Published in final edited form as:

Nature. 2008 August 14; 454(7206): 907–911. doi:10.1038/nature07161.

When the fold fits the function - The external thioesterase of the Surfactin-Synthetase

Alexander Koglin^{1,2}, Frank Löhr¹, Frank Bernhard¹, Vladimir R. Rogov^{1,3}, Dominique P. Frueh², Eric R. Strieter², Mohammad R. Mofid⁴, Peter Güntert^{1,5}, Gerhard Wagner², Christopher T. Walsh², Mohamed A. Marahiel⁴, and Volker Dötsch^{*,1}

¹Institute of Biophysical Chemistry, J-W. Goethe University, Frankfurt/ Main, Germany

²Dept. of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

³Institutes for Protein Research, University Puchino, Moscow, Russia

⁴Dept. of Chemistry/ Biochemistry, Philipps University, Marburg, Germany

⁵Frankfurt Institute for Advanced Studies, Frankfurt/ Main, Germany

Abstract

Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) found in bacteria, fungi and plants utilize two different types of thioesterases for the production of highly active biological compounds^{1, 2}. Type I thioesterases (TEI) catalyze the release step from the assembly line³ of the final product where it is transported from one reaction center to the next as a thioester linked to a 4'-phosphopantetheine cofactor (4'-PP) that is covalently attached to thiolation (T) domains⁴⁻⁹. The second enzyme involved in the synthesis of these secondary metabolites, the type II thioesterase (TEII), is a crucial repair enzyme for the regeneration of functional 4'-PP cofactors of holo T-domains of NRPS and PKS systems¹¹⁻¹³. Mispriming of 4'-PP cofactors by acetyl- and short chain acyl-residues interrupts the biosynthetic system. This repair reaction is very important, since roughly 80% of coenzyme A (CoA), the precursor of the 4'-phosphopantetheine cofactor, is acetylated in bacteria¹⁴. Here we report the first three-dimensional structure of a type II thioesterase free and in complex with a T domain. Comparison with structures of TEI enzymes^{3, 15} shows the basis for substrate selectivity and the different modes of interaction of TEII and TEI enzymes with T domains. In addition, we show that the TEII enzyme exists in several conformations of which only one is selected upon interaction with its native substrate, a modified holo-T domain.

The cyclic lipopeptapeptide surfactin is one of the most potent biosurfactants showing antibacterial and antiviral activity^{2,3,16}. It is synthesized by a complex of three large subunits

Correspondence and requests for materials should be addressed to V.D. (VDoetsch@em.unifrankfurt.de), C.T.W.

(Christopher_Walsh@hms.harvard.edu) and M.A.M. (Marahiel@chemie.unimarburg.de)..

*Corresponding author: vdoetsch@em.uni-frankfurt.de

Author Contributions A.K., V.D., F.B. and M.A.M. designed the experiments and defined the research for the SrfTEII. A.K., V.D. and C.T.W. wrote the manuscript and designed the experiments for the complex structure. A.K., F.B. and F.L. conducted the research including protein expression, data acquisition, resonance assignment and structure calculation. V.V.R. helped with resonance assignment and structure calculation of the SrfTEII. P.G. provided the volumes of the cavities of SrfTEI and of the complex SrfTEII/TycC3-T domain and supported the complex structure calculation. D.F. and G.W. recorded the NMR spectra of the protein complex. A.K. calculated the structure of the isolated SrfTEII, the structure of the protein complex (SrfTEII and TycC3-PCP) and performed all titration experiments. M.A.M. and M.R.M. provided the expression vectors for SrfTEII and TycC3-T domain. M.A.M. provided the purification protocol and biochemical characterizations of the SrfTEII. E.R.S. provided the non-hydrolysable Coenzyme A derivative and synthesized the peptide and amino acid modified substrate analogues.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

that consist of either three modules (subunits SrfA-A and SrfA-B) or one module (subunit SrfA-C) (Fig. 1) with each module being responsible for the addition of one amino acid^{2,3}. The transport of the growing chain between individual modules is achieved by small ~80 amino acid long T domains that interact with the aminoacyl-forming activation A-domain and both the up- and downstream peptide-bond-forming condensation C-domains⁴⁻¹⁰. In addition, some of the T domains of the Surfactin-synthetase have to interact with epimerization domains, located at the C-terminus of subunits SrfA-A and SrfA-B, or with the covalently linked type I thioesterase³. The type I thioesterase catalyzes the macrolactone formation between Leu⁽⁷⁾ and the β -hydroxy fatty acid to release the mature surfactin. Modifications blocking the reactive thiol group of the 4'-PP cofactor attached to any T domain can occur with small molecules present in the cell (acetylation, succinylation and modification with fatty acids) and pose a significant challenge for the organism to keep the assembly line running.

