

Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre

Dom Bellini,^{1,2} Delphine L. Caly,³ Yvonne McCarthy,³ Mario Bumann,⁴ Shi-Qi An,⁵ J. Maxwell Dow,^{3***} Robert P. Ryan^{5**} and Martin A. Walsh^{1,2*}

¹Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK.

²Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0FA, UK.

³Department of Microbiology, Biosciences Institute, University College Cork, Cork, Ireland.

⁴Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland.

⁵Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dow street, Dundee DD1 5EH, UK.

Summary

Bis-(3',5') cyclic di-guanylate (c-di-GMP) is a key bacterial second messenger that is implicated in the regulation of many crucial processes that include biofilm formation, motility and virulence. Cellular levels of c-di-GMP are controlled through synthesis by GGDEF domain diguanylate cyclases and degradation by two classes of phosphodiesterase with EAL or HD-GYP domains. Here, we have determined the structure of an enzymatically active HD-GYP domain protein from *Persephonella marina* (PmGH) alone, in complex with substrate (c-di-GMP) and final reaction product (GMP). The structures reveal a novel trinuclear iron binding site, which is implicated in catalysis and identify residues involved in recognition of c-di-GMP. This structure completes the picture of all domains involved in c-di-GMP metabolism and reveals that the HD-GYP family splits into two distinct subgroups containing bi- and trinuclear metal centres.

Introduction

Bis-(3',5') cyclic di-guanylate (c-di-GMP) is a second messenger utilized by almost all eubacteria that acts to regulate a wide range of functions including developmental transitions, adhesion, biofilm formation, motility and the synthesis of virulence factors (Schirmer and Jenal, 2009; Boyd and O'Toole, 2012). C-di-GMP is synthesized from two GTP molecules by GGDEF domain-containing diguanylate cyclases (DGCs) and degraded by phosphodiesterases (PDEs) with either an EAL or HD-GYP domain (Ryan *et al.*, 2006; Hengge, 2009; Schirmer and Jenal, 2009; Boyd and O'Toole, 2012). Three-dimensional structures have been determined for GGDEF and EAL domains, and have afforded detailed insight into their roles in the turnover of c-di-GMP and regulatory interactions with other proteins (Hengge, 2009; Schirmer and Jenal, 2009; Boyd and O'Toole, 2012). In contrast, enzymatically active HD-GYP domain proteins, such as the paradigm RpfG from the plant pathogen *Xanthomonas campestris* (Ryan *et al.*, 2006), have so far proved intractable to structure determination by X-ray diffraction.

The structure of an unconventional catalytically inactive HD-GYP domain protein from *Bdellovibrio bacteriovorus* (Bd1817) has been determined however (Lovering *et al.*, 2011). This work identified a binuclear iron centre and the role of conserved residues within the HD-GYP family (to include the HD diad) in metal binding. The HD domain superfamily of enzymes, to which the HD-GYP family belongs, has been shown to catalyse phosphomonoesterase and phosphodiesterase reactions depending on their catalytic metal centre being mono- or binuclear respectively (Aravind and Koonin, 1998; Galperin *et al.*, 2001). The determination of the structure of Bd1817 may thus afford some insight into metal binding by enzymatically active HD-GYP domains, but the protein lacks the conserved tyrosine of the GYP motif and has no c-di-GMP phosphodiesterase activity, precluding insights into the role of the other conserved residues.

Here we describe the first crystal structure of an enzymatically active HD-GYP phosphodiesterase protein, PmGH from *Persephonella marina* EX-H1, a thermophilic marine member of the Aquificales. PmGH comprises an

Accepted 29 October, 2013. For correspondence. *E-mail martin.walsh@diamond.ac.uk; Tel. (+44) 12 3577 8518; Fax (+44) 12 3577 8448. **E-mail rpryan@dundee.ac.uk; Tel. (+44) 1382 384272; Fax (+44) 1382 385519. ***E-mail m.dow@ucc.ie; Tel. (+353) 21 4901316; Fax (+353) 21 4275934.

HD-GYP domain fused to a GAF domain. We have also determined structures for this protein in complex with its substrate (c-di-GMP) and final reaction product (GMP). The structures reveal the mode of binding of the dinucleotide and shed light on the catalytic mechanism. A remarkable feature of the structure of *PmGH* was the identification of a trinuclear Fe centre which is buried at the bottom of the cavity forming the c-di-GMP binding site. Adequate space is available for the substrate to bind dynamically and interact with the metal centre to sequentially hydrolyse c-di-GMP to GMP. Knowledge of the conserved residues involved in binding to the novel trinuclear Fe centre together with the analysis of amino acid sequence alignment of a large cohort of HD-GYP domain proteins suggests further classification of this family into two distinct subgroups containing either a bi- or trinuclear metal centre.

