BIOLOGICAL
CRYSTALLOGRAPHY

Volume 71 (2015)

Supporting information for article:

Structural analysis of an oxygen-regulated diguanylate cyclase

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

S1. Dimeric $DosC_{GGDEF}$ structures (form III and IV)

The first of two $DosC_{GGDEF}$ structures displaying dimeric assemblies was determined at 2.2 Å resolution (form III; see Table 2). The final model consists of two chains, with chain A containing residues 297-461 and chain B residues 298-461, arranged in a dimer (Fig. S15a). The conformations of a few C-terminal residues are very different in both chains due to different environments in the crystal lattice. At first glance, the distance between the N-termini seems to be incompatible with linking to the $DosC_{MID}$ domain, but the loops joining these two structural modules may actually provide some additional flexibility. The dimer displays two-fold symmetry and the individual N-terminal A-helices of chain A and B are parallel to each other (Fig. S15a). Very interestingly, most contacts between the GGDEF monomers are made by residues of the I_p-site ($R_{365}SSD_{368}$ in loop C-β2). The situation observed here brings to mind the structures of inhibited dimers cross-linked by c-di-GMP, but without a dinucleotide bridge in this case. The dimer interface has a surface area of about 670 $Å^2$. The most characteristic inter-subunit salt bridges are the two formed by Arg365 from one protomer and Asp368 from its counterpart and *vice versa* (Fig. S15b) Interestingly, Arg365 and Asp368 are conserved in sequence alignments of DosC related proteins (Fig. S1). Other interactions involve symmetric salt bridges between Arg392 and Asp362 (Fig. S15b), and a number of hydrogen bonds. This kind of unusual GGDEF domain dimeric arrangement solely mediated by I_p-sites has not been described previously. Given the involvement of highly conserved residues in dimer formation we hypothesize that this arrangement may play a role in the regulation of DosC activity.

The second $DosC_{GGDEF}$ structure displaying a dimeric arrangement was determined at a resolution of 3.3 Å (form IV; see Table 2). The comparatively low resolution of this structure may be a result of the relatively high solvent content (66%) and spacious solvent channels (about 75 Å in diameter) in the crystals. The final model contains two monomers in the asymmetric unit, which are not related by a two-fold symmetry. The electron density allowed us to model residues 298-460 for chain A and residues 298-458 for chain B. Despite the modest resolution of this structure we could model two sulphate ions (a component of the crystallization solution) in the vicinity of the phosphate binding site of each monomer. These sulphates mimic the γ-phosphate of the substrate. Analysis of crystal packing showed that the two chains are not equivalent (Fig. S16). To obtain a two-fold symmetric dimer, we merged chain B and its symmetry related molecule (*x-y,-y,-z+2/3*) (Fig. S17a). The N-termini in this dimer are about 40 Å apart (plus linker of several residues), which approximately corresponds to the spacing of the C-termini of the middle domain. The C-terminal E helices are almost parallel to each other and are responsible for all contacts between the protomers, most importantly the π-stacking interactions of the two Tyr443 residues and hydrogen bonds between Arg447 from one and Gln436 from the other subunit respectively (Fig. S17b). A tunnel is formed

where the bottom is restricted by the middle module and a ceiling is formed by a pair of E helices. This tunnel connects two half active sites. Superposition of the GTPαS-bound structures onto such a dimeric assembly shows that the distance between α-phosphates of about 12 Å is unsuitable for condensation (Fig. S17b). When the same operation is repeated for the DgcZ GGDEF dimer (PBD code 3TVK) in the product-bound state, the distance between α-phosphates is reduced to about 6 Å (Fig. S17c), which is still not optimal for condensation but is compatible with c-di-GMP binding. Proper positioning of the substrate's 3'-hydroxyl with the α-phosphate of the second substrate molecule can be achieved by adjusting the ribose torsion angles without disrupting the interactions of the guanine base with the protein (Fig. S17c). In such a configuration, the Tyr443/Arg447 pairs are no longer side by side, but face each other (Fig. S17c). Additional contacts can be established by Lys338 and Asp342 with Glu377 and Arg306, respectively. A significant GGDEF subunit rotation and translation would be required to recreate a recently reported active-like dimeric structure of tDGC (Deepthi *et al.*, 2014). We suspect that the Tyr443/Arg447 pair acts as a spacer or guide for GGDEF protomers and suggest that the two arrangements discussed (Fig. S17b and S17c) may represent distinct states of the catalytic cycle.

Supplementary References

Burns, J. L., Douglas Deer, D. & Weinert, E. E. (2014). *Mol. BioSyst.* **10**, 2823-2826.
Deepthi, A., Liew, C. W., Liang, Z. X., Swaminathan, K. & Lescar, J. (2014). *PloS one* **9**, e110912.
Kitanishi, K., Kobayashi, K., Kawamura, Y., Ishigami, I., Ogura, T., Nakajima, K., Igarashi, J., Tanaka, A. & Shimizu, T. (2010). *Biochemistry* **49**, 10381-10393.
Sawai, H., Yoshioka, S., Uchida, T., Hyodo, M., Hayakawa, Y., Ishimori, K. & Aono, S. (2010). *Biochim. Biophys. Acta* **1804**, 166-172.
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.
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.

Table S1 Selected properties of well-characterized globin-coupled diguanylate cyclases.

