Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure of the epimerization domain of tyrocidine synthetase A

Stefan A. Samel,[a] Paul Czodrowski[b] and Lars-Oliver Essen[a]*

[a]Biomedical Research Centre/FB15, Philipps Universität, Hans-Meerwein-Strasse 4, 35032 Marburg, Germany, and [b]Computational Chemistry, Merck KGaA, Frankfurter Strasse 250, 64293 Darmstadt, Germany

Correspondence e-mail: essenl@staff.uni-marburg.de

Received 25 October 2013
Accepted 25 February 2014

PDB reference: E domain of TycA, 2xhg

Tyrocidine, a macrocyclic decapeptide from *Bacillus brevis*, is nonribosomally assembled by a set of multimodular peptide synthetases, which condense two D-amino acids and eight L-amino acids to produce this membrane-disturbing antibiotic. D-Phenylalanine, the first amino acid incorporated into tyrocidine, is catalytically derived from enzyme-bound L-Phe by the C-terminal epimerization (E) domain of tyrocidine synthetase A (TycA). The 1.5 Å resolution structure of the cofactor-independent TycA E domain reveals an intimate relationship to the condensation (C) domains of peptide synthetases. In contrast to the latter, the TycA E domain uses an enlarged bridge region to plug the active-site canyon from the acceptor side, whereas at the donor side a latch-like floor loop is suitably extended to accommodate the αIII helix of the preceding peptide-carrier domain. Additionally, E domains exclusively harbour a conserved glutamate residue, Glu882, that is opposite the active-site residue His743. This active-site topology implies Glu882 as a candidate acid–base catalyst, whereas His743 stabilizes in the protonated state a transient enolate intermediate of the L↔D isomerization.

1. Introduction

Many bacteria and fungi harbour large megasynthetases for producing nonribosomally synthesized peptides (NRPs). These secondary metabolites represent a class of functionally and structurally diverse peptide-derived compounds that are often associated with potent functions ranging from antibiotic (penicillins and vancomycin; Nolan & Walsh, 2009) to cytostatic (epothilone; Goodin *et al.*, 2004) and immunosuppressive (cyclosporine; Survase *et al.*, 2011) activities. This renders NRPs attractive to pharmaceutical and medical research, where they are either used as drugs or at least serve as lead structures during drug development.

The biosynthesis of the antibiotic tyrocidine from *Bacillus brevis* is modularly organized using a cluster consisting of three synthetases: TycA, TycB and TycC (Figs. 1*a* and 1*c*). Like other nonribosomal peptide synthetases (NRPSs), TycA/B/C employ a thiotemplate mechanism that is organized like a molecular assembly line (Marahiel & Essen, 2009; Walsh, 2004) for catalysing the formation of the decapeptidic lactam product. As this process is unidirectional (corresponding to a movement from left to right in Fig. 1*a*), translocations of reaction intermediates and positions of domains relative to each other are referred to as 'upstream' or 'downstream'. NRPSs generally consist of several modules, where each one catalyses the activation, incorporation and, where appropriate, the modification of its substrate. A typical elongation module comprises an adenylation domain for amino-acid activation (A), a peptidyl carrier protein domain (PCP) for shuttling

aminoacyl or peptidyl substrates, which are covalently attached *via* a thioester to the cofactor 4′-phosphopantetheine (4′-Ppan), and a condensation domain for peptide-bond formation (C). General aspects of nonribosomal peptide synthesis, its products as well as mechanisms and structures of domains other than epimerization domains are covered in reviews by Koglin & Walsh (2009), Marahiel & Essen (2009) and Strieker *et al.* (2010).

Besides domains essential for peptide elongation, there are optional domains which modify the substrate and thereby greatly increase the product diversity. For example, one striking characteristic of many nonribosomal peptides is the occurrence of macrocyclic structures (Fig. 1*b*), which are typically introduced in the final step of synthesis. In most cases, as in TycC, this reaction is catalysed by a thioesterase domain, in which the enzyme-bound peptide ester is macrocyclized and cleaved off the enzyme. Another characteristic of nonribosomal peptides is that they often contain unusual building blocks such as D-configured and nonproteinogenic amino acids as well as aryl and fatty acids. Both macrocyclization and unusual building blocks have been the subject of numerous studies and have been shown to be crucial for the bioactivity of the product (Kohli *et al.*, 2002). Obviously, these modifications of NRPs compared with ribosomally synthesized peptides create an enormous structural and functional diversity.

There are three known strategies by which NRPSs incorporate D-configured amino acids. The most common utilizes NRPS epimerization domains. These domains alter the configuration of already activated thioester-bound amino acids. The catalysed reaction itself leads to an equilibrium of two enantiomers or diastereomers. The exclusive incorporation of the D-configured amino acids then depends on the D-specificity of the succeeding, *i.e.* downstream, condensation domain. The biosynthesis pathway of arthrofactin, a biosurfactant from *Pseudomonas* sp. MIS38, exemplifies the second strategy. Here, domains with both condensation and epimerization activity (dual C/E domains) are part of the corresponding NRPS (Balibar *et al.*, 2005). Finally, in cyclosporin synthesis the conversion of the first residue (D-Ala 1) is catalysed by an external epimerase (Hoffmann *et al.*, 1994).

NRPS epimerization domains generally exhibit low substrate specificity. Using excised PCP-E bidomain proteins, epimerization domains were found to accept a rather broad range of noncognate substrates, albeit with lowered efficiency (Stein *et al.*, 2005). Nevertheless, they also show a certain degree of specificity. Whereas the aminoacyl-accepting epimerization domain from the synthetase TycA (TycA-E) is able to accept both PCP-bound aminoacyl as well as peptidyl substrates, the peptidyl-epimerizing domain of TycB fails to epimerize amino-acyl substrates (Stein *et al.*, 2006). The reaction of the cofactor-independent epimerization domain is found to occur by a deprotonation mechanism, most likely *via* a resonance-stabilized enolate intermediate. The C^{α}—H acidity of a thioester-bound amino acid may be comparable to that of the C^{α} atom in *n*-octanoyl-CoA, for which the pK_a was reported to be ∼21 (Vock *et al.*, 1998). Therefore, compared with an amide or an oxoester, this substrate may be considered as pre-activated but still requires further activation by the enzyme, *e.g.* by electrostatic stabilization of the enolate intermediate as exemplified by acyl-CoA dehydrogenases (Nishina *et al.*, 2003; Thorpe, 1989).

We present the first crystal structure of an NRPS epimerization domain at 1.5 Å resolution. The studied epimerization domain of tyrocidine synthetase A catalyses the D/L conversion of 4′-Ppan-bound phenylalanine, the first building block

