Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of the catalytic EAL domain of the *Escherichia coli* direct oxygen sensor

Miroslaw Tarnawski, Thomas R. M. Barends, Elisabeth Hartmann and Ilme Schlichting*

Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, Heidelberg, Germany

Correspondence e-mail: ilme.schlichting@mpimf-heidelberg.mpg.de

Received 12 November 2012
Accepted 14 February 2013

PDB References: DosP, 4hu3; 4hu4

The direct oxygen sensor DosP is a multidomain protein that contains a gas-sensing haem domain and an EAL effector domain. EAL domains are omnipresent signal transduction domains in bacteria. Many EAL domains are active phosphodiesterases and are involved in breakdown of the ubiquitous bacterial second messenger cyclic di-GMP. Despite a great deal of information on the functional and structural aspects of active and inactive EAL domains, little is known about the structural basis of their regulation by their associated sensory domains. Here, two crystal structures of the *Escherichia coli* DosP EAL domain derived from cubic and monoclinic crystal forms that were obtained under tartrate and PEG conditions, respectively, are described. Both of the structures display the typical TIM (triosephosphate isomerase) barrel fold with one antiparallel β-strand. However, unlike other EAL structures, access to the active site in DosP EAL is sterically restricted by the presence of a short helical stretch (Ser637-Ala-Leu-His640) in loop L3 between strand β3 and helix α3. This element, together with an unordered fragment, replaces the short α-helix (named α5 in Tbd1265 EAL) that is found in other EAL-domain structures. Since DosP EAL is an active c-di-GMP phosphodiesterase, the observed inactive conformation is suggested to be of functional relevance for the regulation mechanism of DosP.

1. Introduction

The cyclic di-GMP dinucleotide (c-di-GMP) was discovered as a regulator of cellulose synthesis in *Gluconacetobacter xylinus* (formerly *Acetobacter xylinum*; Ross *et al.*, 1987) more than twenty years ago. Recently, c-di-GMP has emerged as a major bacterial second messenger that controls a wide range of cellular functions at the transcriptional, translational and post-translational levels (Römling, 2011), mediating a variety of functions including developmental transitions, aggregative behaviour, adhesion, biofilm formation and pathogen virulence. In particular, c-di-GMP has been proposed to regulate microbial community behaviour and the transition between motile and sessile lifestyles (Römling & Amikam, 2006).

The intracellular level of c-di-GMP is controlled by both its biosynthesis and degradation. Diguanylate cyclases (DGCs) carrying GGDEF domains catalyze the condensation of two GTP molecules to produce c-di-GMP. Specific phosphodiesterases (PDEs) containing alternative and structurally unrelated EAL or HD-GYP domains catalyze the hydrolysis of c-di-GMP to generate the linear pGpG dinucleotide, which

is subsequently hydrolyzed into two GMP molecules (Krasteva *et al.*, 2012). All of these domains are named after the conserved amino-acid motifs in their active sites. The majority of these domains are found in combination with additional sensory or regulatory domains (*i.e.* PAS, BLUF and GAF; Schirmer & Jenal, 2009).

A large number of GGDEF- and EAL-domain-containing proteins have been identified in bacterial genomes by whole genome-sequencing projects (Galperin *et al.*, 2001). In fact, the EAL domain is one of the most ubiquitous bacterial signal transduction domains and EAL-containing proteins are found in all branches of the bacterial phylogenetic tree, which implies a significant role in bacterial cell physiology and homeostasis. EAL and GGDEF proteins can be grouped into GGDEF proteins, EAL proteins and GGDEF-EAL proteins, and these classes are usually present within the same genome. It is still unclear how the opposing activities of GGDEF-EAL tandem proteins are coordinated. In some cases, the sequence signature motifs are degenerated, rendering one of the domains catalytically inactive. Therefore, these domains are believed to function as allosteric regulators of neighbouring domains (Christen *et al.*, 2005). The *Escherichia coli* genome codes for 12, ten and seven proteins containing a GGDEF domain only, an EAL domain only and GGDEF-EAL tandem domains, respectively (Galperin *et al.*, 2010). One of the widely studied proteins involved in c-di-GMP signalling is the *E. coli* direct oxygen sensor (DosP; also referred to as EcDos), which belongs to the family of haem-based sensors (Delgado-Nixon *et al.*, 2000). DosP is composed of two PAS (Per–ARNT–Sim) domains arranged in tandem at the N-terminal part of the protein followed by a C-terminal GGDEF-EAL tandem. The first PAS domain (PAS1) of DosP is a haem-containing sensory domain that controls the EAL phosphodiesterase activity towards c-di-GMP in an oxygen-dependent fashion, while the adjacent GGDEF domain is inactive. It has been demonstrated that the binding of O_2, CO or NO to the haem-containing PAS1 domain enhanced the phosphodiesterase activity by sixfold to sevenfold (Tanaka *et al.*, 2007). Crystal structures of the sensory DosP PAS1 domain have been determined in reduced and oxygenated forms (Kurokawa *et al.*, 2004; Park *et al.*, 2004). It has been shown that the binding of oxygen (and presumably also other diatomic ligands) to the distal site of the haem cofactor induces significant structural changes by displacing the Met87 side chain (also referred to as Met95) from the haem iron. This triggers rearrangement in the F-helix/G-strand portion of the PAS domain. Despite this insight into the structure and function of the DosP sensory domain, it is not known how the signal is transmitted to the active site of the catalytic domain. Kurokawa and coworkers hypothesized that a scissor-type hinge movement of the PAS-domain dimer may be amplified and may change the subunit–subunit interactions of the EAL domains, thereby affecting their PDE activity.

Here, we report crystal structures of the catalytic EAL domain of DosP in its apo form in both monomeric and dimeric states. These structures constitute a step towards our understanding of the function and mechanism of DosP.

2. Materials and methods

2.1. Protein expression and purification

The DNA fragment encoding the DosP EAL domain (residues 529–799) was PCR-amplified and cloned into pET-28b vector using *Nde*I and *Xho*I restriction sites. The resulting plasmid was transformed into *E. coli* BL21 (DE3) cells and used for overexpression of the EAL domain as an N-terminal His_6-tag fusion.