Protein (Uniprot acc. no.)	Source organism	Proximal/ distal heme ligand	Absorption maxima (nm) and coordination structure	K_d for O_2 (μM)	k_{ox} (min^{-1})	Enzymatically active forms	Enzymatically inactive forms	Oligomeric state/ no. of residues	Reference
DosC (P0AA89)	*Escherichia coli*	His98/ Tyr43	Fe(II): 432, 560 (5cHS) Fe(II)-O_2: 414, 543, 579 (6cLS)	14	0.0092	Fe(III), Fe(II)-O_2, Fe(II)-CO	Fe(II), Fe(II)-NO	N.A./ 460	(Kitanishi *et al.*, 2010)
*Av*GReg (M9YE33)	*Azotobacter vinelandii*	His101/ Tyr44	Fe(II): 430, ~560 (5cHS)* Fe(II)-O_2: 412, 540, 579 (6cLS)*	0.12	<0.001	N.A.	N.A.	N.A./ 464	(Thijs *et al.*, 2007)
*Bpe*GReg (Q7VTL8)	*Bordetella pertussis*	His99/ Tyr43	Fe(II): 431 Fe(II)-O_2: 416	0.64	N.A.	Fe(II)-O_2, Fe(II)-NO, Fe(II)-CO	Fe(II)	Dimer-tetramer/ 475	(Wan *et al.*, 2009; Burns *et al.*, 2014)
HemDGC (Q6ARU5)	*Desulfotalea psychrophila*	His108/ Tyr55	Fe(II): 430, 556 Fe(II)-O_2: 412, 545, 580	N.A.	0.025	Fe(II)-O_2	Fe(II)-NO, Fe(II)-CO, Fe(II), Fe(III)	Tetramer/ 363	(Sawai *et al.*, 2010)
*Pcc*GCS (C6D9C2)	*Pectobacterium carotovorum*	His112/ Tyr57	Fe(II): 433, 562 Fe(II)-O_2: 415, 548, 583	N.A.	N.A.	Fe(II)-O_2, Fe(II)-NO, Fe(II)-CO	Fe(II)	Dimer-tetramer/ 471	(Burns *et al.*, 2014)

*Values for *Av*GReg heme-bound domain are shown, due to some discrepancies observed for probably misfolded full-length *Av*GReg.

Table S2 Absorption spectra of ZZ-$DosC_{full}$ in different ligation states.

Form	Absorption maxima (nm)	Coordination structure
Fe(III)	393.5, 506, 645	5cHS
Fe(II)	432, 561	5cHS
Fe(II)-O_2	412, 542.5, 577.5	6cLS

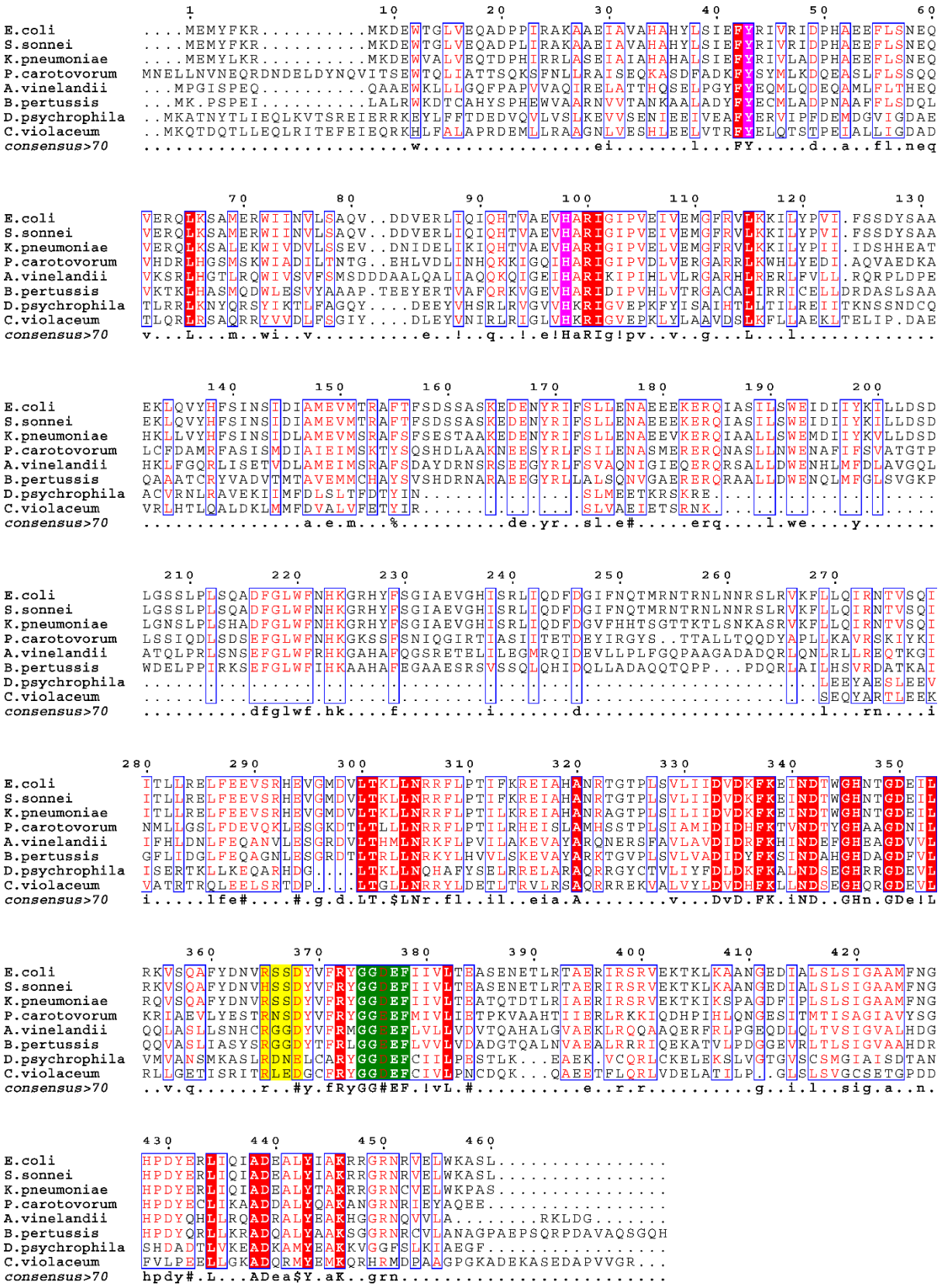

Figure S1 Multiple sequence alignment of DosC from *E. coli* (WP_000426292.1) with GCS-DGCs from other organisms: *S. sonnei* (WP_005146315.1), *K. pneumoniae* (WP_020803662.1), *A.*

vinelandii (WP_012698737.1), *B. pertussis* (NP_882025.1), *P. carotovorum* (WP_015840820.1), *D. psychrophila* (YP_063937.1), *C. violaceum* (NP_899909.1). The sequences were aligned using ClustalW2 and visualized with ESPript3. Residues with strict identity are shown as white characters on a red background, whereas those with high similarity are shown as red characters and framed. The GGDEF motif is coloured green, whereas the primary inhibitory site (I_p-site) is shown in yellow. The proximal and distal heme axial ligands are highlighted in magenta.