The surveillance and repair tasks for the surfactin assembly line are carried out by the stand-alone surfactin type II thioesterase (SrfTEII)¹¹⁻¹³. The importance of this enzyme has been demonstrated by genetic deletions that reduced the production of surfactin by 84%¹⁷. Due to the large variety of acylation modifications and the fact that the SrfTEII has to be able to interact with all seven T domains of the entire assembly line, this TEII has to be - in contrast to the type I thioesterase at the end of the last module - rather non-specific. At the same time, premature cleavage of the correct growing peptide chain has to be avoided. In addition to this repair function, the SrfTEII might also be responsible for loading the β -hydroxy fatty acid onto the first C-domain.

SrfTEII shows the typical α/β -hydrolase fold, with a central 7-stranded β -sheet surrounded by 8 helices (Fig. 2). Comparison with the structure of the *in cis*-acting SrfTEI reveals a similar overall architecture³, but with significant differences (Supplementary Fig. 5). The most obvious alteration is the N-terminal truncation of α -helix II (Asp60 - Leu67). Further modifications include the insertion of an additional helix (helix VII; Cys193 - Trp200) between the active site residues Asp189 and His216 and repositioning of the helix-turn-helix motif canopying the active site in SrfTEI. This lid is moved from its central position observed in the SrfTEI crystal structure³, towards the β -strands VI and VII in SrfTEII. Consequently, the active site residues in the SrfTEII structure (Ser86, Asp189, and His216)^{11,12} are just partially covered by a short loop (Gln138 - Ala 144) and become more accessible compared to SrfTEI. The space surrounding the active site is further enlarged by a kink in one of the helices of the helix-turn-helix motif.

Many of the amide proton resonances of the SrfTEII enzyme show significant line broadening or even two separate peaks, suggesting the existence of multiple conformations (Supplementary Fig. 2). Recently we had described the existence of multiple conformations in the TycC3-T domain⁵ (third module of the tyrocidine C subunit, a substrate for SrfTEII) and had suggested that this conformational exchange is an important driving force for the selective interaction with other domains of the NRPS system⁴⁻¹⁰.

NMR titration experiments¹⁸ of ¹⁵N-enriched SrfTEII were performed with CoA, myristoyl-CoA, apo- and holo-TycC3-T domains. The resulting chemical shift perturbations (Supplementary Fig. 3) revealed interaction with the holo form and virtually no interaction with the apo form in agreement with previous experiments in which the chemical shift differences of the different T domain forms were analyzed⁵. Interestingly, myristoyl-CoA also binds to SrfTEII, suggesting recognition of this β -hydroxy-fatty acid. This might support the previously assumed second function of SrfTEII as the starter enzyme, responsible for loading the β -hydroxy-fatty acid onto the first C-domain¹⁶. A comparison of the interaction interface of the SrfTEII to the *in cis* T recognition of EntF T-TEI¹⁵ (Frueh *et al.*, this issue) reveals that

in particular the sequentially diverse helix-turn-helix motif and its connecting loops are important for selecting the specific substrate.

