LETTERS

Structural basis for the selectivity of the external thioesterase of the surfactin synthetase

Alexander Koglin[1,2], Frank Löhr[1], Frank Bernhard[1], Vladimir V. Rogov[1,3], Dominique P. Frueh[2], Eric R. Strieter[2], Mohammad R. Mofid[4], Peter Güntert[1,5], Gerhard Wagner[2], Christopher T. Walsh[2], Mohamed A. Marahiel[4] & Volker Dötsch[1]

Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) found in bacteria, fungi and plants use two different types of thioesterases for the production of highly active biological compounds[1,2]. Type I thioesterases (TEI) catalyse the release step from the assembly line[3] of the final product where it is transported from one reaction centre to the next as a thioester linked to a 4′-phosphopantetheine (4′-PP) cofactor that is covalently attached to thiolation (T) domains[4–9]. 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 systems[10–12]. Mispriming of 4′-PP cofactors by acetyl- and short-chain acyl-residues interrupts the biosynthetic system. This repair reaction is very important, because roughly 80% of CoA, the precursor of the 4′-PP cofactor, is acetylated in bacteria[13]. Here we report the three-dimensional structure of a type II thioesterase from *Bacillus subtilis* free and in complex with a T domain. Comparison with structures of TEI enzymes[3,14] shows the basis for substrate selectivity and the different modes of interaction of TEII and TEI enzymes with T domains. Furthermore, we show that the TEII enzyme exists in several conformations of which only one is selected on interaction with its native substrate, a modified holo-T domain.

The cyclic lipoheptapeptide surfactin is one of the most potent biosurfactants showing antibacterial and antiviral activity[2,3,15]. It is synthesized by a complex of three large surfactin synthetase subunits 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 adenylation (A) domain and both the upstream and downstream peptide-bond-forming condensation (C) domains[4–9,16]. In addition, some of the T domains of the surfactin synthetase interact with epimerization domains, located at the carboxy termini of the subunits SrfA-A and SrfA-B, or with the covalently linked TEI[3]. The TEI catalyses 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 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)[10–12]. The importance of this enzyme has been demonstrated by genetic deletions that reduce the production of surfactin by 84%[17]. Owing to the large variety of acylation modifications and the fact that 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 TEI at the end of the last module—rather nonspecific. At the same time, premature cleavage of the correctly growing peptide chain has to be avoided. In addition to this repair function, 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 *cis*-acting SrfTEI shows a similar overall architecture[3], but with notable differences (Supplementary Fig. 5). The most obvious alteration is the amino-terminal truncation of α-helix (Asp 60–Leu 67). Further modifications include the insertion of an extra helix (helix 7; Cys 193–Trp 200) between the active site residues Asp 189 and His 216, 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 structure[3] towards the β-strands 6 and 7 in SrfTEII. Consequently, the active site residues in the SrfTEII structure (Ser 86, Asp 189 and His 216)[10,11] are just partially covered by a short loop (Gln 138–Ala 144) and are 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 SrfTEII show line broadening or even two separate peaks, suggesting that several conformations can occur (Supplementary Fig. 2). Recently, we described the existence of several conformations in the TycC3 T domain[5] (the third module of the tyrocidine A synthetase subunit C, a substrate for SrfTEII) and suggested that this conformational exchange is an important driving force for the selective interaction with other domains of the NRPS system[4–9,16].

NMR titration experiments[18] of ^{15}N-enriched SrfTEII were performed with CoA, myristoyl-CoA, apo- and holo-TycC3 T domains. The resulting chemical shift perturbations (Supplementary Fig. 3) showed 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 altered T domain forms were analysed[5]. Notably, 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 domain[15]. A comparison of the interaction interface of SrfTEII with

[1]Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance and Cluster of Excellence Macromolecular Complexes (CEF), J.W.-Goethe University, 60438 Frankfurt am Main, Germany. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Institute of Protein Research, 142290 Pushchino, Russia. [4]Department of Chemistry/Biochemistry, Philipps University, 35032 Marburg, Germany. [5]Frankfurt Institute for Advanced Studies, 60438 Frankfurt am Main, Germany.