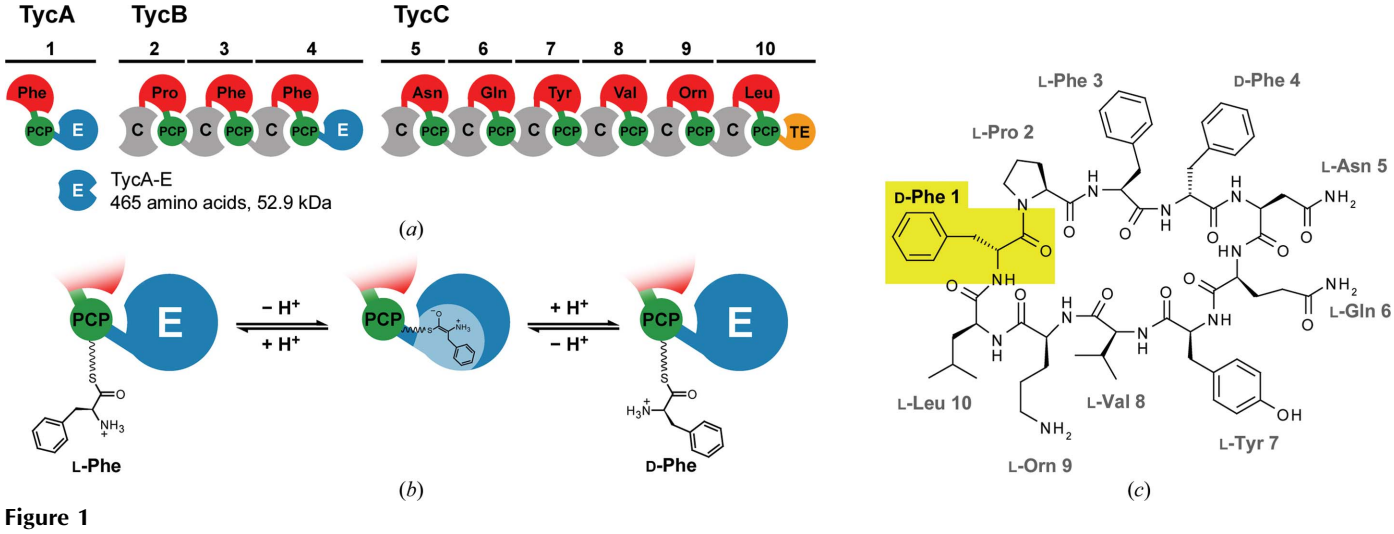


Figure 1
Tyrocidine A biosynthesis and mechanism of the epimerization reaction. (*a*) Schematic depiction of the tyrocidine biosynthesis cluster consisting of three distinct enzymes: TycA, TycB and TycC. In analogous NRP biosynthesis clusters comprising more than one polypeptide, epimerization domains are usually located at the C-termini of the synthetases to mediate mutual recognition with the succeeding synthetase. (*b*) Reaction mechanism of epimerization domains as postulated by Stachelhaus & Walsh (2000). After ATP-dependent activation of the incoming substrate L-Phe and subsequent transesterification onto the 4′-phosphopantetheine cofactor of the PCP domain, the C^{α} proton is abstracted by the epimerization domain. The resulting resonance-stabilized enolate intermediate is then reprotonated to give a mixture of L-Phe and D-Phe. (*c*) Structure of the peptide antibiotic tyrocidine. The cyclic decapeptide adopts a β-hairpin-like structure and contains two residues in the D-configuration: D-Phe 1 and D-Phe 4.

of tyrocidine. D-Amino acids introduced during NRP assembly not only affect the final conformation of the product peptide but also the acceptance of intermediate downstream processing units such as the thioesterase domain. Accordingly, epimerization domains such as TycA-E are crucial for the proper processing of NRP intermediates during synthesis as well as for the bioactivity of the peptidic product.

2. Materials and methods

2.1. Cloning of a C-terminally truncated variant of the TycA E domain

For crystallization, a truncated version of the TycA E domain was prepared that lacks 35 C-terminal amino-acid residues including the COM helix required for interaction with TycB (Hahn & Stachelhaus, 2004) while preserving the vector-encoded C-terminal His_6 tag. For this deletion, inverse PCR mutagenesis was applied to the pQE60-based expression plasmid coding for TycA-E (Linne *et al.*, 2001) using the primers 5′-cca ata **ctg cag** TCT CAT CAC CAT CAC CAT CAC TAA GC-3′ and 5′-cta aat **ctg cag** GAC GCT GAA ATC GCT GGG C-3′ (homologous sequence in capital letters, restriction sites in bold); the parental DNA was then degraded by *Dpn*I and the newly introduced *Pst*I sites were digested and religated before transformation. The obtained expression plasmid coding for TycA-E (Swiss-Prot entry TYCA_BREPA) from Lys604 to Cys1053 and followed by the C-terminal affinity tag LQSHHHHHH was verified by dideoxy sequencing.

2.2. Overexpression and purification of the recombinant TycA E domain

The expression plasmid for TycA-E was transformed into *Escherichia coli* M15[pREP4] cells (Qiagen). Transformed cells were grown in LB medium at 30°C. Overexpression of the recombinant protein was induced by addition of 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{595} of 0.6 and further shaking for 3 h. Cells were harvested by centrifugation and resuspended in buffer *A* (25 m*M* HEPES, 50 m*M* NaCl pH 8.0), frozen in liquid nitrogen and stored at −80°C. Frozen cells were thawed on ice and lysed with an Emulsiflex-C5 fluidizer (Avestin). To remove cell debris, the raw lysate was centrifuged (JA-20; 15 000 rev min^{-1}; 2 × 15 min; 4°C). The C-terminally His_6-tagged protein was purified by Ni^{2+}–NTA chromatography (Qiagen) using buffer *A* and buffer *B* (25 m*M* HEPES, 50 m*M* NaCl, 500 m*M* imidazole pH 8.0) to apparent homogeneity as judged by SDS–PAGE analysis. Residual imidazole was removed using a PD-10 column (GE Healthcare) and SEC buffer (10 m*M* HEPES, 100 m*M* NaCl pH 7.5). Protein-containing fractions were pooled and concentrated to a final protein concentration of 12 mg ml^{-1}. Aliquots of the sample were flash-frozen in liquid nitrogen and stored at −80°C.

2.3. Crystallization and data collection

Prior to crystallization, protein samples were sterile-filtered (Ultrafree-MC, Millipore). Initial crystallization trials utilized a set of nine screens from Sigma–Aldrich and NeXtal (Qiagen) and a Microsys 4004 crystallization robot (Cartesian). After 4 d, several conditions yielded small crystals. Subsequent reproduction showed that those crystals grown in 0.1 *M* potassium phosphate, 20% PEG 8000 appeared to be most suitable for optimization. The best crystals were obtained from 0.05 *M* potassium phosphate, 22% PEG 8000 in hanging drops at 4°C.

Selenomethionine (SeMet)-labelled protein was produced according to the method of Van Duyne and coworkers (Samel *et al.*, 2007; Van Duyne *et al.*, 1993). The quantitative exchange of all six methionines to selenomethionine was confirmed by mass-spectrometric analysis. The SeMet derivative crystallized under similar conditions as native TycA-E at 4°C.

Before flash-cooling in liquid nitrogen, the crystals were briefly transferred into a drop of mother liquor additionally containing 20%(*v*/*v*) glycerol as a cryoprotectant. The best crystals of native TycA-E generally diffracted to 2.7 Å resolution at the in-house diffractometer (rotating-anode Bruker Nonius FR591 operated at 50 kV and 80 mA, MAR Research mar345dtb imaging plate) with mosaicities of about 0.6°. On beamline ID-29 at ESRF, Grenoble, France a native data set was recorded to a maximal resolution of 1.50 Å. Multiple-wavelength anomalous dispersion (MAD) data for SeMet-labelled TycA-E were collected to a resolution of 1.85 Å on beamline PSF14.2, BESSY II, Berlin, Germany.

2.4. Structure determination and crystallographic refinement

Initial attempts to solve the crystal structure by molecular replacement using available C-domain structures (Keating *et al.*, 2002; Samel *et al.*, 2007; Tanovic *et al.*, 2008) failed because of low sequence identities of 13–18% and differing subdomain arrangements. Accordingly, the TycA-E structure was solved by MAD techniques. Data were indexed and processed using *MOSFLM* (Leslie, 1992). The structure was initially determined using the *Auto-Rickshaw* suite at EMBL, Hamburg, Germany (Panjikar *et al.*, 2005). Here, *SHELXD* identified the positions of all six Se atoms and *SHELXE* improved the initial phases by density modification (Schneider & Sheldrick, 2002). Automated model building by *ARP*/*wARP* (Perrakis *et al.*, 2001) generated a model that contained ∼80% of the residues. This initial model was further refined in cycles of manual model building by *Coot* (Emsley & Cowtan, 2004) alternating with refinement by *REFMAC*5 (Winn *et al.*, 2011). Anisotropic displacement parameters for the two subdomains of TycA-E were obtained by translation/libration/screw analysis (TLS) refinement using Lys604–Asn792/Asp971–Ser1002 and Glu793–Phe970/Glu1003–His1057 as the TLS blocks. The refinement converged without significant difference electron densities left at R_{work} = 0.161 and R_{free} = 0.180. Superpositions and calculations of root-mean-square deviations were calculated using the *SSM* algorithm as implemented in *Coot* (Emsley & Cowtan, 2004).