To obtain more detailed insight into the interaction between the SrfTEII and the TycC3-T domain we carried out further NMR titration experiments with the acetyl-holo-form of the T domain and the inactive point mutation Ser86Ala of SrfTEII (Supplementary Fig. 4). Based on the magnitude of the chemical shift differences the acetyl-holo-T domain shows a significantly stronger interaction with SrfTEII than the holo-T domain, again in agreement with previous experiments⁵. To determine the relative orientation of both domains, we measured NOEs between selectively protonated methyl (Ile, Leu, Val) and aromatic (Phe) groups of SrfTEII and amide protons of the TycC3-T domain in otherwise perdeuterated proteins^{19, 20}. These NOEs and the chemical shift perturbations allowed us to calculate the structure of the complex between the TEII enzyme and the T domain (Fig. 3b). Previous investigations had revealed that the holo-form of the TycC3 T domain exists in two different conformations that differ in the relative orientation of helix α II and the positioning of the 4'-PP cofactor⁵. They also demonstrated that interaction with the SrfTEII enzyme selects the more open conformation, the H-state⁵. Interestingly, modification of the 4'-PP cofactor with an acetyl group also shifts the equilibrium and locks the T domain in the H-state. Consequently, we used the H-state of TycC3-T in our structure calculations. The interface between both proteins involves helix α II, parts of helix α I and the C-terminus of the T domain, in agreement with the results reported previously⁵. The interface of the SrfTEII enzyme includes the N-terminal loop, the loops between β -strand II and α -helix I (Phe19 - Gly23), between β -strand IV and α -helix III (Gly88 - Met90), β -strand VI and α -helix VII (Asp189, Asp190) and the loop N-terminal to α -helix VIII (Met217, Phe218 and Gln222) as well as the helix-turn-helix "lid" region located between β -strands V and VI (Fig. 3). This arrangement positions the thiol-group of the 4'-PP cofactor in close proximity to the active site residues of the SrfTEII. The rest of the cofactor is surrounded by hydrophobic amino acids (Fig. 3)⁶. The helix-turn-helix motif is involved in the recognition of helix α II of the T domain, with close contacts to the peptide sequence surrounding the active site Ser45, as suggested previously²¹ and observed as well in the EntF T-TEI interaction (Frueh *et al.*, this issue).

The titration experiments further demonstrated that interaction with the acetyl-holo-T domain selects one of the conformations of the thioesterase as indicated by the disappearance of the double peaks in the HSQC spectrum of the complex (Supplementary Fig. 4). Mapping the sites of the double peaks onto the structure of the TEII reveals that amino acids in the helix-turn-helix motif, loops surrounding the active site, residues located at the N-termini of the first and last helix are affected (Supplementary Figs. 1, 4). Unfortunately, the amino acid sequences that show double peaks are separated by stretches that display only one chemical shift value, making an unambiguous assignment to one particular conformation impossible. Thus, the structure of the SrfTEII represents an intermediate between two distinct conformations. The identity of the regions showing this conformational exchange combined with the selection of one state by binding of the T-domain, however, allows us to build a model of the exchange process. Most likely the two different states represent an open and a closed conformation of the enzyme (Supplementary Fig. 7). A similar plasticity of the lid region has been observed in the crystal structure of the SrfTEI enzyme in which two molecules with different conformations are located in the asymmetric unit³. The main difference between both molecules is the position of the lid, with one conformation representing a more open and the other a more closed state. Since both molecules, however, also form a non-native dimer in the crystal structure, the significance of this observation was not clear. The detection of an equilibrium of distinct conformations in the SrfTEII enzyme shifted towards one conformation by the interaction with a modified holo-T domain is the first confirmation of such a conformational exchange process in solution. Similar exchange processes are also observed in the structure of the EntF T-TEI di-domain (Frueh *et al.* this issue).

Inspection of the complex structure further reveals the structural basis for the recognition of short acyl groups attached to the 4'-PP thiol. Comparison of the active sites of the TEII and TEI enzymes shows that, despite the wide opening of the active site of TEII, the space available to accommodate a group attached to the 4'-PP cofactor is rather limited (Fig. 3, Supplementary Figs. 6 and 8). The open conformation of the SrfTEI enzyme is characterized by a pronounced active site cavity with a volume of 630 Å³. Modeling of a heptapeptide based on residual electron density has shown that this cavity is large enough to accommodate the entire peptide and to enforce a conformation that allows cyclization. In contrast to the deep and bowl-shaped cavity of the TEI enzyme the active site of SrfTEII is embedded in a shallow groove that can accommodate only small acyl substituents on the 4'-PP cofactor (Supplementary Fig. 8). This specificity of TEIIs for small acyl substrates was previously demonstrated in kinetic studies with TycF, the TEII of the Tyrocidin synthetase¹³ and with SrfTEII²². To further investigate this selectivity we have performed titration experiments with ¹⁵N-labeled SrfTEII and unlabeled TycC3-T loaded with either a single amino acid (Ala) or a tri-peptide (Phe-Pro-Phe). Whereas the titration with the Ala-loaded T-domain showed results very similar to the titration experiments with the acetyl-holo-T domain with the active site and surrounding residues showing chemical shift changes or significant line broadening, the titration with the tri-peptide loaded T-domain resulted in only minor chemical shift differences, limited to parts remote from the active site (Fig.4). The titration results also indicate differences in the dynamic behavior between acetyl-holo-T and Ala-holo-T on the one hand and holo-T and peptide-loaded holo-T on the other. While titrations with the first group indicate the formation of a stable complex by selection of one of the conformations represented by double peaks, titrations with the second group show a less pronounced selection of one conformation and mainly result in limited chemical shift differences.

