research papers

CrossMark

Structural analysis of an oxygen-regulated diguanylate cyclase

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

Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany. *Correspondence e-mail: miroslaw.tarnawski@mpimf-heidelberg.mpg.de

Cyclic di-GMP is a bacterial second messenger that is involved in switching between motile and sessile lifestyles. Given the medical importance of biofilm formation, there has been increasing interest in understanding the synthesis and degradation of cyclic di-GMPs and their regulation in various bacterial pathogens. Environmental cues are detected by sensing domains coupled to GGDEF and EAL or HD-GYP domains that have diguanylate cyclase and phosphodiesterase activities, respectively, producing and degrading cyclic di-GMP. The *Escherichia coli* protein DosC (also known as YddV) consists of an oxygen-sensing domain belonging to the class of globin sensors that is coupled to a C-terminal GGDEF domain *via* a previously uncharacterized middle domain. DosC is one of the most strongly expressed GGDEF proteins in *E. coli*, but to date structural information on this and related proteins is scarce. Here, the high-resolution structural characterization of the oxygen-sensing globin domain, the middle domain and the catalytic GGDEF domain in apo and substrate-bound forms is described. The structural changes between the iron(III) and iron(II) forms of the sensor globin domain suggest a mechanism for oxygen-dependent regulation. The structural information on the individual domains is combined into a model of the dimeric DosC holoprotein. These findings have direct implications for the oxygen-dependent regulation of the activity of the cyclase domain.

1. Introduction

The dinucleotide cyclic di-GMP (c-di-GMP) is currently recognized as one of the most widespread and significant bacterial second-messenger molecules (Römling *et al.*, 2013). It has been shown that cyclic di-GMP regulates surface adhesion, biofilm formation and dispersal, motility, virulence, the cell cycle, differentiation and other physiological processes. Cyclic di-GMP plays an important role in the transition between motile and sessile states, which is necessary for the development and persistence of the multicellular microbial communities implicated in bacterial pathogenesis (Kalia *et al.*, 2013; Tamayo *et al.*, 2007).
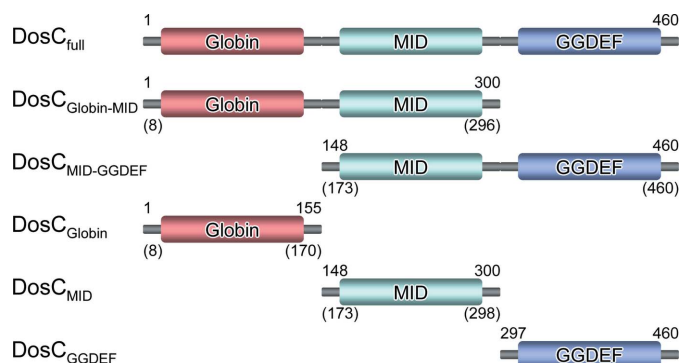
Cyclic di-GMP is produced by diguanylate cyclases (DGCs) and is degraded by phosphodiesterases (PDEs), which contain GGDEF domains and EAL or HD-GYP domains, respectively (Krasteva *et al.*, 2012). The vast majority of bacterial species have numerous proteins with single GGDEF and EAL domains, but also those with GGDEF–EAL modules arranged in tandem. In fact, GGDEF and EAL domains belong to the most abundant protein families encoded in prokaryotic genomes (Galperin *et al.*, 2010). Very often, enzymatic GGDEF and EAL domains are associated with accessory REC, PAS and GAF signal transduction domains (Schirmer & Jenal, 2009). This combination enables modulation of the enzymatic activity of the GGDEF and EAL output domains
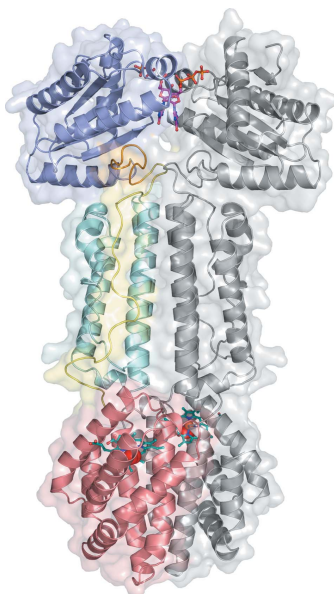
by upstream input domains, which are frequently part of two-component systems. PAS domains detect changes in light, redox potential, oxygen and a variety of other signals (Henry & Crosson, 2011). Enzymes involved in c-di-GMP metabolism possessing such sensory domains can monitor internal or environmental cues and alter the rate of c-di-GMP synthesis or hydrolysis in response to changes. Cyclic di-GMP effectors include diverse transcription factors, riboswitches, PilZ domains, degenerate DGCs and PDEs (Boyd & O'Toole, 2012).

Recently, it has become apparent that small diatomic gaseous molecules (O_2, CO and NO) also serve as signalling molecules affecting transcription, chemotaxis, second-messenger signalling and other physiological processes within the bacterial cell (Aono, 2013; Farhana et al., 2012). Gas sensing plays a crucial role in maintaining homeostasis and is responsible for the cellular adaptive response. Bacterial gas-sensor proteins utilize haem and Fe–S clusters as sensors for diatomic gases (Green et al., 2009). Haem-based sensors constitute a large superfamily in which six major families can be distinguished (Gilles-Gonzales & Gonzales, 2008). Such haem-binding domains are coupled to the transmitters involved in signal transduction, which include histidine protein kinases, methyl-carrier proteins and proteins engaged in nucleotide second-messenger signalling and DNA binding (Germani et al., 2013; Martínková et al., 2013). Further information can be obtained from previous reviews (Igarashi et al., 2011; Weinert & Marletta, 2011).

The cyclic di-GMP-mediated signalling network in *Escherichia coli* has recently been reviewed by Povolotsky & Hengge (2012); c-di-GMP degradation by the PDE activity of the well characterized haem-based oxygen sensor named DosP, belonging to the haem-binding PAS-domain sensor family, is part of the system. DosP phosphodiesterase activity is regulated in an oxygen-dependent fashion (Tanaka et al., 2007). The genomic context suggested and the transcriptional data confirmed that the *dosP* gene is part of a *dosC–dosP* bicistronic operon in which these two genes are co-transcribed. The *dosC* (also referred to as *yddV*) gene encodes a functional

diguanylate cyclase (Méndez-Ortiz et al., 2006). Over-expression of the *dosC* gene resulted in enhanced biofilm formation, decreased motility and an abnormal cell-division process.

Consequently, it became clear that DosC (YddV) is an oxygen sensor belonging to the class of globin-coupled sensors (GCSs) consisting of an N-terminal globin domain coupled to a C-terminal GGDEF domain (Fig. 1). Surprisingly, diguanylate cyclase activity could only be detected when DosC was co-purified with DosP. Incubation of the DosC–DosP complex with GTP ultimately resulted in the production of linear di-GMP (pGpG), indirectly proving DGC activity (Tuckerman et al., 2009). Kitanishi and coworkers employed an HPLC method to monitor the time-dependent increase and decrease in c-di-GMP and GTP concentrations, respectively. The study also showed that the iron(II)–O_2, iron(II)–CO and iron(III) forms of DosC are active, whereas the iron(II) and iron(II)–NO forms are inactive, demonstrating that O_2 binding to the iron(II) form of the DosC haem domain stimulates its DGC activity (Kitanishi et al., 2010). The authors proposed His98 and Tyr43 as proximal and distal haem ligands, respectively. Tyr43 was suggested to be important for O_2 recognition and stability of the iron(II)–O_2 complex, while Leu65 was suggested to restrict water access to the haem distal site of DosC (Nakajima et al., 2012).

DosC is one of the most highly expressed GGDEF proteins in *E. coli* (Sommerfeldt et al., 2009). The gene encoding DosC is under the control of the general stress response master regulator σ^S (RpoS; Weber et al., 2006). It has been shown that DosC, together with some other DGCs and PDEs, is part of the mechanism regulating motility in *E. coli* by modulating the intracellular c-di-GMP level (Boehm et al., 2010). Moreover, Tuckermann and coworkers demonstrated that the DosC–DosP complex from *E. coli* co-purified with a large ribonucleoprotein complex (RNP) containing polynucleotide phosphorylase (PNPase), a component of the RNA degradosome (Tuckerman et al., 2011). It was shown that DosC regulates the expression of genes encoding curli fibre subunits (Tagliabue, Maciag et al., 2010) as well as the genes encoding the enzymes responsible for extracellular polysaccharide poly-*N*-acetylglucosamine (PNAG) biosynthesis (Tagliabue, Antoniani et al., 2010).

In addition to DosC, four other bacterial sensory globin-coupled diguanylate cyclases have been characterized to date, namely *Av*GReg (Thijs et al., 2007), *Bp*GReg (Wan et al., 2009), HemDGC (Sawai et al., 2010) and *Pcc*GCS (Burns et al., 2014). They all share the same architecture as DosC and consist of an N-terminal globin domain linked to a C-terminal GGDEF domain (Fig. 1). Some of their properties are summarized in Supplementary Table S1. Proteins homologous to DosC have been identified in a variety of bacterial species (Freitas et al., 2005), including human pathogens such as *Klebsiella pneumoniae* and *Shigella sonnei* (Supplementary Fig. S1). To date, there are no structural data available for any of these globin-coupled diguanylate cyclases.

Here, we present a comprehensive structural study of DosC, a bacterial oxygen-regulated diguanylate cyclase from *E. coli*.

Figure 1
The domain organization of full-length DosC and other variants used in this study. The first and last residues of each construct are indicated (alternative versions are in parentheses). The individual domains are coloured as follows: sensory globin domain, red; middle domain, cyan; catalytic GGDEF domain, blue. The same colouring scheme is used throughout the other figures.

This is the first report describing the structure of an O_2-sensory diguanylate cyclase and provides novel and considerable insight into the architecture and the function of this important class of proteins.

2. Materials and methods

2.1. Cloning

The DNA fragments encoding full-length DosC (DosC$_{full}$; residues 1–460) from *E. coli*, as well as various truncations (see Fig. 1), were PCR-amplified and cloned into the bacterial expression vector pET-24d (EMD Biosciences) using NcoI and XhoI restriction sites. The resulting constructs contained a C-terminal hexahistidine (His$_6$) tag with the sequence LEHHHHHH. Additionally, several constructs (ZZ-DosC$_{full}$, residues 1–460; ZZ-DosC$_{Globin}$, residues 8–170; ZZ-DosC$_{Globin-MID}$, residues 8–296; ZZ-DosC$_{MID-GGDEF}$, residues 173–460) were cloned into the pET-ZZ vector (a gift from Gunter Stier, BZH Heidelberg; Bogomolovas *et al.*, 2009) using the same restriction sites as before. This resulted in fusion proteins containing an N-terminal His$_6$ tag followed by the solubility-enhancing ZZ-tag and a *Tobacco etch virus* (TEV) protease cleavage site.

2.2. Protein expression and purification

Proteins were expressed in *E. coli* strain BL21 Star (DE3) (Life Technologies). All proteins containing the DosC globin domain were expressed using an auto-induction protocol as described previously (Studier, 2005). Briefly, the cells were precultured in non-inducing MDG medium in the presence of kanamycin (50 µg ml^{-1}) and subsequently grown in ZYM-5052 medium supplemented with kanamycin (50 µg ml^{-1}) and 5-aminolevulinic acid (125 µg ml^{-1}) for 66–72 h at 291 K. Nonhaem constructs were expressed using isopropyl β-D-1-thiogalactopyranoside (IPTG) for induction. The cells were grown in LB medium supplemented with kanamycin (50 µg ml^{-1}) at 310 K. At an optical density at 600 nm of 0.8–1.0, the temperature was reduced to 291 K and protein expression was induced with 0.1 mM IPTG for 16–18 h at 291 K. The cells were harvested by centrifugation and stored at 193 K until further use. All subsequent purification steps were performed at 277 K.

The cell pellet was resuspended in buffer *A* [50 mM NaH$_2$PO$_4$ pH 8.0, 300 mM NaCl, 10 mM imidazole, 10%(v/v) glycerol] supplemented with cOmplete EDTA-free protease inhibitor (Roche), 2.5 mM dithioerythritol (DTE), 1 mg ml^{-1} lysozyme and 10 µg ml^{-1} DNaseI. The cells were lysed using a microfluidizer (Microfluidics) operated at a pressure of 0.7 MPa and the lysates were clarified by centrifugation at 47 800g for 1 h at 277 K. The cleared supernatant was loaded onto a His-Trap FF column (GE Healthcare) pre-equilibrated with buffer *A*. The column was washed with buffer *A* and the protein was eluted in buffer *A* with 500 mM imidazole. Subsequent purification of the various proteins was performed as follows.

Fractions containing DosC$_{Globin}$ (residues 1–155) were dialyzed overnight against buffer *B* (25 mM Tris–HCl pH 8.0, 50 mM NaCl). After dialysis, the sample was concentrated using Amicon Ultra centrifugal units (Millipore) and purified by gel filtration on a Superose 6 column (GE Healthcare) equilibrated with buffer *B*. The eluted protein was finally concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

ZZ-DosC$_{Globin}$ (residues 8–170) was dialyzed against buffer *B* and then incubated for 16–18 h at 277 K with His$_6$-TEV protease (a gift from Gunter Stier, BZH Heidelberg) using a 1:10 molar ratio of TEV:substrate. The sample was loaded onto an Ni–NTA column (Qiagen) to remove the cleaved His$_6$-ZZ-tag and His$_6$-TEV. The flowthrough was further concentrated and subjected to gel-filtration chromatography on a Superdex 200 column (GE Healthcare) in buffer *B*. The protein-containing fractions were pooled, concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

The DosC$_{Globin-MID}$ (residues 1–300) purification was performed essentially as described above for DosC$_{Globin}$ (residues 1–155). ZZ-DosC$_{Globin-MID}$ (residues 8–296) was purified similarly as described above for ZZ-DosC$_{Globin}$ (residues 8–170), except that the buffer used for gel filtration was 25 mM Tris–HCl pH 8.0, 50 mM NaCl, 2.5 mM EDTA, 3 mM DTE, 10%(v/v) glycerol.