Results

Target selection and structure determination

Initial attempts to crystallize the archetypal HD-GYP domain protein RpfG from *Xanthomonas campestris* (Ryan *et al.*, 2006) failed. An extensive search of structural homologues with standalone HD-GYP domains and/or in combination with different sensor and ligand binding domains was then made; this extensive list was rationalized through bioinformatics analysis using criteria known to increase the success rate of crystallization (Slabinski *et al.*, 2007). Of a cohort of 15 proteins, a GAF/HD-GYP domain-containing protein from *P. marina* EX-H1 (*PmGH*) gave crystal hits that were further optimized for structure solution. *PmGH* was shown to have c-di-GMP phosphodiesterase activity both in the crystalline form (see below) and in solution (Fig. 1). The structure was solved by the single wavelength anomalous diffraction (SAD) technique using the anomalous signal from the bound iron (Table 1).

Protein architecture

PmGH forms a dimer with each monomer consisting of an N-terminal GAF domain connected to a C-terminal HD-GYP domain by an approximately 42 residue-long helix ($\alpha 5$ in Fig. 2A). Assembly of the head-to-head dimer relies exclusively on the GAF domain and the long $\alpha 5$ helix with the HD-GYP domain playing no role in the dimeric interface (Fig. 2A). The overall topology of the GAF domain is similar to that of other GAF domain structures (Ho *et al.*, 2000; Kanacher *et al.*, 2002; Martinez *et al.*, 2002), consisting of a six-stranded antiparallel β -sheet ($\beta 3$ - $\beta 2$ - $\beta 1$ - $\beta 6$ - $\beta 5$ - $\beta 4$) sandwiched by a three-helix bundle ($\alpha 1$, $\alpha 2$ and $\alpha 5$) on one side and two short helices ($\alpha 3$ and $\alpha 4$) on the other.

The catalytic HD-GYP domain contains the characteristic HD domain superfamily 5-helix core formed by

helices $\alpha 6$ to $\alpha 10$ that in turn provides the scaffold for sequestering the tri-iron centre through eight conserved protein side-chain ligands (Fig. 2B and Fig. 3). The signature HD motif forms part of this octet (H221 and D222) and is located on $\alpha 7$ at a kink close to the C-terminal end of the helix. Another four conserved histidines provide metal ligands through H189 located at the start of $\alpha 6$, H250 from $\alpha 8$ and H276/277 at the end of $\alpha 9$. D305 from $\alpha 10$ completes the protein metal ligands of the tri-iron centre. In addition to the HD domain 5-helix core, the HD-GYP domain contains two extra C-terminal helices, a short helix $\alpha 11$ and finally $\alpha 12$ which is significantly bent and allows it to pack against helices $\alpha 6$ and $\alpha 10$ (Fig. 2B).

The HD-GYP domain of *PmGH* in its entirety resembles an opened two-clawed chela. One of the chela's claws is comprised of the loop (L7/8) connecting $\alpha 7$ and $\alpha 8$ as well as the start of $\alpha 8$, whereas the other claw is entirely formed by the loop connecting helices $\alpha 10$ and $\alpha 11$ (L10/11) (Fig. 2B). The loop region connecting helices $\alpha 9$ and $\alpha 10$ contains the signature GYP motif and forms a well-ordered structure made up of two orthogonally orientated U-turns as observed in the structure of an inactive HD-GYP domain protein (Lovering *et al.*, 2011). The conserved Y285 of the GYP motif points towards the metal binding centre which is buried in the cavity formed by the chela. The 'GYP' loop forms a barrier on one side of the opened chela, whereas the other side is unobstructed. A sequence-based conserved motif in HD-GYP proteins mapping to loop L9/10 was previously identified to be HHExxDGxGYP (Ryan *et al.*, 2006); however, the *PmGH* structure reveals that L9/10 is composed of 19 residues suggesting an extension of the consensus sequence to HHExxDGxGYPxxxxxxxi, which includes a conserved isoleucine residue (I294 in *PmGH*) that stabilizes the closure of the second U-turn by hydrophobic interactions with G284 from the GYP motif.

Trinuclear Fe binding site

The HD-GYP domain of *PmGH* contains a trinuclear metal centre that is located at the bottom of the cavity formed by the two claws of the open chela (Fig. 2B; Fig. 3A and B). Anomalous diffraction difference maps showed all three metal sites to be occupied by iron atoms, with anomalous peak heights in excess of 40σ separated by approximately 3.5 Å (Fig. 3C). The bond valence sum method (Brown and Altermatt, 1985) was used to estimate the oxidation states of the three Fe ions which assigned the two peripheral sites as occupied by Fe(II) and the middle site by Fe(III). Over-expression of protein from bacteria grown in minimal media and supplementing with divalent metal ions such as manganese resulted in mixed metal occupancies in the peripheral metal sites but the middle metal site was always observed to be occupied by Fe (Fig. 3C and Fig. S3).