In addition to shape and volume of the active site, the interaction of the enzymes with the 4'-PP cofactor might also differ. While in the structure of the SrfTEI the 4'-PP cofactor is almost completely surrounded by a channel formed by the TEI³, it seems to be sandwiched between the SrfTEII and the T domain in the *in trans* complex, demonstrating again the wider opening of the active site.

The structures reported here and their comparison with TEI enzymes show how modulation of the conserved thioesterase fold is used to change the function of the enzyme from one that recognizes the final product of the assembly line to one with a shallow but easily accessible active site that provides a rather unspecific but indispensable repair function.

Recently crystal structure determinations of type I fatty acid synthetases (FAS)^{6,23} have revealed the complex interaction between individual domains in these multi-enzymatic assemblies. These structures underscore the central role that the T domains and different orientations of their cofactors play in the iterative substrate shuttling between active sites.

Methods Summary

The C-terminal His₆-tagged wild type SrfTEII and the inactive Ser86Ala mutant proteins were heterologously expressed in *E. coli* and purified by Ni-chelation affinity chromatography. All isotope enriched protein samples were produced in supplemented M9 minimal media with selectively labeled carbon and nitrogen sources. NMR spectra²⁴⁻²⁹ for backbone and side chain resonance assignment and structure calculation were recorded on Bruker Avance800 and Avance900 spectrometers. All NMR titration experiments¹⁸ of labeled protein samples with small molecules and unlabeled proteins were performed on a Bruker Avance700 spectrometer. Bruker XWINNMR or Topspin 1.3 was used for data processing and UCSF SPARKY 3.111 for resonance assignment and NOE integration. The structure of SrfTEII was calculated on the basis of 2442 NOE upper distance limits and constraints for 92 hydrogen bonds in regular

secondary structure elements using the simulated annealing program CYANA 2.1^{30,31}. The 20 conformers with lowest target function values were energy-refined in explicit water using the RECOORD scripts³² and the CNS 1.1 protocol³³, to represent the solution structure of SrfTEII. All distance constraint violations were smaller than 0.2 Å. The *in trans* complex structure was calculated using the CNS 1.1 simulated annealing protocol. The contact surfaces of both proteins were identified by NMR titration experiments of labeled protein samples with the corresponding unlabeled interacting protein. Constraints to describe the relative domain orientation were obtained from a 3D-¹⁵N-NOESY-TROSY spectrum of a complex of ¹⁵N-labeled, perdeuterated holo-TycC3-T domain and fully perdeuterated, selectively Phe, Ile, Leu and Val protonated SrfTEII^{19,20}. All structural figures were prepared using UCSF CHIMERA 1.2470³⁴ from the Computer Graphics Laboratory, University of California, San Francisco.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Matthias Strieker^{2,4} for editing the manuscript and Birgit Schaefer¹ for her great help and support in sample preparation. We thank Chi Scientific Inc. (Maynard, MA, USA) for the fast and high quality supply of the substrate peptides. The research was funded by the research grant BE-19/11 of the Deutsche Forschungsgemeinschaft (FB, MAM), an enclosed fellowship (AK), the Centre for Biomolecular Magnetic Resonance at the University Frankfurt (BMRZ) and the Cluster of Excellence Frankfurt (Macromolecular Complexes). AK thanks the Human Frontier Science Program Organization for a long-term fellowship awarded in Apr 2007.

Both structure are deposited at the RCSB Protein Data Bank with the accession codes 2RON (structure of SrfTEII) and 2K2Q (complex structure of SrfTEII and H-state TycC3-PCP).