the *in cis* T recognition of EntF T–TEI[14,19] shows 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 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). On the basis of the magnitude of the chemical shift differences, the acetyl-holo-T domain shows a stronger interaction with SrfTEII than the holo-T domain, again in agreement with previous experiments[5]. To determine the relative orientation of both domains, we measured nuclear Overhauser effects (NOEs) between selectively protonated methyl (Ile, Leu and Val) and aromatic (Phe) groups of SrfTEII and amide protons of the TycC3 T domain in otherwise perdeuterated proteins[20,21]. 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 α-helix 2 and the positioning of the 4′-PP cofactor[5]. They also demonstrated that interaction with the SrfTEII enzyme selects the more open conformation, the holo-form specific H state[5]. Notably, 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 the TycC3 T domain in our structure calculations.

The interface between both proteins involves α-helix 2, parts of α-helix 1 and the C terminus of the T domain, in agreement with the results reported previously[5]. The interface of the SrfTEII enzyme includes the N-terminal loop, the loops between β-strand 2 and α-helix 1 (Phe 19–Gly 23), between β-strand 4 and α-helix 3 (Gly 88–Met 90), between β-strand 6 and α-helix 7 (Asp 189, Asp 190), the loop preceding α-helix 8 (Met 217, Phe 218 and Gln 222) and the helix–turn–helix 'lid' region located between β-strands 5 and 6 (Fig. 3). This arrangement positions the thiol-group of the 4′-PP cofactor in close proximity to the active site residues of 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 α-helix 2 of the T domain, with close contacts to the peptide sequence surrounding the active site Ser 45, as suggested previously[22] and observed as well in the EntF T–TEI interaction[19].

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 heteronuclear single-quantum coherence (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 and residues located at the N termini of the first and last helix are affected (Supplementary Figs 1 and 4). Unfortunately, the regions with double peaks are separated by stretches that display only one chemical