The transformed cells were grown aerobically in TB medium supplemented with kanamycin (50 µg ml^{-1}) at 310 K until the optical density at 600 nm (OD_{600}) reached 0.4. Protein expression was induced with 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16–18 h at 293 K. The bacterial cells were harvested by centrifugation, resuspended in buffer *A* (50 m*M* NaH_2PO_4 pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole, 10% glycerol) and lysed using a cell disruptor (Microfluidics) operated at a pressure of 0.7 MPa. The crude lysate was centrifuged to remove cell debris and insoluble fractions; the resulting supernatant was then applied onto an Ni–NTA column (Qiagen). After washing off the unbound fraction, the DosP EAL protein was eluted using buffer *A* supplemented with 500 m*M* imidazole. Fractions containing recombinant protein were pooled and the protein was precipitated with 3.0 *M* ammonium sulfate, pelleted by centrifugation and dissolved in buffer *B* (25 m*M* Tris–HCl pH 7.5, 50 m*M* KCl, 5 m*M* $MgCl_2$, 2.5 m*M* EDTA, 5 m*M* DTE, 10% glycerol). The final purification step was gel filtration on a Superdex 75 10/300 GL column (GE Healthcare) in buffer *B*. The eluted protein was concentrated using a centrifugal ultrafiltration device (Vivaspin). The final EAL-domain protein corresponds to residues 529–799 of full-length DosP, with the following N-terminal vector-derived sequence: MGSSHHHHHHSSGLVPRGSHM. For the expression of a selenomethionine derivative, minimal medium supplemented with seleno-L-methionine (SeMet; 50 mg l^{-1}) and appropriate antibiotics was used. The protein was purified as described for the native form.

2.2. Crystallization, data collection and processing

Native and SeMet-labelled proteins were crystallized at 293 K using the vapour-diffusion method by mixing equal volumes of protein (43 mg ml^{-1} for the native protein and 10 mg ml^{-1} for the SeMet-labelled protein) and precipitant (20% PEG 6000, 0.2 *M* calcium chloride, 0.1 *M* HEPES pH 7.0 for the native protein and 0.5 *M* potassium sodium tartrate, 0.1 *M* MES pH 6.3 for the SeMet-labelled protein) solutions. Thin needle-like (native) and cubic (SeMet-labelled) crystals appeared within 3 d. The crystals were briefly washed in cryoprotectant solution consisting of the reservoir solution supplemented with 20% glycerol (native) or 20% ethylene glycol (SeMet-labelled) before flash-cooling in liquid nitrogen. X-ray diffraction data were collected from crystals maintained at 100 K during data collection on the X10SA beamline at the SLS synchrotron (Villigen, Switzerland). All data were processed with *XDS* (Kabsch, 2010). Full data-collection statistics are given in Table 1.

Table 1
Data-collection, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell.

	SeMet (peak)	SeMet (remote)	SeMet (monomeric form)	Native (dimeric form)
Data collection				
Space group	*I*23	*I*23	*I*23	$P12_11$
Unit-cell parameters				
a (Å)	140.3	140.3	139.8	35.6
b (Å)	140.3	140.3	139.8	171.2
c (Å)	140.3	140.3	139.8	48.8
$\alpha = \gamma$ (°)	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	90.00	90.03
Radiation source	PXII-X10SA, SLS	PXII-X10SA, SLS	PXII-X10SA, SLS	PXII-X10SA, SLS
Wavelength (Å)	0.97896	0.97139	0.97139	0.97936
Temperature (K)	100	100	100	100
Resolution range (Å)	40–3.50 (3.60–3.50)	40–3.70 (3.80–3.70)	40–3.30 (3.40–3.30)	20–2.20 (2.30–2.20)
No. of observed reflections	264243 (21518)	218945 (16921)	41380 (3618)	85673 (5951)
No. of unique reflections	11276 (918)	9515 (734)	13227 (1139)	28963 (3455)
Multiplicity	23.4 (23.4)	23.0 (23.1)	3.1 (3.2)	2.9 (1.6)
Completeness (%)	100.0 (100.0)	99.9 (100.0)	99.4 (100.0)	99.1 (94.7)
$R_{\rm merge}$ (%)	9.6 (47.2)	8.6 (33.1)	5.4 (34.8)	7.8 (53.7)
$\langle I/\sigma(I)\rangle$	29.33 (8.67)	31.28 (11.76)	13.85 (4.00)	9.55 (1.77)
Phasing				
Figure of merit (acentric; before/after DM)	0.52/0.84			
Isomorphous phasing power (acentric/centric)		0.85/0.86		
Anomalous phasing power (acentric)	2.4	3.2		
Isomorphous $R_{\rm cullis}$ (acentric/centric)		0.83/0.83		
Anomalous $R_{\rm cullis}$ (acentric)	0.48	0.52		
Refinement				
No. of reflections in working set			6969	22645
No. of reflections in test set			332	1224
Resolution range (Å)			37.37–3.30	20.0–2.40
No. of non-H atoms				
Total			2125	4040
Protein			2125	3924
Solvent			0	116
R (%)			26.39	23.28
$R_{\rm free}$ (%)			32.31	28.01
R.m.s.d. from ideal				
Bonds (Å)			0.002	0.003
Angles (°)			0.660	0.742
Mean *B* factor (Å^2)				
Overall			90.37	35.83
Protein			90.37	35.94
Solvent			0	31.93
Ramachandran statistics (%)				
Favoured regions			92.5	92.0
Allowed regions			7.1	7.0
Disallowed regions			0.4	1.0

2.3. Structure determination and refinement

The DosP EAL structure was determined by multiple-wavelength anomalous dispersion (MAD) phasing using data collected from an SeMet derivative. Phases were calculated from a two-wavelength (peak and remote) data set using *autoSHARP* (Vonrhein *et al.*, 2007); the phasing statistics are given in Table 1. Density modification and initial model building were performed using *RESOLVE* (Terwilliger, 2003). The crystals of the native protein were pseudomerohedrally twinned with twin-law operator $(h, -k, -l)$ and a twin fraction of 0.416 as estimated from Britton analyses by *phenix.xtriage* (Adams *et al.*, 2010). Phases for the native twinned data set were determined by molecular replacement using the program *Phaser* (McCoy *et al.*, 2007) and the coordinates of the SeMet-derivative protein model. The models of the native and the SeMet-labelled EAL domain were refined by alternating rounds of manual model rebuilding using the graphics program *Coot* (Emsley & Cowtan, 2004) and refinement as implemented in *REFMAC*5 (Murshudov *et al.*, 2011) and *phenix.refine* (Adams *et al.*, 2010). The refinement statistics for both structures are given in Table 1. The structural figures were generated with *PyMOL* (v.1.2r2; http://www.pymol.org) and the superpositions were calculated with *SSM* (Krissinel & Henrick, 2004). Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4hu3 and 4hu4.

2.4. Phosphodiesterase activity assay

Activity measurements were performed essentially as described previously (Barends *et al.*, 2009). The reaction mixture consisted of 1.1 µ*M* protein and 100 µ*M* c-di-GMP.