Despite numerous attempts to obtain information about the substrate-binding mode of the E domain by means of soaking or co-crystallization, no experimental results are available.

Table 1
Data-collection and refinement statistics for TycA-E.

Values in parentheses are for the highest resolution shell

	Native	SeMet peak	SeMet inflection	SeMet remote
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Wavelength (Å)	1.03946	0.97965	0.97984	0.91841
Resolution (Å)	47.7–1.50	36.1–1.87	36.1–1.87	34.0–1.75
Unit-cell parameters (Å)	a = 47.7, b = 74.6, c = 124.8	a = 48.4, b = 74.4, c = 124.0	a = 48.4, b = 74.4, c = 124.0	a = 48.5, b = 74.6, c = 124.2
No. of reflections	299872	136544	135970	166901
Unique reflections $\|F\|$	72029	38150	38184	46465
Completeness	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$\langle I\rangle/\langle\sigma(I)\rangle$	12.3 (1.8)	17.8 (7.7)	17.6 (7.2)	15.6 (3.9)
$R_{\rm merge}$†	0.070 (0.868)	0.055 (0.173)	0.055 (0.190)	0.059 (0.242)
R factor/$R_{\rm free}$‡	0.161 (0.257)/ 0.180 (0.241)			
Used reflections	70891 (5164)			
Average B factor (Å^2)	12.9			
No. of atoms	3704			
No. of waters	441			
No. of heterogens	31			
R.m.s.d., bonds (Å)	0.01			
R.m.s.d., angles (°)	1.20			

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ R factor $= \sum_{hkl}||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$; $R_{\rm free}$ is calculated with a randomly selected 1.5% of the data.

Similar problems in obtaining peptidyl/NRPS domain complexes by either cocrystallization or soaking with peptidyl/amino-acid analogues have been encountered before for other NRPS domains (Bloudoff *et al.*, 2013; Bruner *et al.*, 2002; Samel *et al.*, 2006, 2007; Tanovic *et al.*, 2008), indicating their low intrinsic affinity for bound substrate intermediates.

2.5. Modelling and molecular-dynamics simulation of the substrate-loaded PCP-E bidomain

The modelling of a substrate-loaded TycA PCP-E bidomain was based on (i) the structure of trichothecene 3-*O*-acetyltransferase TRI101 in complex with coenzyme A (PDB entry 2zba; Garvey *et al.*, 2008), (ii) the structure of the PCP domain of the fifth module of TycC (PDB entry 2jgp; Samel *et al.*, 2007) and (iii) sterical constraints for PCP-E domain arrangements. For details refer to the Supporting Information[1], where pK value calculations of catalytic residues as well as the setup and analysis of molecular-dynamics studies based on the substrate-loaded model are also described. Protein interfaces were analysed and quantified by the *PISA* service of the European Bioinformatics Institute (Krissinel & Henrick, 2007).

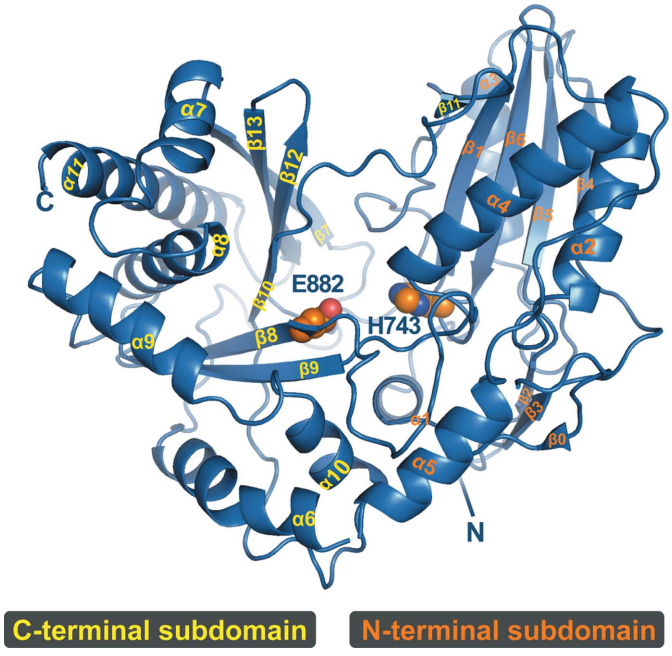


Figure 2
Overall structure of the epimerization domain of tyrocidine synthetase A at 1.50 Å resolution. The TycA E domain consists of two structurally similar subdomains, referred to as the N-terminal and the C-terminal subdomain. Each subdomain has a chloramphenicol acetyltransferase (CAT)-like fold. The conserved residue His743 was shown to be crucial for catalysis. Along with another conserved residue, Glu882, it is shown in orange and marks the position of the active site. Apart from an additional N-terminal β-strand (β0) the secondary-structure assignment for the TycA E domain corresponds to that described before for the TycC6 C domain (Samel *et al.*, 2007).

3. Results and discussion

3.1. Overall structure of the TycA epimerization domain

The crystal structure of the recombinant TycA E domain (Lys604–Cys1053) was determined by multiple-wavelength anomalous dispersion (MAD) using SeMet-labelled protein at 1.85 Å resolution and was subsequently refined to 1.5 Å resolution using the native data set (Table 1). The epimerization domain crystallized in space group $P2_12_12_1$ with one polypeptide of TycA-E and 399 water molecules per asymmetric unit. In the crystal structure of TycA-E (Fig. 2), the epimerization domain is well defined from Lys604 to Cys1053 of TycA and additionally shows an elongation of the C-terminal helix α11 by seven residues which were derived from the vector-encoded affinity tag. Structurally, the TycA E domain comprises two subdomains, each of which exhibits a chloramphenicol acetyltransferase (CAT)-like fold (Lewendon *et al.*, 1994). The active site is in the centre of the domain, being buried at the interface of the two subdomains.

Phylogenetic studies have indicated that epimerization domains are closely related to condensation domains (Rausch *et al.*, 2007; Roongsawang *et al.*, 2005). A structural comparison of the TycA E domain against the PDB using the *SSM* server (Krissinel & Henrick, 2004) indeed shows that this domain is mostly related to NRPS condensation domains such as VibH (Keating *et al.*, 2002), TycC6 (Samel *et al.*, 2007),

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: MH5117).

Table 2
SSM alignment of the TycA epimerization domain.

Protein†	Organism	PDB code	R.m.s.d. (Å)	N_{align}‡	$\%_{seq}$§	Subtype
TycA-E	*Bacillus brevis*	2xhg	0	465	100.0	NRPS
VibH	*Vibrio cholerae*	1l5a	3.283	376	13.6	NRPS
TycC6	*Bacillus brevis*	2jgp	2.611	365	18.4	NRPS
SrfA-C	*Bacillus subtilis*	2vsq	3.047	361	14.4	NRPS
CDA1	*Streptomyces coelicolor*	4jn3	3.488	378	14.3	NRPS
TlmII	*Streptoalloteichus hindustanus*	4hvm	3.458	335	12.8	NRPS
PapA5	*Mycobacterium tuberculosis*	1q9j	3.810	328	13.7	PKS
TRI3 trichothocene-15-*O*-acetyltransferase	*Fusarium sporotrichioides*	3fot	3.790	334	13.5	PKS
Vinorine synthase	*Rauvolfia serpentina*	2bgh	3.780	291	8.2	PKS
TRI101 3-*O*-acetyltransferase–CoA–T-2	*Fusarium sporotrichioides*	2zba	4.256	317	8.5	PKS
HCT	*Coffea canephora*	4g0b	3.760	297	9.1	CAT
HCT	*Sorghum bicolor*	4kec	4.083	310	10.6	CAT
DmAT–malonyl-CoA	*Dendrathema morifolium*	2e1t	4.569	301	8.3	CAT
Choline acetyltransferase	*Homo sapiens*	2xhg	3.951	284	6.0	CAT
Choline acetyltransferase	*Rattus norvegicus*	1q6x	3.899	278	5.8	CAT
Carnithin palmitoyltransferase II	*Rattus norvegicus*	2fyo	4.059	287	8.7	CAT
Carnithin palmitoyltransferase–CoA	*Mus musculus*	1ndi	4.201	282	7.4	CAT
Carnithin palmitoyltransferase	*Homo sapiens*	1nm8	4.299	281	7.8	CAT

† Structure alignment was carried out using the *SSM* server (http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver) with default settings (Krissinel & Henrick, 2004). Initial hits were manually sorted to minimize redundancy for improved significance. ‡ Length of alignment N_{align}: query and target structures are aligned in three-dimensions on the basis of spatial closeness, minimizing r.m.s.d. and the maximizing the number of aligned residues. § The sequence identity $\%_{seq}$ is the percentage of pairs of identical residues N_{ident} among all aligned: $\%_{seq} = N_{ident}/N_{align}$.

