio,L.G. (1948) *Proc. Soc. Exp. Biol. Med.*, **67**, 539ff.
Leenders,F., Stein,T., Kablitz,B., Franke,P. and Vater,J. (1999) *Rapid Commun. Mass Spectrom.*, **13**, 943–949.
Linne U. and Marahiel,M.A. (2000) *Biochemistry*, **39**, 10439–10447.
Marahiel,M.A., Stachelhaus,T. and Mootz,H.D. (1997) *Chem. Rev.*, **97**, 2651–2673.
Menkhaus,M., Ullrich,C., Kluge,B., Vater,J., Vollenbroich,D. and Kamp,R.M. (1993) *J. Biol. Chem*., **268**, 7678–7684.
Mootz,H. and Marahiel,M.A. (1997) *Curr. Opin. Chem. Biol.*, **1**, 543–551.
Mootz,H.D. and Marahiel,M.A. (1999) *Curr. Opin. Biotechnol.*, **10**, 341–348.
Mootz,H., Schwarzer,D. and Marahiel,M.A. (2000) *Proc. Natl Acad. Sci. USA*, **11**, 5848–5853.
Nakano,M.M. and Zuber,P. (1989) *J. Bacteriol.*, **171**, 5347–5353.
Nakano,M.M., Magnuson,R., Myers,A., Curry,J., Grossman,A.D. and Zuber,P. (1991) *J. Bacteriol.*, **173**, 1770–1778.
Ohno,A., Ano,T. and Shoda,M. (1992) *Biotechnol. Lett.*, **14**, 1165–1168.
Peypoux,F. and Michel,G. (1992) *Appl. Microbiol. Biotechnol.*, **36**, 515–517.
Peypoux,F., Bonmatin,J.M., Labbé,H., Das,B.C., Ptak,M. and Michel,G. (1991) *Eur. J. Biochem.*, **202**, 101–106.
Ranganathan,A. *et al.* (1999) *Chem. Biol*., **6**, 731–741.
Rodriguez,E. and McDaniel,R. (2001) *Curr. Opin. Microbiol*., **4**, 526–534.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual.* 2nd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbor, NY.
Schauwecker,F., Pfennig,F., Grammel,N. and Keller,U. (2000) *Chem. Biol*., **7**, 287–297.
Schneider,A., Stachelhaus,T. and Marahiel,M.A. (1998) *Mol. Gen. Genet.*, **257**, 308–318.
Stachelhaus,T., Schneider,A. and Marahiel,M.A. (1995) *Science*, **269**, 69–72.
Stachelhaus,T., Mootz,H.D. and Marahiel,M.A. (1999) *Chem. Biol.*, **6**, 493–505.

Staunton,J. and Wilkinson,B. (2001) *Curr. Opin. Chem. Biol.*, **5**, 159–164.
Stein,T., Vater,J., Kruft,V., Otto,A., Wittmann-Liebold,B., Franke,P., Panico,M., McDowell,R. and Morris,H.R. (1996) *J. Biol. Chem.*, **271**, 15428–15435.
Symmank,H., Saenger,W. and Bernhard,F. (1999) *J. Biol. Chem.*, **274**, 21581–21588.
Ullrich,C., Kluge,B., Palacz,Z. and Vater,J. (1991) *Biochemistry*, **30**, 6503–6508.
von Döhren,H., Keller,U., Vater,J. and Zocher,R. (1997) *Chem. Rev.*, **97**, 2675–2705.
von Döhren,H., Dieckmann,R. and Pavela-Vrancic,M. (1999) *Chem. Biol.*, **6**, 273–279.
Weber,T., Baumgartner,R. Renner,C., Marahiel,M.A. and Holak,T.A. (2000) *Struct. Fold. Des.*, **15**, 407–418.
Weinreb,P.H., Quadri,L.E.N., Walsh,C.T. and Zuber,P. (1998) *Biochemistry*, **37**, 1575–1584.
Yakimov,M.M., Giuliano,L., Timmis,K.N. and Golyshin,P.N. (2000) *J. Mol. Microbiol. Biotechnol.*, **2**, 217–224.

Received February 20, 2002; revised August 2, 2002; accepted August 26, 2002