Published in final edited form as:
Nature. 2008 August 14; 454(7206): 903–906. doi:10.1038/nature07162.

Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase

Dominique P. Frueh[1], **Haribabu Arthanari**[1], **Alexander Koglin**[1], **David A. Vosburg**[1,†], **Andrew E. Bennett**[1], **Christopher T. Walsh**[1], and **Gerhard Wagner**[1]

1*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.*

Abstract

Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) produce numerous secondary metabolites with various therapeutic/antibiotic properties[1]. Like fatty acid synthases (FAS), these enzymes are organized in modular assembly lines in which each module, made of conserved domains, incorporates a given monomer unit into the growing chain. Knowledge about domain or module interactions may enable reengineering of this assembly line enzymatic organization and open avenues for the design of new bioactive compounds with improved therapeutic properties. So far, little structural information has been available on how the domains interact and communicate. This may be because of inherent interdomain mobility hindering crystallization, or because crystallized molecules may not represent the active domain orientations[2]. In solution, the large size and internal dynamics of multidomain fragments (>35 kilodaltons) make structure determination by nuclear magnetic resonance a challenge and require advanced technologies. Here we present the solution structure of the apo-thiolation–thioesterase (T–TE) di-domain fragment of the *Escherichia coli* enterobactin synthetase EntF NRPS subunit. In the holoenzyme, the T domain carries the growing chain tethered to a 4′-phosphopantetheine whereas the TE domain catalyses hydrolysis and cyclization of the iron chelator enterobactin. The T–TE di-domain forms a compact but dynamic structure with a well-defined domain interface; the two active sites are at a suitable distance for substrate transfer from T to TE. We observe extensive interdomain and intradomain motions for well-defined regions and show that these are modulated by interactions with proteins that participate in the biosynthesis. The T–TE interaction described here provides a model for NRPS, PKS and FAS function in general as T–TE-like di-domains typically catalyse the last step in numerous assembly-line chain-termination machineries.

The *Escherichia coli* enterobactin synthetase (Ent) synthesizes the siderophore enterobactin, a virulence factor used by *E. coli* to infect iron-limited microenvironments of vertebrate hosts. Enterobactin is obtained by three iterative condensations of dihydroxybenzoate (DHB) with serine, followed by macrocyclizing release of the macrolactone that can form high-affinity hexadentate complexes with ferric iron[3,4] (Fig. 1). Ent is a two-module (EntB/EntE and EntF), three-protein assembly line that combines the features of both a type I NRPS, in which all domains interact *in cis* in a single protein chain (EntF), and a type II NRPS, in which the

Correspondence and requests for materials should be addressed to G.W. (gerhard_wagner@hms.harvard.edu) C.T.W. (christopher_walsh@hms.harvard.edu), or D.P.F. (dominique_frueh@hms.harvard.edu).
†Present address: Department of Chemistry, Harvey Mudd College, 301 Platt Boulevard, Claremont, California 91711-5901, USA.
Author Contributions D.P.F., G.W. and C.T.W. designed research and wrote the manuscript. D.P.F. conducted the research including protein expression, resonance assignment and structure calculation. H.A. prepared various samples and helped during the acquisition of NMR experiments. D.A.V. prepared various samples. A.K. helped during the structure calculation and prepared the C domains EntD and Sfp samples. A.E.B. participated in early stages of the project.
Author Information Atomic coordinates have been deposited in the Protein Data Bank under accession code 2ROQ. Reprints and permissions information is available at www.nature.com/reprints.

domains interact *in trans* as free-standing units (EntB/EntE). The dedicated phosphopantetheinyl transferase (PPTase), EntD, loads the apo form of the T domain with 4′-phosphopantetheine (4′-PP), which tethers the growing acyl chain while allowing the bound substrates to navigate between the catalytic sites (Fig. 1a). The EntE and EntF adenylation (A) domains select the appropriate substrates (DHB or serine) and activate them as acyl-adenosine-monophosphates. These are then loaded on the 4′-PP arms attached to Ser 245 of the holo-EntB aryl carrier protein (ArCP) and to Ser 48 of holo-EntF T domains, respectively (Fig. 1b). The EntF condensation (C) domain catalyses the DHB–Ser amide-bond formation (Fig. 1c). The DHB–Ser chains are next transferred to the active site Ser 180 of the thioesterase (TE) domain, thereby making the 4′-PP arm of the T domain available for the next cycle of condensation (Fig. 1d). In this step of the reaction, the 4′-PP arm (wavy line in Fig. 1d) has to reach from Ser 48 of the T domain to Ser 180 of the TE domain, in which the triad Ser 180, Asp 207 and His 313 catalyses the oxoester bond formation between the substrate and Ser 180. The steps shown in Fig. 1b–d are repeated twice to produce the tethered linear DHB–Ser trimer before its intramolecular cyclization by the TE domain of EntF, which constitutes the release of the enterobactin trilactone (Fig. 1e). The structural basis of communication between the T and TE domains in chain-release steps at the end of NRPS assembly lines has been so far unknown.