SrfA-C (Tanovic *et al.*, 2008), CDA1 (Bloudoff *et al.*, 2013) and TlmII followed by some acyltransferases (ATs) derived from polyketide synthases (Table 2). The observed sequence identities, 13–18% for NRPS C domains, 8–14% for PKS AT domains and <11% for stand-alone CAT domains, indicate a common evolutionary origin for paired CAT-like subdomains in megasynthetases, as exemplified by the NRPS C and PKS AT domains. The close relationship of the TycA E domain to NRPS C domains is corroborated by the superposition with the C domain from module 6 of the tyrocidine synthetase C, which shows an r.m.s.d. of 2.61 Å for 365 C^{α} positions (Fig. 3).

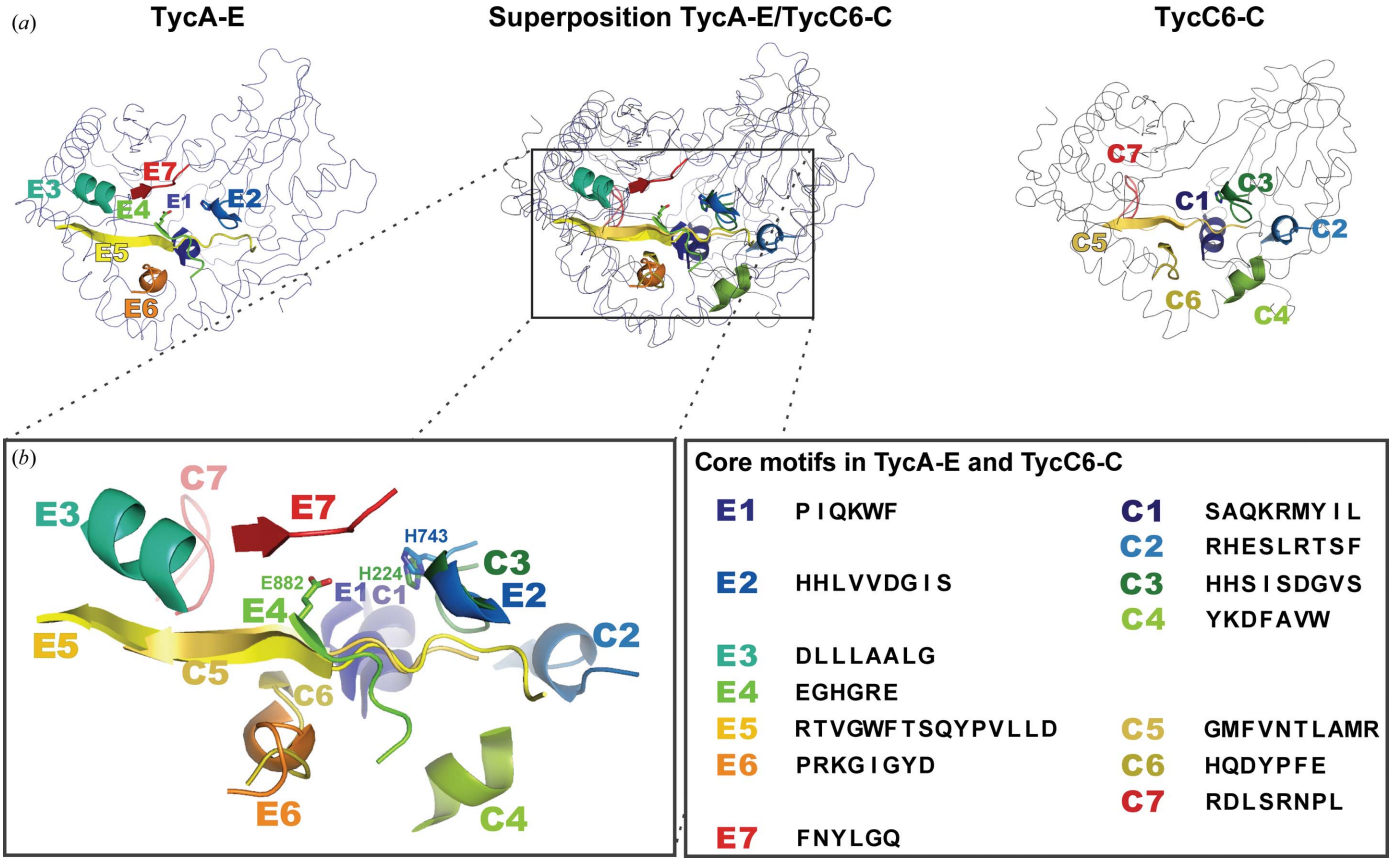


Figure 3
Core motifs of the TycA epimerization domain compared with those of the TycC6 condensation domain. (*a*) Most of the core motifs are located close to the active site, which is in the centre of the domain. For better orientation the side chains of two conserved residues are shown: His743 (light blue) and Glu882 (light green). These active-site residues are themselves parts of the core motifs E2 and E4, respectively. (*b*) Detailed view of the superposition of core motifs as defined in the E- and C-domain structures. Core-motif sequences that are horizontally aligned in the table on the right have corresponding structures and positions, respectively.

Extensive sequence analyses of NRPS condensation and epimerization domains identified several conserved core motifs, which fostered the discrimination of the different catalytic units within NRPSs (Marahiel *et al.*, 1997; Rausch *et al.*, 2007). Despite their common evolutionary heritage and structural similarity, condensation and epimerization domains share only one of the core motifs as defined by Marahiel and coworkers (Konz & Marahiel, 1999; Marahiel *et al.*, 1997): the so-called His-motif (E2/C3). Although the sequences of these two core motifs differ, their corresponding conformations are almost identical (Fig. 3). The E2/C3 motifs contain the histidine residue (His743 in TycA-E) that is crucial for catalytic activity in NRPS. The only exception, the free-standing condensation domain VibH, utilizes a soluble substrate and still shows an unusual ∼10% residual catalytic activity for its H→A mutant compared with the wild-type enzyme. Despite significantly lower sequence similarity, there are three additional pairs of core motifs (E1/C1, E5/C5 and E6/C6) which