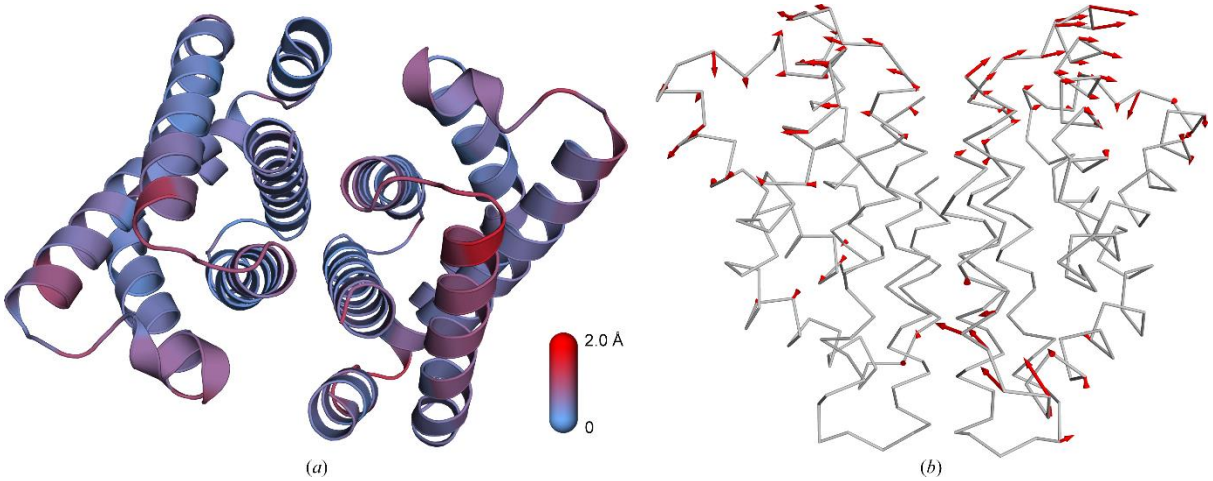

Figure S2 Structural changes between the Fe(III) and Fe(II) states of DosC globin domain. (*a*) Cartoon representation of $DosC_{Globin}$ Fe(III) structure coloured according to the distances between aligned Cα atoms pairs of two forms (chains A and B). Colour spectrum from blue to red covers the displacement range of 0-2.0 Å. (*b*) Side view of $DosC_{Globin}$ Fe(III) structure in backbone representation. Red arrows indicate direction and magnitude of movements exceeding 0.5 Å.

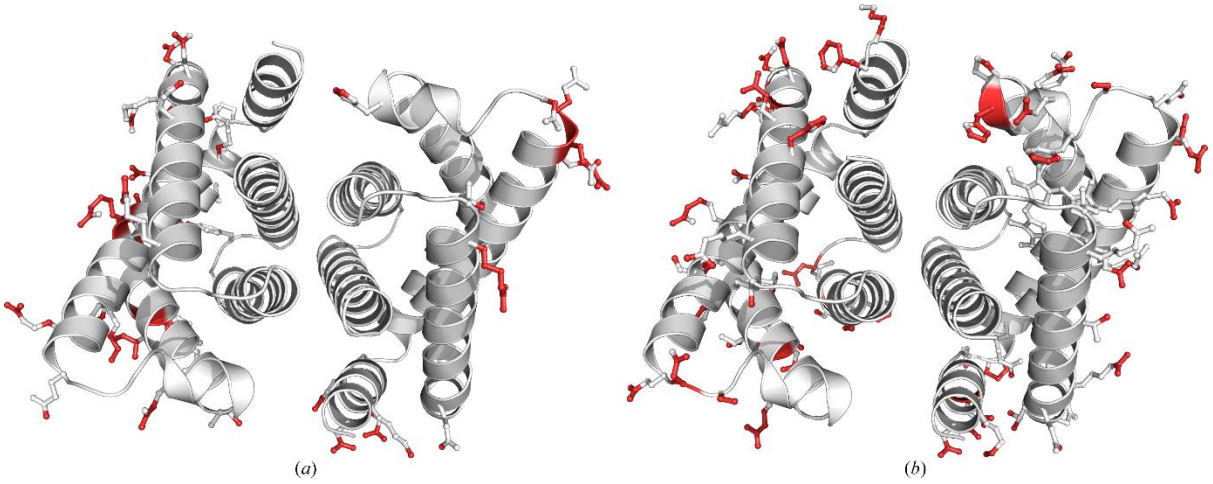

Figure S3 Intermolecular contacts of $DosC_{Globin}$ structures with symmetry-related molecules in the crystal. Residues involved in contacts with symmetry mates are shown in stick representation, all atoms which make contacts within 4.0 Å radius are highlighted in red. The $DosC_{Globin}$ Fe(III) form is shown in (*a*) and chains A and B of the Fe(II) form in (*b*).