Appendix

Methods

Protein expression and purification

The SrfTEII protein (residues 1-241) and its inactive S86A mutant were heterologously expressed in the *E. coli* strain BL21 (DE3) using a pET-expression vector system containing a C-terminal His₆-tag. They were purified by Ni-chelation affinity chromatography and subsequent gel filtration chromatography using a Pharmacia Superdex-75 column. The purity of all protein samples was validated by SDS PAGE analysis. All labeled samples of the SrfTEII protein were produced in supplemented M9 media with stable isotope enriched glucose (¹³C or ¹³C/²H; Cambridge Isotopes Laboratories, Andover, USA) as the only carbon source and ¹⁵N ammonium chloride (Cambridge Isotopes Laboratories, Andover, USA) as the nitrogen source. For the preparation of fully perdeuterated samples the aqueous solvent was replaced by D₂O and perdeuterated glucose was used as the carbon source. The stable isotope enriched (²H, ¹⁵N, ¹²C) and unlabeled samples of the TycC3-T domain (1-87) were expressed and purified as described previously⁵.

NMR data acquisition and assignment

All NMR spectra for backbone and side chain resonance assignment and structure calculation of the SrfTEII protein were recorded on an Avance800 or an Avance900 spectrometer equipped with a 5 mm triple resonance, z-gradient cryogenic probe at 298 K. The resonance assignment was based on standard triple-resonance experiments, following standard protocols²⁴⁻²⁹. An additional 4-dimensional ¹⁵N/¹⁵N-resolved NOESY spectrum of ¹⁵N-labeled and 70% perdeuterated SrfTEII was recorded to verify the sequential resonance assignment and to define unambiguous amide-amide based distance constraints. DSS (4, 4-dimethyl-4-silapentane-1-

sulfonate) was used as an internal chemical shift reference. All ^{15}N -HSQC based NMR titration experiments¹⁸ were performed on an Avance700 spectrometer at 296 K. The spectral width was set to 12.5 ppm in the proton dimension and 35 ppm in the nitrogen dimension. A total of 2048 points in the direct and 1024 points in the indirect dimension were collected for all HSQC spectra. The TROSY-based ^{15}N -edited NOESY experiments to measure interface distance constraints in the complex of Ser86Ala SrfTEII with the acetyl-holo-TycC3-T domain (1:1 ratio) and wild type SrfTEII with the holo-TycC3-T domain (1:2.5 ratio) were recorded on an Avance 900 with a mixing time of 182 ms. XWINNMR 3.1 or Topspin 1.3 (Bruker) were used for processing and SPARKY 3.111 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) for resonance assignment and NOE peak integration. The SrfTEII protein could be assigned to 92% completeness of all 227 non-proline backbone resonances and 91.5% of all resonances.

Structure calculation and refinement

Structures of SrfTEII were calculated by simulated annealing in torsion angle space with the program CYANA 2.130³¹. Refinements in explicit water (TIP3P) were performed using the RECOORD scripts³² and the CNS 1.1 protocol³³. In all calculations 210 unambiguously identified backbone amide to amide contacts from a 4D $^{15}\text{N}/^{15}\text{N}$ -resolved NOESY experiment²⁷ were applied with upper distance bounds of 6.5 Å. The secondary structure was defined by 184 distance constraints for 92 backbone hydrogen-bonds identified on the basis of unambiguous NOEs in a 3D- ^{15}N -resolved NOESY spectrum and by 278 torsion angle constraints derived from chemical shift values with the program TALOS³⁵. Furthermore 2442 distance constraints were generated based on 7802 NOESY cross peaks from the same 3D- ^{15}N -resolved NOESY and three 3D- ^{13}C -resolved NOESY spectra recorded for aliphatic, methyl and aromatic side chains separately. An initial structure was calculated from 2851 manually assigned NOEs and the torsion angle constraints with CYANA 2.1³¹. The calculation yielded a structural bundle (20 out of 100 calculated structures, sorted by lowest energy) with a precision (RMSD) of 3.8 Å and an accuracy (target function) of 1.9 Å². The CYANA script with automatic NOE assignment was used with the resulting 1706 distance constraints of the previous calculation and additional 94 distance constraints obtained from manually assigned CH-NOEs selective for aromatic side chains. The interpretation of 4951 ambiguous NOEs resulted in an additional 642 distance constraints and in a structural bundle with an RMSD of 1.37 Å for all heavy atoms and an averaged target function value of 4.9 Å² for the 20 structures with lowest CYANA target function values out of 100 calculated conformers. The final set of 20 out of 150 calculated structures does not show distance or van-der-Waals violations larger than 0.20 Å, no angle violations > 2.6° and 75% of all dihedral angles are located in most favored regions and additional 24.8% in additionally allowed regions. Energy-refinement in explicit water (TIP3P) was performed using the RECOORD scripts³² and the CNS 1.1 protocol³³, after transforming the CYANA derived structures into the CNS/XPLOR format. Pockets in SrfTEI3 and SrfTEII were identified and their sizes computed using the CASTp algorithm^{37,38} with a probe radius of 2.0 Å.