3. Results and discussion

3.1. Structure overview

The structure of the apo form of the DosP EAL domain was determined from selenium multiple-wavelength anomalous dispersion data to 3.3 Å resolution. The asymmetric unit of the cubic crystal form contains one monomer. The model of the DosP EAL structure consists of residues 531–799; the N-terminal fusion tag together with the two N-terminal residues (529–530) is not visible in the electron-density maps. The quality of the electron density is poorer in the N- and C-terminal parts and also in the region corresponding to loop L5, where it is continuous but lacks some of the main-chain and side-chain details. The temperature factors are distributed such that the highest values are found in the loops linking $\beta4$ and $\alpha4$ (loop L4) and $\beta5$ and $\alpha5$ (loop L5) as well as in helix $\alpha5$ itself.

The DosP EAL domain adopts a $(\beta/\alpha)_8$-barrel fold containing two antiparallel β-strands (Fig. 1*a*). Two extended antiparallel strands $\beta1$ and $\beta2$ are connected by a β-hairpin. Strand $\beta1$ runs antiparallel to the other seven strands, lining the $\alpha\beta$-barrel from the inside. Thus, the specific arrangement of the secondary-structure elements is best described as $\alpha\beta(\beta\alpha)_7$ (Fig. 2). The arrangement of the barrel observed here differs from the canonical TIM-barrel composed of exclusively parallel β-strands, but since it was also found in the structures of other EAL domains it appears to be a characteristic feature. Helix $\alpha1$, together with a linker containing a short helical insertion (αL2) between strand $\beta2$ and helix $\alpha2$, forms a lobe positioned aside the top of the barrel (Fig. 1*b*). The region between strands $\beta6$ and $\beta7$ comprises two helices: $\alpha6$a and $\alpha6$b.

We determined the structure of the native apo form of DosP EAL by molecular replacement to a resolution of 2.4 Å (Fig. 1*c*). The monoclinic crystal form contains two monomers (chains *A* and *B*) in the asymmetric unit. The r.m.s.d. value for their backbone C^{α} atoms is 0.95 Å (over 245 C^{α} positions), with the highest deviations observed in loop L4 between strand $\beta4$ and helix $\alpha4$ and in the linker between strand $\beta2$ and helix $\alpha2$. The electron-density map of native DosP EAL allows the tracing of residues 541–798 of polypeptide chains *A* and *B* of the EAL dimer. Again, the N-terminal fusion tag together with several N-terminal residues (529–540) is not visible in the electron-density maps; the same is true for the very last C-terminal residue (799). Additionally, in the dimeric structure there is a gap in the polypeptide chain owing to a lack of interpretable electron density for residues 701–711. This region corresponds to the compound helix $\alpha5$. It is unclear why this fragment is so highly disordered, given that helix $\alpha5$ is most likely to be involved in the interactions at the dimer interface. The distribution of temperature factors along the polypeptide chain of the dimeric structure shows a comparable situation: the highest *B* factors are observed in structural elements forming the apical lobe as well as in loops L3 linking $\beta3$ and $\alpha3$ and L4 linking $\beta4$ and $\alpha4$.

When comparing the monomeric DosP EAL structure with each chain of the DosP EAL dimer, we obtain r.m.s.d. values of 1.27 Å (over 233 C^{α} positions) and 1.24 Å (over 236 C^{α} positions) for chain *A* and chain *B*, respectively. The differences are mainly observed within

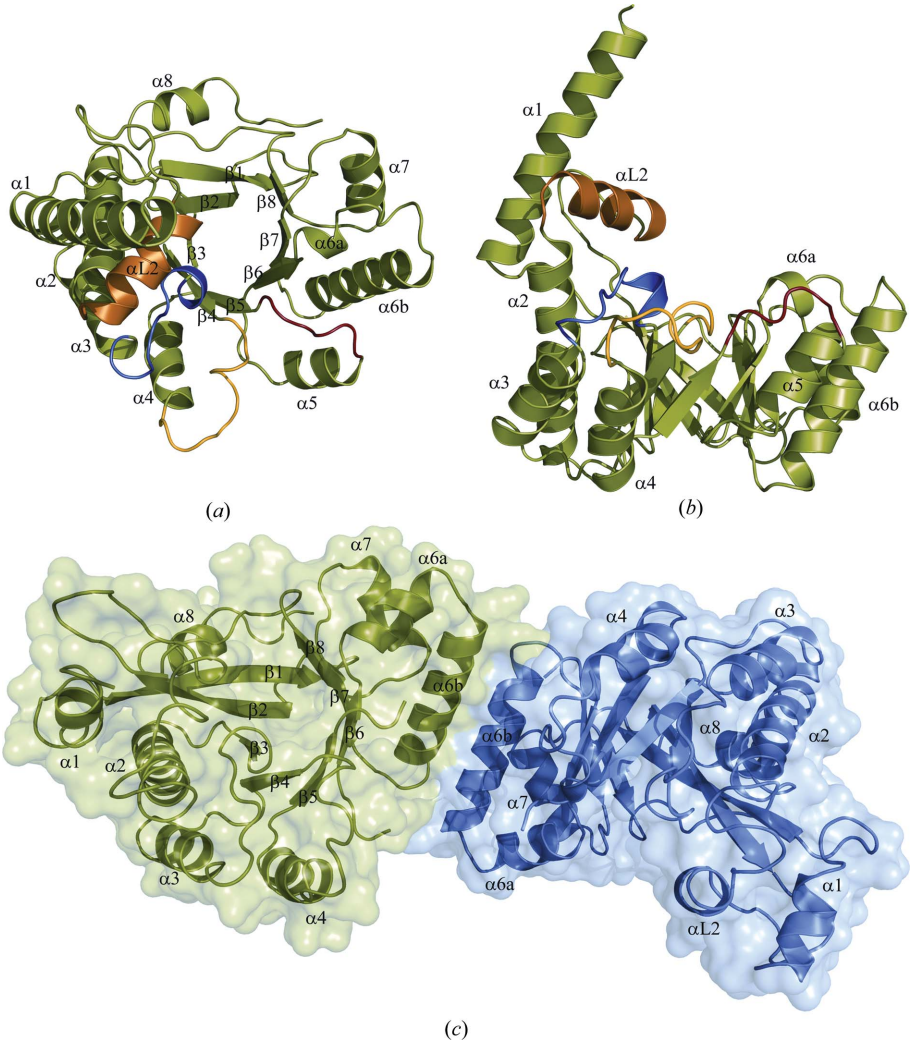

Figure 1
Overall structure of the EAL domain of DosP. The monomeric form is shown as (*a*) a top view along the TIM-barrel axis depicting the arrangement of secondary-structure elements and (*b*) a side view after 90° rotation around the *x* axis, with the lobe on the left side. The loops are coloured as follows: L2 together with αL2, orange; L3, blue; L4, yellow; L5, red. (*c*) The dimeric form composed of chain *A* (shown in green) and chain *B* (shown in blue) together with a surface representation.