Here we describe the structure of the 37 kDa apo-EntF T–TE di-domain in which the active site Ser 48 (of the T domain) was replaced with Ala to enable production of a homogeneous sample. This mutation is isomorphous as the NMRs of the T–TE di-domain are unchanged except for the immediate environment of residue 48, and the interdomain contacts are identical (see Methods and Supplementary Fig 1 and Supplementary Fig 2). It represents the physiological structure before CoA addition as indicated at the top of Fig. 1a. The mutated apoprotein allows us to distinguish domain interactions from domain/substrate contributions (including the 4′-PP arm), a crucial knowledge for successful assembly line reprogramming by domain swap. The T and TE domains have a well-defined relative orientation with a primarily hydrophobic interface (Fig. 2a, c). This results in a 1,300 $Å^2$ buried surface area not including the linker, which interacts with both domains and contributes to a further 200 $Å^2$. The globular T domain forms a three-helical bundle that is wedged between the globular core of the thioesterase and two α-helices ($\alpha4_{TE}$–$\alpha5_{TE}$) protruding from this core. These helices resemble two webbed fingers that form a lid covering the T and TE active sites. Ser48Ala and Ser 180 are found 17 Å apart at both ends of a canyon formed by loop $L12_{TE}$ and helix $\alpha4_{TE}$. This allows the 4′-PP arm, which is tethered to Ser 48 of the T domain in the holoprotein and can span up to ~20 Å, to reach Ser 180 of the TE domain. Dynamics data indicate that the two helices ($\alpha4_{TE}$–$\alpha5_{TE}$) form a mobile flap that is able to open to accommodate the 4′-PP arm (see below). Thus, the structure should not be interpreted as rigid but rapidly fluctuating to allow loading of the 4′-PP arm and the growing enterobactin chain. This resembles many other proteins that have to undergo major opening fluctuations to allow binding of target proteins or substrates, such as HIV protease[5].

The T domain contacts TE via the beginning of helix $\alpha2_T$ that contains Ser48Ala, the end of loop $L3_T$, and the single turn helix $\alpha1'_T$ (Fig. 2c). The region near Ser48Ala interacts with both the core of the TE domain, via $L2_{TE}$, and the tips of the helical fingers, which thus cover the T active site (Fig. 2b). In EntF, helix $\alpha1'_T$ stabilizes the hydrophobic core of the T domain with a phenylalanine side chain (Phe 41) anchored in the heart of the helical bundle. An adjacent phenylalanine (Phe 42) interacts with hydrophobic residues of $\beta1_{TE}$. Point mutations of either phenylalanines lead to a disruption of the T/TE interaction (here, T/TE denotes a domain interaction, whereas T–TE indicates two covalently linked domains) and an unfolded T domain (data not shown). $\alpha1'_T$ is absent in the only other structure of an (excised) T domain[6,7], but has been observed in related FAS and PKS acyl carrier proteins (ACP) and may contribute to domain recognition[8–10] (see Supplementary Information). Loop $L3_T$, which interacts with $\beta1_{TE}$, features helical characteristics and is mobile as indicated by the low intensities of the

corresponding NMR signals (Supplementary Fig. 8). In general, the T domain shows internal mobility as further supported by fast NH exchange (Supplementary Fig. 9). Such dynamics have been observed for single-domain ACP-type T domains in FAS systems[11], and may be a general property of acyl and peptidyl carrier proteins (see Supplementary Information).

The TE domain adopts the overall fold of α/β-hydrolases. However, the 'finger' region differs in the two systems for which TE structures have been determined[12,13]. Helix $\alpha4_{TE}$ faces the catalytic triad, located in a conserved region, and delimits part of the 'cyclization bucket' previously described for substrate macrocyclization in surfactin type I thioesterase[12] (TEI; Fig. 2b, grey and white). For surfactin TEI, two conformers have been observed for this helix[12], whereas the electron density was too low in fengimycin TEI to allow any structural interpretation[13] (see Supplementary Information). Moreover, NMR studies of the free standing surfactin type II thioesterase (TEII) indicated two conformations in the corresponding lid[14]. Accordingly, many residues in this region of the EntF TE domain show reduced intensities in their NMR signals, which is indicative of slow modulations of their environments (Supplementary Fig. 8). Furthermore, unlike the core of the TE domain, all amide protons of the finger region exchange rapidly with solvent (Supplementary Fig. 9), indicative of breathing/opening motions. Such mobility may be necessary to allow for access to both T and TE active sites, which would be occluded in a more rigid canyon structure (see later). The mobility also provides conformational plasticity to the TE active site to accommodate the various intermediates stored on Ser 180 and allow the 4′-PP arm to navigate in and out of the canyon in the holoprotein, which would not be possible in a static interpretation of our solution structure.

In addition to the intradomain dynamics reported above, we observed an open form of the T–TE molecule, $T–TE^0$, in which TE and T interact little or not at all (see Supplementary Information and Fig. 3). Indeed, although ACP-type T domains have recently been observed in FAS[15,16], they have previously been reported to be invisible in the crystal structures of FAS or PKS due to low electron density[17–19], in agreement with dynamic domain interactions. To investigate the significance of these dynamic events, the T–TE fragment was titrated with binding partners that participate in different steps of enterobactin synthesis. To become functional as carrier proteins, T domains first need to be primed with 4′-PP arms by a PPTase (Fig. 1a). When the broad-specificity PPTase Sfp[20] was added to a solution of T–TE, a binding site was identified around the active-site residue 48, as expected[6] (Fig. 3a, green). In addition, however, the subset of signals of the $T–TE^0$ conformer increased, suggesting that Sfp drives the equilibrium towards the dislodged form of the di-domain (Fig. 3). This is reasonable because Ser48Ala is partially covered by the tips of the fingers in the di-domain (Fig. 2b). The simultaneous observation of new signals and shifts of resonances (Supplementary Fig. 11) indicates the presence of complex dynamic processes involved in the function of this enzyme. The EntF-specific EntD enzyme[20] caused similar selection of the open state; in contrast, the ACP-specific AcpS PPTase[21] did not (Supplementary Fig. 12), highlighting the specificity of the mechanism.