After elution from the HisTrap FF column, DosC$_{MID}$ (residues 148–300) was dialyzed overnight against buffer *B*1 [25 mM Tris–HCl pH 8.0, 100 mM NaCl, 5%(v/v) glycerol]. The dialysed sample was concentrated using Amicon Ultra centrifugal units (Millipore) and stored at 193 K. The protein sample was then dialyzed against 25 mM sodium acetate pH 4.6, 10%(v/v) glycerol, diluted to 5 mg ml^{-1} and incubated with 100 µg ml^{-1} chymotrypsin at 310 K. After 4 h, the protease was inactivated by the addition of phenylmethylsulfonyl fluoride (PMSF) to 1 mM and the sample was loaded onto a Mono Q column (GE Healthcare) pre-equilibrated with buffer *C* [25 mM Tris–HCl pH 8.5, 10%(v/v) glycerol]. The column was washed with buffer *C* and the protein was subsequently eluted using a linear gradient of 0–500 mM NaCl in the same buffer. Selected fractions were combined, buffer-exchanged into 25 mM Tris–HCl pH 8.5, 50 mM NaCl, 10%(v/v) glycerol, concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

Fractions containing DosC$_{GGDEF}$ (residues 297–460) were dialyzed against buffer *B*2 [25 mM bis-tris pH 6.0, 100 mM NaCl, 5%(v/v) glycerol]. After dialysis, the sample was concentrated using Amicon Ultra centrifugal units (Millipore) and purified by gel filtration on a Superdex 200 column (GE Healthcare) in buffer *B*2. The eluted protein was concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

ZZ-DosC$_{MID-GGDEF}$ (residues 173–460) was dialyzed against buffer *B*2, concentrated and subjected to gel filtration on a Superdex 200 column (GE Healthcare) equilibrated with buffer *B*2. Owing to insufficient solubility in buffer *B*2, the sample was finally buffer-exchanged into buffer *B*1 [25 mM Tris–HCl pH 8.0, 100 mM NaCl, 5%(v/v) glycerol], concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

ZZ-DosC$_{full}$ (residues 1–460) was dialyzed against buffer $B1$ and purified by gel filtration on a Superdex 200 column (GE Healthcare) in the same buffer. Fractions containing ZZ-DosC$_{full}$ were concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

For purification of DosC$_{full}$ (residues 1–460), the lysis buffer was additionally supplemented with 0.3 M arginine. The cells were lysed and the lysate was clarified as described above. The protein was precipitated from the clarified lysate at 50% ammonium sulfate saturation, pelleted by centrifugation and redissolved in buffer A. The sample was loaded onto a HisTrap FF column (GE Healthcare) pre-equilibrated with buffer A. The column was washed with buffer A and the protein was eluted in buffer A with 500 mM imidazole. Immediately after elution, arginine was added to a final concentration of 50 mM. The protein was again precipitated at 50% ammonium sulfate saturation, pelleted by centrifugation and stored at 253 K. Finally, the pellet was dissolved in buffer D [20 mM CHES pH 9.5, 20 mM NaCl, 2.5 mM EDTA, 2.5 mM DTE, 25 mM arginine, 25 mM glutamic acid, 10%(v/v) glycerol] and further purified by gel filtration on a Superdex 200 column (GE Healthcare) equilibrated with the same buffer. The eluted protein was finally concentrated, aliquoted, frozen in liquid nitrogen and stored at 193 K.

2.3. Protein crystallization

Crystallization was performed at 293 K using the vapour-diffusion method. Crystals of DosC$_{Globin}$ (residues 8–170) were prepared under aerobic conditions using 'as-isolated' protein in the iron(III) form. Crystals with a hexagonal rod morphology (DosC$_{Globin}$ form I) were grown by vapour diffusion against the precipitant, mixing protein solution (20 mg ml^{-1}) and precipitant solution consisting of 0.1 M trisodium citrate pH 5.6, 0.2 M ammonium acetate, 27%(w/v) PEG 4000. The concentration of PEG 4000 in the crystallization drop was increased to 44%(w/v) for cryoprotection. After 20 min of incubation at 293 K, crystals were harvested and flash-cooled in liquid nitrogen.

Crystals of DosC$_{Globin}$ (residues 1–155) were prepared under anaerobic conditions in a glove box (Belle, UK). Anaerobic conditions were maintained by continuous recirculation of the glove-box atmosphere through metal powder cartridges according to the manufacturer's instructions. All solutions used were degassed. Firstly, the protein was reduced with 10 mM sodium dithionite in the glove box and the buffer was then exchanged to 25 mM Tris–HCl pH 8.0, 50 mM NaCl to remove the reducing agent. The expected spectral shift characteristic of the iron(II) form was confirmed using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) and finally the protein was concentrated to 20 mg ml^{-1}. Small plate-shaped crystals (DosC$_{Globin}$ form II) growing in clusters were obtained in the glove box by mixing protein solution and a reservoir solution consisting of 0.1 M sodium phosphate–citrate pH 4.2, 0.2 M sodium chloride, 16%(w/v) PEG 3000. Single crystals were harvested in the glove box and briefly washed in cryoprotectant solution consisting of the reservoir solution supplemented with 25%(v/v) ethylene glycol before flash-cooling in liquid nitrogen.

Crystals of DosC$_{MID}$ (residues 173–298) with a plate-like morphology (DosC$_{MID}$ form I) were grown by mixing protein solution at 8.9 mg ml^{-1} and reservoir solution composed of 0.1 M bis-tris pH 5.5, 0.2 M magnesium chloride, 17%(w/v) PEG 3350, 0.1 M Gd-HPDO3A [gadolinium complex of 10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid]. Before flash-cooling in liquid nitrogen, the crystals were briefly washed for cryoprotection in a solution consisting of 0.1 M bis-tris pH 5.5, 0.2 M magnesium chloride, 35%(w/v) PEG 3350. For phasing, the same crystal form was derivatized with potassium iodide. Prior to flash-cooling in liquid nitrogen, the crystals were soaked for 80 s in cryoprotectant solution supplemented with 2 M potassium iodide. Additionally, a different crystal form (DosC$_{MID}$ form II) of DosC$_{MID}$ was obtained using 0.2 M ammonium dihydrogen phosphate, 20%(w/v) PEG 3350. The crystals were briefly rinsed in cryoprotectant solution consisting of the corresponding precipitant solution supplemented with 20%(v/v) glycerol before flash-cooling in liquid nitrogen.

Crystals of DosC$_{GGDEF}$ (residues 297–460) in the apo form (DosC$_{GGGDEF}$ form I) were obtained by mixing protein solution at 7.4 mg ml^{-1} and reservoir solution consisting of 0.1 M trisodium citrate pH 5.6, 0.2 M ammonium acetate, 30%(w/v) PEG 4000. The crystals were briefly washed in cryoprotectant solution consisting of the reservoir solution with ethylene glycol added to a final concentration of 20%(v/v) prior to flash-cooling in liquid nitrogen. For co-crystallization, 1 mM Rp-GTPαS (Biolog Life Science Institute, Germany) and 10 mM MgCl$_2$ were added to the DosC$_{GGDEF}$ protein solution. Crystals of Rp-GTPαS-bound DosC$_{GGDEF}$ (DosC$_{GGGDEF}$ form II) grew in a mixture of protein solution and precipitant solution consisting of 0.1 M imidazole pH 8.0, 0.2 M calcium acetate, 22%(w/v) PEG 1000. Prior to flash-cooling in liquid nitrogen, the crystals were briefly incubated in a solution consisting of the precipitant solution with ethylene glycol added to a final concentration of 25%(v/v). Very thin needle-shaped crystals (DosC$_{GGGDEF}$ form III) growing in clusters were obtained by mixing protein solution (6.8 mg ml^{-1}) and reservoir solution consisting of 0.2 M potassium sodium tartrate, 22%(w/v) PEG 3350. Individual crystals were briefly rinsed in reservoir solution supplemented with 25%(v/v) ethylene glycol before flash-cooling in liquid nitrogen. Clusters of rod-shaped crystals (DosC$_{GGGDEF}$ form IV) were obtained by mixing protein solution (6.5 mg ml^{-1}) and precipitant solution consisting of 0.1 M Tris–HCl pH 8.5, 0.2 M lithium sulfate, 21%(w/v) PEG 3350, 0.3 M Gly-Gly-Gly. For cryoprotection, the crystals were briefly washed in a solution consisting of 0.1 M Tris–HCl pH 8.5, 0.2 M lithium sulfate, 21%(w/v) PEG 3350, 25%(v/v) ethylene glycol and were subsequently flash-cooled in liquid nitrogen.

2.4. Data collection, structure determination and refinement

Single-crystal X-ray diffraction data were collected at 100 K on the X10SA beamline at the Swiss Light Source, Paul

Table 1
Data-collection, phasing and refinement statistics for the DosC$_{Globin}$ and DosC$_{MID}$ structures.

Values in parentheses are for the highest resolution shell.

	DosC$_{Globin}$			DosC$_{MID}$			
	Native–Fe	Iron(III) (form I)	Iron(II) (form II)	Native	KI derivative	Form I	Form II
Data collection							
Space group	$P6_122$	$P6_122$	$P1$	$P12_11$	$P12_11$	$P12_11$	$P2_12_12_1$
Unit-cell parameters							
a (Å)	98.7	97.4	48.0	49.0	49.1	49.2	41.9
b (Å)	98.7	97.4	58.3	43.4	43.6	43.6	51.3
c (Å)	151.6	152.2	59.4	52.3	52.7	52.4	92.7
α (°)	90.00	90.00	79.52	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	75.69	100.63	100.16	100.94	90.00
γ (°)	120.00	120.00	75.29	90.00	90.00	90.00	90.00
Radiation source	Rigaku Micro-Max-007 HF	PXII-X10SA, SLS	PXII-X10SA, SLS	Rigaku Micro-Max-007 HF	Rigaku Micro-Max-007 HF	PXII-X10SA, SLS	PXII-X10SA, SLS
Wavelength (Å)	1.5418	0.97627	1.73450	1.5418	1.5418	0.99999	0.99999
Temperature (K)	100	100	100	100	100	100	100
Resolution range (Å)	50–2.5 (2.6–2.5)	50–2.0 (2.1–2.0)	50–2.4 (2.5–2.4)	50–2.0 (2.1–2.0)	50–2.6 (2.7–2.6)	50–1.5 (1.6–1.5)	50–1.9 (2.0–1.9)
No. of observed reflections	325519 (36244)	314863 (43265)	226090 (22345)	313211 (23704)	364594 (31544)	227848 (38436)	204428 (30041)
No. of unique reflections	28408 (3191)	29494 (3920)	22142 (2367)	14023 (1226)	13139 (1358)	35038 (6119)	16326 (2282)
Multiplicity	11.5 (11.4)	10.7 (11.0)	10.2 (9.4)	22.3 (19.3)	27.7 (23.2)	6.5 (6.3)	12.5 (13.2)
Completeness (%)	99.2 (99.9)	99.9 (99.7)	94.6 (88.3)	94.5 (61.3)	99.4 (97.3)	99.7 (99.8)	100.0 (100.0)
R_{merge} (%)	9.4 (47.5)	5.0 (68.4)	13.0 (75.2)	6.9 (17.8)	12.5 (37.8)	5.9 (38.0)	6.2 (62.4)
$\langle I/\sigma(I)\rangle$	25.20 (6.25)	26.44 (3.52)	12.36 (2.64)	39.63 (19.45)	28.53 (12.50)	18.57 (5.37)	25.22 (4.48)
CC$_{1/2}$† (%)	99.9 (96.4)	100.0 (90.9)	99.7 (85.0)	99.9 (99.6)	99.9 (98.4)	99.9 (94.4)	100.0 (93.6)
Phasing							
Figure of merit (acentric; before/after DM)	0.280/0.870				0.224/0.893		
Isomorphous phasing power (acentric/centric)					1.101/1.023		
Anomalous phasing power (acentric)	0.628				0.861		
Isomorphous R_{cullis} (acentric/centric)					0.792/0.820		
Anomalous R_{cullis} (acentric)	0.915				0.863		
Refinement							
No. of reflections		29493	22138			35037	16326
No. of reflections in test set		1475	1107			1752	817
Resolution range (Å)		48.71–2.00	45.34–2.40			48.33–1.50	46.33–1.90
No. of non-H atoms							
Protein		2396	4858			1943	1867
Ligand/ion		86	172			9	0
Water		137	68			255	62
Total		2619	5098			2207	1929
R (%)		22.21	21.73			14.74	23.68
R_{free} (%)		25.74	25.05			18.86	28.27
R.m.s. deviations from ideal							
Bond lengths (Å)		0.003	0.002			0.007	0.004
Angles (°)		0.856	0.698			0.978	0.723
B factors (Å^2)							
Protein		45.14	43.79			20.66	36.48
Ligand/ion		46.29	39.43			20.84	0
Water		45.63	38.48			31.20	39.51
Average		45.21	43.57			21.88	36.58
Ramachandran statistics (%)							
Favoured regions		99.32	98.81			98.74	99.10
Allowed regions		0.68	1.19			1.26	0.90
Disallowed regions		0	0			0	0

† As implemented in *XDS* (Kabsch, 2010).

Scherrer Institute, Villigen, Switzerland. For phasing of the DosC$_{Globin}$ and DosC$_{MID}$ structures, data were collected using an in-house MicroMax-007 HF rotating-anode X-ray generator (Rigaku) operated at 40 kV and 30 mA with Osmic VariMax HF mirrors with the slit sizes set to 0.7/0.5 mm. Crystals were maintained at 100 K using a 700 Series Cryostream (Oxford Cryosystem) and diffraction images were recorded using a MAR345 image plate. All data were processed with *XDS* (Kabsch, 2010).

The DosC$_{Globin}$ (residues 8–170) structure in the iron(III) form was determined by iron single-wavelength anomalous dispersion (Fe-SAD) phasing using data collected at the Cu $K\alpha$ wavelength. Experimental phases were calculated using *autoSHARP* (Vonrhein *et al.*, 2007) with six heavy-atom

Table 2
Data-collection and refinement statistics for DosC$_{GGDEF}$ structures.

Values in parentheses are for the highest resolution shell.