Table 2
Properties of EAL domains from different proteins.

Protein name	Organism	PDE activity	Class from Rao *et al.* (2009)	Monomer/dimer	Ligand	PDB code	Reference
DosP	*Escherichia coli*	Yes	1	Yes/yes	None	4hu3, 4hu4	This work
Tbd1265	*Thiobacillus denitrificans*	Yes	1	No/yes	c-di-GMP	2r6o, 3n3t	Tchigvintsev *et al.* (2010)
PdeA (CC3396)	*Caulobacter crescentus*	Yes	1	Yes/yes	pGpG, c-di-GMP	3u2e, 3s83, 4hjf	Midwest Center for Structural Genomics (unpublished work)
RocR	*Pseudomonas aeruginosa*	Yes	1	No/yes	None	3sy8	Chen *et al.* (2012)
BlrP1	*Klebsiella pneumoniae*	Yes	1	No/yes	c-di-GMP	3gfx, 3gfy, 3gfz	Barends *et al.* (2009)
YkuI	*Bacillus subtilis*	No	2	No/yes	c-di-GMP	2bas, 2w27	Minasov *et al.* (2009)
Lmo0111	*Listeria monocytogenes*	n/a	2	No/yes†	None	3kzp	Midwest Center for Structural Genomics (unpublished work)
FimX	*Pseudomonas aeruginosa*	No	3	Yes/yes†	c-di-GMP	3hvb, 3hv8, 3hv9	Navarro *et al.* (2009)
FimX	*Xanthomonas campestris* pv. *campestris*	No	3	Yes/no	c-di-GMP	4f3h	Chin *et al.* (2012)
LapD	*Pseudomonas fluorescens* (Pf0-1)	No	3	Yes/yes	c-di-GMP	3pjt, 3pju	Navarro *et al.* (2011)
LapD	*Pseudomonas fluorescens* (Pf-5)	No	3	Yes/no	None	3pfm	Midwest Center for Structural Genomics (unpublished work)

† The dimerization mode is different from that observed for active EAL domains.

loops L2, L3 and loop L4, which shows a different conformation in the dimeric and monomeric structures.

3.2. Comparison with other EAL-domain structures

Several structures of EAL domains have recently been reported. They share a common TIM-barrel fold but differ in some other features such as the presence or absence of ligand (c-di-GMP or pGpG), oligomeric state (monomers or dimers) and dimerization mode, as summarized in Table 2. DosP EAL shows the highest sequence identity to Tbd1265 (35%) and the PdeA (CC3396) EAL domain from *Caulobacter crescentus* (34%); the r.m.s.d. values are 3.24 Å (over 344 C^{α} positions) and 3.01 Å (over 411 C^{α} positions) for the corresponding structures (PDB entries 2r6o and 3u2e, respectively; Tchigvintsev *et al.*, 2010; Midwest Center for Structural Genomics, unpublished work). The observed divergence is a consequence of different rotation angles between subunits constrained by the dimer geometry. Accordingly, when comparing single protomers (chain *A*) the r.m.s.d. values are 1.80 Å (over 215 C^{α} positions) and 1.82 Å (over 221 C^{α} positions) for the structures with PDB codes 2r6o and 3u2e, respectively. The most striking differences are observed in loop L5 and the loops L3 and L4 between β3 and α3 and between β4 and α4, respectively.

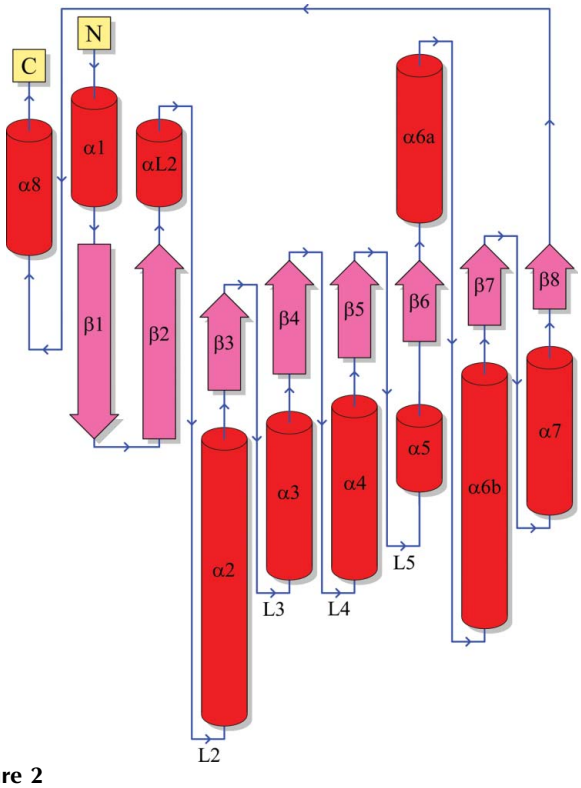

Figure 2
Topology scheme based on the DosP EAL-domain structure. The α-helices are shown in red and the β-strands are shown in magenta; the diagram was prepared using *PDBsum* (Laskowski, 2009).

3.3. Dimerization

DosP EAL was found to be dimeric by gel-filtration chromatographic analysis (Supplementary Fig. S1), which is in line with the previously reported dimerization of full-length DosP (Lechauve *et al.*, 2009) and other full-length EAL-domain proteins such as YkuI (Minasov *et al.*, 2009) and BlrP1 (Barends *et al.*, 2009). In the crystal structure of the native DosP EAL domain the dimer interface is formed mainly by interactions between a pair of antiparallel α-helices (α5 and α6b), which were named the compound helix and the dimerization helix, respectively, in BlrP1 (Barends *et al.*, 2009). The rotation angle required to superpose the *A* and *B* monomers is very close to 180°. The fact that the DosP EAL domain exists both as a monomer (SeMet-labelled protein) and a dimer (native protein) is likely to be caused by the crystallization conditions and not to be of functional relevance. We surmise that the interactions between subunits may be disrupted by the high ionic strength of the crystallization solution for the SeMet-labelled protein as the dimer interface area is not very extensive, but authentic parameters cannot be calculated as the dimer model is lacking helix α5, which is typically part of the interface. The reconstruction of a dimeric structure by aligning the two EAL monomers on the dimer scaffold results

in a number of atomic clashes in the region of helix α5 (Supplementary Fig. S2). This suggests that the actual conformation of the compound helices in the dimer must be slightly different from that observed in the monomer.