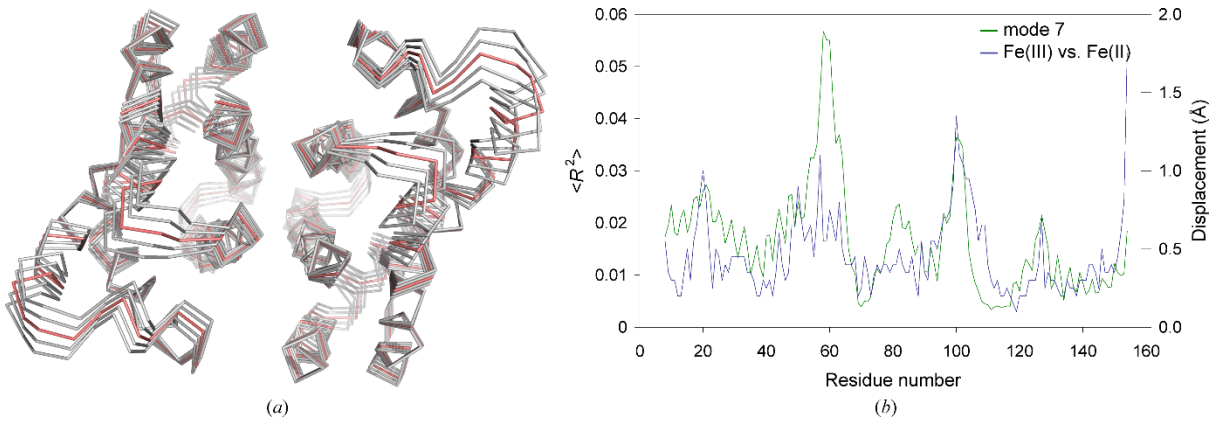

Figure S4 Normal-mode analysis of $DosC_{Globin}$ domain. (*a*) Four perturbed structures following mode 7 are shown in grey, whereas the crystal structure of the globin domain in the Fe(III) state is shown in red. (*b*) Distribution of fluctuations as described by the normalized mean square displacement $<R^2>$ of Cα atoms in the lowest frequency normal mode (mode 7) (in green) and actual displacements of Cα atoms observed between the Fe(III) and Fe(II) structures (in blue). Both parameters were averaged over chains A and B.

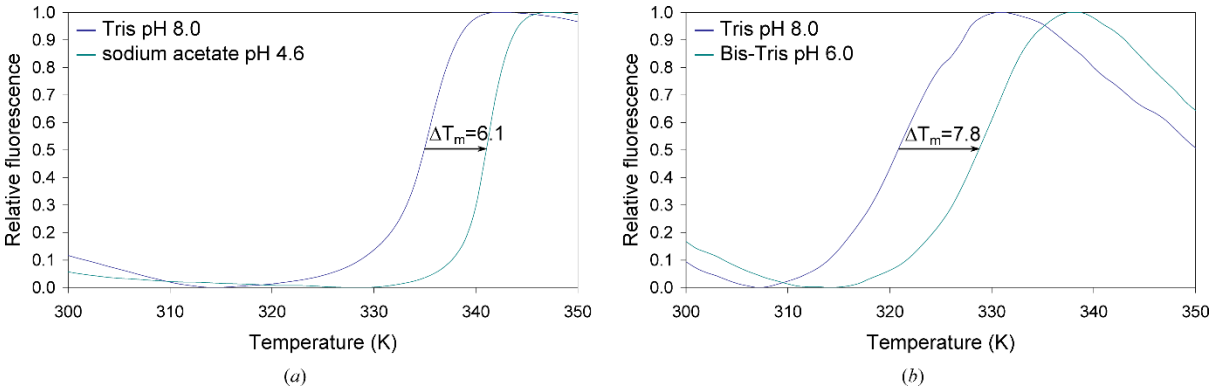

Figure S5 Thermal stability shift assay using differential scanning fluorimetry for buffer optimization. Melting curves of $DosC_{MID}$ (*a*) and $DosC_{GGDEF}$ (*b*) in different buffers: 100 mM Tris-HCl pH 8.0 and 100 mM sodium acetate pH 4.6 (*a*); 100 mM Tris-HCl pH 8.0, 50 mM NaCl and 100 mM Bis-Tris pH 6.0, 100 mM NaCl (*b*). Calculated increase in T_m is shown in the plots.

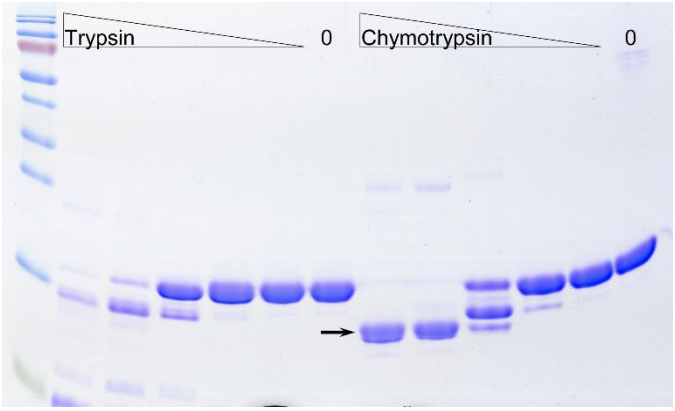

Figure S6 Limited proteolysis of $DosC_{MID}$ (residues 148-300). SDS-PAGE analysis of proteolytic fragments obtained by trypsin and chymotrypsin digestion at different protease concentrations (100-0.05 μg ml^{-1}). The arrow indicates a protease stable fragment corresponding to residues 173-298, observed at the highest chymotrypsin concentrations.