	Apo form (form I)	GTPαS-bound (form II)	Form III	Form IV
Data collection				
Space group	$P4_1$	$P12_11$	$P2_12_12_1$	$P3_121$
Unit-cell parameters				
a (Å)	70.8	28.5	53.8	130.1
b (Å)	70.8	52.5	58.9	130.1
c (Å)	35.0	50.9	113.6	59.1
α (°)	90.00	90.00	90.00	90.00
β (°)	90.00	106.08	90.00	90.00
γ (°)	90.00	90.00	90.00	120.00
Radiation source	PXII-X10SA, SLS	PXII-X10SA, SLS	PXII-X10SA, SLS	PXII-X10SA, SLS
Wavelength (Å)	0.91162	0.97627	0.97627	1.00002
Temperature (K)	100	100	100	100
Resolution range (Å)	50.0–1.2 (1.3–1.2)	50.0–1.15 (1.2–1.15)	50.0–2.2 (2.3–2.2)	50.0–3.3 (3.4–3.3)
No. of observed reflections	698375 (138857)	145322 (9050)	107025 (11758)	86577 (7313)
No. of unique reflections	54503 (11517)	48048 (4460)	18580 (2201)	8829 (745)
Multiplicity	12.8 (12.1)	3.0 (2.0)	5.8 (5.3)	9.8 (9.8)
Completeness (%)	100.0 (99.8)	93.6 (72.8)	97.9 (95.3)	99.3 (99.6)
R_{merge} (%)	6.7 (37.9)	4.2 (41.6)	9.9 (78.3)	10.2 (80.8)
$\langle I/\sigma(I)\rangle$	20.99 (6.28)	13.07 (2.14)	11.36 (2.54)	21.33 (3.71)
$CC_{1/2}$† (%)	99.9 (96.5)	99.8 (82.7)	99.8 (79.3)	99.9 (91.6)
Refinement				
No. of reflections	54503	48043	18580	8829
No. of reflections in test set	2726	2403	929	442
Resolution range (Å)	35.42–1.20	35.81–1.15	48.61–2.20	43.73–3.30
No. of non-H atoms				
Protein	1403	1377	2660	2622
Ligand/ion	21	42	0	10
Water	185	172	78	0
Total	1609	1591	2738	2632
R (%)	12.31	13.80	20.76	20.00
R_{free} (%)	14.55	16.48	25.17	25.16
R.m.s. deviations from ideal				
Bond lengths (Å)	0.008	0.006	0.008	0.004
Angles (°)	1.237	1.177	1.132	0.737
B factors (Å^2)				
Protein	17.27	17.65	46.14	82.98
Ligand/ion	18.49	19.97	0	85.58
Water	32.19	31.97	43.76	0
Average	19.03	19.26	46.07	82.99
Ramachandran statistics (%)				
Favoured regions	98.88	98.84	98.46	98.12
Allowed regions	1.12	1.16	1.54	1.88
Disallowed regions	0	0	0	0

† As implemented in *XDS* (Kabsch, 2010).

positions identified by *SHELXD* (Sheldrick, 2010). The quality of the electron-density map obtained after solvent flattering enabled the autobuilding of an initial model using *ARP/wARP* (Langer *et al.*, 2008). This model was further optimized by iterative cycles of manual building using the graphics program *Coot* (Emsley *et al.*, 2010) and refinement as implemented in *phenix.refine* (Afonine *et al.*, 2012) and was used to determine phases for the other native data sets [and DosC$_{Globin}$ residues 1–155 in the iron(II) form] by molecular replacement (MR) using *Phaser* (McCoy *et al.*, 2007). The resulting models were further refined by alternating rounds of manual rebuilding and refinement as described above.

Experimental phases for the DosC$_{MID}$ (residues 173–298) structure were obtained by the single isomorphous replacement with anomalous signal (SIRAS) method using data collected with Cu $K\alpha$ radiation from native and potassium iodide-derivatized crystals (form I). *SHELXD* found 42 sites for anomalous scatterers and phase information was obtained by *autoSHARP* using these sites. The phases allowed the calculation of a readily interpretable electron-density map after density modification, which enabled automated building of the model with *ARP/wARP*. The initial model was used to phase all subsequently determined structures by molecular replacement using *Phaser*. All final models were optimized by alternating cycles of manual rebuilding and refinement performed using *Coot* and *phenix.refine*.

The DosC$_{GGDEF}$ (residues 297–460) structure in the apo form was determined by molecular replacement using *Phaser* and coordinates from PDB entry 3ign as a search model. The molecular-replacement solution was used to automatically build an initial model using *ARP/wARP*. This initial model was used to determine all subsequent structures by molecular replacement. Model rebuilding and refinement of all structures was again carried out using *Coot* and *phenix.refine*, respectively. Noncrystallographic symmetry (NCS) restraints were included during refinement where appropriate.

The simulated-annealing OMIT maps for ligands were generated using the composite OMIT tool in *PHENIX* (Adams *et al.*, 2010). Data-collection, phasing and refinement statistics are summarized in Tables 1 and 2. The model quality was validated using the validation methods implemented in *PHENIX*.

The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 4zva (DosC$_{Globin}$ form I), 4zvb (DosC$_{Globin}$ form II), 4zvc (DosC$_{MID}$ form I), 4zvd (DosC$_{MID}$ form II), 4zve (DosC$_{GGGDEF}$ form I), 4zvf (DosC$_{GGGDEF}$ form II), 4zvg (DosC$_{GGGDEF}$ form III) and 4zvh (DosC$_{GGGDEF}$ form IV).

2.5. Analytical ultracentrifugation and multi-angle light scattering

The protein size distribution in solution was assessed by sedimentation-velocity experiments using a Beckman

ProteomeLab XL-I analytical ultracentrifuge equipped with an An-60 Ti rotor using absorbance optics. The measurements were performed at 50 000 rev min^{-1} (201 600g) and 293 K in a two-sector cell with a 1.2 cm optical path length at a protein concentration of 10 μM in 20 mM CHES pH 9.5, 20 mM NaCl, 2.5 mM EDTA, 2.5 mM DTE, 25 mM arginine, 25 mM glutamic acid, 10%(v/v) glycerol. Radial absorbance scans were obtained at 411 nm using a radial step size of 0.003 cm in a continuous scanning mode. The data were evaluated using *SEDFIT* v.14.1 (Schuck, 2010) and the sedimentation coefficient was reduced to that in water at 293 K ($s_{20,w}$). The frictional ratio f/f_{min} was calculated as described previously (Erickson, 2009). These results were used to calculate the molecular weight.

The molar mass in solution was determined by multi-angle light scattering coupled with size-exclusion chromatography (SEC-MALS). Gel filtration was carried out using a Waters liquid-chromatography system on a Superdex 200 10/300 GL column (GE Healthcare) connected to a refractive-index detector (Waters), a photodiode-array detector (Waters) and a MALS detector (DAWN HELEOS, Wyatt Technology). Isocratic elution was performed at 293 K at a flow rate of 0.3 ml min^{-1} using a mobile phase consisting of 20 mM CHES pH 9.5, 20 mM NaCl, 2.5 mM EDTA, 2.5 mM DTE, 25 mM arginine, 25 mM glutamic acid, 10%(v/v) glycerol and 300 µl of 67 μM protein solution. The molar mass was calculated from the light-scattering data using the *ASTRA* software (Wyatt Technology).

2.6. Enzymatic activity assay

Real-time activity measurements by circular-dichroism (CD) spectroscopy were performed essentially as described previously (Stelitano *et al.*, 2013). In brief, c-di-GMP can be detected and quantified based on the specific CD signal of the intercalated c-di-GMP dimer at 282 nm. The method is based on the selective ability of manganese ions to enhance the formation of an intercalated c-di-GMP dimer in solution and the linear dependence of the CD signal on the c-di-GMP concentration. The condensation reaction was initiated by the addition of 70 μM GTP to a 10 μM solution of DosC variant proteins in assay buffer [20 mM Tris–HCl pH 8.0, 100 mM NaCl, 10 mM MgCl$_2$, 2.5 mM MnCl$_2$, 5%(v/v) glycerol]. A Jasco J-810 spectropolarimeter was used to monitor the time course of the reaction by following the CD signal at 282 nm using a 1 cm quartz cuvette. A sample in the iron(II) state was prepared in a glove box by reducing the protein with 10 mM sodium dithionite; all solutions used were degassed. Next, the buffer was exchanged to 25 mM Tris–HCl pH 8.0, 100 mM NaCl, 5%(v/v) glycerol to remove the reducing agent. The shift of the Soret band from 394 to 431 nm was confirmed using a NanoDrop ND-1000 spectrophotometer. An enzymatic reaction under anaerobic conditions was prepared in a glove box in a 1 cm sealable quartz cuvette (Hellma, Germany). After measuring the CD trace prior to substrate addition, the cuvette was again transferred to the glove box, where the reaction was started by the addition of GTP. In this

case, the progress of the reaction could not be monitored for the first 2 min after substrate addition. All measurements were performed in duplicate at 293 K. For DosC feedback-inhibition measurements, a 10 μM protein sample was pre-incubated with 40 μM c-di-GMP prior to starting the enzymatic reaction by the addition of GTP to a final concentration of 70 μM.

2.7. UV–vis spectroscopy

UV–vis spectroscopic measurements were performed using a Jasco V-650 spectrophotometer with a 1 cm quartz cuvette. All measurements were performed at 293 K. The protein was diluted to 10 μM in 20 mM Tris–HCl pH 8.0, 100 mM NaCl, 5%(v/v) glycerol. The spectrum of the iron(II)–O$_2$ form was recorded using this air-saturated buffer additionally containing 6 mM DTE. The spectrum of the iron(II) form was obtained after reduction with 1 mM sodium dithionite.

2.8. Solubility optimization, limited proteolysis and mass spectrometry

Differential scanning fluorimetry (DSF) was used to identify buffer/salt conditions and additives that stabilize purified proteins. The measurements were performed according to established protocols (Fedorov *et al.*, 2012) using in-house and commercial screens. In brief, purified protein was diluted to 50–100 µg ml^{-1} and SYPRO Orange was added to 5× the final concentration. 10 µl protein solution was mixed with 10 µl screen solution in an optical 96-well PCR plate. Fluorescence data were collected on an Mx3005p qPCR system (Agilent Technologies) with excitation and emission wavelengths of 492 and 610 nm, respectively. The temperature was increased from 298 to 368 K with 30 s incubation per kelvin. The mid-point temperature of the melting curve (T_m) was determined using the *Excel*-based tools available from ftp://ftp.sgc.ox.ac.uk/pub/biophysics. Fitting of the data to the Boltzmann equation was performed using *GraphPad Prism* 5 (GraphPad Software). The solubility of DosC$_{full}$ was screened similarly to previously described approaches (Izaac *et al.*, 2006; Jancarik *et al.*, 2004) and the potential of stabilizing additives during cell lysis was additionally explored (Leibly *et al.*, 2012).

The identification of protease-resistant domains suitable for crystallization was performed by limited proteolysis using trypsin and chymotrypsin. Briefly, protein solution at 5 mg ml^{-1} was incubated at 310 K for 60 min with various concentrations of a protease (100–0.05 µg ml^{-1}). The reaction was terminated and proteolysis products were separated by 17% SDS–PAGE. The identity of the stable proteolytic fragments was verified by peptide mass fingerprinting with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) on an Axima TOF^2 Performance mass spectrometer (Shimadzu Biotech) after tryptic digests of gel slices. N-terminal protein sequence information was obtained by MALDI-TOF MS using the in-source decay (ISD) mode of the instrument. Purified protein samples were analysed under denaturing conditions by electrospray

ionization time-of-flight mass spectrometry (ESI-TOF MS) on a maXis spectrometer (Bruker Daltonics).

2.9. Bioinformatics

The protein sequences of GCS-DGCs from various bacterial species were aligned using *ClustalW*2 (Larkin *et al.*, 2007) and were visualized with *ESPript*3 (Robert & Gouet, 2014). Secondary-structure predictions were made using *JPred* (Cole *et al.*, 2008) and *PSIPRED* (Buchan *et al.*, 2013). A three-dimensional model of the full-length dimeric DosC (residues 6–460) was constructed using the crystal structures of the individual domains as templates. The crystal structures of DosC$_{Globin}$ and DosC$_{MID}$ both represent dimeric arrangements. A protein–protein docking approach was used to predict the relative orientations of these two individual modules. For this purpose, the *ClusPro* 2.0 server (Kozakov *et al.*, 2013), implementing *PIPER* (an FFT-based docking program), was used. DosC$_{Globin}$ was set as the receptor and DosC$_{MID}$ was the docked ligand. The resulting models were manually assessed and a single model preserving the twofold symmetry was selected and its symmetry was further optimized. The genuine active DosC$_{GGDEF}$ dimer is unavailable; to circumvent this, two monomers of DosC$_{GGDEF}$ were superposed with the GGDEF domains of DgcZ (YdeH) from *E. coli* (PDB entry 4h54; Zähringer *et al.*, 2013) to recreate its dimeric arrangement. Next, one of the GGDEF monomers was shifted to position both substrates at reaction distance. The model of the catalytically competent GGDEF dimer was manually positioned on the globin–MID module similarly as observed in DgcZ. The conformation of the loop region (residues 296–305) was modelled based on the DgcZ structure. Finally, the missing residues in the linker regions (residues 155–176 and 291–295) were modelled using the *ModLoop* server (Fiser & Sali, 2003), relying on the *MODELLER* loop-modelling routine. The secondary structure of the models was analysed with *DSSP* (Kabsch & Sander, 1983). *SSM* was used for structural alignments (Krissinel & Henrick, 2004). The packing of coiled-coil regions was evaluated with *SOCKET* (Walshaw & Woolfson, 2001). *NACCESS* was used to calculate the solvent-accessible surface area (Hubbard & Thornton, 1993). Macromolecular interfaces were calculated with *PISA* (Krissinel & Henrick, 2007). Cavities were analysed using *DoGSiteScorer* (Volkamer *et al.*, 2012). Protein–ligand interactions were evaluated with *LigPlot*+ (Laskowski & Swindells, 2011). Poisson–Boltzmann electrostatics were calculated using *PDB*2*PQR* (Dolinsky *et al.*, 2004) and *APBS* (Baker *et al.*, 2001). The protein sequences homologous to DosC$_{MID}$ (residues 173–298) and with full coverage were identified by a *BLAST* search (Altschul *et al.*, 1997) and redundant hits were removed using *EMBOSS SkipRedundant* (Rice *et al.*, 2000); 78 sequences were retained for the analysis. Mapping of sequence conservation onto the protein structure was performed by *ConSurf* (Celniker *et al.*, 2013). The sequence logo was generated using *WebLogo*3 (Crooks *et al.*, 2004). Crystal contacts were evaluated using *NCONT*, which is part of the *CCP*4 suite (Winn *et al.*, 2011). Normal modes of DosC globin

domain in the iron(III) form were calculated using the *elNémo* web server (Suhre & Sanejouand, 2004). The structural figures were generated with *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Overview

DosC from *E. coli* was predicted to contain an N-terminal globin-based sensory domain and a C-terminal GGDEF domain. The region between these two domains could not be assigned to any particular family, and for the purpose of this study we termed it the middle domain (abbreviated MID). Based on domain and secondary-structure predictions and using naturally occurring methionines as start codons, we designed a number of DosC variants comprising individual domains and two domains as well as full-length DosC (Fig. 1). When we attempted to overexpress and purify all variants for subsequent structural studies, we found that all three single domains (DosC$_{Globin}$, DosC$_{MID}$ and DosC$_{GGDEF}$) as well as DosC$_{Globin-MID}$ were expressed as soluble proteins and could be purified. In contrast, both DosC$_{MID-GGDEF}$ and DosC$_{full}$ were expressed mostly as inclusion bodies, displayed very limited solubility and were prone to aggregation. To overcome this issue, they were overexpressed in fusion with an N-terminal solubility-enhancing ZZ-tag (Bogomolovas *et al.*, 2009). Alternatively, we also undertook a combined approach to improve the solubility and aggregation behaviour of full-length DosC, which proved to be partially successful, as discussed later. Importantly, we succeeded in crystallizing and solving the structures of all three individual domains of full-length DosC. Despite numerous attempts, we were unable to obtain crystals of any of the dual modules, *i.e.* globin–MID and MID–GGDEF, or the full-length protein.