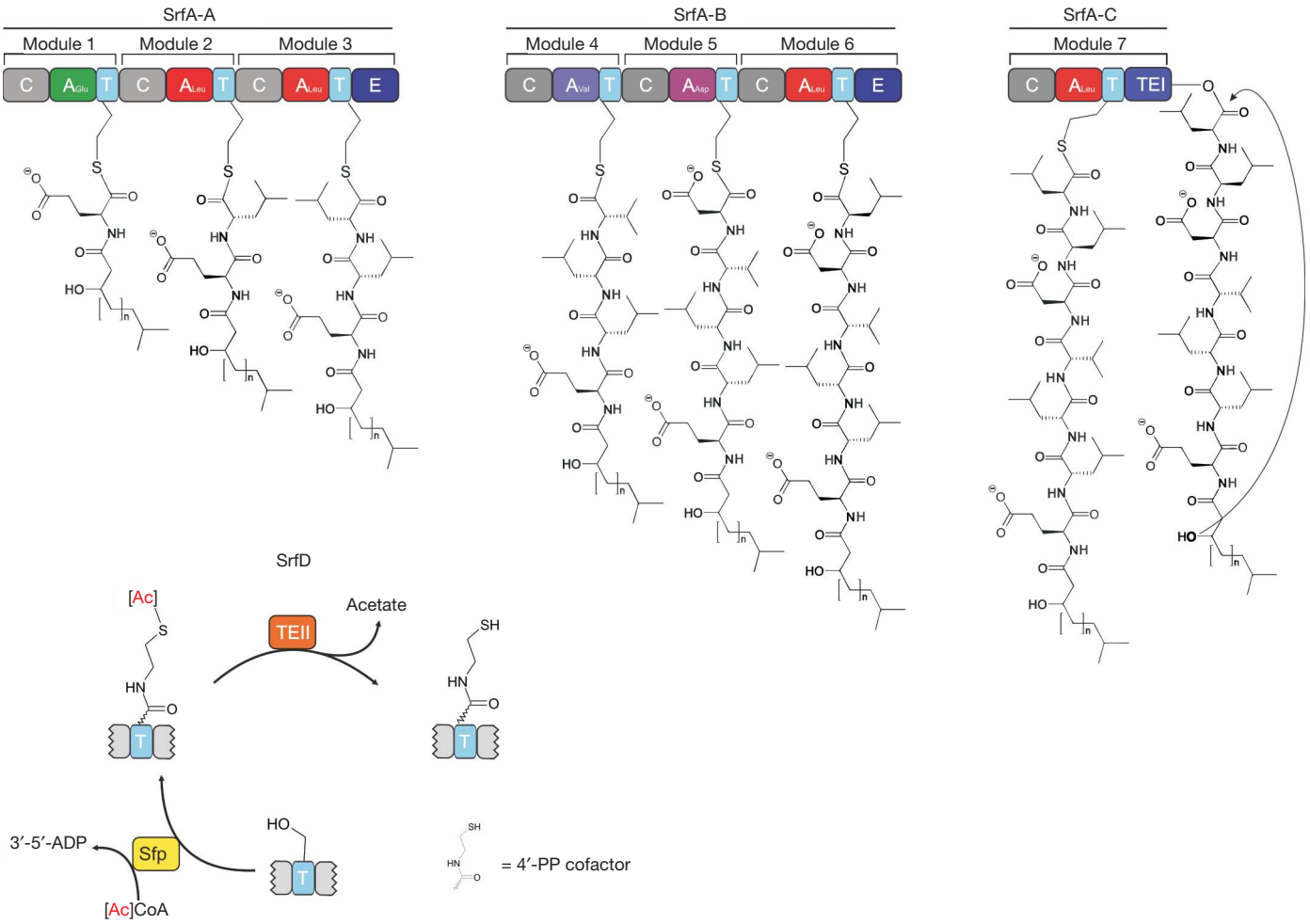

Figure 1 | Assembly line of the surfactin non-ribosomal peptide synthetase. The surfactin synthetase consists of the three subunits SrfA-A, SrfA-B and SrfA-C that together comprise seven modules, each being responsible for the incorporation of one amino acid residue. The release of the lipoheptapeptide is catalysed by 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. Ac, acetyl.