The EntF C domain[22] condenses the substrates that are covalently attached as thioesters to 4′-PP arms on the EntB and EntF T domains (Fig. 1c). Specific chemical shift changes on addition of the EntF C domain to T–TE showed the C-domain-binding face on T (Fig. 4), and only minor effects were observed on TE. Because the T/C surface does not overlap with the TE interaction surface, no disruption of the T/TE interface is necessary and only a modulation of their interaction is observed. Notably, no interaction was observed when the TycC5 (the fifth module of the tyrocidine synthetase C subunit) C domain was added to a T–TE solution, suggesting that this T–C recognition is EntF specific (see Supplementary Fig. 13). For both C-domain and Sfp interactions, perturbations on TE are not limited to the interface, but affect the finger region and the TE active site, showing that the binding events probably influence the

dynamics and/or the conformations in these areas. These observations apply to the apo form of the molecule and thus only report on domain–domain interactions. It is probable that during biosynthesis the 4′-PP arm, loaded with various intermediates, also affects these dynamic effects.

The structure of the T–TE di-domain provides a starting point for elucidating the communication between domains in NRPS systems. Understanding how the central carrier domains interact with their partner catalytic domains is a prerequisite for understanding the NRPS assembly logic in chain initiation, elongation and termination steps. Our data document a well-defined, yet dynamic, interface between the T and TE domains, consistent with the idea that the T domain must interact with the upstream C and A domains as well as this downstream TE domain. The PPTase and C domain, both of which must visit the T domain, modulate the interaction between the T and TE domains as well as the conformational and/or dynamic properties of each domain. Notably, the *in trans* interaction between surfactin TEII and the TycC3 T domain uses a different domain orientation but places the two active sites at a comparable distance[14]. Moreover, the *in trans* domain interaction selects subsets of conformational states similarly to the *in cis* situation described here. NRPS domains need to communicate with each other at specific points during the peptide synthesis, hence natural product assembly lines must be dynamic entities. The observed malleability may serve to regulate this interplay by stabilizing a given interaction during the course of the corresponding reaction.

METHODS SUMMARY

Recombinant protein was produced in an *E. coli* expression system. The Ser48Ala mutation was used because it was difficult to produce exclusively holo- or apo-T–TE, rather than a mixture, in bacterial expression systems tested. Moreover, wild-type apo-T–TE exists in two inter-converting conformational states for the T domain (data not shown); a similar phenomenon is observed for the apo-T domain of tyrocidin synthetase, TycC3, in which the Ser48Ala mutation locks the protein in a single conformation[6]. Out of the 337 backbone resonances and the associated side-chain resonances, 313 were assigned with transverse relaxation optimized spectroscopy (TROSY) versions of conventional experiments[23] sometimes with non-uniform sampling[24]. A double TROSY hNcaNH was recorded to facilitate the assignment[25]. Most distance restraints were obtained from time-shared nuclear Overhauser enhancement spectroscopy (NOESY) experiments[26]. High-resolution structures were obtained with the program CYANA[27] (backbone root mean square deviations (r.m.s.d.) of 1.2 Å for residues in secondary structure elements, with corresponding r.m.s.d. for the single domains at 0.6 Å (T domain) and 1.0 Å (TE domain) (Fig. 2 and supporting Fig. 3)). The presence of complex dynamic process prevented the use of residual dipolar couplings[28,29]. Titrations were conducted with a constant T–TE concentration and with decreasing concentration of the titrant, to ensure that no degradation was induced by the titration. The determination of solvent accessible residues was obtained by changing the solvent from H_2O to D_2O with a PD10 column (GE healthcare).

METHODS

Protein expression and purification

A pET30a+ plasmid containing the EntF T–TE gene, with a C-terminal His_6 tag, was transformed into *E.coli* BL21(DE3) cells. Isotope enrichment of protonated samples was produced by overexpression in M9 minimal media in H_2O containing $^{15}NH_4Cl$ or ^{13}C glucose or both. Deuterated samples (including methyl labelled samples) were prepared in D_2O containing ^{2}H-^{13}C glucose. Partially deuterated samples were prepared in recycled D_2O with ^{13}C glucose. ‘ILV methyl labelled’ samples, which are selectively protonated on methyls

of Leu, Val and Ile (δ1), were prepared as described in[26,30], with α-ketobutyrate and α-ketoisovalerate precursors either uniformly enriched in ^{13}C, for assignments of side-chain signals, or only ^{13}C enriched at the methyls of the precursors, for the NOE experiments. Samples selectively enriched in ^{15}N for a given amino acid were prepared by supplying the medium with a single ^{15}N-enriched amino acid and 19 unlabelled amino acids[31]. The growth times were 4 h in H_2O, 6 h in recycled D_2O and 12 h in D_2O. Induction times were 2 h in H_2O, 4 h in recycled D_2O and 6 h in D_2O. For all samples, the cells were lysed by sonication and the protein was purified using Ni–NTA resin (Qiagen) followed by FPLC using a Sephadex gel-filtration column (S75 or S200). The final sample concentrations were approximately 300 μM in 20 mM phosphate buffer (pH = 6.7) with 150 mM NaCl, and 1 mM EDTA and dithiothreitol.