3.2. Sensory domain

3.2.1. Globin-domain structure determination. Initially, we crystallized the isolated DosC$_{Globin}$ sensory domain (residues 1–155), but owing to low sequence similarity no suitable model for molecular replacement was available. Unfortunately, experimental phasing approaches with native iron and other heavy-atom derivatives were also unsuccessful. After designing a new DosC$_{Globin}$ construct (residues 8–170) based on limited proteolysis results of the DosC$_{Globin-MID}$ dual-domain protein, we obtained a new crystal form in the hexagonal space group $P6_122$, which in turn allowed us to determine the structure using the anomalous signal of the haem Fe atom by the single-wavelength anomalous dispersion (SAD) method (Table 1). The structure of the shorter construct in the triclinic space group $P1$ was subsequently solved by molecular replacement (Table 1).

3.2.2. Iron(III) form structure. The crystal structure of the sensory domain of DosC (residues 8–170) in the iron(III) form was determined at 2.0 Å resolution with two molecules per asymmetric unit (Table 1). The final model consists of chains *A* and *B*, each containing residues 8–154 (plus two extra N-terminal vector-derived residues). No electron density was

observed for residues 155–170, suggesting they are unordered. Despite low sequence similarity, the $DosC_{Globin}$ monomer displays the typical predominantly (83%) α-helical globin fold. There are eight helices in each monomer; they are named Z, A, B, C, E, F, G and H according to the classical globin nomenclature (Fig. 2a). The short N-terminal part (Val81–Asp83) of the F helix has a 3_{10}-helix conformation. Chains A and B of $DosC_{Globin}$ are virtually identical to each other (Table 3). Closely related globin structures deposited in the PDB are HemAT from *Bacillus subtilis* (PDB entry 1or6; Zhang & Phillips, 2003) and *Gs*GCS from *Geobacter sulfurreducens* (PDB entry 2w31; Pesce *et al.*, 2009). A comparison of the DosC globin monomer and dimer with the structures of

HemAT and *Gs*GCS shows that the alignment of monomers results in lower r.m.s.d. values than obtained for the dimeric arrangements (Table 3). The two $DosC_{Globin}$ monomers assemble in a homodimer with an interface surface area of about 1300 Å^2, which is lower than the buried surface area in the HemAT GCS dimer (1800 Å^2). The dimerization interface is formed exclusively by residues belonging to helices G and H (Fig. 2a), unlike in HemAT, where the N-terminal Z helices provide some additional contacts. In $DosC_{Globin}$ the Z helix is arranged almost parallel to the H helix. The classical knobs-into-holes packing in the coiled-coil assembly of the H-helices involves Ala130, Lys133, Leu134, Tyr137, His138 and Ile141. The symmetrical four-helix bundle builds a dimer core; however, the twofold symmetry of the observed dimeric arrangement is imperfect and the rotation angle between the subunits deviates by 2.0° from 180°. The unliganded HemAT GCS structure (PDB entry 1or6) displays a higher degree of symmetry, deviating by only 0.3° from 180°. This difference may be especially important because it was previously proposed for HemAT that disruption of this symmetry may play a crucial role in signal transduction (Zhang & Phillips, 2003). Gel-filtration experiments revealed that $DosC_{Globin}$ exists as a dimer in solution which, together with the crystal structure, confirms that the homodimer is a functional quaternary assembly. In contrast, despite a larger dimerization interface, the HemAT sensory domain was found to be monomeric at low protein concentrations during isolation (Zhang & Phillips, 2003). Our *in vitro* activity studies show that DosC in the iron(III) form resembles the iron(II)–O_2 form with regard to DGC activity. Based on this, we are convinced that these two forms also share some similarities in structure.

3.2.3. $DosC_{Globin}$ haem-binding pocket. The haem is localized between helices E and F, which form the 'roof' and 'floor' of the hydrophobic haem-binding pocket, with helices B, C and G completing it from the sides and behind, respectively. The calculated volume of the resulting cavity is about 685 Å^3 on average, which is less than that for HemAT, which has a cavity volume of about 995 Å^3 (Zhang & Phillips, 2003). The cavity depth is also reduced, at about 17 and 20 Å for the $DosC_{Globin}$ and HemAT sensory domains, respectively. The strictly conserved histidine residue His98, which was shown to be essential for haem binding (Kitanishi *et al.*, 2010), is the proximal haem axial ligand. The His98 $N^{\varepsilon 2}$ atom coordinates the haem iron, giving a pentacoordinated complex. As
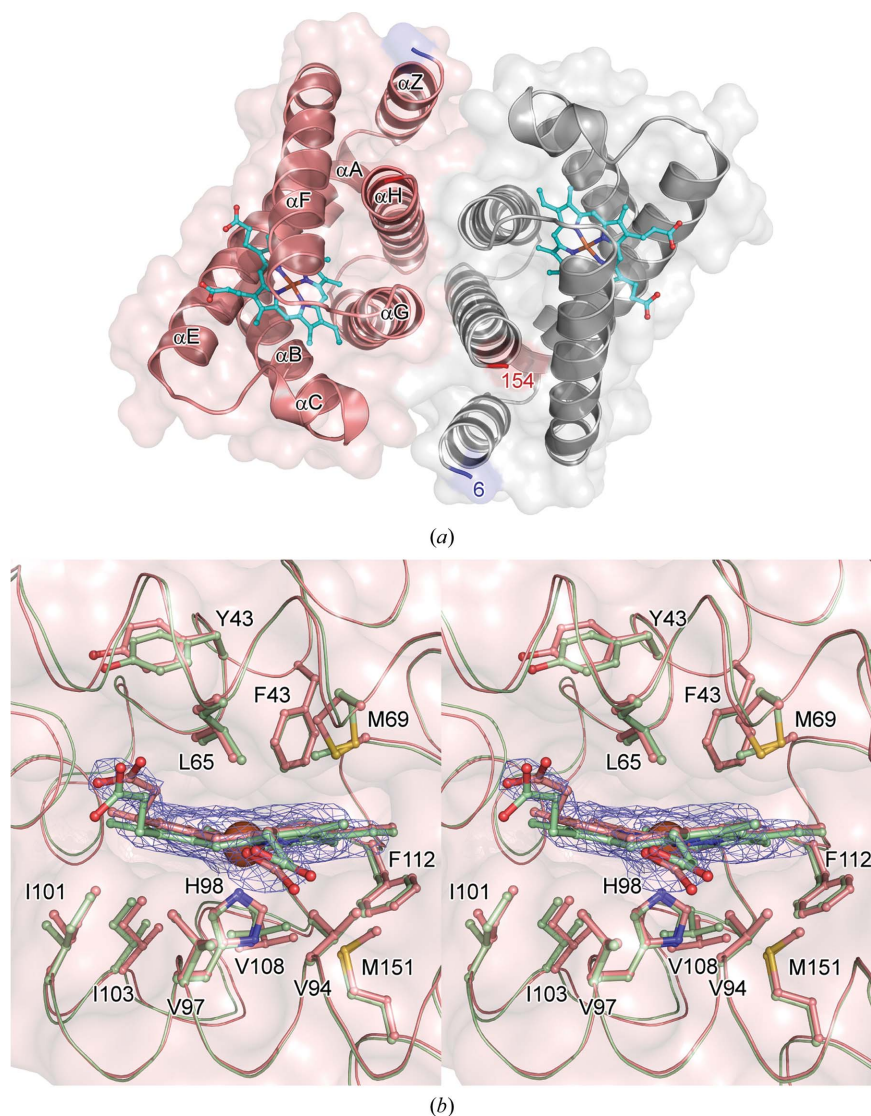
Figure 2
Globin-domain crystal structure. (a) Overview of the $DosC_{Globin}$ dimer as viewed along the twofold symmetry axis. Protomer A, with secondary-structure elements labelled, is coloured red and protomer B is shown in grey. The N- and C-termini of both protomers are labelled and are coloured blue and red, respectively. (b) Close-up stereoview of the $DosC_{Globin}$ haem-binding pocket showing comparison between iron(III) (red) and iron(II) (green) structures. Side chains of the residues of the proximal and distal haem sides are overlaid and labelled. The simulated-annealing $2mF_{obs} - DF_{calc}$ OMIT electron-density map of the haem ligand in the iron(III) structure is contoured at 1.0σ.

Table 3
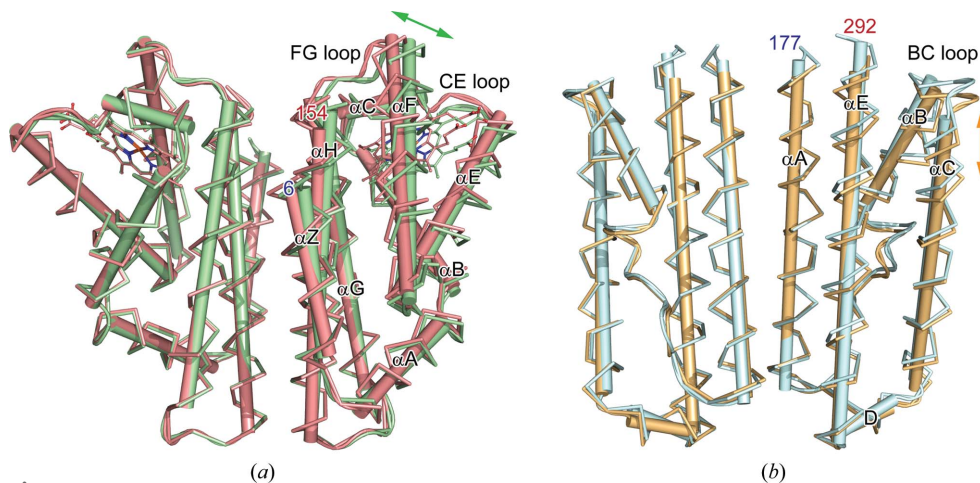Comparison of backbone r.m.s.d. values for individual domains of DosC and their related structures.

Values were obtained with *SSM* (Krissinel & Henrick, 2004).

Structure 1	Structure 2	R.m.s.d. for C^α atoms (Å)	No. of aligned residues
DosC$_{Globin}$, iron(III), A	DosC$_{Globin}$, iron(III), B	0.34	149
DosC$_{Globin}$, iron(III), A	HemAT, 1or6, A	1.66	136
DosC$_{Globin}$, iron(III), AB	HemAT, 1or6, AB	2.79	266
DosC$_{Globin}$, iron(III), A	GsGCS, 2w31, A	1.95	138
DosC$_{Globin}$, iron(III), AB	GsGCS, 2w31, AB	3.03	242
DosC$_{Globin}$, iron(II), AB	DosC$_{Globin}$, iron(II), CD	0.21	299
DosC$_{MID}$, form I, A	DosC$_{MID}$, form I, B	0.21	115
DosC$_{MID}$, form II, A	DosC$_{MID}$, form II, B	0.70	113
DosC$_{MID}$, form I, AB	DosC$_{MID}$, form II, AB	0.73	226
DosC$_{MID}$, form I, A	DgcZ CZB, 3t9o, A	2.78	94
DosC$_{MID}$, form I, AB	DgcZ CZB, 3t9o, AB	3.78	129
DosC$_{GGDEF}$, apo, A	Maqu_2607 GGDEF, 3ign, A	1.57	155
DosC$_{GGDEF}$, apo, A	DgcZ GGDEF, 3tvk, A	1.53	148
DosC$_{GGDEF}$, apo, A	XCC4471 GGDEF, 3qyy, A	1.61	145
DosC$_{GGDEF}$, apo, A	YfiN GGDEF, 4iob, A	1.30	149
DosC$_{GGDEF}$, apo, A	DosC$_{GGDEF}$, GTPαS, A	0.82	161

expected, there are no water molecules present on either the proximal or the distal side of the haem. However, there is a single water molecule at hydrogen-bonding distance close to the proximal His98. Owing to the confined haem pocket, the calculated average solvent-accessible surface area of the haem cofactor is lower in DosC$_{Globin}$ than in HemAT and *Gs*GCS, at 103, 107 and 123 Å^2, respectively. The porphyrin ring is slightly bent with respect to the C_2'' axis between the A and D pyrrole groups and deviates by about 20° from planarity. Furthermore, the haem A- and D-propionates are solvent-exposed and hydrogen-bonded to solvent water molecules. At the proximal side of the haem several hydrophobic residues (Val94, Val97, Ile101, Ile103, Val108, Phe112 and Met151) are arranged in a ring with His98 in the centre (Fig. 2b). The architecture of the haem pocket at the distal site is dominated by Phe42, Leu65, Met69 and Tyr43, with the latter pointing outwards in the unliganded iron(III) form (Fig. 2b). The side chain of Leu65 restricts access to the haem iron, while Tyr43 is expected to swing in and stabilize O_2 in the complexed form, similarly to as in the HemAT cyanide complex (PDB entry 1or4; Zhang & Phillips, 2003). A contribution of Gln60 to oxygen binding (Kitanishi *et al.*, 2010) can be definitely ruled out as this residue is surface-exposed and located outside the haem pocket.

3.2.4. Iron(II) form structure. The structure of the DosC$_{Globin}$ domain (residues 1–155) in the iron(II) state was determined at a resolution of 2.4 Å (Table 1). Crystals of the iron(II) form were grown and harvested under anaerobic conditions, but the iron(II) state was not analysed spectroscopically *in crystallo*. The final model contains four monomers arranged in two functional dimers. Similarly to the hexagonal crystal form, the electron-density map allowed us to trace residues 5–154 for chains A and C and residues 6–154 for chains B and D. The DosC$_{Globin}$ dimers (AB and CD) are practically identical (Table 3). The only region showing noticeable differences between individual monomers is the loop connecting helices C and E. The largest differences are seen in chain C owing to the distinct local environment. Similarly to the iron(III) form, the interface surface area of each homodimer is about 1300 Å^2. The dimeric arrangement of the iron(II) form is even less symmetric, with the rotation angle linking the two subunits deviating by 2.5° from 180° on average. In the iron(II) DosC$_{Globin}$ structure the haem is five-coordinated, with His98 serving as the proximal axial ligand (Fig. 2b) as in the iron(III) form. The haem environment also resembles the iron(III) state, with the exception of the side chain of Met69, which adopts a different rotamer (Fig. 2b). This results in average interatomic Fe–S$^\delta$ and Fe–C$^\varepsilon$ distances of about 6.6 and 5.0 Å [5.5 and 7.1 Å in the iron(III) form], respectively. The side chain of Met69 is in the *gauche* conformation along the C$^\beta$–C$^\gamma$ bond, with an average C$^\alpha$–C$^\beta$–C$^\gamma$–S$^\delta$ torsion angle of 62° [−64° in the iron(III) form]. Additionally, the side chain of Leu65 is located slightly closer to the porphyrin ring, with an average Fe–C$^{\delta1}$ distance of 3.6 Å, instead of 3.8 Å as found for the iron(III) form. Superposition on the porphyrin ring atoms shows that the haem-plane conformations are very similar for the iron(III) and iron(II) structures, with only some differences in the conformations of the propionate groups.