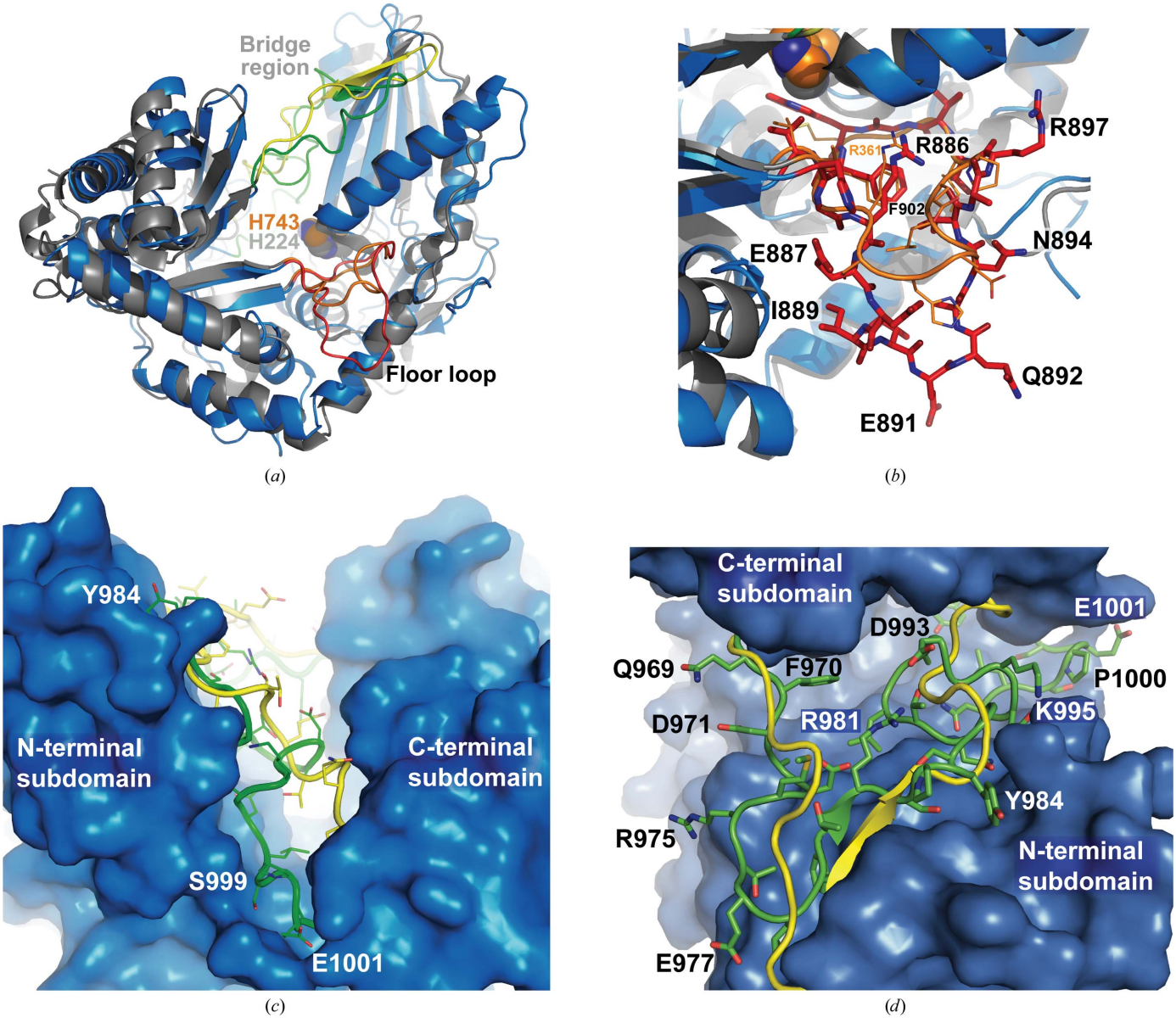


Figure 4
Structural comparison of E and C domain specific floor loops and bridge regions. (*a*) Superposition of the epimerization domain of TycA (blue/red/green) and the condensation domain of TycC6 (grey/orange/yellow) as viewed from the donor side of the domains. Besides their structural similarity, the most prominent differences between the two domains are located in the floor loop and the bridge region. (*b*) Detailed view of the floor-loop structures. In the epimerization domain (red) this loop is extended and reaches latch-like further onto the 'donor side' of the domain. (*c*) View of the superimposed 'bridge regions' from the acceptor side. In the epimerization domain the peptide chain of the posterior bridge region (residues Asp993–Glu1001) fills the canyon by running in between the two CAT-like subdomains. Consequently, the channel that is observed in the condensation domain is blocked at its acceptor side in the epimerization domain. (*d*) Top view of the superimposed bridge regions. Compared with the bridge region of the C domain (yellow), the bridge region of the TycA E domain (green) is more twisted and thereby closes the entrance on the acceptor side.

correspond structurally to each other and are hence discussed below in terms of their potential contribution to substrate recognition and catalysis.

3.2. E-domain specific structural features

Besides the overall similarity between E and C domains, there are two regions in which profound structural differences are observed. The first is a loop formerly named the 'floor loop' (Samel *et al.*, 2007) and the second is the 'bridge region' (Fig. 4*a*).

Compared with the TycC6 condensation domain, the floor loop is extended by five residues in the epimerization domain (Fig. 4*b*). It protrudes latch-like onto the donor side of the domain and may hence interact with the incoming PCP domain. The enlarged floor loop is embedded between the core motifs E4 and E5. At least one residue of the core motif E4, Glu882, is crucial for the catalytic mechanism (see below) as this conserved glutamate is located within the active site and points towards the presumed C^{α} position of the bound substrate. The residues of the E4 motif following Glu882 can be considered as the anterior region of the floor loop, while residues of the succeeding core motif, E5, form the posterior region of the floor loop by continuing as strand β9. As the upstream PCP domain approaches the epimerization domain from this side, the latch-like floor loop may play an important role in the specific PCP-E domain recognition.

The second region of the epimerization domain showing remarkable structural differences compared with condensation domains is the bridge region. This part of the polypeptide chain corresponds to the acceptor site of the C domain and is likewise extended by the insertion of 11 residues following strand β11. The bridge region adopts a twisted conformation in the cleft formed by the two CAT-like subdomains (Fig. 4*d*) and thereby covers not only the active site on top of the epimerization domain but also blocks its acceptor side (Fig. 4*c*). Consequently, the active site is only accessible from the side that corresponds to the donor side of condensation domains (Samel *et al.*, 2007; Tanovic *et al.*, 2008). We therefore propose that the PCP domain preceding the epimerization domain will interact analogously with the epimerization domain as it would do with the donor side of its downstream C domain. This enhanced versatility of PCP domains in terms of interacting not only with upstream and downstream C domains as well as A domains, but also with E domains, underpins the central role the PCP domains play as a shuttle for intermediates during NRP synthesis (Samel *et al.*, 2007; Tanovic *et al.*, 2008). Interestingly, a similar blockage of the acceptor-like region as in the TycA E domain has been reported previously for PapA5, an acyltransferase associated with polyketide synthesis (Buglino *et al.*, 2004). In PapA5 the channel for substrate access is formed from the same region that corresponds to the donor side of C domains such as VibH and TycC6. At its other end this canyon-like channel is blocked like the TycA E domain by a short α-helical segment. Buglino and coworkers suggested for PapA5 that in the course of its catalysed reaction the channel may be closed owing to conformational rearrangement of the enzyme and that the resulting cavity may accommodate one of the substrates. The increased atomic displacement parameters observed for part of the bridge region of TycA-E facing the portal entry (in C domains, the donor side) around the active-site channel (Gln969–Glu977) may likewise allow some conformational adaptation upon docking of a substrate-loaded PCP domain.

3.3. The active site of the TycA epimerization domain

Previous mutagenesis studies on the epimerization domain of GrsA (Stachelhaus & Walsh, 2000) identified residues which are crucial for catalytic function. Accordingly, catalytic mechanisms were predicted in which either one or two residues participate in general base/general acid catalysis during the L↔D isomerization of bound substrate. One of them, the strictly conserved residue His743, points towards the substrate-binding site in the TycA-E structure. This residue is part of core motif E2 and is involved in catalysis, as the histidine residue corresponding to His743 in GrsA-E was indispensable for catalysis. Furthermore, His743 corresponds to the active-site histidine residue in condensation domains (His147 in VibH and His224 in TycC6, respectively), which are part of the core motif of the condensation domain C3 (H**H***xxx*DSG; the catalytic histidine is shown in bold). Interestingly, in the epimerization domain a likewise strictly conserved glutamate residue, Glu882, faces His743 from the opposite side of the active site. This glutamate residue is part of core motif E4 (consensus sequence **E**GHGRE; Glu882 in bold) and lacks a counterpart in NRPS condensation domains. Accordingly, a model for NRPS E domains can be suggested in which the active-site residue Glu882 acts as a catalytic general acid–base. The cationic side chain of protonated His743, the main-chain amide group of Gly747 and the dipole momentum of helix α4 may then stabilize the transiently occurring enolate intermediate and therefore favour its formation.