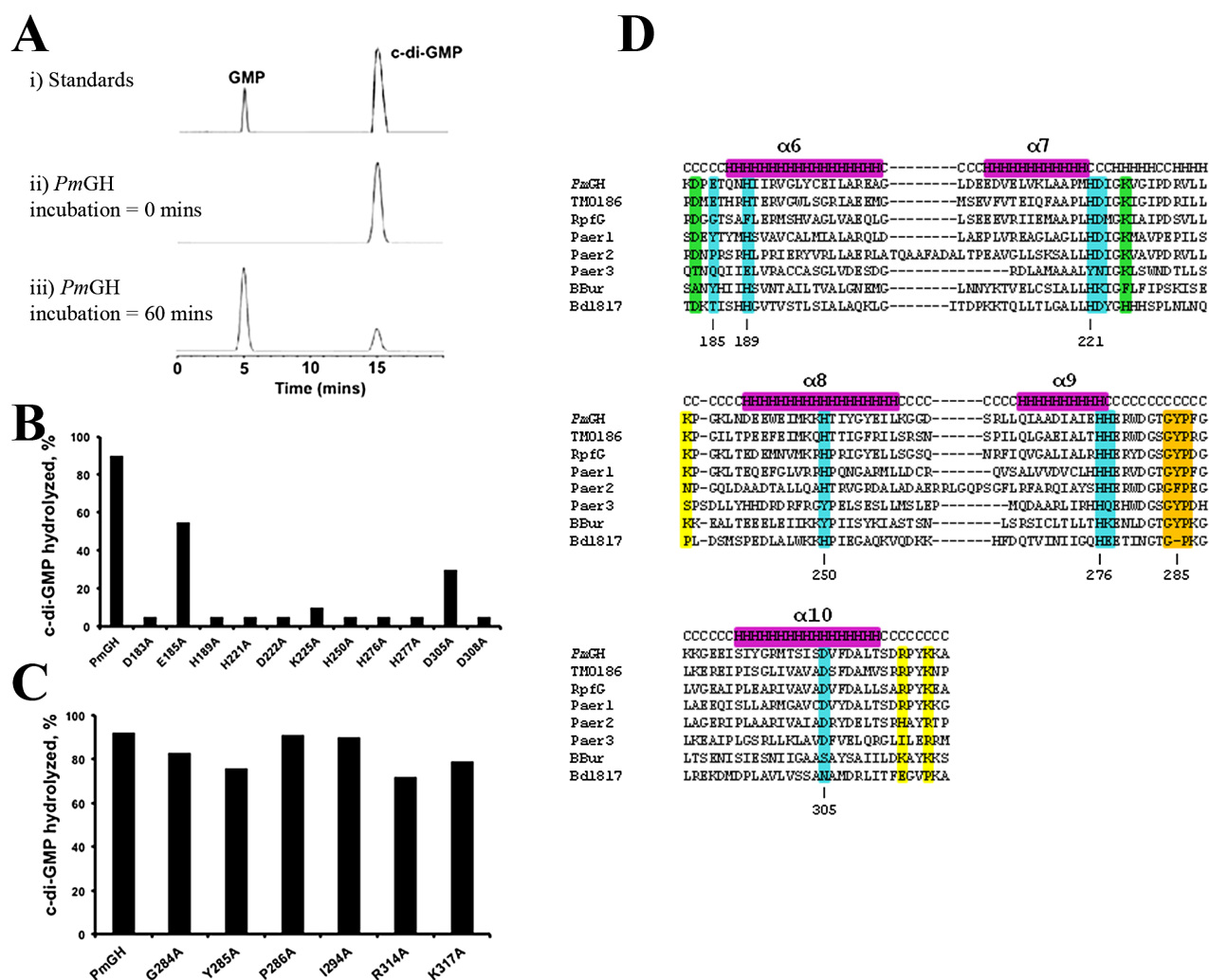


Fig. 1. PDE activity of purified HD-GYP domain *PmGH* and variants with alanine substitutions. **A.** Representative HPLC traces showing standards (i), aliquots of reaction mixtures boiled at 0 min (ii) and after 60 min (iii) incubation with the *PmGH* protein. The identity of the product was confirmed by mass spectrometry. **B.** Effects of alanine substitutions in metal ligands (E185, H189, H221, D222, H250, H276, H277 and D305), a strongly conserved residue in the family (D308) and the two putative catalytic residues (D183 and K225) on cyclic di-GMP hydrolysis. **C.** Effects of alanine substitutions in residues in the GYP motif (G284, Y285 and P286), the conserved I294 position and other residues involved in substrate binding (R314 and K317) on cyclic di-GMP hydrolysis. **D.** Primary sequence alignment of HD-GYP domains of *PmGH* with some of the most well-characterized HD-GYP proteins, such as TM0186 (*Thermotoga maritima*), RpfG (*Xanthomonas campestris* pv. *campestris*), Paer1-3 (*Pseudomonas aeruginosa*), BBur (*Borrelia burgdorferi*) and Bd1817 (*Bdellovibrio bacteriovorus*). Metal ligands, catalytic residues, substrate ligands and GYP motif, based on the *PmGH* structure, are highlighted in cyan, green, yellow and orange respectively.

These data support the oxidation state assignment and indicate that the middle Fe site of the trinuclear centre is specific for Fe(III), while the peripheral sites can accommodate other metals under Fe deficient conditions. The tri-iron centre delineates the floor of the open chela with the two peripheral metal sites being in close proximity to Y285 of the GYP motif on one side (the G-site) and H221 of the HD motif on the other side (the H-site). The middle iron binding site (M-site) is sandwiched between the G- and H-sites with Fe-Fe bond distances of 3.40 Å and 3.67 Å

respectively (Fig. 2B and Fig. 3). All three metals are octahedrally co-ordinated (Fig. 3A and B). In *PmGH* the HD-GYP domain conserved residues E185, H189, H221, D222, H250, H276, H277 and D305 contribute to ligand binding with the metal co-ordination sphere completed by succinate and imidazole bound from the crystallization buffer as well as two solvent molecules which bridge the G-M and M-H metal pairs (Fig. 3 and Fig. S1). Both pairs of Fe sites, G-M and M-H, are triply bridged: the G-M pair by the carboxylate groups of D222 and a succinate from the