Assignment of backbone resonances

TROSY versions of backbone experiments[23,32,33] were used in a 70% deuterated ^{2}H–^{15}N–^{13}C sample. This led to the assignment of 50% of the signals, 30% of which were assigned with the automated assignment program IBIS[34]. HN-TROSY-HSQC[35] spectra were recorded on samples selectively enriched in ^{15}N for Ile, Leu, Val, Lys, Phe, Tyr and Gly (that partially scrambles into Ser). These spectra were combined with the ^{1}H, ^{15}N projections of a CCONH spectrum of a ILV methyl-labelled sample and with the ^{1}H, ^{15}N projections of an HNCOCACB spectrum. These combinations enabled the identification of Ile, Leu, Val, Lys, Phe, Tyr, Gly and Ser residues that were neighbours of Ile, Leu, Val and Gly residues. Similarly, a non-uniformly sampled[24] HNCO spectrum was recorded on a sample uniformly labelled in ^{15}N and selectively labelled on carbonyl carbons of prolines. In addition, a non-uniformly sampled double-TROSY-hNcaNH experiment[25] was designed to correlate triplets of amide nitrogens belonging to contiguous residues. All of the information together led to 70% of the backbone assignment. A different strategy was used to assign residues with non-exchangeable amide protons, which constitute 30% of the TE domain. Non-uniform sampling was used to record high-resolution HNCO, HNCA and HNCOCA spectra of a ^{15}N–^{13}C-labelled sample at 600 MHz with a good signal-to-noise ratio in spite of the fast relaxation due to the sample protonation. A ^{13}C-detected MQ-HACACO[36] was then used to assign Ha signals. These experiments were used together with a ^{15}N-dispersed NOESY-TROSY-HSQC leading to the assignment of 93% of the backbone resonances (313 out of 337 residues).

Assignment of side-chain resonances

HCCONH and CCONH[37] experiments were recorded on a 70% deuterated, uniformly ^{15}N- and ^{13}C-enriched sample. These experiments were also recorded on a ILV-methyl-labelled sample together with HCCH-total correlation spectroscopy (TOCSY)[38]. An HCCH-COSY[39] was recorded on a ^{15}N–^{13}C-enriched sample to assign the signals of residues with non-exchangeable amide protons. Spin systems obtained with this experiment were related to backbone assignments by the MQ-HACACO experiment. Further assignments were obtained with NOESY spectra described below, after initial structure calculations.

Measurement of NMR constraints

Torsion angles constraints were obtained with the program TALOS[40]. Most of the distance constraints were measured with time-shared three-dimensional (3D) TS-NOESY-HN-TROSY/NOESY-HC-PEP-HSQC and 4D HN-TROSY/HC-HSQC-NOESY-HN-TROSY/HC-PEP-HSQC experiments[26] with a mixing time (τ_m) of 200 ms, recorded on a ILV-methyl-labelled sample that also contained protonated phenylalanine[41]. A time-shared HSQC-NOESY (D.P.F., unpublished experiment), with high resolution for the dimension featuring the cross-peaks, was recorded to provide accurate integration of the NOE signals. A NOESY-

HN-TROSY spectrum (τ_m = 90 ms) was recorded on a protonated sample to measure restraints involving non-exchangeable amide protons.

NMR data acquisition

All experiments were conducted on Bruker spectrometers equipped with cryoprobes. For the assignment of backbone resonances, experiments were recorded on a 900 MHz spectrometer, except for those involving ^{13}C–^{15}N samples and selectively labelled samples, which were measured at 600 MHz, and the MQ-HACACO experiment, which was recorded on a 500 MHz spectrometer equipped with a TXI cryoprobe. Side-chain experiments were measured at 600 MHz, except for the HCCH-COSY, which was recorded at 900 MHz. NOESY spectra were recorded at 750 MHz (3D and 4D time-shared experiments) and 900 MHz (3D time-shared HSQC-NOESY and 3D NOESY-TROSY). All spectra were processed with nmrPipe[42] and analysed with CARA[43].

Structure calculation

The structures of each domain were first solved individually, with constraints measured on the T–TE fragment. For the T domain, a homology model with the TycC3 T domain (PDB code 1JMK[20]) could be produced[44]. For the TE domain, sequential alignment with the surfactin TEI (1DNY12) provided erroneous results. However, we could use secondary structure alignment, on the basis of TALOS predictions and NOE patterns, to produce a model with the program DeepView[44,45]. The program CYANA[27] was used to predict NOEs on the basis of these models. The quality of the model could immediately be inspected in the various NOESY spectra and numerous corrections were made, in particular with help from 4D spectra. This resulted in a first low-resolution structure (r.m.s.d. of about 4 Å in each domain and 10 Å for the T–TE fragment). This structure was then used to identify new distance constraints and to make further corrections. This was repeated for each new structure. One-hundred-and-eighty-one hydrogen bond constraints were added for residues in secondary structured regions, once the local r.m.s.d. were shown to be under 0.5 Å, and 497 torsion angle constraints were included when they supported the existing structure. In the end, 4,383 distance constraints (1,163 intraresidues, 2,546 medium range and 935 long range) with 34 interdomain distance constraints were used to produce a structure with a global r.m.s.d. of 2.2 Å. The structural refinement in explicit water was performed using the RECOORD[46] scripts, after converting the CYANA derived structures into the CNS/XPLOR[47] format.