3.4. The E domain within the context of the TycA synthetase

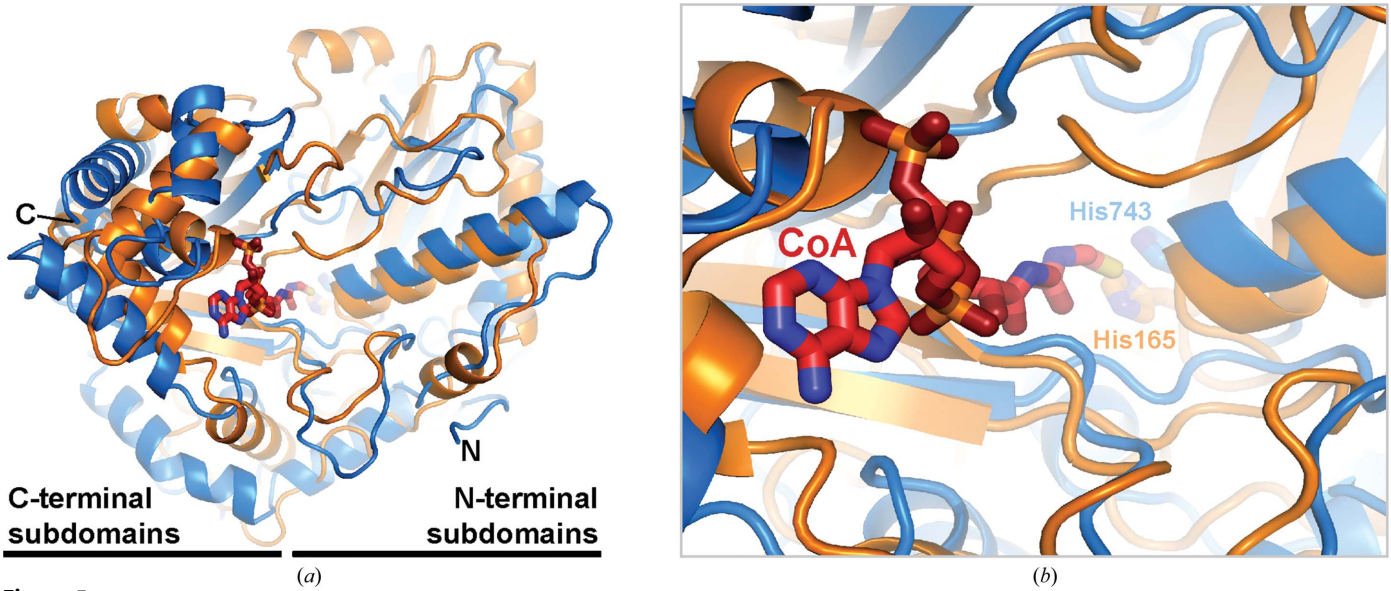
Owing to the absence of further structural data on complexes of the TycA E domain, we had to rely on structure-based modelling to predict the mode of substrate binding and interaction with the preceding PCP domain of the TycA CATE module. Structural comparison of the TycA E domain with the PDB showed an intimate relationship to other catalytic domains harbouring paired CAT-like subdomains (Table 2). One of these related structures, the ternary complex of the trichothecene 3-*O*-acetyltransferase TRI101 with coenzyme A and the inhibitor T-2 (PDB entry 2zba; Garvey *et al.*, 2008), represents a valid model of how the 4′-phosphopantetheine cofactor invades the active-site cleft of NRPS/PKS-like condensation/AT domains. Its interaction appears to be a conserved feature among this class of enzymes, because the structurally related hydroxycinnamoyl transferase from *Sorghum bicolor* (Walker *et al.*, 2013) shows an almost identical binding mode for CoA despite its low pairwise sequence identity to TRI101 (18%). The large overall r.m.s.d. of 4.26 Å (for 317 C^{α} atoms) between the epimerization domain of TycA

and TRI101 is caused by a 27° swivelling motion of the two CAT-like domains relative to each other. Interestingly, swivelling motions indicative of some structural plasticity within this domain type have been found in different NRPS condensation domains with offsets of up to 25° (Bloudoff *et al.*, 2013). Accordingly, the individual N-terminal and C-terminal CAT-like subdomains of TycA-E and TRI101 superimpose much better, with only 2.11 Å (for 70 C^{α} positions) and 2.40 Å (for 68 C^{α} positions) deviation, respectively. The superposition of the two structures shows that the 4′-phosphopantetheinyl moiety of CoA can be placed between the subdomains in an elongated fashion, which is sterically feasible with the active-site portal of the epimerization domain (Figs. 5*a* and 5*b*). The thus modelled orientation of the phosphopantetheine cofactor positions the thioester bond and consequently the C^{α} atom of a 4′-Ppan-bound phenylalanine substrate in between the catalytic residues His743 and Glu882.

The TycA PCP domain (Gln523–Arg603) was modelled by the *MODELLER* 9 suite (Eswar *et al.*, 2008) using the structure of the TycC5 PCP domain in its A/H conformation (PDB entry 2vsq; Tanovic *et al.*, 2008) as a template. Subsequent docking of the TycA PCP domain onto the TycA E domain (Lys604–Cys1053) allowed only one orientation to fulfil the constraints by (i) forming a continuous polypeptide chain and (ii) suitably placing the E-domain docked 4′-Ppan cofactor to be anchored at Ser563 of the PCP domain (Fig. 6*a*, Supplementary Fig. 1). The resulting PCP-E bidomain model (see Supporting Information) shows a domain interface with an area of 837 $Å^2$, where approximately two thirds of the interactions are mediated *via* helices αIII and αIV and approximately one third *via* helix αII and the αI–αII loop of the PCP domain. Interestingly, the αIII helix is intimately packed against the latch-like floor loop of the epimerization domain, suggesting that the E-domain specific extension of the floor loop acts as a recognition zone for PCP-domain interaction. A second specific docking site may be provided by the αI–αII loop harbouring Ser563, the 4′-Ppan attachment site of the TycA PCP domain (Supplementary Fig. S1*a*). As in other PCP domains, which precede epimerization domains in NRPS, the TycA PCP domain carries a conserved sequence motif, GGD**S**I, instead of the GGH**S**L motif found in condensation domains (Linne *et al.*, 2001). According to the PCP-E model the aspartate of the GGDSI motif, Asp562, is capable of forming salt bridges to Lys296 and Arg930 of helix α9 from the E domain.

3.5. Implications for the mechanism of the NRPS epimerization reaction

Our current understanding of the epimerization catalysed by NRPS E domains describes a sequence of proton-abstraction and proton-donation reactions. Assuming that a catalytic dyad consisting of a glutamate and a histidine residue catalyses the reaction, this situation is reminiscent of that observed in triosephosphate isomerase (TIM) from chicken muscle (*Gallus gallus*) and yeast (*Saccharomyces cerevisiae*). This enzyme catalyses the interconversion of dihydroxyacetone phosphate and *R*-glyceraldehyde 3-phosphate and has been intensively studied by Knowles and coworkers (Knowles, 1991). Here, the interaction between His95 and Glu165 in TIM apparently corresponds to that provided by His743 and Glu882 in the TycA E domain (Fig. 6*b*). TIM utilizes the



(*a*) (*b*)

Figure 5
Structural comparison of the epimerization domain of TycA and trichothecene 3-*O*-acetyltransferase (TRI101) in complex with CoA (PDB entry 2zba). (*a*) Total view of the superposed crystal structures from the donor side (E, blue; TRI101, orange). The molecule of CoA observed in the complex structure of TRI101 is shown as mainly red sticks. As shown by subdomain-wise structural superposition between TycA-E and TRI101, the central region including β-sheets 1–6–5–4 and 8–9–10–12–13 as well as the catalytic helix α4 is well preserved. (*b*) Detailed view of the presumed site of substrate entry to the active site. The active-site residues of the epimerization domain and TRI101, His743 and His165, respectively, are shown. Despite the slight differences in the structures, which are commented on in the text, the general match with respect to the arrangement of the active sites positioned the CoA molecule onto the donor side of the epimerization domain.