3.2.5. Different arrangement of iron(III) and iron(II) DosC$_{Globin}$. The most striking differences, however, are
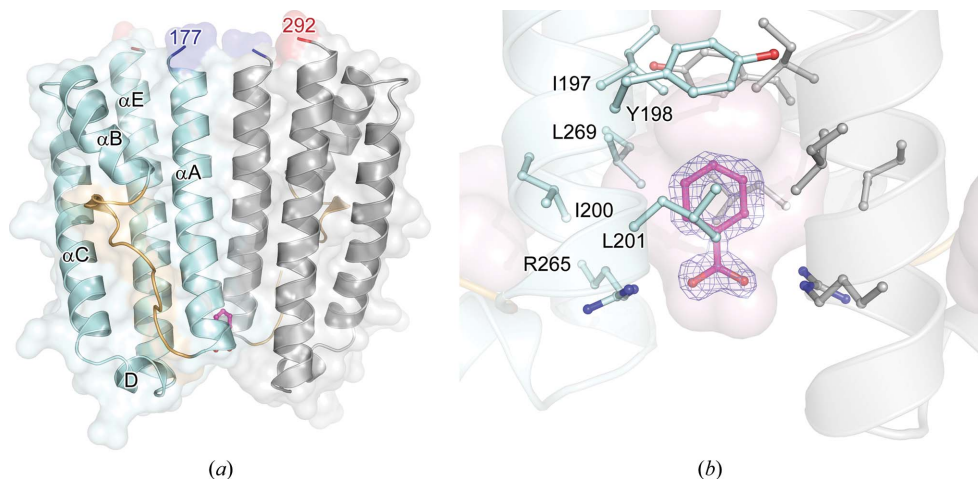

Figure 3
Differences in the arrangement of the DosC$_{Globin}$ and DosC$_{MID}$ dimers. The crystallographic dimers were aligned according to one chain to depict changes in the relative orientation between protomers. The DosC$_{Globin}$ crystal structures of the iron(III) (coloured red) and iron(II) (coloured green) forms are overlaid in (a); the DosC$_{MID}$ structures of the monoclinic (coloured cyan) and orthorhombic (coloured orange) forms are compared in (b). The N- and C-termini and secondary-structure elements are labelled. The observed shifts are indicated by arrows.

observed when we align DosC$_{Globin}$ dimers in the iron(III) and iron(II) forms by superimposing the single subunits in order to emphasize structural changes. This comparison shows that several elements located in the 'corner region' are displaced by about 1.0–2.0 Å towards the outside of the dimer in the iron(II) form (Fig. 3a). The F helix is shifted most strongly in its N-terminal part, but considerable changes are also seen for helices C and E as well as loops CE and FG (Fig. 3a), while helices Z, A, B and H do not change so much. The fact that there is nearly no difference in the distance (∼0.5 Å) between the C-termini strongly suggests that the 'upper' corners (the F helix, C helix and FG loop) of the dimer participate in the signal transduction. The extent of rearrangement between the iron(III) and iron(II) structures of DosC$_{Globin}$ is much larger than that described for the unliganded and liganded structures of HemAT (Zhang & Phillips, 2003). We investigated the crystal packing of the two forms to test whether the constraints imposed by the crystal lattice may account for the observed differences. For this purpose, we analysed the differences between the iron(III) and iron(II) structures in the context of intermolecular contacts with symmetry-related molecules. As shown in Supplementary Fig. S2, displacement of the 'corner regions' (helix C and loops CE and FG) is observed in both subunits. The same qualitative changes are seen in both monomers, although they differ noticeably in the intermolecular contacts formed within the crystal (Supplementary Fig. S3). Moreover, not only do the single monomers form different contacts with symmetry mates, but the dimers of the iron(III) and iron(II) forms also differ in this respect (Supplementary Fig. S3). The same structural changes are observed when the two dimers of the iron(II) form present in the asymmetric unit are compared with the iron(III) structure. We cannot completely rule out an effect of crystal packing on the structures, but we are convinced that crystal contacts do not have a significant influence on the structural elements of the 'corner regions' exhibiting the most pronounced changes. In addition to analysing possible crystal contact effects, we also performed normal-mode analysis of the DosC globin domain. We found that the change in conformation between the iron(III) and iron(II) structures is described well by the first nontrivial lowest frequency normal mode (mode 7; Supplementary Fig. S4a). The main regions involved in the collective motions identified by normal-mode analysis are predominantly localized within the 'upper corner regions', as are the conformational changes observed between iron(III) and iron(II) structures (Supplementary Fig. S4b). We propose that this rearrangement of the globin homodimer is part of the signalling mechanism leading to GGDEF activation.

3.3. Middle domain

3.3.1. Defining the middle domain and its structure determination. The central domain of DosC does not belong to a classified protein family. Freitas *et al.* (2003) termed the domain between the N-terminal globin sensor and the C-terminal GGDEF domain 'ERERQR', reflecting a conserved sequence motif. In DosC this motif is degenerate and we named this part of the protein the 'middle domain'. The DosC$_{MID}$ protein (Fig. 1) was expressed and purified, but displayed limited solubility in the initial buffer system. We used DSF to optimize the buffer composition. This approach, also called a thermal shift assay, measures the increase or decrease in protein melting temperature (T_m) under a variety of buffer conditions or upon ligand binding. Sodium acetate pH 4.6 instead of Tris pH 8.0 increased the T_m by 6 K and shifted it to 341 K (Supplementary Fig. S5a). Since the initial DosC$_{MID}$ (residues 148–300) crystallization attempts utilizing a protein sample in the optimized buffer were unsuccessful, we used limited proteolysis to identify a stable domain that was amenable to crystallization. To this end, we tested different proteases and discovered that chymotrypsin, even at a relatively high concentration, yields a single proteolysis product (Supplementary Fig. S6), in contrast to trypsin, which resulted in nearly complete DosC$_{MID}$ degradation (Supplementary Fig. S6). The DosC$_{MID}$ chymotrypsin-stable proteolytic fragment was produced on a preparative scale and purified by anion-exchange chromatography. The protein was sequenced by ISD MALDI mass spectrometry, showing that the first amino acid is Ser173. To fully analyse the DosC$_{MID}$ stable domain boundaries, the protein was further analysed using high-

Figure 4
Crystal structure of the DosC middle domain. (a) The side view of the dimer with protomer A coloured in cyan; protomer B is shown in grey. The convoluted stretch following the A helix is highlighted in yellow; the N- and C-termini of both protomers are labelled and are coloured blue and red, respectively. The benzoate moiety bound at the dimer interface is shown in magenta. (b) Enlarged view of the bound benzoate ligand; the side chains of the residues forming the central cavity are labelled and the inner surface of the cavity is coloured pink. The simulated-annealing $2mF_{obs} - DF_{calc}$ OMIT electron-density map of the benzoate ligand is contoured at 1.0σ.

resolution ESI mass spectrometry. This resulted in a molecular weight of 14 660 Da, clearly indicating that the proteolytic $DosC_{MID}$ product contains the Ser173–Asp298 fragment with a calculated molecular weight of 14 660.6 Da. This fragment yielded well diffracting crystals in the monoclinic space group $P12_11$. Molecular replacement was not possible owing to the lack of a suitable search model. The $DosC_{MID}$ (residues 173–298) structure was determined by the SIRAS method using data from native and halide-soaked crystals grown in the presence of Gd-HPDO3A (Girard *et al.*, 2002) (Table 1). The structure of $DosC_{MID}$ from crystals grown without the gadolinium complex in the orthorhombic space group $P2_12_12_1$ was subsequently determined by molecular replacement (Table 1).

3.3.2. $DosC_{MID}$ structures. The crystal structure of the DosC middle domain was determined at a resolution of 1.5 Å with two equivalent monomers in the asymmetric unit (Table 1). The $DosC_{MID}$ model consists of chains *A* and *B*, containing residues 177–292 and 177–291, respectively. The monomers are arranged in a highly symmetric twofold homodimer (Fig. 4*a*), and in contrast to $DosC_{Globin}$ the rotation angle between the monomers deviates by only 0.17° from 180°. Chains *A* and *B* are extremely similar to each other (Table 3). The $DosC_{MID}$ monomer is almost entirely α-helical (78%), containing five helices (A–E). Four of these helices (A, B, C and E) form an antiparallel four-helix bundle (Fig. 4*a*). Helices A, C and E are approximately 38–43 Å long, with the C-terminal helix E being the longest. The N-terminal helix A is followed by a segment (residues 204–215) that adopts a convoluted, irregular conformation, yet looks exactly the same in both chains. The middle part of the relatively short helix B (residues 216–228) has a π-helix conformation (residues 221–225). The short 3_{10}-helix D situated at one end of the bundle spans residues Thr256–Asn260. The $DosC_{MID}$ dimer interface area is about 950 Å^2. It is formed in part by the E helix and to a large extent by the A helices, forming a coiled-coil structure. A number of residues from the A helix are involved in a knobs-into-holes packing: Glu183, Ile186, Ala187, Leu190, Ser191, Glu193, Ile194 and Ile197. Unexpectedly, both the N- and C-termini are located on the same side of the helical bundle about 11 and 28 Å apart, respectively. A symmetric cavity was identified at the dimer interface at the opposite end of the molecule (Fig. 4*b*). This central cavity is about 13.5 Å deep and its volume is about 275 Å^3.

Additionally, we obtained a different crystal form of $DosC_{MID}$ and solved its structure at 1.9 Å resolution by molecular replacement (Table 1). The final model of this form also consists of two chains, each spanning residues 178–290. The monomers assemble in a parallel homodimer with a slightly higher interface surface area of about 1000 Å^3. Essentially, this structure displays the same features as the structure of the monoclinic form described above. However, the almost perfect twofold symmetry of the dimeric arrangement observed in the monoclinic form is broken and the rotation angle relating the two protomers deviates by 1.39° from 180°. The volume of the central cavity is nearly the same (~264 Å^3), but the r.m.s.d. value for chains *A* and *B* is slightly higher (Table 3).

Analysis of the sequence conservation of $DosC_{MID}$ showed that elements proximal to the adjacent GGDEF domain (the N-terminal part of helix A, the C-terminal part of helix D and helix B) contain more conserved residues (Supplementary Fig. S7).

3.3.3. Unusual ligand-binding site. In the monoclinic crystal form, the cavity in the centre of the dimer unexpectedly displays a very clear positive electron density. It was modelled as a benzoate moiety with a high degree of confidence owing to the high resolution of the data (Fig. 4*b*). The O atoms of the benzoic acid carboxylic group interact with the N^ε and $N^{\eta 2}$ atoms of symmetrically arranged Arg265 side chains: one from each monomer. These side chains form an orifice, while the side chains of Ile197, Tyr198, Ile200, Leu201 and Leu269 line the hydrophobic bottom of the cavity (Fig. 4*b*). Interestingly, since the protein was not intentionally exposed to benzoic acid, it must originate either from the lysate or, more likely, it was present as an impurity in one of the chemicals used. Benzoic acid was only observed in the structures derived from crystals obtained by co-crystallization with Gd-HPDO3A, which suggests Gd-HPDO3A to be the source of the benzoic acid. Functional relevance of the bound benzoate molecule is questionable as access to the cavity in the full-length DosC is likely to be occluded by the $DosC_{Globin}$ dimer.

3.3.4. Differences in the $DosC_{MID}$ dimeric arrangements. A comparison of the $DosC_{MID}$ dimeric structures derived from two different crystal forms shows minor differences between the dimers (Table 3). However, when aligning monomers of the two structures in order to highlight quaternary-structure changes, a rotation of about 1.5° and a 1.0–2.0 Å 'downshift' of the E–D–C–B helical elements with respect to the neighbouring subunit is observed (Fig. 3*b*). Helix A, which is engaged in a coiled-coil structure, is less affected, with the most pronounced differences being observed for the C-terminal helix E and the short helix B together with helix C (Fig. 3*b*). It is unlikely that the observed change is caused by the absence of the benzoate ligand, as the cavity is hardly affected. Some of the differences may be attributed to the different crystal packing of the two forms; however, analysis of the crystal packing suggests that at least some parts showing a change in conformation are not involved in close contacts. We are convinced that the observed differences reflect intrinsic flexibility of the middle domain, which may have some functional relevance.

The $DosC_{MID}$ domain shows distant similarity (Table 3) to the zinc-binding domain [assigned to the chemoreceptor zinc-binding domain (CZB) according to Pfam classification] of DgcZ from *E. coli* (PDB entry 3t9o; Zähringer *et al.*, 2013). The proteins have a similar topology and are characterized by the presence of an irregular segment preceding the second helix (helix B) in their four-helix bundle fold (Supplementary Fig. S8*a*). A mechanism relying on the flexibility modulation of the CZB domain upon zinc binding/release has previously been proposed for the activation of DgcZ (Zähringer *et al.*, 2013).

The ligand-binding region (LBR) domains of bacterial chemoreceptors also contain four-helix bundles arranged in a

dimer and they seem to be more distant relatives of DosC$_{MID}$ (Supplementary Fig. S8b). Signal transmission by LBR has been studied in detail and it has been shown that ligand binding generates a 1 Å piston-like movement of the terminal α-helix (Yu & Koshland, 2001).

3.4. Cyclase domain

3.4.1. Structure determination. The DosC$_{GGDEF}$ domain also exhibited very limited solubility in the standard purification buffer; therefore, we decided to optimize the buffer conditions using DSF. The use of bis-tris pH 6.0 instead of Tris–HCl pH 8.0 shifted the T_m from 321 to 329 K (Supplementary Fig. S5b). DosC$_{GGDEF}$ exhibited increased stability and solubility in the optimized buffer. This was essential for the crystallization of DosC$_{GGDEF}$: we obtained crystals in tetragonal ($P4_1$), monoclinic ($P2_1$), orthorhombic ($P2_12_12_1$) and trigonal ($P3_121$) space groups, including the apo form, the GTPαS-bound form and two previously undescribed dimeric forms (forms III and IV). All DosC$_{GGDEF}$ structures were solved by molecular replacement (Table 2).

3.4.2. DosC$_{GGDEF}$ monomer. The crystal structure of the GGDEF catalytic domain of DosC in the apo form was determined at atomic resolution (1.2 Å; Table 2). The electron density was interpretable for 161 of 172 residues in the one monomer present in the asymmetric unit, corresponding to residues 298–458 of full-length DosC. Similar to adenylyl cyclases (Pei & Grishin, 2001), the overall fold of the DosC$_{GGDEF}$ catalytic domain consists of a predominantly antiparallel five-stranded β-sheet core (β2–β3–β1–β4–β5) surrounded by five α-helices (A–E) (Fig. 5a). DosC$_{GGDEF}$ displays a unique short N-terminal 3$_{10}$-helix (Val299–Lys302) and a β-hairpin in the D–β4 region, which is also present in other GGDEF-domain structures.