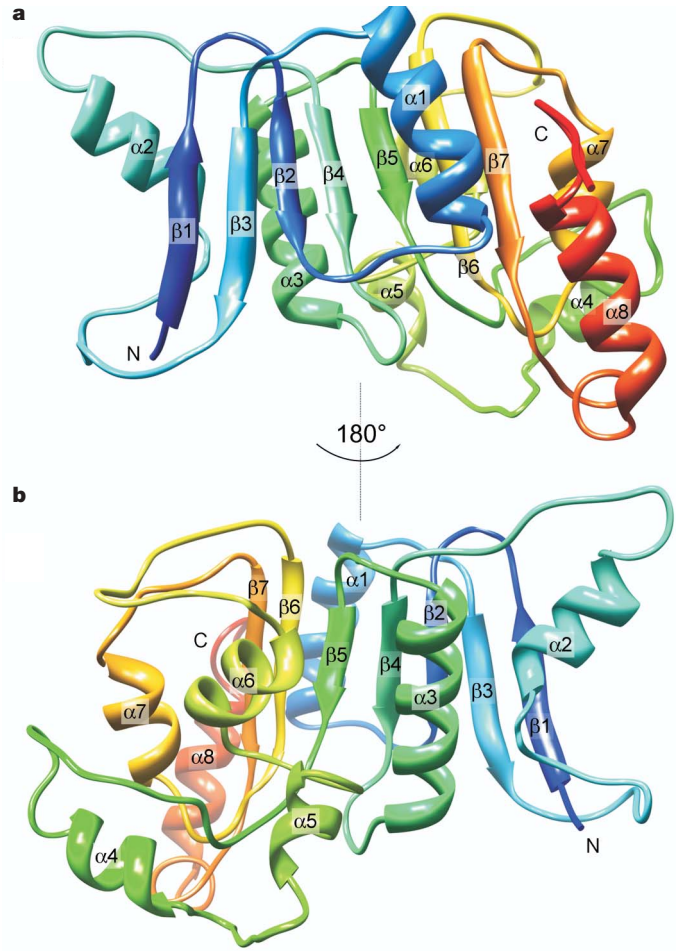

Figure 2 | Mean structure of SrfTEII. **a**, **b**, Ribbon diagrams of the front (**a**) and back (**b**) views are shown. The mean structure is derived from a final ensemble of the best 20 structures out of 150 structures calculated with CYANA and refined in explicit water using the RECOORD scripts and CNS1.1. The structure is colour coded from blue (N terminus) to red (C terminus).

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, allowed us to build a model of the exchange process. The two different states probably 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 unit[3]. 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. Because both molecules, however, also form a non-native dimer in the crystal structure, the importance of this observation was not clear. The detection of an equilibrium of distinct conformations in the SrfTEII enzyme, which 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[19].

The complex structure further shows 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 limited (Fig. 3 and 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 Å^3. Modelling of a heptapeptide on the basis of 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, 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 synthetase[13], and with SrfTEII[23]. To investigate this selectivity further, we performed titration experiments with ^{15}N-labelled SrfTEII and the unlabelled TycC3 T domain loaded with either a single amino acid (Ala) or a tripeptide (Phe-Pro-Phe). Titration with the Ala-loaded T

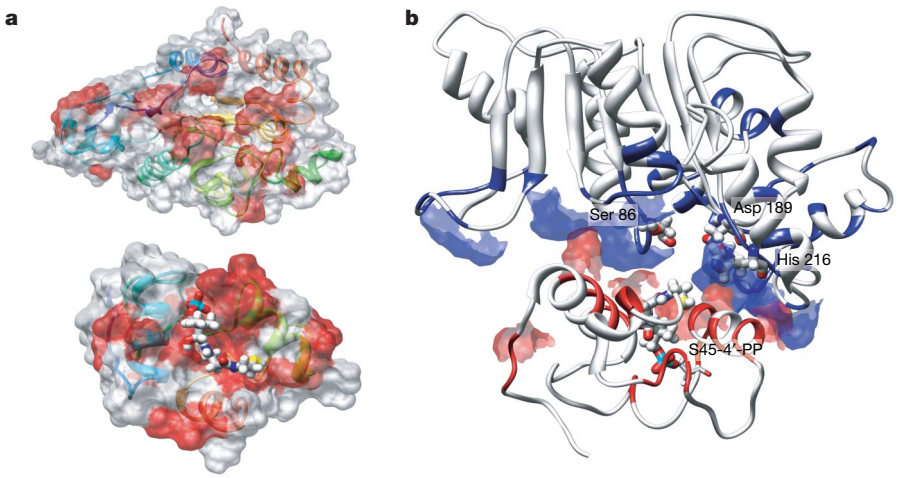

Figure 3 | Complex structure of SrfTEII and the TycC3 T domain. **a**, Interaction surfaces (red) for SrfTEII (top) and the TycC3 T domain (bottom) are based on chemical shift perturbations observed in NMR titration experiments of ^{15}N-labelled Ser86Ala SrfTEII with unlabelled acetyl-holo-T domain and vice versa. The interaction surface identified on the T domain is identical to results published previously[5]. **b**, Ribbon diagram of the refined average structure of the complex of SrfTEII and the 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 domain (red) showing intermolecular NOEs. Residues showing chemical shift changes are coloured in the ribbon diagrams accordingly. Some of the active site residues of the TEII triad (Ser 86 and His 216) and the modified T domain active site (Ser 45-4′-PP) showed very severe line broadening (Supplementary Figs 3 and 4). The position of the 4′-PP cofactor shown is based on the position in the free H state of the TycC3 peptidyl carrier protein (TycC3–PCP)[5].