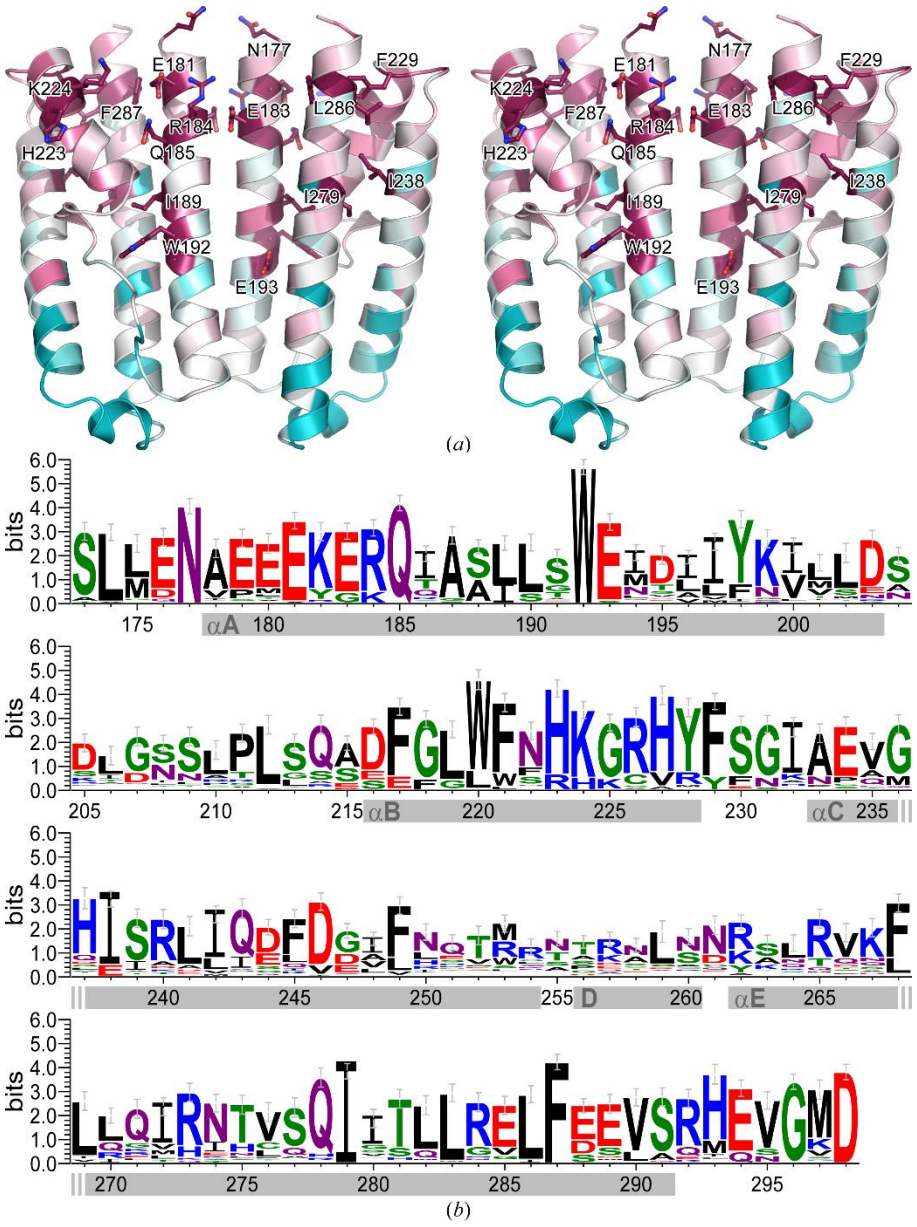

Figure S7 Conserved regions in DosC$_{MID}$ domain structure. (*a*) Stereoview of DosC$_{MID}$ coloured by sequence conservation of 78 homologues. The most conserved residues are shown in magenta (side-chains in stick representation and labelled) and the least conserved residues in cyan. (*b*) A sequence logo representation of DosC MID domain based on alignment of 78 homologues. A logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of the symbols within the stack indicates the relative frequency of each amino at that position. Amino acids are coloured according to chemical properties: polar (green), neutral (purple), basic (blue), acidic (red) and hydrophobic (black). Helical elements are marked by grey rectangles and are labelled.

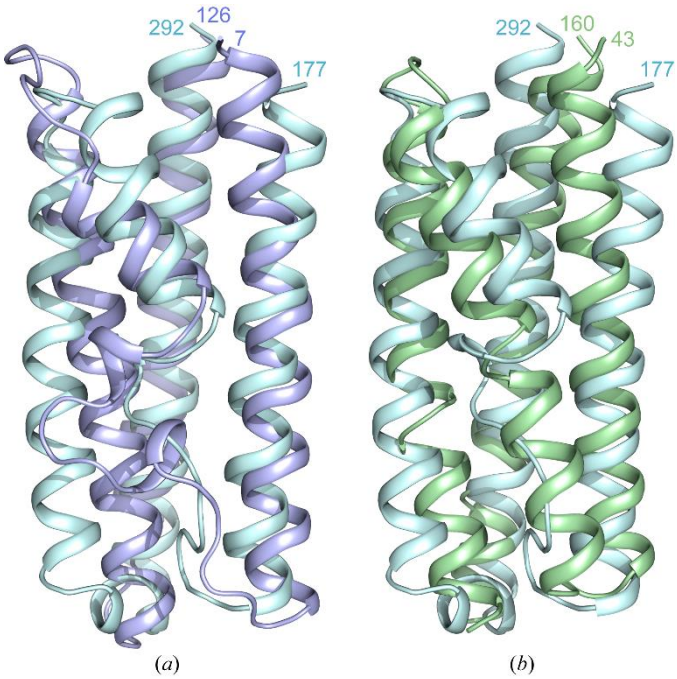

Figure S8 Superimposition of a $DosC_{MID}$ single protomer (shown in cyan) with (*a*) CZB domain of DgcZ diguanylate cyclase from *E. coli* (PDB code 3T9O; blue) and (*b*) proximal module of the LBR of McpS chemoreceptor from *Pseudomonas putida* (PDB code 2YFB; green). N- and C-termini are labelled.