The catalytic site (A site) bearing the conserved GGDEF motif (G$_{374}$GDEF$_{378}$) is located in the β2–β3 loop, while the canonical RXXD (R$_{365}$SSD$_{368}$) primary inhibitory site (I$_p$ site) is located in the loop between β2 and the C helix (Fig. 5a). The c-di-GMP product feedback inhibition of DGCs has been described either by an allosteric mechanism (Christen et al., 2006) or by immobilization of two GGDEF domains in a nonproductive conformation such that the A sites are kept apart from each other (Wassmann et al., 2007). Such inactive arrangements are mediated by the I$_p$ site and the secondary inhibitory site (I$_s$ site)

and include homodomain and heterodomain cross-links bridged by intercalated c-di-GMP dimers. We demonstrated that DosC undergoes product feedback inhibition, but we can only speculate what the inhibited conformation would look like. The arginine residue (Arg322) in the I$_s$ site is conserved in DosC and formation of an inhibited GGDEF dimer bridged by c-di-GMP dimer seems to be possible; however, this would require quite extensive rearrangements. Another possibility, which cannot be excluded, is that the I$_s$ site is located in the adjacent middle domain. The arginine residue (Arg226) within the conserved helix B of the DosC$_{MID}$ domain could be a potential candidate for such an alternative I$_s$ site. Despite several trials, we were not able to co-crystallize the DosC$_{GGDEF}$ domain with c-di-GMP, since addition of the ligand resulted in protein precipitation.

The monomeric structure of DosC$_{GGDEF}$ is highly similar to the GGDEF domains of the Maqu_2607 protein from Marinobacter aquaeolei (PDB entry 3ign; Vorobiev et al., 2012) and DgcZ from E. coli (PDB entry 3tvk; Zähringer et al., 2013) containing canonical inhibitory sites (Table 3). The r.m.s.d.s for the GGDEF domains of XCC4471 from Xanthomonas campestris (PDB entry 3qyy; Yang et al., 2011) and YfiN from Pseudomonas aeruginosa (PDB entry 4iob; Giardina et al., 2013), which are missing the conserved residues in the I$_p$ site, are also comparable (Table 3). The most variable regions are loops D–β4 and β4–E.

3.4.3. Citrate and substrate binding to the active site. In the apo form of the DosC$_{GGDEF}$ structure, a citrate molecule (which is a crystallization-buffer constituent) binds in the proximity of the active site in a fashion mimicking a
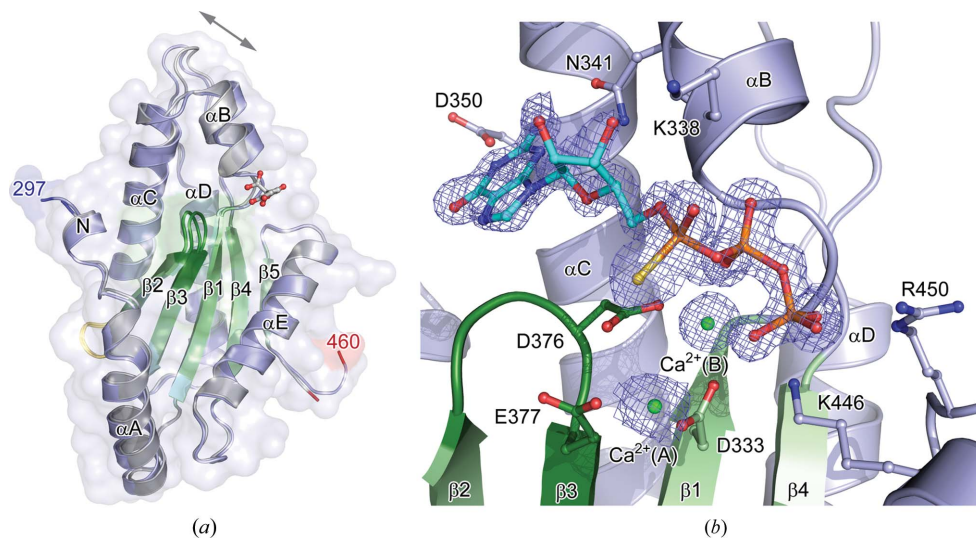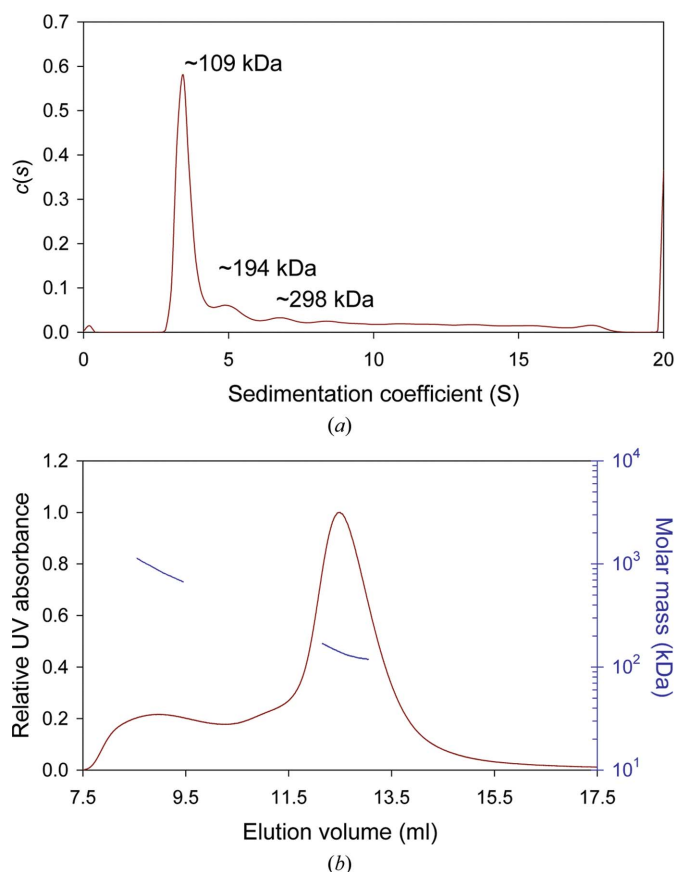

Figure 5
Monomeric DosC$_{GGDEF}$ crystal structures. (a) Comparison of the GTPαS-bound (blue/green) and apo (grey/cyan) forms. Secondary-structure elements are labelled and the N- and C-termini are labelled and coloured blue and red, respectively. The A site containing the GGDEF motif is shown in dark green, while the I$_p$ site is coloured yellow. The citrate moiety bound close to the phosphate-binding site of the apo-form structure is shown in grey and the change in conformation is indicated by the grey arrow. (b) Close-up view of the active site with the substrate analogue GTPαS bound; the colouring is the same as in (a). The side chains of the residues involved in catalysis are labelled. Two metal ions (sites A and B) are displayed as green spheres. Water molecules and the alternate conformation of Glu377 are not shown for clarity (see Supplementary Fig. S10). The simulated-annealing $2mF_{obs} - DF_{calc}$ OMIT electron-density map of the GTPαS ligand is contoured at 1.0σ.

triphosphate-binding mode (Fig. 5a). The carboxylic groups of the citrate interact with the Lys336 and Lys338 ε-amino groups as well as with the backbone amides of these two residues and Glu339 (Supplementary Fig. S9b). These residues form the phosphate-binding site. We believe that citrate binding does not affect the conformation of the GGDEF domain.

To check the effect of substrate binding on the DosC$_{GGDEF}$ domain structure, we co-crystallized it with a nonhydrolyzable substrate analogue. The crystal structure of DosC$_{GGDEF}$ in complex with GTPαS was determined at a resolution of 1.15 Å (Table 2, Fig. 5a). The final model contains one monomer in the asymmetric unit and the electron density allowed the building of all DosC$_{GGDEF}$ residues (297–460). The residues of the C-terminal expression tag were unresolved. The electron density clearly showed the presence of a GTPαS molecule and two hexacoordinated metal ions in the active site (Fig. 5b, Supplementary Fig. S10). These metal sites are termed A and B in analogy to adenylyl cyclases. Initially, both metal ions were modelled as Mg^{2+} ions, but inspection of the $mF_{obs} - DF_{calc}$ difference electron-density map showed a strong positive peak (12.4σ) at the position of metal B (Supplementary Fig. S10a). Analysis of the anomalous electron-density map revealed two peaks of 15.5σ and 7.0σ (Supplementary Fig. S10b) at positions corresponding to metals B and A. This difference suggests that site B is fully occupied and site A is partially occupied by a Ca^{2+} ion present in the crystallization drop. This would also explain the somewhat distorted octahedral geometry observed for site B (Supplementary Fig. S10b), as Ca^{2+} coordination has a greater angular flexibility than that of Mg^{2+}. The substrate-binding site is composed of a subsite for the guanine base and the β- and γ-phosphates of GTP (Supplementary Fig. S9a), respectively. The side chains of Asn341 and Asp350 are responsible for substrate specificity, while Asp333, Asp376 and Glu377 coordinate two metal ions (sites A and B). The model shows that the GTP β- and γ-phosphates form hydrogen bonds to the backbone amides of residues within the phosphate-binding site, similar to the interaction observed with citrate in the absence of substrate. Moreover, the GTP phosphates are also stabilized by additional interactions with Lys446 and Arg450 (Fig. 5b). The Ca^{2+} ion in site B interacts with the β- and γ-phosphates of GTPαS and is additionally coordinated by the backbone carbonyl of Val334 in the phosphate-binding site and the carboxylates of Asp333 and Asp376, as well as a water molecule (Fig. 5b, Supplementary Fig. S10b). The second Ca^{2+} ion (site A) coordinated by Asp333, Glu377 and three water molecules does not interact with the thiol-modified α-phosphate in GTPαS (Fig. 5b, Supplementary Fig. S10b). The Glu377 side chain adopts two alternate conformations, but both of them coordinate metal A, unlike in the previous structure of PleD in complex with GTPαS (PDB entry 2v0n; Wassmann et al., 2007). As a result of the shorter side chain of Asp376 (corresponding to Glu370 in PleD), the positions of both metal ions differ by about 1 Å. The torsion angles for the GTPαS phosphates show some differences when compared with the 2v0n structure. The side chain of Lys338 is proposed (Wassmann et al., 2007) to stabilize the pentacoordinate

phosphoryl transition state and the pyrophosphate leaving group. It is generally believed that DGCs use a two-metal-aided mechanism, as suggested for adenylyl cyclases (Tesmer et al., 1999). Comparison of the models of DosC$_{GGDEF}$ in the apo and GTPαS-bound forms results in an intermediate r.m.s.d. value (Table 3). The conformational changes observed upon ligand binding include a displacement of about 1.0 Å of the β2–β3 loop containing the GGDEF motif, but the location of the I$_p$ site remains unaltered (Fig. 5a). Otherwise, the conformational changes evident in the GGDEF motif propagate to more distant elements, resulting in about 1.5 and 2.5 Å shifts of the B helix and loop D–β4, respectively (Fig. 5a).

3.4.4. Two novel GGDEF-domain dimeric arrangements. In addition to the monomeric DosC$_{GGDEF}$ structures in the apo and GTPαS-bound forms, we determined two structures presenting dimeric assemblies. The first dimeric DosC$_{GGDEF}$ structure (form III) was determined at 2.2 Å resolution (Table 2) and shows an unusual symmetric dimer mediated by I$_p$ sites. The second dimeric DosC$_{GGDEF}$ structure (form IV) was determined at a resolution of 3.3 Å (Table 2). This structure contains two GGDEF monomers not related by

Figure 6
The oligomeric state of full-length DosC. (a) The sedimentation-coefficient distribution $c(s)$ obtained from analytical ultracentrifugation sedimentation-velocity analysis of DosC$_{full}$. Data were fitted with an r.m.s.d. of 0.007, yielding a sedimentation coefficient $s_{20,w}$ of 5.43 S and a calculated molecular weight of 109 kDa for the major peak. (b) Analytical gel filtration with multi-angle light-scattering analysis. The elution profile of DosC$_{full}$ at 280 nm is shown as a solid red line. The calculated molecular masses of aggregate and dimeric species (plotted in blue) were found to be 865 and 136 kDa, respectively.

twofold symmetry; however, a dimeric arrangement is formed by one of the chains and its symmetry mate. So far, it is not clear whether these structures are of functional relevance; both are described in more detail in Supplementary §S1.

3.5. Oligomeric state of full-length DosC

As mentioned previously, recombinantly expressed, full-length DosC was found to be predominantly insoluble and extremely susceptible to aggregation. We undertook a combined approach to improve the solubility of DosC and to reduce its aggregation. We found out that the use of CHES buffer at pH 9.5 significantly increases the solubility of DosC. Additionally, the inclusion of arginine in the buffers during cell lysis and an equimolar mixture of arginine and glutamic acid in subsequent purification steps improved the stability of the full-length protein and facilitated its characterization.

The oligomeric state of DosC was investigated by analytical centrifugation. Sedimentation coefficients for the various oligomers and the frictional ratio for the major species were calculated. The sedimentation coefficient for the major species of $DosC_{full}$ was $s_{20,w} = 5.43$ S; this species accounted for almost 60% of the protein (Fig. 6a). This sedimentation coefficient is consistent with a molecular weight of 109 kDa, which is in very good agreement with a homodimeric assembly (108.5 kDa). For this species, a frictional ratio f/f_{min} of 1.5 was found, suggesting that the homodimer has a moderately elongated shape. In addition to the predominant DosC dimer, a very broad range of multiple oligomeric species, including a partially resolved double dimer (194 kDa) and triple dimer (298 kDa), up to a molecular weight of 1.19 MDa were observed as minor components of the DosC preparation (Fig. 6a). These species probably correspond to aggregated dimers of DosC present in solution under the experimental conditions employed. They are unlikely to represent multimers of biological relevance.

To corroborate these results, we used SEC-MALS. We used the same buffer conditions as for the analytical centrifugation analysis and observed two consecutive peaks in the elution profile: a minor wide peak and unresolved smear corresponding to high-molecular-weight oligomeric species and a major peak representing the most prominent DosC oligomer (Fig. 6b). The elution volume of the main peak corresponds to the apparent molecular weight of 128 kDa suggestive of a DosC homodimer. The molecular masses for the minor and major peaks were determined by SEC-MALS to be 865 and 136 kDa, respectively (Fig. 6b). The SEC-MALS results clearly demonstrate that DosC forms a stable dimer in solution. While the effects have not been quantified, these and other results show that the inclusion of a mixture of L-Arg and L-Glu in the sample buffer strongly suppresses the aggregation of $DosC_{full}$, but does not eliminate it completely.