A comparison of the available EAL-domain structures (see Table 2) shows that most of them (*i.e.* PDB entries 3gfx, 2r6o, 3u2e, 3sy8, 2bas, 3pjt and 3pfm) share the same dimerization mode through a conserved interface, which is very similar to the situation observed for the dimeric DosP EAL domain described above. The structure of the putative EAL domain from *Listeria monocytogenes* (PDB entry 3kzp; Midwest Center for Structural Genomics, unpublished work) presents a more compact dimer with some additional interactions between helices α2 and α4 (Supplementary Fig. S3). A completely different arrangement with two monomers contacting each other by the tops of the barrels is observed in the structure of the FimX EAL domain (PDB entry 3hvb; Navarro *et al.*, 2009) from *Pseudomona aeruginosa*: the nucleotide-binding sites are buried in an inner cavity and are thus inaccessible to the substrate. However, the biological significance of this dimer remains unclear, especially since the EAL domain of FimX has been shown to be monomeric in solution (Navarro *et al.*, 2009) and the structure of the FimX EAL domain from *Xanthomonas campestris* (PDB entry 4f3h) showed no dimerization (Chin *et al.*, 2012).

Thus, it would appear that enzymatically inactive EAL proteins such as FimX and likely Lmo0111 do not adopt the typical dimer arrangement observed in all structures of active EAL domains (BlrP1, Tbd1265, DosP *etc.*) and which seems to be one of the prerequisites for EAL phosphodiesterase activity.

3.4. Active and inactive EAL domains

Rao *et al.* (2009) categorized EAL domains into three different classes. Class 1 EALs are active and function as PDEs; they have conserved catalytic residues and a conserved loop 6 (linking β5 and α5, and here referred to as loop L5). Class 2 EAL domains contain conserved catalytic residues and a degenerated loop 6; they are most likely to be inactive, but their potential to be activated cannot be excluded. Class 3 EALs lack one or more of the catalytic residues, have a degenerated loop 6 and are predicted to be catalytically inactive.

Based on this classification, we infer that from the set of ten structures containing EAL domains available in the PDB, only four (besides DosP EAL) are enzymatically active EALs: BlrP1, Tbd1265, PdeA and RocR (Fig. 3, Table 2). The rest of the structures presumably represent catalytically inactive EAL domains, with YkuI and Lmo0111 belonging to class 2 and FimX and LapD belonging to class 3 (Fig. 3, Table 2). The functional studies agree with this classification, as BlrP1 (Barends *et al.*, 2009), Tbd1265 (Tchigvintsev *et al.*, 2010), PdeA (CC3396; Christen *et al.*, 2005) and RocR (Rao *et al.*, 2008) have been confirmed to be active PDEs *in vitro*. Additionally, it has also been shown that the DosP EAL domain is also a cyclic di-GMP-specific phosphodiesterase (Schmidt *et al.*, 2005). We have measured the rate of hydrolysis of cyclic di-GMP to pGpG for the DosP EAL domain and the k_{cat} was determined to be $1.43 \pm 0.07\ s^{-1}$ at room temperature (Supplementary Fig. S4).

3.5. The catalytic site

Despite the chemical diversity of reactions catalyzed by TIM-barrel proteins, the active site is usually found at the C-terminal end of the β-barrel. Similarly, in the case of EAL proteins such as BlrP1 and Tbd1265, it has previously been shown that c-di-GMP binds to the active site located at the top of the barrel in an extended conformation.

The same situation is expected for the DosP EAL domain. It has been shown that the glutamic acid residue in the EAL signature motif is essential for enzymatic activity (Christen *et al.*, 2005). The EAL motif (Glu576-Ala-Leu578) is located close to the middle of the antiparallel strand β2, with the Glu576 and Leu578 side chains forming the bottom of the groove at the binding site, while the Ala577 side chain is buried in the core of the domain.

In several previous biochemical studies of EAL-domain proteins it has been shown that Mg^{2+} or Mn^{2+} is required for the hydrolysis of c-di-GMP, while Ca^{2+} was found to be a strong inhibitor of enzymatic activity (Christen *et al.*, 2005; Schmidt *et al.*, 2005). The mutational analysis of RocR from *P. aeruginosa* suggested that the EAL domains use a one-metal-ion catalytic mechanism (Rao *et al.*, 2008). Since the structure of RocR (PDB entry 3sy8; Chen *et al.*, 2012) does not contain c-di-GMP it does not help to discriminate between one-metal-ion or two-metal-ion mechanisms, as the second metal ion may only bind in the presence of the nucleotide.

Recently, the structures of the YkuI, BlrP1, FimX and Tbd1265 EALs have been determined in complex with c-di-GMP. However, only the BlrP1 and Tbd1265 EALs provided insight into the catalytic mechanism of active EAL domains, as FimX was found to be inactive and activity of YkuI could be neither confirmed nor ruled out. The BlrP1 and Tbd1265 EAL structures showed two metal ions coordinated by several conserved carboxylates and one asparagine at the active site. The two metal ions (Me1 and Me2) together with a conserved lysine residue were proposed to be responsible for activation of the catalytic water molecule and the production of the nucleophilic hydroxide (Barends *et al.*, 2009; Tchigvintsev *et al.*, 2010). Another water molecule was proposed to serve as a proton donor.

Taken together, these results support a two-metal-ion catalytic mechanism of c-di-GMP hydrolysis by EAL domains, which appears to also apply to DosP EAL. From the comparison of the active sites of DosP EAL and Tbd1626 EAL we can clearly see that all of the residues coordinating c-di-GMP and the two metal ions are conserved, namely Glu576, Arg580, Asn635, Glu667, Asp697, Asp698, Lys718 and Glu754, and display virtually identical conformations.

Nevertheless, there are some exceptions; the Asn635 side chain adopts an alternate rotamer conformation and most notably both aspartic acid residues (Asp697 and Asp698) are

strongly displaced towards the outside of the barrel (Fig. 4*a*). The average distance between corresponding carboxyl positions in Tbd1265 EAL (Asp646 and Asp647) and DosP EAL is increased by 1.5 Å for Asp697 and ∼2.7 Å for Asp698. Since the substrate and the metal ions are not present in the DosP EAL structure, we can only speculate about the exact interactions at the active site. Presumably, the mode of c-di-GMP binding would be similar to that observed for Tbd1265 EAL. The proposed model assumes interactions of the Asn635 side chain with the c-di-GMP phosphate P1 (forming the scissile bond) and of phosphate P2 with Arg580. In this model, the two divalent metal cations are coordinated by Glu576, Asn635, Glu667, Asp697, Asp698 and Glu754. The side chain of Lys718 is proposed to contribute to the activation of the bound water molecule.

The structural differences between the Tbd1265 and DosP EAL domains cannot be ascribed to a lack of substrate in the active site, as no significant structural rearrangement upon c-di-GMP binding was reported in the case of the enzymatically active Tbd1265 EAL and the inactive YkuI.