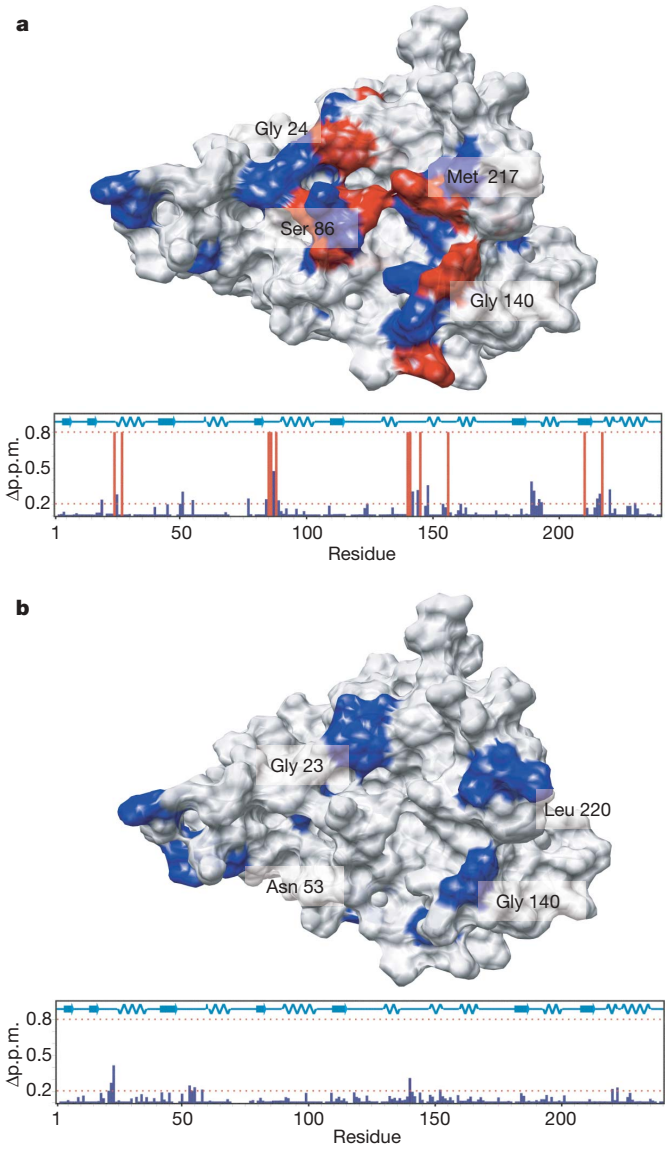

Figure 4 | Size limitation in substrate recognition by SrfTEII. The combined ^{15}N and ^{1}H chemical shift perturbation (Δp.p.m.) of ^{15}N-HSQC-based NMR titration experiments of ^{15}N-enriched SrfTEII with the unlabelled TycC3 T domain loaded either with an alanine residue non-hydrolysably bound to the 4′-PP cofactor (**a**) or with the tripeptide Phe-Pro-Phe (**b**) show a selectivity for small substrates by SrfTEII. The interaction of SrfTEII with the Phe-Pro-Phe–holo-TycC3 T domain is limited to residues also involved in holo-T domain recognition, which supports the described selection of small substrates driven by the limited size of the cavity. Residues with resonances broadened beyond detection in the 1:1 SrfTEII:Ala-holo-TycC3 T domain complex are indicated by red bars. Chemical shift differences of amino acids showing chemical shift perturbations are shown as blue bars. The surface of the SrfTEII protein is coloured accordingly.

domain showed results very similar to those with the acetyl-holo-T domain with the active site and surrounding residues showing chemical shift changes or line broadening. In contrast, the titration with the tripeptide-loaded T domain resulted in only minor chemical shift differences, far from the active site (Fig. 4). The titration results also indicate differences in the dynamic behaviour between acetyl-holo-T and Ala-holo-T on the one hand, and holo-T and peptide-loaded holo-T on the other. Whereas 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 only limited chemical shift differences.