Oligomerization studies have previously been performed on other members of the GCS-DGCs family, including HemDGC, BpGReg and PccGCS. HemDGC was reported to form a tetramer regardless of the haem ligation state (Sawai et al., 2010). In contrast, both PccGCS and BpGReg were found

to exist in multiple oligomeric states, with the tetrameric assemblies being the most active form (Burns et al., 2014). Oligomerization-dependent cyclase activity modulation was suggested to be another level of regulation. However, our findings do not favour such a mechanism for DosC. Although DosC shares considerable sequence similarity with PccGCS and BpGReg, we never observed significant amounts of tetrameric DosC assemblies. A trivial explanation might be that the aggregation behaviour of full-length DosC might mask or disrupt such higher-order interactions. However, based on our detailed structural analysis, we believe that the homodimer is the functional unit of DosC and that oligomeric state-dependent regulation is unlikely to play a major role.

3.6. Spectral properties and enzymatic activity

In order to measure the enzymatic activity of the different DosC constructs ($DosC_{full}$, $DosC_{MID-GGDEF}$ and $DosC_{GGDEF}$), we decided to find an alternative way of improving the solubility of full-length and MID-GGDEF constructs. To this end, we used the solubility-enhancing ZZ-tag (Bogomolovas et al.,
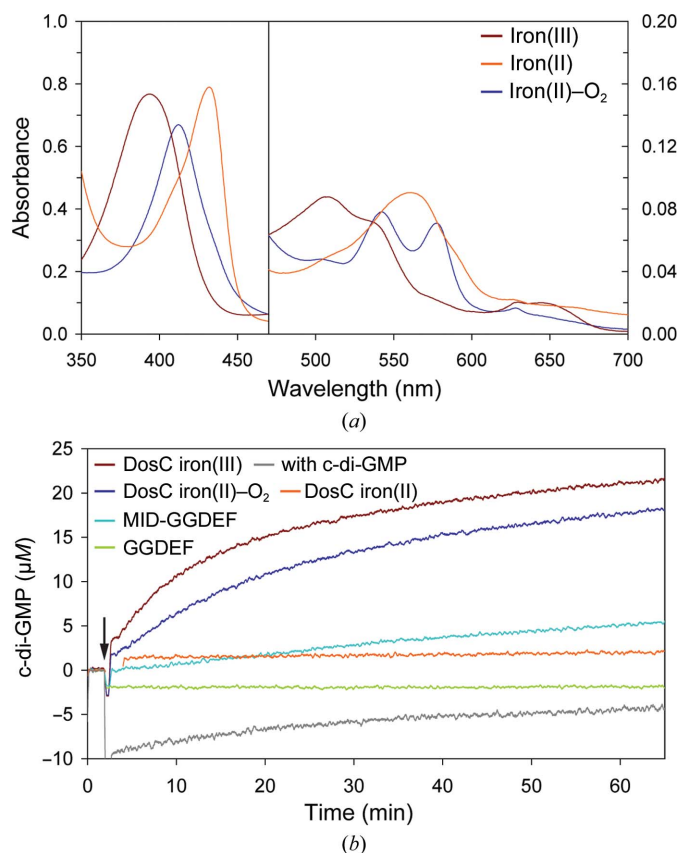
Figure 7
Spectroscopic properties and enzymatic activity of DosC. (a) Optical absorption spectra of ZZ-$DosC_{full}$ in various states: iron(III) (red), iron(II) (orange) and iron(II)–O$_2$ (blue). The absorption maxima are summarized in Supplementary Table S2. (b) Catalytic activity of different DosC constructs measured in real time by CD spectroscopy at 282 nm. The arrow indicates the moment when GTP was added to the enzymatic reaction. The plotted traces correspond to ZZ-$DosC_{full}$ iron(III) (red), ZZ-$DosC_{full}$ iron(II) (orange), ZZ-$DosC_{full}$ iron(II)–O$_2$ (blue), ZZ-$DosC_{MID-GGDEF}$ (teal), $DosC_{GGDEF}$ (green) and ZZ-$DosC_{full}$ iron(III) with c-di-GMP (grey).

2009) as a fusion partner for these proteins. This resulted in a considerable improvement in the stability and solubility; both proteins could be purified without strong aggregation. The spectral properties of ZZ-DocC$_{full}$ were investigated and did not show differences from our DosC$_{GSC}$ spectra and previously reported data (Nakajima *et al.*, 2012). The spectrum of the 'as-isolated' iron(III) form of ZZ-DocC$_{full}$ contained a Soret band at 393.5 nm, a broader peak around 500 nm and a charge-transfer band around 650 nm, suggesting a five-coordinated high-spin complex (Fig. 7a, Supplementary Table S2). This agrees with the DosC$_{Globin}$ iron(III) crystal structure and indicates that neither water nor hydroxide are axial ligands of the distal haem site in the iron(III) state. Upon reduction with sodium dithionite, the Soret band shifts to 432 nm (Fig. 7a, Supplementary Table S2). The iron(II) form is a five-coordinated high-spin complex, as observed in the DosC$_{Globin}$ iron(II) structure. We found that the iron(II)–O$_2$ complex can by formed by exposing previously reduced protein to air-saturated buffers in the presence of low-millimolar concentrations of DTE. The redox potential of the isolated haem domain of DosC was reported to be -17 mV (Kitanishi *et al.*, 2010), and DTE with its redox potential of -330 mV at pH 7.0 (Cleland, 1964) maintains the haem iron in the reduced state. The ZZ-DocC$_{full}$ iron(II)–O$_2$ complex is stable over hours at room temperature under aerobic conditions. Oxygen binding to the iron(II) form shifts the Soret band from 432 to 412 nm (Fig. 7a, Supplementary Table S2) and the iron(II)–O$_2$ complex corresponds to a six-coordinated low-spin state.

We measured the enzymatic activity using a recently developed circular-dichroism-based assay (Stelitano *et al.*, 2013) by quantifying the c-di-GMP product formed in real time. As expected, we could not detect any enzymatic activity for the DosC$_{GGDEF}$ construct, as it is monomeric in solution and condensation of two GTP molecules requires a dimeric assembly (Fig. 7b). The ZZ-DosC$_{MID-GGDEF}$ construct lacking the sensory globin domain exhibited an approximately four-fold reduced activity compared with the full-length ZZ-DocC$_{full}$ iron(III) form, which was previously described to be the most active DosC form (Kitanishi *et al.*, 2010). We also observed conversion of GTP to c-di-GMP by the ZZ-DocC$_{full}$ iron(II)–O$_2$ form; however, its activity was somewhat diminished (about 1.2-fold) relative to the iron(III) form (Fig. 7b). A slightly modified experimental setup was used to measure the activity of the unliganded iron(II) form; the reaction was prepared and started in a glove box under anaerobic conditions. DosC in the iron(II) form was found to be practically inactive (Fig. 7b); however, a very minute increase in the c-di-GMP concentration ($\sim$1 μM after $\sim$60 min) could be detected. The stability of the iron(II) form against oxidation was positively verified over the time course of the experiment. Additionally, the iron(II)–O$_2$ form also appeared to be stable, with only trace oxidation detectable. The k_{cat} values for the iron(III) and iron(II)–O$_2$ forms were determined to be 0.124 and 0.066 min^{-1} at 293 K, respectively (Supplementary Fig. S11). It should be noted that these values are higher compared with those reported previously (0.066 and 0.022 min^{-1}

measured at 303 K; Kitanishi *et al.*, 2010). The reason for this may be the much decreased enzyme aggregation in our study. As mentioned previously, GGDEF domains can undergo negative-feedback inhibition by c-di-GMP binding to the conserved inhibitory site (I$_p$ site). DosC contains such a canonical I$_p$ site (R$_{365}$SSD$_{368}$), so we decided to measure its product feedback inhibition. The full-length protein in the iron(III) form was incubated with a fourfold molar excess of c-di-GMP before starting the condensation reaction. The ZZ-DocC$_{full}$/c-di-GMP activity was similar to that observed for ZZ-DosC$_{MID-GGDEF}$ (Fig. 7b), but the mechanism behind this observation must be different. Interestingly, in this case we observed an inexplicable initial drop in signal (approximately equivalent to the protein concentration used in the assay).

3.7. Proposed domain arrangement in full-length DosC

The results of the hydrodynamic experiments performed in this study showed that full-length DosC forms a stable homodimer with a moderately elongated shape in solution. Furthermore, all purified DosC constructs, except DosC$_{GGDEF}$, were found to be dimeric in gel-filtration assays. To further confirm such an assembly, we used small-angle X-ray
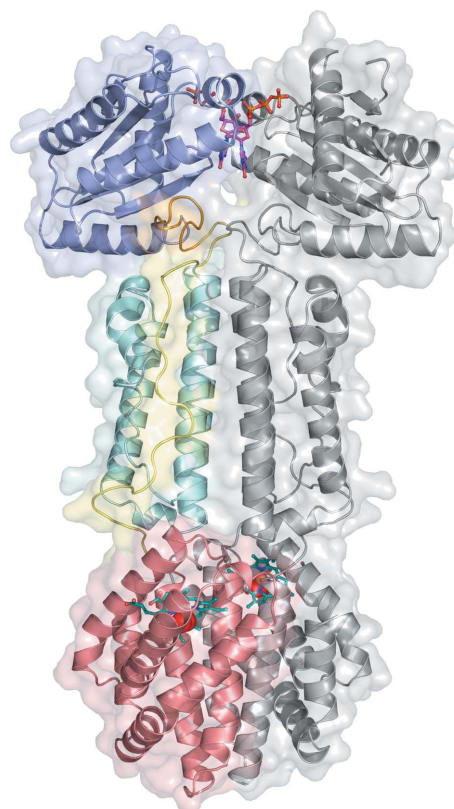
Figure 8
Proposed domain organization of full-length DosC from *E. coli*. The globin, MID and GGDEF domains of protomer *A* are coloured red, cyan and blue, respectively. Protomer *B* is shown in grey. The GGDEF domains were positioned at the condensation distance, with two GTP molecules shown in magenta. Oxygen molecules bound to the haem cofactors are shown in red. The loop fragment (shown in orange) was modelled based on the DgcZ structure and other missing linker residues were modelled and are shown in yellow; see the text for details.

scattering (SAXS) for DosC$_{full}$ shape reconstruction and domain placement. However, it became evident from the solution scattering data that aggregation was still present even at the lowest protein concentration tested, which made *ab initio* shape determination unfeasible. Unfortunately, we were also unable to obtain any structural information for the dual modules DosC$_{Globin-MID}$ and DosC$_{MID-GGDEF}$, and combining them to reconstruct the full-length assembly was thus impossible. We therefore decided to create a model of full-length DosC based on a bottom-up approach using the crystal structures of the individual domains as templates (Fig. 8). Both DosC$_{Globin}$ and DosC$_{MID}$ show dimeric assemblies in all of our crystal structures. Protein–protein docking using the *ClusPro*2.0 server was performed to align the globin and MID modules with each other. Remarkably, docking of the DosC$_{MID}$ domain onto the sensory domain without any prior knowledge resulted in a subset of very similar models (Supplementary Fig. S12*a*). The same binding mode was independently confirmed by the use of *HexServer* (Macindoe *et al.*, 2010) for protein–protein docking (Supplementary Fig. S12*b*). A catalytically competent DosC$_{GGDEF}$ dimer was generated by aligning two GTPαS-bound DosC$_{GGDEF}$ monomers with the GGDEF domains of DgcZ (PDB entry 4h54) and by shifting one of the monomers into reaction distance as described previously (Zähringer *et al.*, 2013). Next, the catalytic GGDEF dimer was manually positioned onto the globin–MID assembly to resemble the arrangement found in the DgcZ structure (PDB entry 4h54). The conserved fragment N-terminal to the GGDEF domain forms a loop showing the same conformation in DgcZ and WspR (De *et al.*, 2009). In DosC$_{GGDEF}$ the corresponding fragment (residues 297–305) forms a helical turn, but its conformation was modelled based on the DgcZ structure. Finally, for the complete structure the *MODELLER* loop-modelling routine was used to position an extended unstructured linker (residues 155–176) and a loop (residues 291–295) connecting the middle module to the cyclase domain. These structural elements are supposed to be particularly flexible; thus, their actual conformation may differ somewhat from that presented in the model (Fig. 8). The presence of such a long and disordered surface-exposed loop could possibly be one explanation for the aggregation behaviour of full-length DosC. A recent SAXS study showed that a mixture of L-Arg and L-Glu improves protein stability by collapsing flexible loops onto the protein core (Blobel *et al.*, 2011). While unproven for our system, such a loop-compaction mechanism would be fully compatible with the DosC model proposed here.

Moreover, analysis of the surface electrostatics of the proposed interdomain interfaces also supports our full-length DosC model (Supplementary Fig. S13). The surface potential of DosC$_{MID}$ clearly shows a positively charged basal portion and a negatively charged apical half (Supplementary Fig. S13*b*). In addition, there are negatively charged patches on the sides of each DosC$_{MID}$ protomer in the regions where the globin-MID linker is proposed to be. The negatively charged interface on the DosC$_{Globin}$ dimer is complemented by a positively charged basal part of the middle domain (Supple-

mentary Fig. S13*a*). The apical region of the DosC$_{MID}$ with a negative charge is complemented from the underside by the positively charged GGDEF dimer (Supplementary Fig. S13*c*).

The reasonable assumption of linear connectivity of all three modules would imply that the N- and C-termini would be expected to be at opposite ends of the middle domain, directly facing the globin and GGDEF domains, respectively. Surprisingly, we found that the N- and C-termini of the middle module are both at one end of the DosC$_{MID}$ protomer. However, this fact can be easily explained since residues 155–176 are not present in the structures taken into consideration. These residues are believed to form an extended linker which folds back to make a connection with the globin domain. The absence of this fragment as a well defined element in the DosC$_{Globin}$ (8–170) crystal structure also supports a disordered conformation.