Titrations

See Supplementary Information. The reported chemical shift differences were calculated by $\Delta\delta(^1H, ^{15}N) = ((\Delta\omega_H)^2 + ((1/10)\Delta\omega_N)^2)^{1/2}$, where $\Delta\omega_i$ is the chemical shift difference between the two species for nucleus i. A red dashed line on the resulting figures indicates a threshold of 1 s.d. that was used to single out outliers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank M. Sastry and A. Marintchev for advice in sample preparation; Z. Zhou for providing the Ser48Ala plasmid; Z.-Y. Sun for participating in the development of new NMR techniques; and P. Selenko for comments on the manuscript. This work was supported by NIH grants GM47467, EB 002026, and GM066360 and a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund (D.A.V.). A.K. thanks the Human Frontier Science Program for a long-term fellowship awarded in April 2007.

References

1. Walsh CT. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science 2004;303:1805–1810. [PubMed: 15031493]
2. 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;15:781–792. [PubMed: 17637339]
3. Gehring AM, Bradley KA, Walsh CT. Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. Biochemistry 1997;36:8495–8503. [PubMed: 9214294]
4. 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]
5. Wagner G. The importance of being floppy. Nature Struct. Biol 1995;2:255–257. [PubMed: 7796257]
6. Koglin A, et al. Conformational switches modulate protein interactions in peptide antibiotic synthetases. Science 2006;312:273–276. [PubMed: 16614225]
7. Weber T, Baumgartner R, Renner C, Marahiel MA, Holak TA. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure 2000;8:407–418. [PubMed: 10801488]
8. 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]
9. Johnson MA, Peti W, Herrmann T, Wilson IA, Wuthrich 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]
10. Zornetzer GA, Fox BG, Markley JL. Solution structures of spinach acyl carrier protein with decanoate and stearate. Biochemistry 2006;45:5217–5227. [PubMed: 16618110]
11. Kim Y, Prestegard JH. A dynamic model for the structure of acyl carrier protein in solution. Biochemistry 1989;28:8792–8797. [PubMed: 2690950]
12. Bruner SD, et al. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure 2002;10:301–310. [PubMed: 12005429]
13. Samel SA, Wagner B, Marahiel MA, Essen LO. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. J. Mol. Biol 2006;359:876–889. [PubMed: 16697411]
14. Koglin A, et al. Structural basis for the selectivity of the external thioesterase of the surfactin–synthetase. Nature. (this issue)
15. 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]
16. Lomakin IB, Xiong Y, Steitz TA. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. Cell 2007;129:319–332. [PubMed: 17448991]
17. Maier T, Jenni S, Ban N. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. Science 2006;311:1258–1262. [PubMed: 16513975]
18. Jenni S, Leibundgut M, Maier T, Ban N. Architecture of a fungal fatty acid synthase at 5 Å resolution. Science 2006;311:1263–1267. [PubMed: 16513976]
19. Jenni S, et al. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. Science 2007;316:254–261. [PubMed: 17431175]
20. Lambalot RH, et al. A new enzyme superfamily—the phosphopantetheinyl transferases. Chem. Biol 1996;3:923–936. [PubMed: 8939709]
21. Lambalot RH, Walsh CT. Cloning, overproduction, and characterization of the Escherichia coli holo-acyl carrier protein synthase. J. Biol. Chem 1995;270:24658–24661. [PubMed: 7559576]
22. Roche ED, Walsh CT. Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. Biochemistry 2003;42:1334–1344. [PubMed: 12564937]

23. Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc. Natl Acad. Sci. USA 1998;95:13585–13590. [PubMed: 9811843]
24. Rovnyak D, et al. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. J. Magn. Reson 2004;170:15–21. [PubMed: 15324754]
25. Frueh DP, et al. Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. J. Am. Chem. Soc 2006;128:5757–5763. [PubMed: 16637644]
26. Frueh DP, Vosburg DA, Walsh CT, Wagner G. Determination of all NOes in 1H–13C-Me-ILV-U-2H–15N proteins with two time-shared experiments. J. Biomol. NMR 2006;34:31–40. [PubMed: 16505962]
27. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol 1997;273:283–298. [PubMed: 9367762]
28. Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 1997;278:1111–1114. [PubMed: 9353189]
29. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc. Natl Acad. Sci. USA 1995;92:9279–9283. [PubMed: 7568117]
30. Gardner KH, Kay LE. Production and incorporation of 15N, 13C, 2H (1H-□1 methyl) isoleucine into proteins for multidimensional NMR studies. J. Am. Chem. Soc 1997;119:7599–7600.
31. Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. Methods Enzymol 1989;177:44–73. [PubMed: 2691846]
32. Ferentz AE, Wagner G. NMR spectroscopy: a multifaceted approach to macromolecular structure. Q. Rev. Biophys 2000;33:29–65. [PubMed: 11075388]
33. 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 1999;34:93–158.
34. Hyberts SG, Wagner G. IBIS – a tool for automated sequential assignment of protein spectra from triple resonance experiments. J. Biomol. NMR 2003;26:335–344. [PubMed: 12815260]
35. Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T2 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 1997;94:12366–12371. [PubMed: 9356455]
36. Pervushin K, Eletsky A. A new strategy for backbone resonance assignment in large proteins using a MQ-HACACO experiment. J. Biomol. NMR 2003;25:147–152. [PubMed: 12652123]
37. Grzesiek S, Anglister J, Bax A. Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. J. Magn. Reson. B 1993;101:114–119.
38. Bax A, Clore M, Gronenborn AM. 1H-1H Correlation via isotropic mixing of 13C magentization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. J. Magn. Reson 1990;88:425–431.
39. Ikura M, Kay LE, Bax A. Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. J. Biomol. NMR 1991;1:299–304. [PubMed: 1841700]
40. 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]
41. Medek A, Olejniczak ET, Meadows RP, Fesik SW. An approach for high-throughput structure determination of proteins by NMR spectroscopy. J. Biomol. NMR 2000;18:229–238. [PubMed: 11142513]
42. Delaglio F, et al. NMRPipe a Multidimensional Spectra Processing System Based on UNIX Pipes. J. Biomol. NMR 1995;6:277–293. [PubMed: 8520220]
43. Keller, RLJ. The Computer Aided Resonance Assignment Tutorial. Cantina: 2004.
44. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381–3385. [PubMed: 12824332]

45. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723. [PubMed: 9504803]

46. Nederveen AJ, et al. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 2005;59:662–672. [PubMed: 15822098]

47. Brunger AT, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D 1998;54:905–921. [PubMed: 9757107]

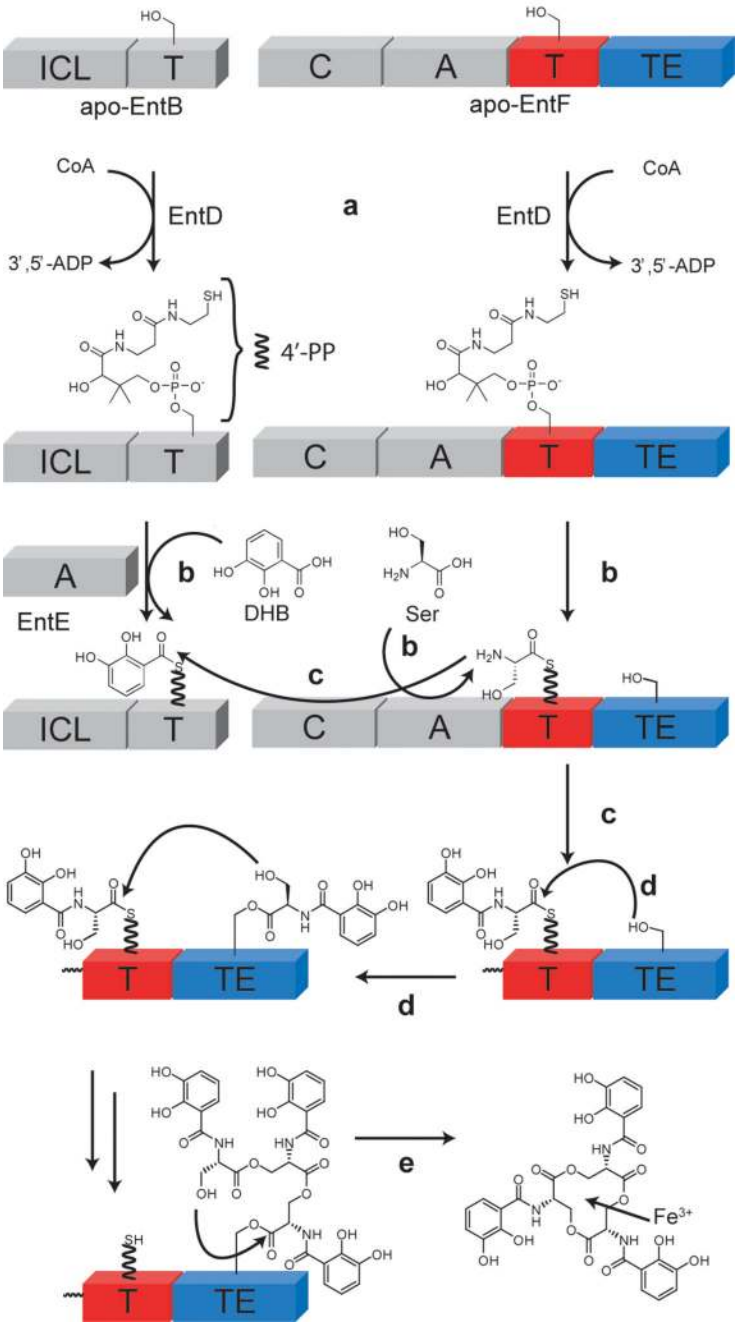

Figure 1. Summary of enterobactin synthesis

The grey, red and blue boxes represent individual domains; the C, A, T and TE domains are described in the text. The EntB ICL domain is an isochorismate ligase that participates in the conversion of chorismate to DHB, a preliminary step to the synthesis of enterobactin. Initially, EntD loads the phosphopantetheinyl arm (4′-PP, represented as a wavy line) onto the active-site serines of apo-EntB and apo-EntF. The synthesis steps (**a**–**e**) are described in the text.