NMR titrations and complex structure calculation

Chemical shift perturbations were measured in ^{15}N -HSQC based NMR spectra for titration experiments¹⁸ of ^{15}N -labeled holo- and acetyl-holo-TycC3-T domain with unlabeled SrfTEII and vice versa. The chemical shift perturbations and the line shapes of all titration experiments were analyzed using MestReC 4.9.9.6³⁶. It has been already demonstrated that the H-state of TycC3-T domain is recognized as a substrate by the SrfTEII⁵. Interaction surfaces of the carrier protein and the thioesterase were identified by the titration experiments and the structures of both proteins were used to calculate the structure of the enzymatically active complex. Constraints to describe the relative domain orientation were obtained from a 3D- ^{15}N -resolved NOESY-TROSY spectrum of a ^{15}N -labeled completely perdeuterated holo-TycC3-T domain

and fully perdeuterated selectively Phe, Ile, Leu and Val protonated SrfTEII^{19,20}. Nine unambiguous and 17 additional ambiguous constraints between TycC3-T domain amide protons and SrfTEII FILV-side chain protons were identified. Distance constraints based on these NOEs were applied to the structural calculation of the *in trans* di-domain complex using the CNS 1.1 protocol. All structural figures were prepared using UCSF CHIMERA 1.2470³⁴.

In-vitro modification of TycC3-T domain

The in-vitro modification of unlabeled apo-TycC3-T domain was carried out in 2 ml reaction mixtures of 0.25 mM apo-TycC3-T domain, 0.5 mM acetyl-CoA (SIGMA), 20 μ M Sfp and 5 mM MgCl₂, buffered in 100 mM sodium phosphate at pH 8.0 for 45 min at room temperature. The reaction mixture was subsequently purified by desalting (Econo-pac 10 DG (BIORAD) desalting column) and concentrated using an Amicon Ultra 4 Ultracell - 5k (MILLIPORE) filter device with a molecular weight cut off (MWCO) of 5 kD.

Synthesis of Peptidyl-amino CoA substrates

The general procedure for the synthesis of peptidyl-amino CoA substrates was based on a synthesis described previously³⁹⁻⁴⁰. Briefly, 10 mmol of amino-CoA⁴¹, 15 mmol of PyBOP, and 40 mmol of potassium carbonate were added to 10 mmol of the Boc-protected peptide. The solids were subsequently dissolved in a 1:1 THF/water (500 mL total) mixture and allowed to stir at room temperature overnight. The reaction mixture was directly purified by preparative high-performance liquid chromatography (HPLC) using a single injection on a Phenomenex C18 250 $\times$ 21.2 mm, 10 mm, 100 Å column and eluting using a gradient of 0 to 60 % acetonitrile containing 0.1 % trifluoroacetic acid (TFA) over 30 min and a flow rate of 10 mL/min while monitoring at 260 nm. The identities of the Boc-protected peptidyl-amino CoA substrates were verified by high-performance liquid chromatography-mass spectrometry (HPLC-MS) and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS. Next, cleavage of the Boc-protecting group was carried out by dissolving the Boc-protected peptidyl-amino CoA in a 95:2.5:2.5 mixture of TFA, trifluoroethanol, and water and allowing this mixture to stir at room temperature for 2h. The deprotected peptidyl-amino CoA substrates were purified by preparative HPLC using the same conditions as described above. HPLC-MS and MALDI-TOF MS were used to confirm the identity of the peptidyl-amino CoA substrates.