In conclusion, we believe in the functional relevance of the DosC dimer since the domain structure is consistent with the various constraints imposed by the twofold symmetry of individual domains, an assumption of a non-intertwined and symmetric dimer of the full-length protein, linker lengths, shape complementarity, electrostatics *etc.*

3.8. Possible signalling mechanism

Based on the structural differences between the iron(III) and iron(II) forms of the DosC globin domain and the fact that the iron(III) and iron(II)–O$_2$ species are both catalytically active, we predict that similar changes would occur owing to oxygen binding to the iron(II) form and that the iron(II)–O$_2$ state would structurally resemble the iron(III) form. Therefore, it is likely that the observed structural differences of the DosC$_{Globin}$ subunits in the iron(III) and iron(II) forms (Fig. 3*a* and Supplementary Fig. S2; see §3.2.5) mirror the authentic signal-transduction mechanism. Our proposed signal-transduction pathway, derived from insights obtained from the crystal structures presented in this study, includes several events which are most likely to occur sequentially. The first event is oxygen binding and the formation of a hexacoordinated iron(II)–O$_2$ complex stabilized by interaction with the hydroxyl group of Tyr43. Next, the globin dimer is predicted to rearrange (see §3.2.5), and repositioning of the globin domain 'corner regions' would directly affect the arrangement of the middle domain and/or its flexibility. The shift or change in flexibility of the middle-domain C-terminal helices may affect the adjacent loop region, consequently enabling the productive positioning of two GGDEF domains in the proper location and at the correct distance required for catalysis. DosC in the unliganded iron(II) form may be considered to be autoinhibited and upon oxygen binding this inhibition is released. However, removal of the globin sensory domain did not result in full DGC activation (Fig. 7*b*), so perhaps a defined conformational change is additionally required. It has previously been proposed that zinc binding to the CZB domain of DgcZ rigidifies it and impedes the mobility of the GGDEF domains, thereby preventing a productive encounter of the substrates (Zähringer *et al.*, 2013).

It is worth mentioning that the B helix (Asp216–Tyr228) of DosC$_{MID}$, the sequence of which is quite conserved when compared with other related proteins (Supplementary Figs. S1 and S7*b*) may play an important role in this allosteric regulation. The mutation of His225 to alanine in *Bp*GReg (corresponding to His223 in DosC) resulted in an inactive phenotype when expressed in *S. typhimurium* (Wan *et al.*, 2009). A similar mutation (His223 to alanine) in DosC resulted in a wild-type phenotype when overexpressed in *E. coli* (Kitanishi *et al.*, 2010). Currently, the role of His223 is not entirely clear, but it may be involved in GGDEF-domain activation.

It is very tempting to speculate that the signal transmitted from the sensory domain could result in important rearrangements of the architecture of the DosC$_{MID}$ dimeric assembly. In the isolated DosC$_{MID}$ dimer, the N-terminal A helices pack close to each other (about 10 Å) and the C-terminal E helices are separated by about 30 Å (Supplementary Fig. S14*a*), whereas in the case of the CZB dimer of DgcZ we find quite the opposite situation: the C-terminal helices are closer and are only about 15 Å apart (Supplementary Fig. S14*b*). Even when assuming that some additional flexibility is provided by the linker region, this difference in the distance between the C-termini of the MID and CZB domains suggests that the MID–GGDEF linkage will probably be arranged somewhat differently to the CZB–GGDEF interdomain connection in DgcZ. Such an extreme reorganization as seen in the CZB dimer clearly requires large-scale rotation and the formation of a completely new dimer interface (Supplementary Fig. S14*c*). However, we were not able to confirm such a configuration by molecular-docking predictions, but rather the genuine binding mode observed in the DosC$_{MID}$ crystal structure. Another possibility might be that in response to the activation signal the unstructured segment preceding helix B of the middle domain together with the long globin–MID linker may become partially structured and thus complement the DosC$_{MID}$ helical bundle. Currently, both of these scenarios remain speculative as there is no direct experimental evidence to select one over the other.

4. Conclusions

Multidomain sensory proteins are notoriously difficult or impossible to crystallize, and full-length DosC proved to be no exception. Nevertheless, by using a 'divide-and-conquer' strategy, we have developed a credible model of full-length DosC (Fig. 8) and propose a possible regulatory mechanism for its enzymatic activity. Our work clearly demonstrates that the previous model proposed for *Bp*GReg (Wan *et al.*, 2009), in which dimerization is mediated only by the globin domain, can no longer be regarded as relevant for DosC. DosC from *E. coli* was shown to be a constitutive dimer and in this form its enzymatic activity is regulated by oxygen binding. We were able to structurally characterize the middle domain of DosC, which connects the sensory and effector modules of this protein, for the first time. This previously underestimated element is very likely to be both essential for and directly involved in the intramolecular signal transduction in DosC. Sequence alignments and secondary-structure predictions showed that it is conserved among other oxygen-dependent diguanylate cyclases (Supplementary Fig. S1). This suggests that the majority of the members of this protein family (*Av*GReg, *Bp*GReg, *Pcc*GCS *etc.*) probably share the structural features described here for DosC.

5. Related literature

The following reference is cited in the Supporting Information for this article: Deepthi *et al.* (2014).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Aono, S. (2013). *Adv. Microb. Physiol.* **63**, 273–327.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Blobel, J., Brath, U., Bernadó, P., Diehl, C., Ballester, L., Sornosa, A., Akke, M. & Pons, M. (2011). *Eur. Biophys. J.* **40**, 1327–1338.
Boehm, A., Kaiser, M., Li, H., Spangler, C., Kasper, C. A., Ackermann, M., Kaever, V., Sourjik, V., Roth, V. & Jenal, U. (2010). *Cell*, **141**, 107–116.
Bogomolovas, J., Simon, B., Sattler, M. & Stier, G. (2009). *Protein Expr. Purif.* **64**, 16–23.
Boyd, C. D. & O'Toole, G. A. (2012). *Annu. Rev. Cell Dev. Biol.* **28**, 439–462.
Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. (2013). *Nucleic Acids Res.* **41**, W349–W357.
Burns, J. L., Deer, D. D. & Weinert, E. E. (2014). *Mol. Biosyst.* **10**, 2823–2826.
Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T. & Ben-Tal, N. (2013). *Isr. J. Chem.* **53**, 199–206.
Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M. & Jenal, U. (2006). *J. Biol. Chem.* **281**, 32015–32024.
Cleland, W. W. (1964). *Biochemistry*, **3**, 480–482.

Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). *Genome Res.* **14**, 1188–1190.

De, N., Navarro, M. V., Raghavan, R. V. & Sondermann, H. (2009). *J. Mol. Biol.* **393**, 619–633.

Deepthi, A., Liew, C. W., Liang, Z.-X., Swaminathan, K. & Lescar, J. (2014). *PLoS One*, **9**, e110912.

DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). *Nucleic Acids Res.* **32**, W665–W667.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Erickson, H. P. (2009). *Biol. Proced. Online*, **11**, 32–51.

Farhana, A., Saini, V., Kumar, A., Lancaster, J. R. Jr & Steyn, A. J. (2012). *Antioxid. Redox Signal.* **17**, 1232–1245.

Fedorov, O., Niesen, F. H. & Knapp, S. (2012). *Methods Mol. Biol.* **795**, 109–118.

Fiser, A. & Sali, A. (2003). *Bioinformatics*, **19**, 2500–2501.

Freitas, T. A., Hou, S. & Alam, M. (2003). *FEBS Lett.* **552**, 99–104.

Freitas, T. A., Saito, J. A., Hou, S. & Alam, M. (2005). *J. Inorg. Biochem.* **99**, 23–33.

Galperin, M. Y., Higdon, R. & Kolker, E. (2010). *Mol. Biosyst.* **6**, 721–728.

Germani, F., Moens, L. & Dewilde, S. (2013). *Adv. Microb. Physiol.* **63**, 1–47.

Giardina, G., Paiardini, A., Fernicola, S., Franceschini, S., Rinaldo, S., Stelitano, V. & Cutruzzolà, F. (2013). *PLoS One*, **8**, e81324.

Gilles-Gonzales, M.-A. & Gonzales, G. (2008). *The Smallest Biomolecules: Diatomics and their Interactions with Heme Proteins*, edited by A. Ghosh, pp. 18–65. Amsterdam: Elsevier.

Girard, É., Chantalat, L., Vicat, J. & Kahn, R. (2002). *Acta Cryst.* D**58**, 1–9.

Green, J., Crack, J. C., Thomson, A. J. & LeBrun, N. E. (2009). *Curr. Opin. Microbiol.* **12**, 145–151.

Henry, J. T. & Crosson, S. (2011). *Annu. Rev. Microbiol.* **65**, 261–286.

Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS: A Computer Program for Calculating Accessibilities.* Department of Biochemistry and Molecular Biology, University College London.

Igarashi, J., Kitanishi, K. & Shimizu, T. (2011). *Handbook of Porphyrin Science*, Vol. 15, *Biochemistry of Tetrapyrroles*, edited by K. M. Kadish, K. M. Smith & R. Guilard, pp. 399–460. Singapore: World Scientific.

Izaac, A., Schall, C. A. & Mueser, T. C. (2006). *Acta Cryst.* D**62**, 833–842.

Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst.* D**60**, 1670–1673.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kalia, D., Merey, G., Nakayama, S., Zheng, Y., Zhou, J., Luo, Y., Guo, M., Roembke, B. T. & Sintim, H. O. (2013). *Chem. Soc. Rev.* **42**, 305–341.

Kitanishi, K., Kobayashi, K., Kawamura, Y., Ishigami, I., Ogura, T., Nakajima, K., Igarashi, J., Tanaka, A. & Shimizu, T. (2010). *Biochemistry*, **49**, 10381–10393.

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R. & Vajda, S. (2013). *Proteins*, **81**, 2159–2166.

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.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

Laskowski, R. A. & Swindells, M. B. (2011). *J. Chem. Inf. Model.* **51**, 2778–2786.

Leibly, D. J., Nguyen, T. N., Kao, L. T., Hewitt, S. N., Barrett, L. K. & Van Voorhis, W. C. (2012). *PLoS One*, **7**, e52482.

Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M.-D. & Ritchie, D. W. (2010). *Nucleic Acids Res.* **38**, W445–W449.

Martínková, M., Kitanishi, K. & Shimizu, T. (2013). *J. Biol. Chem.* **288**, 27702–27711.

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.

Méndez-Ortiz, M. M., Hyodo, M., Hayakawa, Y. & Membrillo-Hernández, J. (2006). *J. Biol. Chem.* **281**, 8090–8099.

Nakajima, K., Kitanishi, K., Kobayashi, K., Kobayashi, N., Igarashi, J. & Shimizu, T. (2012). *J. Inorg. Biochem.* **108**, 163–170.

Pei, J. & Grishin, N. V. (2001). *Proteins*, **42**, 210–216.

Pesce, A., Thijs, L., Nardini, M., Desmet, F., Sisinni, L., Gourlay, L., Bolli, A., Coletta, M., Van Doorslaer, S., Wan, X., Alam, M., Ascenzi, P., Moens, L., Bolognesi, M. & Dewilde, S. (2009). *J. Mol. Biol.* **386**, 246–260.

Povolotsky, T. L. & Hengge, R. (2012). *J. Biotechnol.* **160**, 10–16.

Rice, P., Longden, I. & Bleasby, A. (2000). *Trends Genet.* **16**, 276–277.

Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.

Römling, U., Galperin, M. Y. & Gomelsky, M. (2013). *Microbiol. Mol. Biol. Rev.* **77**, 1–52.

Sawai, H., Yoshioka, S., Uchida, T., Hyodo, M., Hayakawa, Y., Ishimori, K. & Aono, S. (2010). *Biochim. Biophys. Acta*, **1804**, 166–172.

Schirmer, T. & Jenal, U. (2009). *Nature Rev. Microbiol.* **7**, 724–735.

Schuck, P. (2010). *Eur. Biophys. J.* **39**, 1261–1275.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.

Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N. & Hengge, R. (2009). *Microbiology*, **155**, 1318–1331.

Stelitano, V., Brandt, A., Fernicola, S., Franceschini, S., Giardina, G., Pica, A., Rinaldo, S., Sica, F. & Cutruzzolà, F. (2013). *Nucleic Acids Res.* **41**, e79.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Suhre, K. & Sanejouand, Y.-H. (2004). *Nucleic Acids Res.* **32**, W610–W614.

Tagliabue, L., Antoniani, D., Maciag, A., Bocci, P., Raffaelli, N. & Landini, P. (2010). *Microbiology*, **156**, 2901–2911.

Tagliabue, L., Maciag, A., Antoniani, D. & Landini, P. (2010). *FEMS Immunol. Med. Microbiol.* **59**, 477–484.

Tamayo, R., Pratt, J. T. & Camilli, A. (2007). *Annu. Rev. Microbiol.* **61**, 131–148.

Tanaka, A., Takahashi, H. & Shimizu, T. (2007). *J. Biol. Chem.* **282**, 21301–21307.

Tesmer, J. J., Sunahara, R. K., Johnson, R. A., Gosselin, G., Gilman, A. G. & Sprang, S. R. (1999). *Science*, **285**, 756–760.

Thijs, L., Vinck, E., Bolli, A., Trandafir, F., Wan, X., Hoogewijs, D., Coletta, M., Fago, A., Weber, R. E., Van Doorslaer, S., Ascenzi, P., Alam, M., Moens, L. & Dewilde, S. (2007). *J. Biol. Chem.* **282**, 37325–37340.

Tuckerman, J. R., Gonzalez, G. & Gilles-Gonzalez, M. A. (2011). *J. Mol. Biol.* **407**, 633–639.

Tuckerman, J. R., Gonzalez, G., Sousa, E. H., Wan, X., Saito, J. A., Alam, M. & Gilles-Gonzalez, M. A. (2009). *Biochemistry*, **48**, 9764–9774.

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F. & Rarey, M. (2012). *J. Chem. Inf. Model.* **52**, 360–372.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.

Vorobiev, S. M., Neely, H., Yu, B., Seetharaman, J., Xiao, R., Acton, T. B., Montelione, G. T. & Hunt, J. F. (2012). *J. Struct. Funct. Genomics*, **13**, 177–183.

Walshaw, J. & Woolfson, D. N. (2001). *J. Mol. Biol.* **307**, 1427–1450.

Wan, X., Tuckerman, J. R., Saito, J. A., Freitas, T. A., Newhouse, J. S., Denery, J. R., Galperin, M. Y., Gonzalez, G., Gilles-Gonzalez, M. A. & Alam, M. (2009). *J. Mol. Biol.* **388**, 262–270.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U. & Schirmer, T. (2007). *Structure*, **15**, 915–927.

Weber, H., Pesavento, C., Possling, A., Tischendorf, G. & Hengge, R. (2006). *Mol. Microbiol.* **62**, 1014–1034.

Weinert, E. E. & Marletta, M. A. (2011). *Handbook of Porphyrin Science*, Vol. 15, *Biochemistry of Tetrapyrroles*, edited by K. M. Kadish, K. M. Smith & R. Guilard, pp. 123–157. Singapore: World Scientific.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Yang, C.-Y., Chin, K.-H., Chuah, M. L.-C., Liang, Z.-X., Wang, A. H.-J. & Chou, S.-H. (2011). *Acta Cryst.* D**67**, 997–1008.

Yu, E. W. & Koshland, D. E. Jr (2001). *Proc. Natl Acad. Sci. USA*, **98**, 9517–9520.

Zähringer, F., Lacanna, E., Jenal, U., Schirmer, T. & Boehm, A. (2013). *Structure*, **21**, 1149–1157.

Zhang, W. & Phillips, G. N. Jr (2003). *Structure*, **11**, 1097–1110.