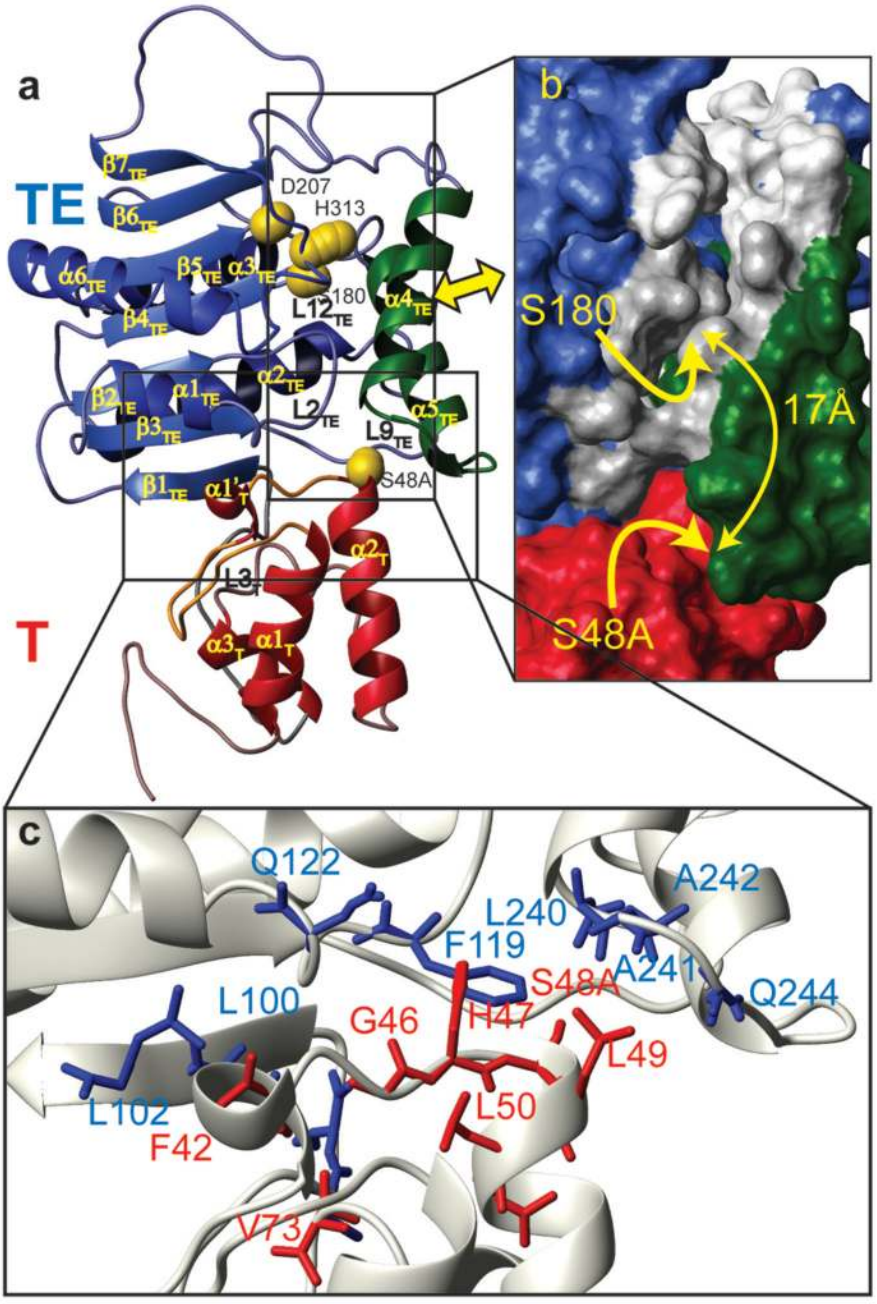

Figure 2. Structure of the EntF T–TE fragment

a, A ribbon diagram is shown. The T domain (red) is wedged between the core (blue) and the lid (green) of the TE domain. Active sites are shown as yellow spheres. The double-headed arrow emphasizes that the flap formed by the $\alpha4_{TE}$ and $\alpha5_{TE}$ helices is relatively mobile, opening frequently, which seems to be necessary to accommodate the 4′-PP arm in the processes depicted in Fig. 1. **b**, Surface representation of the region containing the canyon (grey) which must open to accommodate the 4′-PP arm in the holoprotein. The cyclization bucket[12] is shown in white and grey. In this conformation, Ser48Ala and Ser 180 are buried. **c**, The domain interface is shown. The side chains of all residues giving rise to interdomain nuclear Overhauser effects (NOEs) are shown.