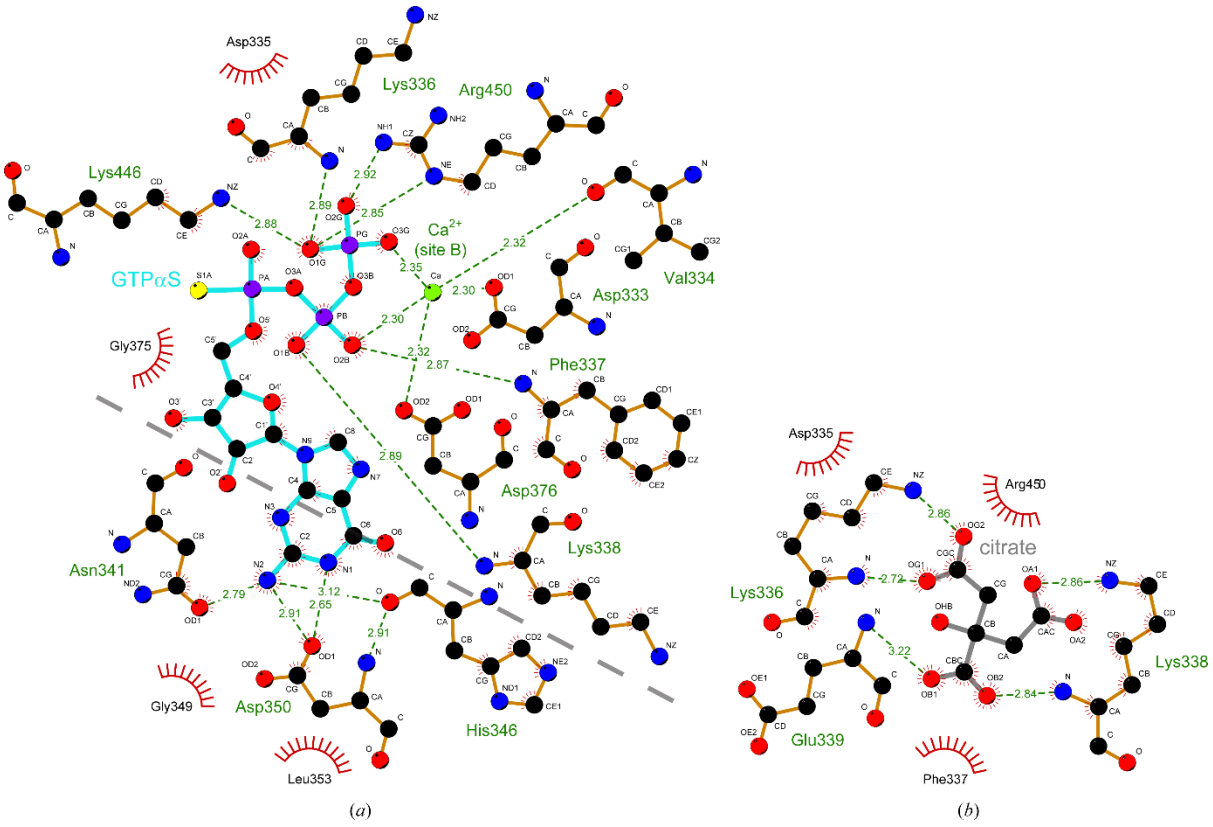

Figure S9 Schematic diagrams depicting interactions between $DosC_{GGDEF}$ and (*a*) GTPαS or (*b*) citrate, based on the ligand-bound and apo-form GGDEF domain crystal structures, respectively. The

dashed line in (*a*) delineates two sub-sites for the guanine base and for the phosphates of GTP prepared with LigPlot+.

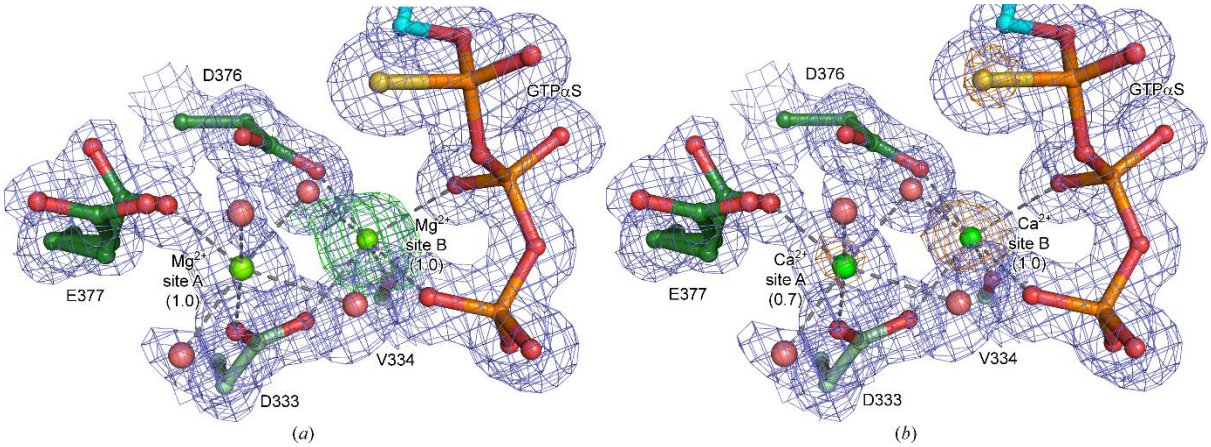

Figure S10 Two metal ions at the active site of the GTPαS-bound GGDEF domain structure. Side-chains and main-chain atoms of the residues coordinating two metal ions (site A and B) are labelled. Metal ions and water molecules are displayed as green and pink spheres, respectively. The $2mF_{obs}$-DF_{calc} electron density maps contoured at 1.0 σ are shown in blue. (*a*) Model with Mg^{2+} ions – the mF_{obs}-DF_{calc} difference electron density map contoured at 3.0 σ (green) is shown. (*b*) Model with Ca^{2+} ions – the anomalous electron density map contoured at 3.0 σ (orange) is shown and the occupancies are given in brackets.

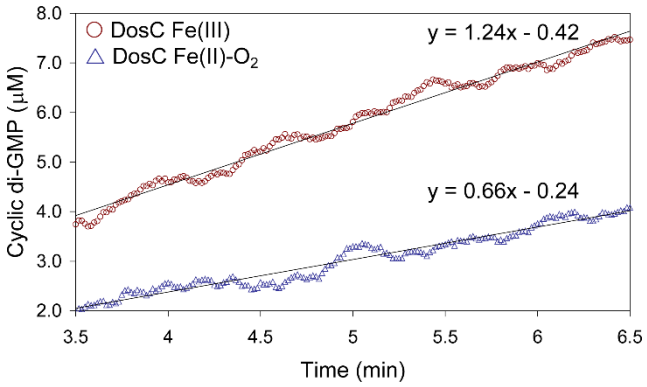

Figure S11 Turnover rate of GTP condensation for ZZ-$DosC_{full}$ measured under initial velocity conditions was obtained by a linear fitting of the increase in c-di-GMP with time. The slope of the linear range of the progress curve after including the enzyme concentration provides k_{cat} values of 0.124 and 0.066 s^{-1} at 293 K, for the Fe(III) and Fe(II)-O_2 forms, respectively.