In addition to the shape and volume of the active site, the interaction of the enzymes with the 4′-PP cofactor might also differ. In the SrfTEI structure the 4′-PP cofactor is almost completely surrounded by a channel formed by the TEI[3], whereas it seems to be sandwiched between 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 structures of type I fatty acid synthetases[6,24] have shown the complex interaction between individual domains in these multi-enzymatic assemblies, confirming the central role that the T domains and different orientations of their cofactors have in the iterative substrate shuttling between active sites.

METHODS SUMMARY

The C-terminal His_6-tagged wild-type SrfTEII and the inactive Ser86Ala mutant proteins were heterologously expressed in *Escherichia coli* and purified by Ni-chelation affinity chromatography. All isotope enriched protein samples were produced in supplemented M9 minimal media with selectively labelled carbon and nitrogen sources. NMR spectra[25–27]for backbone and side chain resonance assignment, and structure calculation, were recorded on Bruker Avance800 and Avance900 spectrometers. All NMR titration experiments[18] of labelled protein samples with small molecules and unlabelled 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 2,442 NOE upper distance limits and constraints for 92 hydrogen bonds in regular secondary structure elements using the program CYANA 2.1 (ref. 28). The 20 conformers with the lowest target function values were energy refined in explicit water using the RECOORD scripts[29] and the CNS 1.1 protocol[30], 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 labelled protein samples with the corresponding unlabelled interacting protein. Constraints to describe the relative domain orientation were obtained from a three-dimensional (3D)-^{15}N-nuclear Overhauser enhancement spectroscopy (NOESY)-total correlation spectroscopy (TROSY) spectrum of a complex of ^{15}N-labelled, perdeuterated holo-TycC3 T domain and fully perdeuterated, selectively Phe, Ile, Leu and Val protonated SrfTEII (refs 20 and 21).

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

Received 11 September 2007; accepted 10 June 2008.

1. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303,** 1805–1810 (2004).
2. Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **24,** 735–749 (2007).
3. Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10,** 301–310 (2002).
4. Kim, Y. & Prestegard, J. H. A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **28,** 8792–8797 (1989).
5. Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312,** 273–276 (2006).
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* **316,** 288–290 (2007).
7. Findlow, S. C., Winsor, C., Simpson, T. J., Crosby, J. & Crump, M. P. Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* **42,** 8423–8433 (2003).
8. Johnson, M. A., Peti, W., Herrmann, T., Wilson, I. A. & Wüthrich, K. Solution structure of Asl1650, an acyl carrier protein from *Anabaena* sp. PCC 7120 with a variant phosphopantetheinylation-site sequence. *Protein Sci.* **15,** 1030–1041 (2006).
9. Zornetzer, G. A., Fox, B. G. & Markley, J. L. Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* **45,** 5217–5227 (2006).
10. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl Acad. Sci. USA* **99,** 14083–14088 (2002).