dipole moments of α-helices to tune the protonation states of bound substrates and, in analogy to the findings for the TycC6 condensation domain (Samel *et al.*, 2007), a similar contribution may be expected for the epimerization domain. In the latter, the dipole moment of helix α4 points away from the active site (Fig. 6*c*) and could hence electrostatically favour the formation of a negative charge in an enolate intermediate. Furthermore, structure-based calculations of the protonation states using *MEAD* or the H^{++} server (Anandakrishnan *et al.*, 2012; Bashford & Gerwert, 1992; Gordon *et al.*, 2005) resulted in calculated pK_a values of 7.8–9.0 for the catalytically important residue His743. Most of the pK upshift of the imidazole side chain, which is also found for the catalytic histidine of the condensation domain of TycC6 (the calculated pK_a of His224 is 9.0; Samel *et al.*, 2007), is caused by its local environment, including a hydrogen bond between its ND1 atom and the carbonyl group of Val746 (Ser227 in TycC6), electrostatic interactions with Asp747 (Asp228 in TycC6) and Glu882 and packing between the electron-rich side chains of Gln636 and Trp901 (Met123 and Met372 in TycC6). In the absence of bound substrate and under physiological conditions the catalytic His743 is hence predicted to be protonated, as is its counterpart in the TycC6 condensation domain, whose catalytic activity in peptide-bond formation was suggested to depend on electrostatic interactions with the sp^3 intermediate (Fig. 6*d*) rather than on general acid–base catalysis (Samel

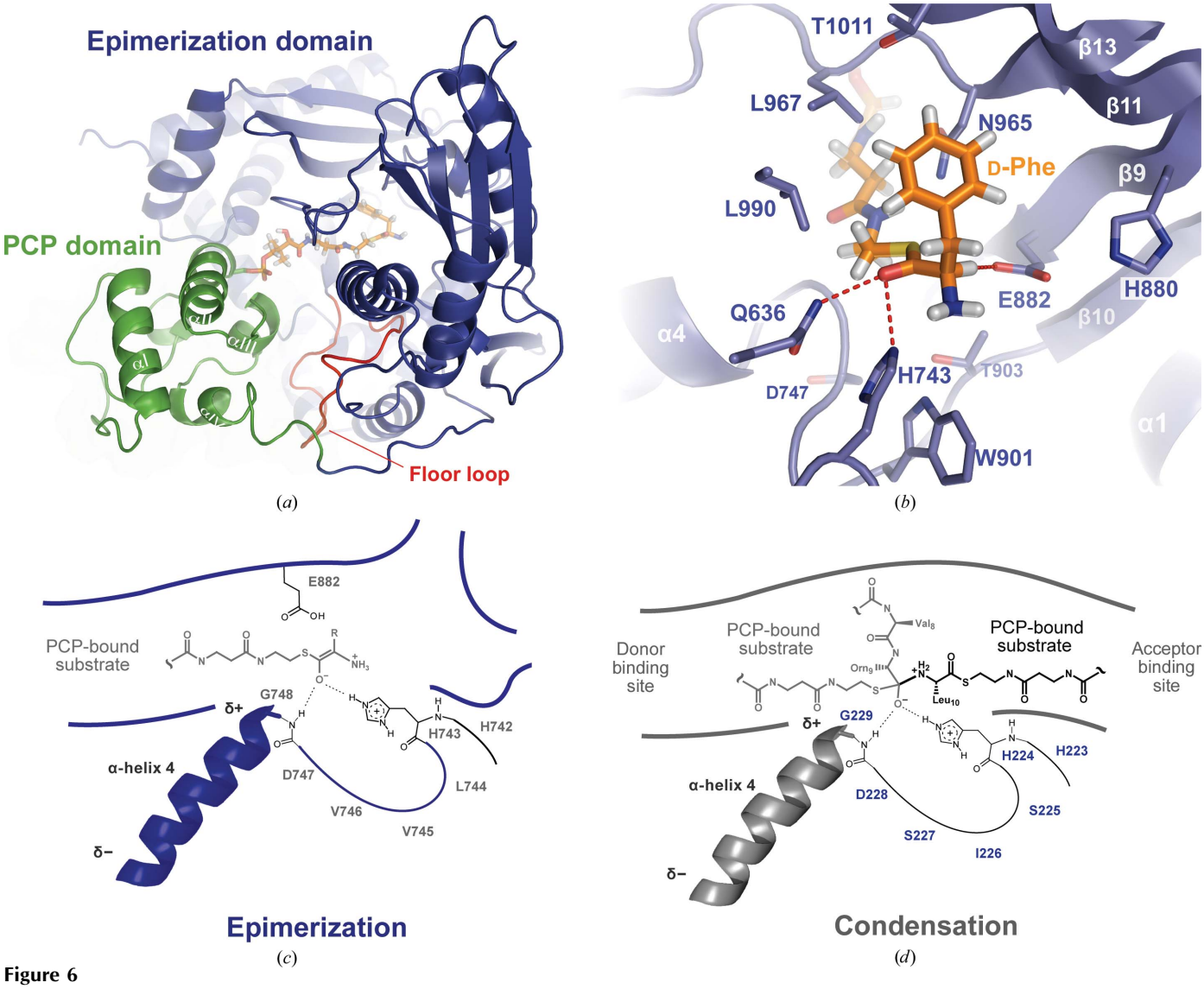


Figure 6
Mechanistic implications for E-domain and C-domain catalysis. (*a*) Overview of the model of a PCP-E bidomain loaded with the cofactor 4′-Ppan and the substrate/product amino acid, D-Phe (PCP, green; E-domain, blue). The floor loop that forms interdomain interactions with the αIII and αIV helices is highlighted in red. The cofactor and the substrate amino acid are shown in orange as a stick model. (*b*) View of the cofactor-tethered product D-Phe bound within the active site. Residues in the vicinity of the D-phenylalanine are shown as sticks. (*c*) Electrostatic stabilization of the enolate intermediate formed during E-domain catalysis. During a D→L conversion Glu882 may act as a catalytic base by abstracting a proton from the C^{α} position of the 4′-Ppan-bound phenylalanine. The conserved histidine residue His743, the main-chain amide of Asp747 and the dipole moment of helix α4 contribute to the electrostatic stabilization of the negative charge formed on the O atom of the enolate intermediate. The C^{α} atom may then reprotonate by proton abstraction from His743. (*d*) Electrostatic stabilization of the sp^3 reaction intermediate during C-domain catalysis. The amino group of the 4′-Ppan-anchored amino acid from the acceptor side has attacked the thioester of the 4′-Ppan-bound peptide intermediate from the donor side.

et al., 2007). Nevertheless, binding of 4′-Ppan tethered L-phenylalanine may still revert the pK_a values of His743 and Glu882 in the active site to allow deprotonation during an L→D conversion by His743 and reprotonation by Gln882. The TycA E domain may hence lower the pK_a of the C^{α} proton to a similar extent as was observed for acyl-CoA dehydrogenases. Spectroscopic studies with this class of enzymes have shown decreases in the pK_a by 10 pK units for the acyl-CoA substrates upon binding to the active site (Vock *et al.*, 1998).