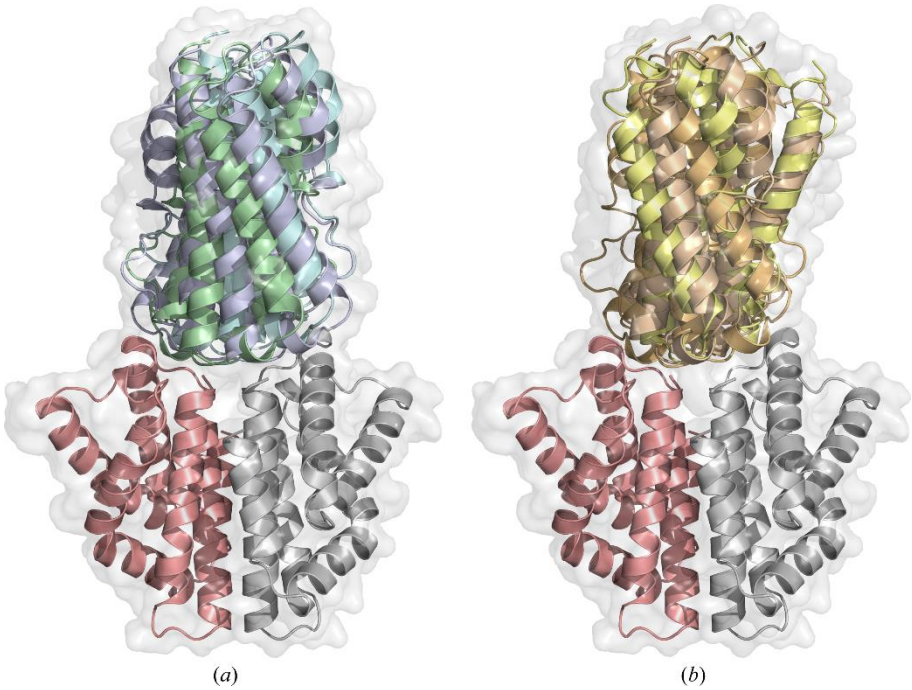

Figure S12 Docking of $DosC_{MID}$ onto $DosC_{Globin}$ dimer (chains A and B shown in red and grey) using (*a*) ClusPro 2.0 and (*b*) Hex server. Superimposition of selected models resulting from independent protein-protein docking predictions shows a very similar binding mode in both cases.

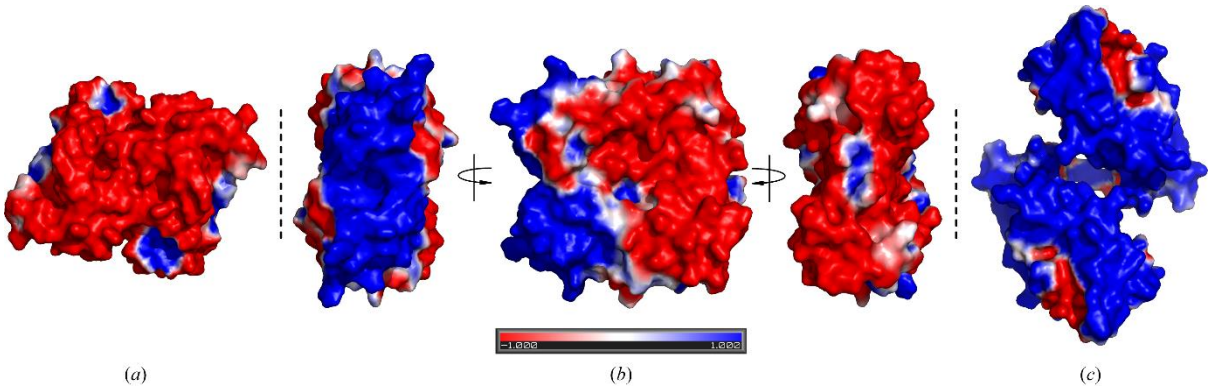

Figure S13 Distribution of the electrostatic surface potential on the globin, MID and GGDEF domains of DosC. The $DosC_{MID}$ domain is shown in the centre (*b*) with corresponding domain-domain interfaces between globin (*a*) and GGDEF (*c*) on the sides. The molecular surface is coloured according to the calculated electrostatic potential on the solvent accessible surface, which ranges from red (-1.0 $k_bTe_c^{-1}$) to blue (+1.0 $k_bTe_c^{-1}$).

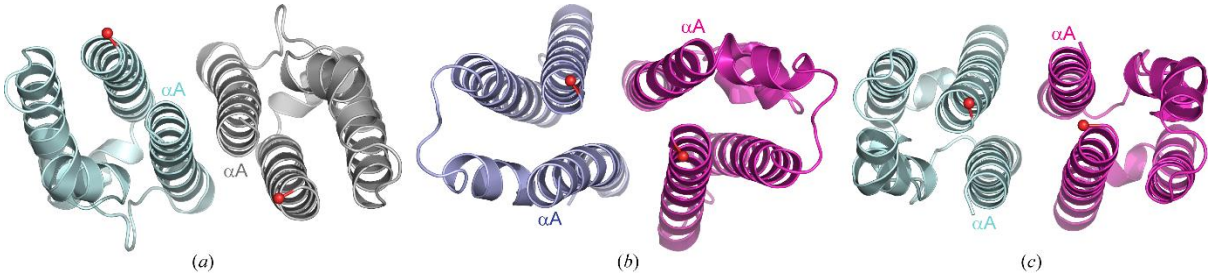

Figure S14 Dimeric arrangement of (*a*) $DosC_{MID}$ and (*b*) CZB domain of DgcZ as viewed along the symmetry axis. In $DosC_{MID}$ the A-helices pack against each other, whereas in the CZB domain they diverge due to a different angle between protomers. A hypothetical alternative $DosC_{MID}$ dimer arrangement resembling the CZB dimer geometry is shown in (*c*). The geometry in (*c*) requires a rotation of about 40°, something seen in neither the crystal structures of $DosC_{MID}$ nor found in the molecular docking predictions. The C-termini are marked with red spheres.