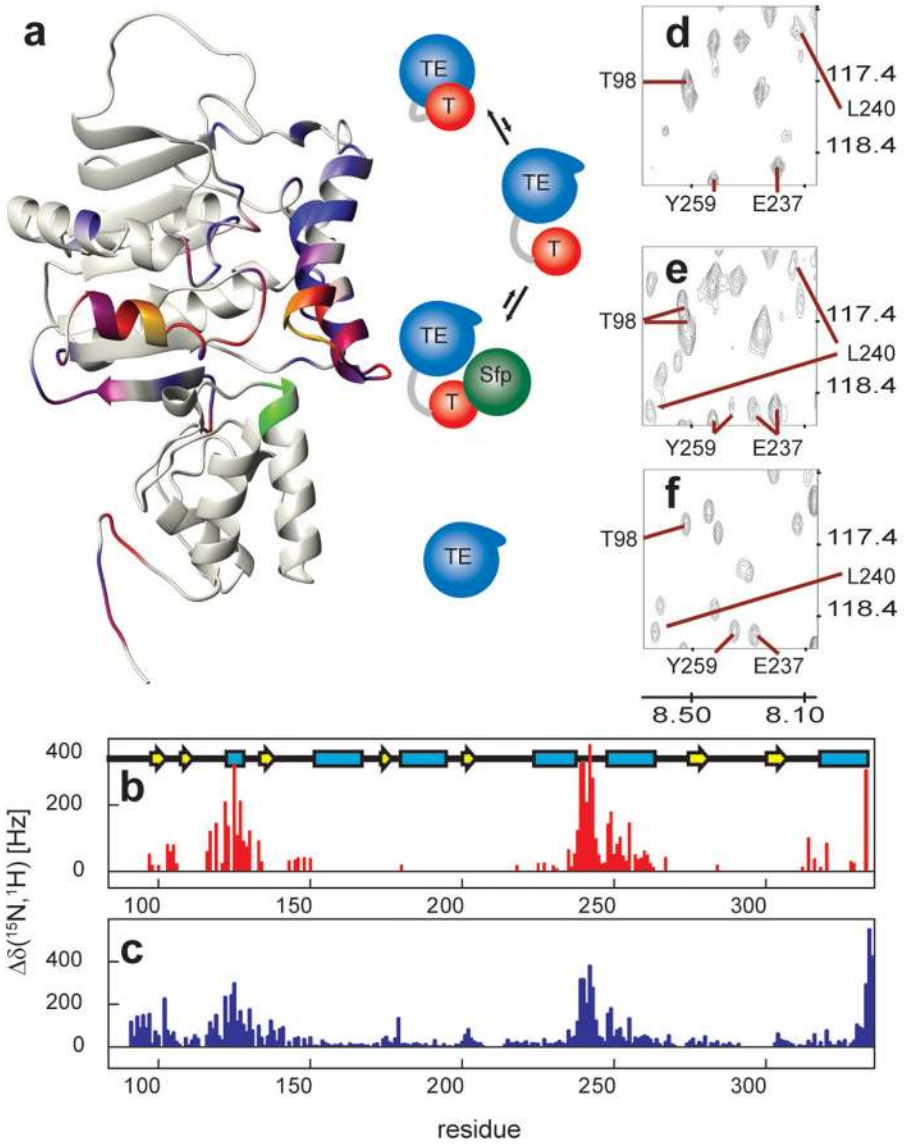

Figure 3. Interaction with Sfp

a, Residues subject to two different environments. The colour code represents small (blue), medium (red) and large (yellow) chemical shift differences between the two forms as reported in **b**. **c**, Chemical shift differences between the major conformer of free T–TE and the excised TE domain. **d**–**f**, Spectral signatures of the Sfp interaction. **d**, The reference spectrum of Ser48Ala. The open form is only visible with longer acquisitions (Supplementary Fig. 3). On addition of Sfp, the signals of the residues in green (see **a**) disappear from the spectrum, and those of the second form appear (**e**). **f**, Corresponding region of the spectrum of the excised TE.

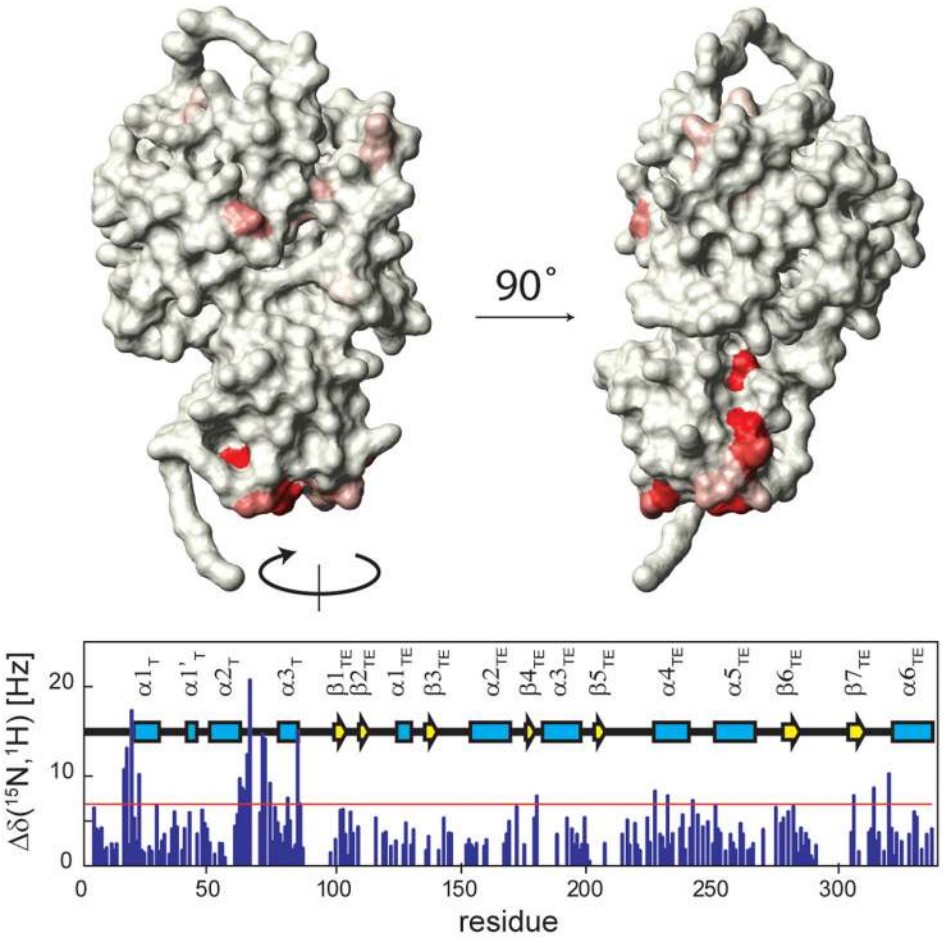

Figure 4. Interaction with the C domain

Many resonances experience shifts on addition of the C domain (bottom), indicating a modulation of the environment of the corresponding residues. Those belonging to the T domain form an interaction surface that does not overlap with the T–TE interface (top). Thus, no disruption of the T–TE interaction is observed. The effects on the TE domain may be due to weak, indirect interactions with the C domain, or to a secondary effect due to fluctuations of the dynamics or modifications of the conformations in these regions. Residues with shifts larger than one s.d. from the mean (red line, bottom) are colour coded with a gradation from white to red.