4. Concluding remarks

Given that the previous crystal structure of a PCP-C bidomain (Samel *et al.*, 2007) yielded only a nonproductive mode for PCP-domain association along the donor side of the C domain, the presented model for a productively docked PCP-E bidomain provides a first glimpse of how PCP domains recognize cognate downstream E and C domains. Obviously, the majority of interactions are provided by the floor loop and the αIII helix. The postulated interface between the PCP and E domains resembles in terms of size the well defined interfaces with A domains as previously described for PA1221 (915 Å^2; Mitchell *et al.*, 2012) and the EntE-EntB chimera (978 Å^2; Sundlov & Gulick, 2013; Sundlov *et al.*, 2012). However, PCP-A domain interactions are mainly formed by helix αII. Interestingly, this helix of the PCP domain also plays a prominent role in interactions with the acceptor side of C domains as shown by the smaller interface between the PCP and condensation domains of SrfA-C (486 Å^2; Tanovic *et al.*, 2008). Concerning the donor side, extensive mutagenesis studies by Walsh and coworkers (Lai *et al.*, 2006) corroborated that helix αIII is intimately involved in interactions with the downstream C domain. This indicates that PCP domains may generally depend on helix αIII to recognize the C and E domains at the floor loops of their donor side portals. Given the structural plasticity observed for PCP domains upon substrate loading (Koglin *et al.*, 2006), the unidirectionality of NRP synthesis may be further fostered by reaction cycle-dependent conformational changes which control the relative accessibility of the αII and αIII helices of PCP domains.

5. Related literature

The following references are cited in the Supporting Information: Bashford (1997), Bashford & Gerwert (1992), Case *et al.* (2010), Czodrowski *et al.* (2006), Deerfield & Pedersen (1995), Dupradeau *et al.* (2010), Emsley & Cowtan (2004), Eswar *et al.* (2008), Garvey *et al.* (2008), Gordon *et al.* (2005), Humphrey *et al.* (1996), Krissinel & Henrick (2007), Linne *et al.* (2001) and Stein *et al.* (2006).

This work was supported by grants from the Deutsche Forschungsgemeinschaft (ES152/4), the DFG-Graduiertenkolleg 'Protein function at the atomic level' (SAS) and the LOEWE Research Centre of Synthetic Microbiology. The authors thank the support staff at synchrotron beamline PSF14-2, BESSY-II (Berlin, Germany) and Christoph Müller-Dieckmann at beamline ID29, ESRF (Grenoble, France), Uwe Linne for mass-spectrometric analysis of SeMet-labelled proteins, Maik Veelders, Wolfgang Grosse and Xenia Grosse for discussion and technical support, Mohamed Marahiel for the plasmid pQE60-TycA and astonishing insights into science, and Andreas Heine for MAD data collection and support during initial structure solution. We thank the BMBF (support code 05ES3XBA/S) for travel grants to BESSY II (Berlin).

References

Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. (2012). *Nucleic Acids Res.* **40**, W537–W541.
Balibar, C. J., Vaillancourt, F. H. & Walsh, C. T. (2005). *Chem. Biol.* **12**, 1189–1200.
Bashford, D. (1997). *Scientific Computing in Object-Oriented Parallel Environments*, edited by Y. Ishikawa, R. R. Oldehoeft, J. V. W. Reynders & M. Tholburn, pp. 233–240. Berlin, Heidelberg: Springer. doi:10.1007/3-540-63827-X_66.
Bashford, D. & Gerwert, K. (1992). *J. Mol. Biol.* **224**, 473–486.
Bloudoff, K., Rodionov, D. & Schmeing, T. M. (2013). *J. Mol. Biol.* **425**, 3137–3150.
Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T. & Stubbs, M. T. (2002). *Structure*, **10**, 301–310.
Buglino, J., Onwueme, K. C., Ferreras, J. A., Quadri, L. E. & Lima, C. D. (2004). *J. Biol. Chem.* **279**, 30634–30642.
Case, D. C. *et al.* (2010). *AMBER* 11. University of California, San Francisco, California, USA
Czodrowski, P., Dramburg, I., Sotriffer, C. A. & Klebe, G. (2006). *Proteins*, **65**, 424–437.
Deerfield, D. W. II & Pedersen, L. G. (1995). *J. Mol. Struct.* **358**, 99–106.
Dupradeau, F. Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W. & Cieplak, P. (2010). *Phys. Chem. Chem. Phys.* **12**, 7821–7839.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Eswar, N., Eramian, D., Webb, B., Shen, M.-Y. & Sali, A. (2008). *Methods Mol. Biol.* **426**, 145–159.
Garvey, G. S., McCormick, S. P. & Rayment, I. (2008). *J. Biol. Chem.* **283**, 1660–1669.
Goodin, S., Kane, M. P. & Rubin, E. H. (2004). *J. Clin. Oncol.* **22**, 2015–2025.
Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. & Onufriev, A. (2005). *Nucleic Acids Res.* **33**, W368–W371.
Hahn, M. & Stachelhaus, T. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 15585–15590.
Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H. & Zocher, R. (1994). *J. Biol. Chem.* **296**, 12710–12714.
Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. (2002). *Nature Struct. Biol.* **9**, 522–526.
Knowles, J. R. (1991). *Nature (London)*, **350**, 121–124.
Koglin, A., Mofid, M. R., Löhr, F., Schäfer, B., Rogov, V. V., Blum, M. M., Mittag, T., Marahiel, M. A., Bernhard, F. & Dötsch, V. (2006). *Science*, **312**, 273–276.
Koglin, A. & Walsh, C. T. (2009). *Nat. Prod. Rep.* **26**, 987–1000.
Kohli, R. M., Takagi, J. & Walsh, C. T. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1247–1252.
Konz, D. & Marahiel, M. A. (1999). *Chem. Biol.* **6**, R39–R48.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lai, J. R., Fischbach, M. A., Liu, D. R. & Walsh, C. T. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 5314–5319.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Lewendon, A., Murray, I. A., Shaw, W. V., Gibbs, M. R. & Leslie, A. G. W. (1994). *Biochemistry*, **33**, 1944–1950.
Linne, U., Doekel, S. & Marahiel, M. A. (2001). *Biochemistry*, **40**, 15824–15834.
Marahiel, M. A. & Essen, L.-O. (2009). *Methods Enzymol.* **458**, 337–351.
Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. (1997). *Chem. Rev.* **97**, 2651–2674.
Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. (2012). *Biochemistry*, **51**, 3252–3263.
Nishina, Y., Sato, K., Tamaoki, H., Tanaka, T., Setoyama, C., Miura, R. & Shiga, K. (2003). *J. Biochem.* **134**, 835–842.
Nolan, E. M. & Walsh, C. T. (2009). *Chembiochem*, **10**, 34–53.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). *Acta Cryst.* D**57**, 1445–1450.
Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. (2007). *BMC Evol. Biol.* **7**, 78.
Roongsawang, N., Lim, S. P., Washio, K., Takano, K., Kanaya, S. & Morikawa, M. (2005). *FEMS Microbiol. Lett.* **252**, 143–151.
Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. (2007). *Structure*, **15**, 781–792.
Samel, S. A., Wagner, B., Marahiel, M. A. & Essen, L.-O. (2006). *J. Mol. Biol.* **359**, 876–889.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Stachelhaus, T. & Walsh, C. T. (2000). *Biochemistry*, **39**, 5775–5787.
Stein, D. B., Linne, U., Hahn, M. & Marahiel, M. A. (2006). *Chembiochem*, **7**, 1807–1814.
Stein, D. B., Linne, U. & Marahiel, M. A. (2005). *FEBS J.* **272**, 4506–4520.
Strieker, M., Tanović, A. & Marahiel, M. A. (2010). *Curr. Opin. Struct. Biol.* **20**, 234–240.
Sundlov, J. A. & Gulick, A. M. (2013). *Acta Cryst.* D**69**, 1482–1492.
Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. (2012). *Chem. Biol.* **19**, 188–198.
Survase, S. A., Kagliwal, L. D., Annapure, U. S. & Singhal, R. S. (2011). *Biotechnol. Adv.* **29**, 418–435.
Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. (2008). *Science*, **321**, 659–663.
Thorpe, C. (1989). *Trends Biochem. Sci.* **14**, 148–151.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Vock, P., Engst, S., Eder, M. & Ghisla, S. (1998). *Biochemistry*, **37**, 1848–1860.
Walker, A. M., Hayes, R. P., Youn, B., Vermerris, W., Sattler, S. E. & Kang, C. (2013). *Plant Physiol.* **162**, 640–651.
Walsh, C. T. (2004). *Science*, **303**, 1805–1810.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.