References for the methods section

24. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U S A* 1997;94:12366–12371. [PubMed: 9356455]
25. Ikura M, Kay LE, Bax A. Improved three-dimensional ^1H - ^{13}C - ^1H correlation spectroscopy of a ^{13}C -labeled protein using constant-time evolution. *J. Biomol. NMR* 1991;1:299–304. [PubMed: 1841700]
26. Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U S A* 1998;95:13585–13590. [PubMed: 9811843]
27. Sattler M, Schleucher J, Griesinger C. Heteronuclear Multidimensional NMR Experiments for the Structure Determination of Proteins in Solution Employing Pulsed Field Gradients. *Prog. NMR Spectrosc* 1999;34:93–158.
28. Ferentz AE, Wagner G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys* 2000;33:29–65. [PubMed: 11075388]
29. Bax A, Clore M, Gronenborn AM. HCCH-TOCSY. *J. Magn. Reson* 1990;88:425–431.
30. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol* 1997;273:283–298. [PubMed: 9367762]

31. Güntert P. Automated NMR structure calculation with CYANA. *Meth. Mol. Biol* 2004;278:353–378.
32. Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AMJJ. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* 2005;59:662–672. [PubMed: 15822098]
33. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* 1998;54:905–921. [PubMed: 9757107]
34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem* 2004;25:1605–1612. [PubMed: 15264254]
35. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 1999;13:289–302. [PubMed: 10212987]
36. Cobas CF, Sardina J. Nuclear magnetic resonance data processing. MestRe-C: A software package for desktop computers. *Concept. Magn. Reson* 2003;19A:80–96.
37. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci* 1998;7:1884–1897. [PubMed: 9761470]
38. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 2006;34:W116–W118. [PubMed: 16844972]
39. Belshaw PJ, Walsh CT, Stachelhaus T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* 1999;284:486–489. [PubMed: 10205056]
40. Sieber SA, Walsh CT, Marahiel MA. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J. Am. Chem. Soc* 2003;125:10862–10866. [PubMed: 12952465]
41. Meier JL, Mercer AC, Rivera H, Burkart MD. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J. Am. Chem. Soc* 2006;128:12174–12184. [PubMed: 16967968]

References

1. Walsh CT. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* 2004;303:1805–1810. [PubMed: 15031493]
2. Kopp F, Marahiel MA. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat Prod Rep* 2007;24(4):735–749. [PubMed: 17653357]
3. Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* 2002;10:301–310. [PubMed: 12005429]
4. Kim Y, Prestegard JH. A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* 1989;28:8792–8797. [PubMed: 2690950]
5. Koglin A, Mofid MR, Löhr F, Schäfer B, Rogov VV, Blum MM, Mittag T, Marahiel MA, Bernhard F, Dötsch V. Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* 2006;312:273–276. [PubMed: 16614225]
6. Leibundgut M, Jenni S, Frick C, Ban N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* 2007;316:288–290. [PubMed: 17431182]
7. Findlow SC, Winsor C, Simpson TJ, Crosby J, Crump MP. Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* 2003;42:8423–8433. [PubMed: 12859187]
8. Johnson MA, Peti W, Herrmann T, Wilson IA, Wüthrich K. Solution structure of Asl1650, an acyl carrier protein from *Anabaena* sp. PCC 7120 with a variant phosphopantetheinylation-site sequence. *Protein Sci* 2006;15:1030–1041. [PubMed: 16597827]