The 'retracted' side chains of Asp697 and Asp698 disrupt the geometry of the active site, which prevents metal-ion binding. Both aspartic acid residues are part of the conserved loop L5 [DDFG(T/A)GYSS], referred to by Rao *et al.* (2009) as loop 6. The dynamic character of loop 6 was demonstrated for RocR from *P. aeruginosa* using an HD-exchange approach and it was proposed to mediate EAL-domain dimerization as well as c-di-GMP and metal-ion binding (Rao *et al.*, 2009). An important regulatory role of loop 6 was also confirmed by the recent structural studies of RocR (Chen *et al.*, 2012).

It is noteworthy that a similar conformation of the residues coordinating metal ions to that observed in DosP EAL is found in the structure of YkuI (PDB entry 2bas; Minasov *et al.*, 2009; Fig. 4*b*). The side chains of Asp152 and Asn153 (corresponding to Asp697 and Asp698 in DosP) are pulled out of the active site by about 2.0–2.8 Å (with respect to the Tbd1265 EAL structure). This could indicate that YkuI might in fact potentially be an active phosphodiesterase but the signal which triggers its activity has not yet been discovered. This hypothesis is consistent with the highly similar quaternary

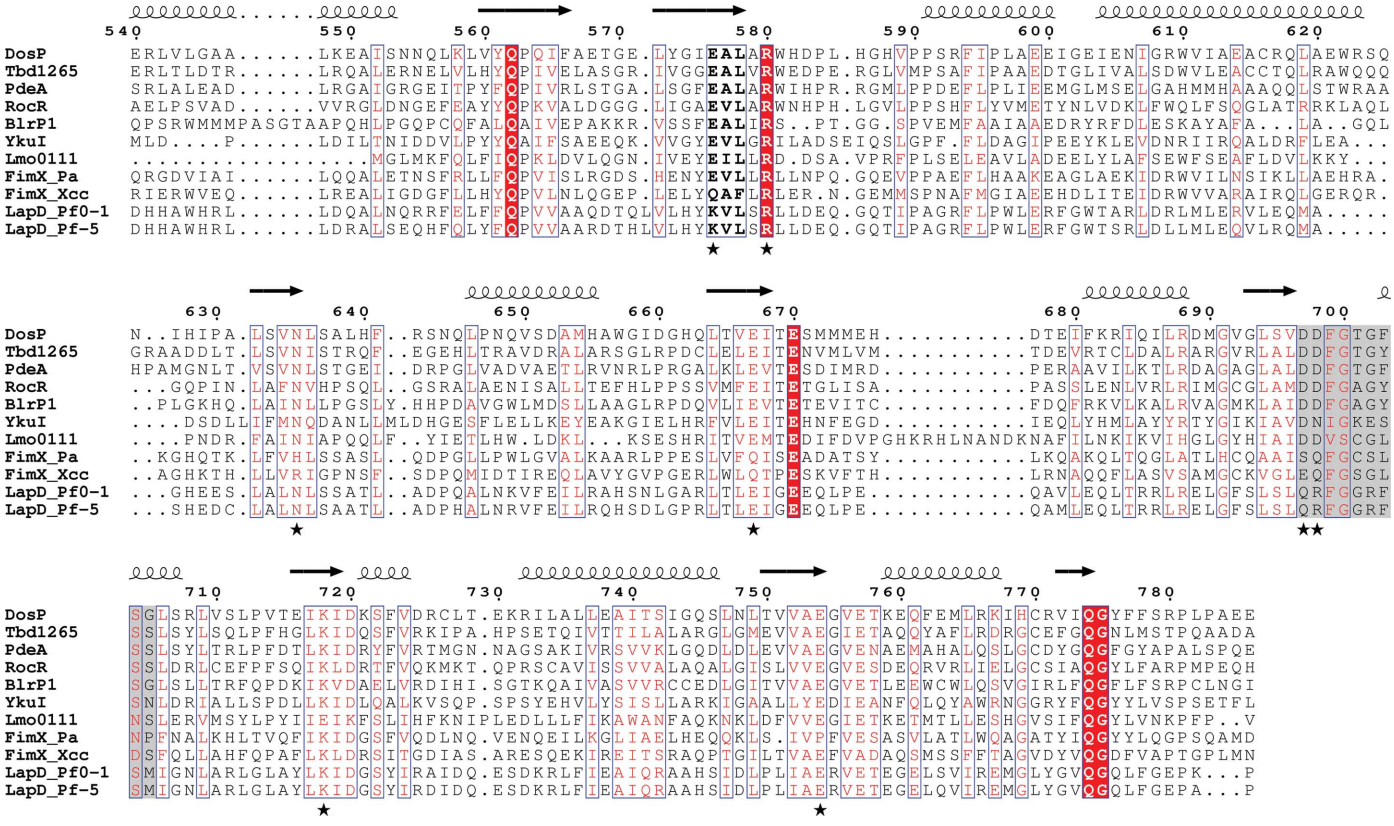

Figure 3
Sequence alignment of EAL domains with known crystal structures. EAL domains were annotated using the SMART 7 database (Letunic *et al.*, 2012) for the following proteins: DosP from *E. coli* (P76129; this study), Tbd1265, a putative DGC/PDE from *T. denitrificans* (Q3SJE6; PDB entry 2r6o), PdeA from *C. crescentus* (Q9A310; PDB entry 3u2e), RocR from *P. aeruginosa* (Q9HX69; PDB entry 3sy8), BlrP1 from *K. pneumoniae* (A6T8V8; PDB entry 3gfx), YkuI from *B. subtilis* (O35014; PDB entry 2bas), Lmo0111, a putative PDE from *L. monocytogenes* (Q8YAK7; PDB entry 3kzp), FimX from *P. aeruginosa* (Q9HUK6; PDB entry 3hvb), FimX from *X. campestris* (B0RS05; PDB entry 4f3h) and LapD from *P. fluorescens* (Q3KK31 and Q4KKF5; PDB entries 3pjt and 3pfm). The sequences were aligned using *T-Coffee* (Di Tommaso *et al.*, 2011) and visualized with *ESPript* (Gouet *et al.*, 1999). Residues with strict identity are shown as white characters on a red background, whereas those with high similarity are shown as red characters and are framed. The EAL motif is shown in bold and loop L5 (also referred to as loop 6) is shown on a grey background. The residues in the active site are indicated by black asterisks. The secondary-structure elements of DosP EAL are depicted above the alignment.

structures of BlrP1 and YkuI, as discussed by Barends *et al.* (2009).