11. Linne, U., Schwarzer, D., Schroeder, G. N. & Marahiel, M. A. Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *Eur. J. Biochem.* **271,** 1536–1545 (2004).
12. Yeh, E., Kohli, R. M., Bruner, S. D. & Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBioChem* **5,** 1290–1293 (2004).
13. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16,** 4174–4183 (1997).
14. Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37,** 2648–2659 (1998).
15. Steller, S. *et al.* Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* **43,** 11331–11343 (2004).
16. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15,** 781–792 (2007).
17. Schneider, A. & Marahiel, M. A. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch. Microbiol.* **169,** 404–410 (1998).
18. Zuiderweg, E. R. P. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41,** 1–7 (2002).
19. Frueh, D. P. *et al.* Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. *Nature* doi:10.1038/nature07162 (this issue).
20. Gardner, K. H. & Kay, L. E. Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ1 methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* **119,** 7599–7600 (1997).
21. Frueh, D. P., Vosburg, D. A., Walsh, C. T. & Wagner, G. Determination of all NOes in ^{1}H- ^{13}C-Me-ILV-U-^{2}H-^{15}N proteins with two time-shared experiments. *J. Biomol. NMR* **34,** 31–40 (2006).
22. Lai, J. R., Koglin, A. & Walsh, C. T. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45,** 14869–14879 (2006).
23. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl Acad. Sci. USA* **99,** 14083–14088 (2002).
24. Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316,** 254–261 (2007).
25. 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. USA* **94,** 12366–12371 (1997).
26. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158 (1999).
27. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33,** 29–65 (2000).
28. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283–298 (1997).
29. Nederveen, A. J. *et al.* RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59,** 662–672 (2005).
30. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Strieker for editing the manuscript and B. Schaefer for her help and support in sample preparation. We thank Chi Scientific Inc. 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 (F.B. and M.A.M.); an enclosed fellowship (A.K.); the Centre for Biomolecular Magnetic Resonance at the University Frankfurt; the Cluster of Excellence Frankfurt (Macromolecular Complexes); and the Volkswagen Foundation (P.G.). A.K. thanks the Human Frontier Science Program Organization for a long-term fellowship awarded in April 2007.

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.R.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.P.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 the TycC3 T domain. M.A.M. provided the purification protocol and biochemical characterization of the SrfTEII. E.R.S. provided the non-hydrolysable CoA derivate and synthesized the peptide and amino acid modified substrate analogues.

Author Information The coordinates have been deposited in the Protein Data Bank under accession number 2RON (structure of SrfTEII) and 2K2Q (complex structure of SrfTEII and H state TycC3–PCP). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to V.D. (VDoetsch@em.uni-frankfurt.de), C.T.W. (Christopher_Walsh@hms.harvard.edu) or M.A.M. (Marahiel@chemie.uni-marburg.de).

METHODS

Protein expression and purification. The SrfTEII protein (residues 1–241) and its inactive Ser86Ala mutant were heterologously expressed in the *E. coli* strain BL21 (DE3) using a pET-expression vector system containing a C-terminal His_6-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 labelled samples of the SrfTEII protein were produced in supplemented M9 media with stable isotope enriched glucose (^{13}C or $^{13}C/^{2}H$; Cambridge Isotopes Laboratories) as the only carbon source and ^{15}N ammonium chloride (Cambridge Isotopes Laboratories) as the nitrogen source. For the preparation of fully perdeuterated samples the aqueous solvent was replaced by D_2O and perdeuterated glucose was used as the carbon source. The stable isotope enriched (^{2}H, ^{15}N, ^{12}C) and unlabelled samples of the TycC3 T domain (1–87) were expressed and purified as described previously[5].