9. Zornetzer GA, Fox BG, Markley JL. Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* 2006;45:5217–5227. [PubMed: 16618110]
10. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen LO. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 2007;15(7):781–792. [PubMed: 17637339]
11. Schwarzer D, Mootz HD, Linne U, Marahiel MA. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* 2002;99:14083–14088. [PubMed: 12384573]
12. Linne U, Schwarzer D, Schroeder GN, Marahiel MA. Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *Eur J Biochem* 2004;271:1536–1545. [PubMed: 15066179]
13. Yeh E, Kohli RM, Bruner SD, Walsh CT. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem* 2004;5:1290–1293. [PubMed: 15368584]
14. Conti E, Stachelhaus T, Marahiel MA, Brick P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* 1997;16:4174–4183. [PubMed: 9250661]
15. Gehring AM, Mori I, Walsh CT. Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* 1998;37:2648–2659. [PubMed: 9485415]
16. Steller S, Sokoll A, Wilde C, Bernhard F, Franke P, Vater J. Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* 2004;43:11331–11343. [PubMed: 15366943]
17. Schneider A, Marahiel MA. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch Microbiol* 1998;169:404–410. [PubMed: 9560421]
18. Zuiderweg ERP. Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry* 2002;41:1–7. [PubMed: 11771996]
19. Gardner KH, Kay LE. Production and incorporation of ^{15}N , ^{13}C , ^2H (^1H -1 methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc* 1997;119:7599–7600.
20. Frueh DP, Vosburg DA, Walsh CT, Wagner G. Determination of all nOes in ^1H - ^{13}C -Me-ILV-U- ^2H - ^{15}N proteins with two time-shared experiments. *J. Biomol. NMR* 2006;34:31–40. [PubMed: 16505962]
21. Lai JR, Koglin A, Walsh CT. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* 2006;45:14869–14879. [PubMed: 17154525]
22. Schwarzer D, Mootz HD, Linne U, Marahiel MA. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U S A* 2002;99:14083–14088. [PubMed: 12384573]
23. Jenni S, Leibundgut M, Boehringer D, Frick C, Mikolásek B, Ban N. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* 2007;316:254–261. [PubMed: 17431175]

Figure 1. Assembly line of the Surfactin non-ribosomal peptide synthetase

The Surfactin synthetase consists of the three subunits SrfA-A, SrfA-B and SrfA-C that comprise seven modules, each being responsible for the incorporation of one amino acid residue. The release of the lipopeptide is catalyzed by the TEI attached to SrfC (module 7). The function of SrfD, the external thioesterase TEII (lower left), is the recycling of misprimed T domains. The 4'-PP cofactor is depicted shortened, attached to the T domains.

Figure 2. Mean structure of the SrfTEII

a, Front view. **b**, Back view. The mean structure is derived from a final ensemble of the best 20 structures out of 150 structures calculated with CYANA and refined in explicit water using the RECOORD scripts and CNS1.1. The structure is color-coded from blue (N-terminus) to red (C-terminus) and vertically rotated by 180° showing the front view (**a**) and the view from the backside (**b**).

Figure 3. Complex structure of SrfTEII and TycC3-T domain

a, Interaction surfaces (red) for SrfTEII (I) and TycC3-T domain (II) are based on chemical shift perturbations observed in NMR titration experiments of ^{15}N -labeled S86A SrfTEII with unlabeled acetyl-holo-T domain and vice versa. The interaction surface identified on the T-domain is identical to results published before⁵. **b**, Ribbon diagram of the refined average structure of the complex of SrfTEII and TycC3-T domain calculated using the simulated annealing protocol of CNS1.1. Surfaces are displayed only for the residues of SrfTEII (blue) and the TycC3-T (red) showing intermolecular NOEs. Residues showing chemical shift changes are colored in the ribbon diagrams accordingly. Some of the active site residues of the TEII triad (Ser86 and His216) and the modified T domain active site (Ser45-4'-PP) showed very severe line broadening (Supplementary Fig 3 and 4). The position of the 4'-PP cofactor shown is based on the position in the free H-state of TycC3-PCP⁵.

Figure 4. Size limitation in substrate recognition by SrfTEII

The combined ^{15}N and ^1H chemical shift perturbation of ^{15}N -HSQC based NMR titration experiments of ^{15}N enriched SrfTEII with unlabeled TycC3-T domain loaded either with an Alanine residues non-hydrolysable bound to the 4'-PP cofactor (**a**) or with the tri-peptide PheProPhe (**b**) reveal a selectivity for small substrates by the SrfTEII. The interaction of SrfTEII with PheProPhe-holo-TycC3-T-domain is limited to residues involved also in holo T-domain recognition, which supports the described selection of small substrates driven by the limited size of the cavity. For residues with resonances broadened beyond detection in the 1:1 SrfTEII, Ala-holo-TycC3-T-domain complex, the chemical shift deviation is set to the maximum and indicated as red bars. Chemical shift differences of amino acids showing chemical shift perturbations are shown as blue bars. The surface of the SrfTEII protein is colored accordingly to the diagrams.