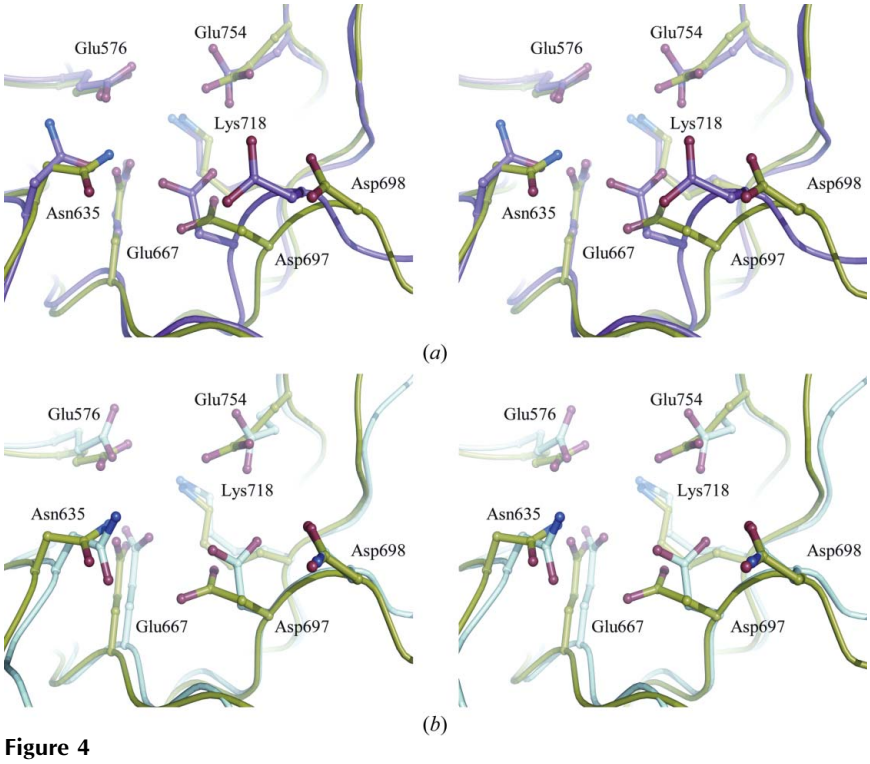

Figure 4
Stereo figures displaying differences in the arrangement of loop L5 and the metal-coordinating residues Asp697 and Asp698. (*a*) Active site and loop L5 of the DosP EAL structure (PDB entry 4hu4, chain *A*; green) superposed with the Tbd1265 EAL (PDB entry 2r6o, chain *B*; magenta); a shift in the backbone results in the displacement of the Asp697 and Asp698 side chains in comparison to their counterparts. (*b*) Superposition with the YkuI EAL domain (PDB entry 2bas, chain *A*; cyan) results in a good overlay of these residues and the backbone.

Interestingly, unlike in other described EAL structures, access to the active site of DosP EAL is sterically restricted by the presence of a short helical stretch (Ser637-Ala-Leu-His640) in loop L3 between strand $\beta3$ and helix $\alpha3$ (Fig. 5*a*). This unusual element, together with an unordered fragment, replaces the short α-helix (referred to as $\alpha5$ in Tbd1265 EAL) found in other EAL-domain structures. The one-turn helix is in close proximity to the active site and would result in steric clashes with one of the guanine rings of c-di-GMP (Fig. 5*a*). It may apparently prevent substrate binding and thus might be part of the enzyme-regulatory mechanism.

In case of the DosP EAL domain, the cavity accommodating one of the guanine-ring moieties of the substrate is occupied by the side chains of Ala638 and Leu639 (Fig. 5*a*) and owing to this fact the hydrophobic pocket is dramatically reduced in size compared with the Tbd1265 EAL structure (Fig. 5*b*). A similar conformation of this region is observed for both dimeric and monomeric DosP EAL structures.

This observation may explain why our attempts to obtain a structure in complex with substrate by cocrystallization in the presence of c-di-GMP and Ca^{2+} ions were unsuccessful. Since both crystal forms show the same conformation of the active site, it is unlikely that this is a result of crystal-packing forces or crystallization conditions.

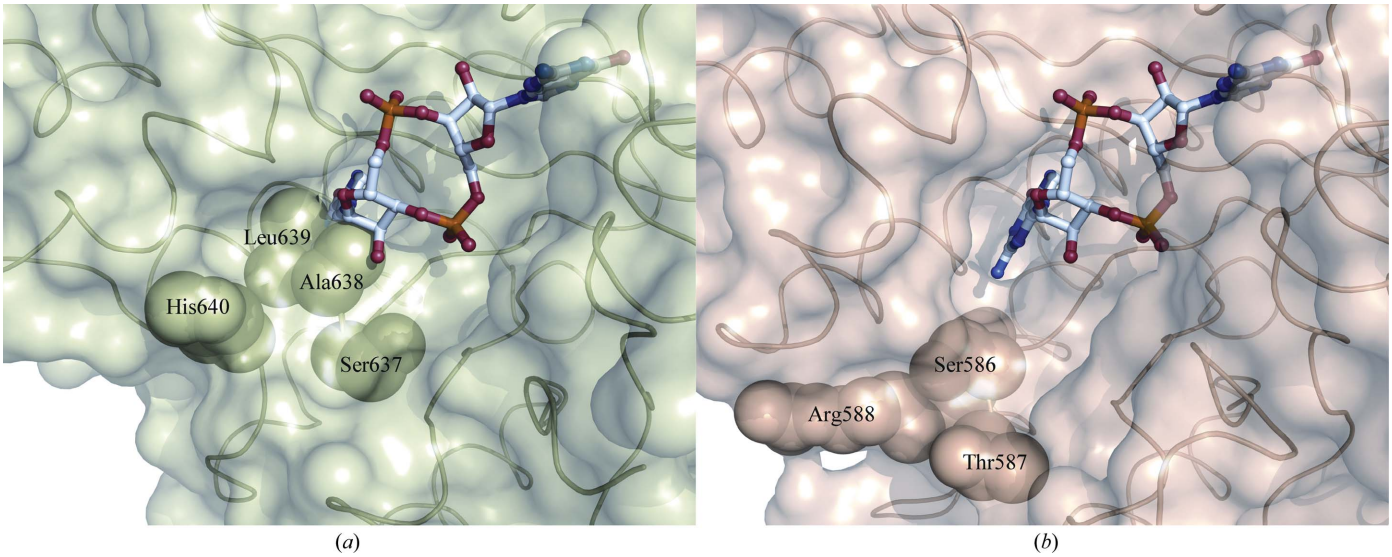

Figure 5
Close-up view of the c-di-GMP binding pocket. (*a*) Chain *A* of the dimeric DosP EAL domain is shown as a C^{α} trace and surface representation (green). The side chains of residues Ser637-Ala-Leu-His640 are blocking the binding site, which would result in clashes with a bound c-di-GMP molecule, which is superimposed here for demonstration. (*b*) The same view in Tbd1265 EAL (PDB entry 3n3t, chain *A*; orange), in which the guanine moiety of the substrate is accommodated in the binding cavity and the corresponding side chains (Ser586-Thr-Arg-Gln589) point away from the pocket (the Gln589 side chain is buried and is not visible).