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 protocols[25–27,31–33]. An additional 4D ^{15}N–^{15}N-resolved NOESY spectrum of ^{15}N-labelled 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-sulphonate) was used as an internal chemical shift reference. All ^{15}N-HSQC based NMR titration experiments[18] were performed on an Avance700 spectrometer at 296 K. The spectral width was set to 12.5 p.p.m. in the proton dimension and 35 p.p.m. in the nitrogen dimension. A total of 2,048 points in the direct and 1,024 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 Avance900 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) 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 2.1 (refs 28 and 34). Refinements in explicit water (TIP3P) were performed using the RECOORD scripts[29] and the CNS 1.1 protocol[30]. In all calculations 210 unambiguously identified backbone amide to amide contacts from a 4-dimemsional $^{15}N/^{15}N$-resolved NOESY experiment[26] 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 TALOS[35]. Furthermore, 2,442 distance constraints were generated on the basis of 7,802 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 2,851 manually assigned NOEs and the torsion angle constraints with CYANA 2.1. The calculation yielded a structural bundle (20 out of 100 calculated structures, sorted by lowest target function value) with a precision (root mean square deviation, r.m.s.d.) of 3.8 Å and an average target function value of 1.9 $Å^2$. The CYANA script with automatic NOE assignment was used with the resulting 1,706 distance constraints of the previous calculation and an extra 94 distance constraints obtained from manually assigned CH-NOEs selective for aromatic side chains. The interpretation of 4,951 ambiguous NOEs resulted in an additional 642 distance constraints and in a structural bundle with an r.m.s.d. of 1.35 Å for all heavy atoms and an average target function value of 4.9 $Å^2$ for the 20 structures with lowest CYANA target function values out of 150 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 $>2.6^\circ$ and 75% of all dihedral angles are located in most favoured regions and 24.8% in additionally allowed regions of the Ramachandran plot. Pockets in SrfTEI (ref. 3) and SrfTEII were identified and their sizes computed using the CASTp algorithm[36,37] 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 experiments[18] of ^{15}N-labelled holo- and acetyl-holo-TycC3 T domain with unlabelled SrfTEII and vice versa. The chemical shift perturbations and the line shapes of all titration experiments were analysed using MestReC 4.9.9.6 (ref. 38). It has been demonstrated previously that the H state of the TycC3 T domain is recognized as a substrate by the SrfTEII (ref. 5). 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-labelled completely perdeuterated the holo-TycC3 T domain and fully perdeuterated selectively Phe, Ile, Leu and Val protonated SrfTEII (refs 20 and 21). Nine unambiguous and seventeen additional ambiguous constraints between the TycC3 T domain amide protons and SrfTEII FILV-side-chain protons were identified. Distance constraints on the basis of 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 1.2470 (ref. 39).

***In vitro* modification of the TycC3 T domain.** The *in vitro* modification of the unlabelled 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 μM Sfp and 5 mM $MgCl_2$, 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 desalting column (BioRad)) and concentrated using an Amicon Ultra 4 Ultracell – 5k (Millipore) filter device with a molecular weight cutoff of 5 kDa.

Synthesis of peptidyl-amino CoA substrates. The general procedure for the synthesis of peptidyl-amino CoA substrates was based on a synthesis described previously[40,41]. In brief, 10 mmol of amino-CoA[42], 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 mixture (500 ml total) 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 × 21.2 mm, 10 mm, 100 Å column and eluting using a gradient of 0% to 60% acetonitrile containing 0.1% trifluoroacetic acid over 30 min and a flow rate of 10 ml min^{-1} while monitoring at 260 nm. The identities of the Boc-protected peptidyl-amino CoA substrates were verified by HPLC-mass spectrometry (MS) and matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry. 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 trifluoroacetic acid, trifluoroethanol and water, and allowing this mixture to stir at room temperature for 2 h. 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.

31. Ikura, M., Kay, L. E. & Bax, A. Improved three-dimensional ^{1}H-^{13}C-^{1}H correlation spectroscopy of a ^{13}C-labeled protein using constant-time evolution. *J. Biomol. NMR* **1,** 299–304 (1991).
32. 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. USA* **95,** 13585–13590 (1998).
33. Bax, A., Clore, M. & Gronenborn, A. M. ^{1}H-^{1}H Correlation via isotropic mixing of ^{13}C magentization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **88,** 425–431 (1990).
34. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278,** 353–378 (2004).
35. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13,** 289–302 (1999).
36. Liang, J., Edelsbrunner, H. & Woodward, C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7,** 1884–1897 (1998).
37. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated resiudes. *Nucleic Acids Res.* **34,** W116–W118 (2006).
38. Cobas, C. F. & Sardina, J. Nuclear magnetic resonance data processing. MestRe-C: A software package for desktop computers. *Concepts Magn. Reson. A* **19,** 80–96 (2003).
39. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
40. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284,** 486–489 (1999).
41. Sieber, S. A., Walsh, C. T. & Marahiel, M. A. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J. Am. Chem. Soc.* **125,** 10862–10866 (2003).
42. Meier, J. L., Mercer, A. C., Rivera, H. & Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for *in vivo* protein modification. *J. Am. Chem. Soc.* **128,** 12174–12184 (2006).