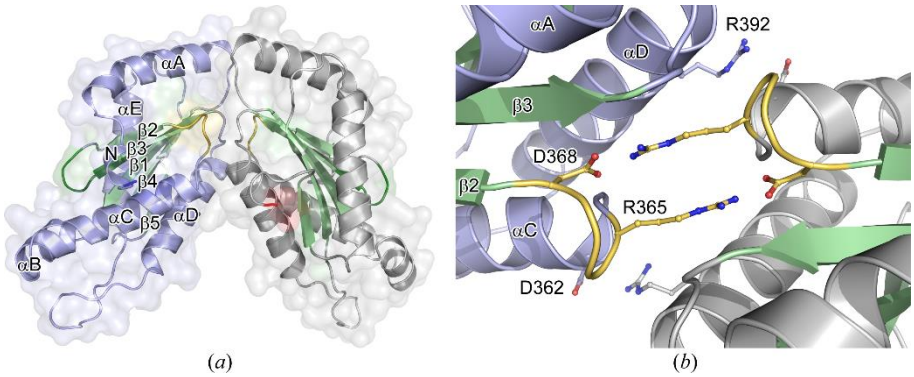

Figure S15 Structure of the presumably inactive crystallographic GGDEF domain dimer mediated by I_p-sites. (*a*) Side view of novel dimeric $DosC_{GGDEF}$ arrangement, secondary structure elements of protomer A, coloured blue/green, are labelled, protomer B is shown in grey/green. I_p-sites ($R_{365}SSD_{368}$) are coloured yellow, while A-sites are shown in dark green. (*b*) Dimer interface formed by symmetric interactions between I_p-sites of both protomers as viewed along dimer symmetry axis.

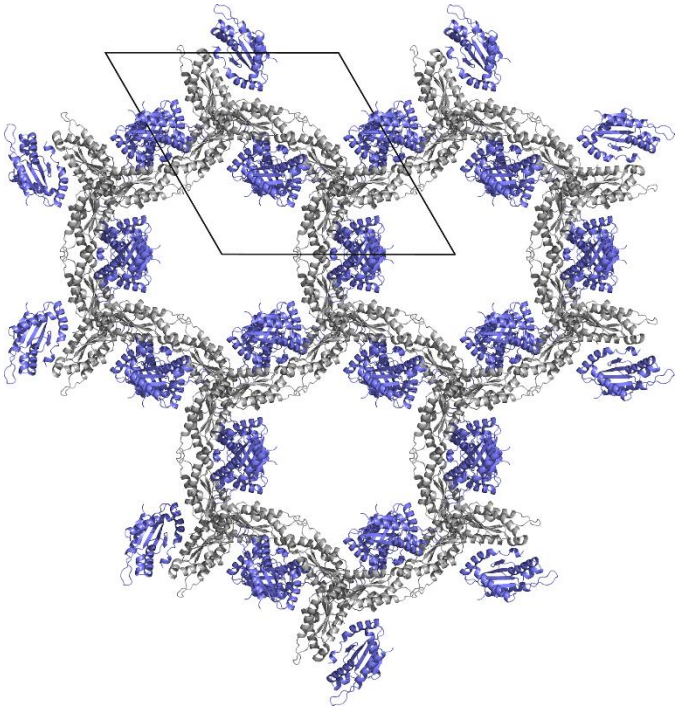

Figure S16 The crystal packing observed in the $DosC_{GGDEF}$ structure (form II) in the trigonal space-group $P3_121$. The crystals contain two chains (A and B) in the asymmetric-unit, chains A are coloured gray and the dimer-forming chains B are coloured blue. The diagram shows a projection along the crystallographic *z*-axis with the unit-cell shown in black.

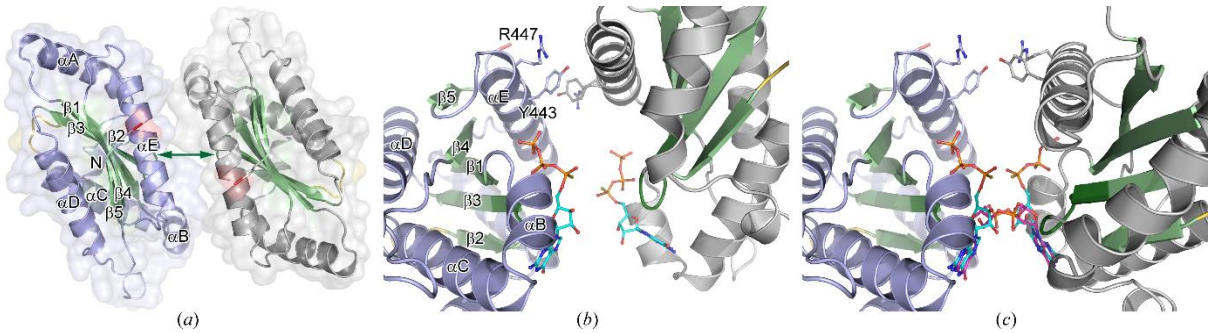

Figure S17 Proposed structure of active-like $DosC_{GGDEF}$ domain dimeric arrangement. (*a*) Dimer viewed along symmetry axis, formed by chain B (blue/green) and its symmetry related molecule (grey/green). Secondary structure elements are labelled, A-sites are coloured dark green (indicated by the green arrow), I_p-sites yellow, N- and C-termini blue and red, respectively. (*b*) Close-up view of two active half-sites in a non-productive orientation, GTPαS molecules (cyan) were overlaid by aligning the GTPαS-bound monomeric GGDEF domain structure. (*c*) The same orientation as in (*b*), but both protomers were aligned to the 3TVK structure representing the product-bound state. GTPαS molecules (cyan) were again overlaid for presentation showing that the distance between them is reduced, at the same time the guanine bases overlap with a c-di-GMP (magenta) as seen in 3TVK structure.