The inhibited DosP EAL domain observed in both crystal structures might possibly reflect a conformation which is not strongly populated in solution and thereby may suggest a regulatory mechanism. Since we and others (Schmidt *et al.*, 2005) have shown that the isolated DosP EAL domain is enzymatically active *in vitro*, the inactive conformation of the enzyme described here is suggested to relate to an inhibited form occurring in the inactive state of holo DosP.

4. Conclusions

Based on the presented structures, we can clearly observe that not only is access to the active site partially blocked by the presence of loop L3, but also important residues anchored to loop L5 are in a nonproductive arrangement.

Both observations strongly imply that the enzymatically active DosP EAL domain in the presented structures adopts an inactive conformation. This in turn would suggest a positive regulatory mechanism for DosP in which a basally inactive EAL domain is activated by a conformational change of the PAS domains, rather than a negative regulation mechanism in which active EAL domains are suppressed by the PAS sensor; however, these mechanisms do not have to be mutually exclusive.

We notice that loop L5 and to some extent also loop L3 are both in proximity to the dimer interface, and thus their conformation may be influenced by changes at the interface. Such coupling would be most likely to involve the compound helix $\alpha5$ in signal propagation from the sensory domains. The disorder of $\alpha5$ may be the reason for the inhibitory arrangement of loop L3, with the conformation of $\alpha5$ possibly being a result of the truncation and removal of neighbouring domains.

In conclusion, this is possibly the first example of a regulated EAL domain with proven catalytic activity in a conformation corresponding to an 'off' state. We hypothesize that a signal transmitted from the sensory domain traverses the dimer interface and causes a conformational rearrangement of loops L3 and L5 that renders the EAL active site competent for substrate binding and catalysis.

The authors are grateful to Robert L. Shoeman, Anton Meinhart and Max Cryle for discussion and support and to Ingrid Vetter for help in solving the SeMet-derivative crystal structure. Diffraction data were collected on beamline X10SA at the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland. We thank the Dortmund–Heidelberg data-collection team and the PXII staff for their support in setting up the beamline and Chris Roome and Ingrid Vetter for expert support with the crystallographic software. We acknowledge financial support by the Deutsche Forschungsgemeinschaft SFB 623.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Barends, T. R., Hartmann, E., Griese, J. J., Beitlich, T., Kirienko, N. V., Ryjenkov, D. A., Reinstein, J., Shoeman, R. L., Gomelsky, M. & Schlichting, I. (2009). *Nature (London)*, **459**, 1015–1018.
Chen, M. W., Kotaka, M., Vonrhein, C., Bricogne, G., Rao, F., Chuah, M. L. C., Svergun, D., Schneider, G., Liang, Z.-X. & Lescar, J. (2012). *J. Bacteriol.* **194**, 4837–4846.
Chin, K.-H., Kuo, W.-T., Yu, Y.-J., Liao, Y.-T., Yang, M.-T. & Chou, S.-H. (2012). *Acta Cryst.* D**68**, 1380–1392.
Christen, M., Christen, B., Folcher, M., Schauerte, A. & Jenal, U. (2005). *J. Biol. Chem.* **280**, 30829–30837.
Delgado-Nixon, V. M., Gonzalez, G. & Gilles-Gonzalez, M. A. (2000). *Biochemistry*, **39**, 2685–2691.
Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F. & Notredame, C. (2011). *Nucleic Acids Res.* **39**, W13–W17.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Galperin, M. Y., Higdon, R. & Kolker, E. (2010). *Mol. Biosyst.* **6**, 721–728.
Galperin, M. Y., Nikolskaya, A. N. & Koonin, E. V. (2001). *FEMS Microbiol. Lett.* **203**, 11–21.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Krasteva, P. V., Giglio, K. M. & Sondermann, H. (2012). *Protein Sci.* **21**, 929–948.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Kurokawa, H., Lee, D.-S., Watanabe, M., Sagami, I., Mikami, B., Raman, C. S. & Shimizu, T. (2004). *J. Biol. Chem.* **279**, 20186–20193.
Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–359.
Lechauve, C., Bouzhir-Sima, L., Yamashita, T., Marden, M. C., Vos, M. H., Liebl, U. & Kiger, L. (2009). *J. Biol. Chem.* **284**, 36146–36159.
Letunic, I., Doerks, T. & Bork, P. (2012). *Nucleic Acids Res.* **40**, D302–D305.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Minasov, G., Padavattan, S., Shuvalova, L., Brunzelle, J. S., Miller, D. J., Basle, A., Massa, C., Collart, F. R., Schirmer, T. & Anderson, W. F. (2009). *J. Biol. Chem.* **284**, 13174–13184.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Navarro, M. V., De, N., Bae, N., Wang, Q. & Sondermann, H. (2009). *Structure*, **17**, 1104–1116.
Navarro, M. V., Newell, P. D., Krasteva, P. V., Chatterjee, D., Madden, D. R., O'Toole, G. A. & Sondermann, H. (2011). *PLoS Biol.* **9**, e1000588.
Park, H., Suquet, C., Satterlee, J. D. & Kang, C. (2004). *Biochemistry*, **43**, 2738–2746.
Rao, F., Qi, Y., Chong, H. S., Kotaka, M., Li, B., Li, J., Lescar, J., Tang, K. & Liang, Z.-X. (2009). *J. Bacteriol.* **191**, 4722–4731.
Rao, F., Yang, Y., Qi, Y. & Liang, Z.-X. (2008). *J. Bacteriol.* **190**, 3622–3631.
Römling, U. (2011). *Environ. Microbiol.* **14**, 1817–1829.
Römling, U. & Amikam, D. (2006). *Curr. Opin. Microbiol.* **9**, 218–228.
Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de Vroom, E., van der Marel, G. A., van Boom, J. H. & Benziman, M. (1987). *Nature (London)*, **325**, 279–281.
Schirmer, T. & Jenal, U. (2009). *Nature Rev. Microbiol.* **7**, 724–735.
Schmidt, A. J., Ryjenkov, D. A. & Gomelsky, M. (2005). *J. Bacteriol.* **187**, 4774–4781.
Tanaka, A., Takahashi, H. & Shimizu, T. (2007). *J. Biol. Chem.* **282**, 21301–21307.
Tchigvintsev, A., Xu, X., Singer, A., Chang, C., Brown, G., Proudfoot, M., Cui, H., Flick, R., Anderson, W. F., Joachimiak, A., Galperin, M. Y., Savchenko, A. & Yakunin, A. F. (2010). *J. Mol. Biol.* **402**, 524–538.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.