Table 1. Data collection and refinement statistics.

	<i>PmGH</i> Fe peak	<i>PmGH</i> native	<i>PmGH</i> c-di-GMP	<i>PmGH</i> GMP
Data collection				
X-ray source	Diamond I24	Diamond I02	Diamond I02	Diamond I04-1
Wavelength (Å)	1.74	0.95	0.98	0.92
Resolution range (Å) ^a	71–2.70 (2.8–2.70)	72–2.03 (2.1–2.03)	66–2.68 (2.75–2.68)	71–2.55 (2.62–2.55)
Space group	I222	I222	I222	I222
Unit cell parameters				
a, b, c (Å)	69.7, 180.8, 232.2	69.8, 182.4, 232.5	70.2, 183.0, 233.5	70.1, 181.1, 231.5
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
No. of observations	102 579	417 588	189 599	254 040
No. of unique observations	38 405	95 384	42 290	48 066
<i>R</i> _{merge} (%)	12.2 (72.1)	4.1 (56.0)	7.7 (68)	8.6 (68)
Completeness (%) ^a	98.7 (97.3)	99.8 (98.0)	99.4 (99.6)	99.2 (98.7)
Mean <i>I</i> /σ ^B	28.3 (5.3)	15.2 (2.3)	14.3 (1.9)	12.7 (2.2)
B wilson	–	52.42	69.14	60.06
Refinement				
<i>R</i> _{work} / <i>R</i> _{free} (%)	–	19/21	18/23	18/22
Rms deviations, bonds/angles	–	0.011/1.36°	0.013/1.72°	0.010/1.42°
Average B factor (Å ²)	–	52.8	45.2	37.3
Ramachandran favoured	–	96.8%	96.1%	96.4%

a. Values in parentheses refer to the high resolution shell. Anomalous pairs were kept separate during merging of all datasets.

crystallization buffer in a bidentate fashion and by a monodentate bridged solvent molecule; whereas the M-H pair is bidentately co-ordinated by the carboxylate groups of E185 and a second succinate ion and by another monodentate bridged solvent molecule (Fig. 3A and B). Analysis of metal-ligand bond lengths for the bridging solvent molecules is consistent with the bridging ligand being a hydroxide ion. This correlates with other HD domains containing diiron centres (Brown *et al.*, 2006; Lovering *et al.*, 2011).

C-di-GMP and GMP binding

Structures of complexes of *PmGH* with GMP and c-di-GMP were determined from crystal soaking experiments with both nucleotides. Crystal soaks with GMP and c-di-GMP gave identical structures of *PmGH* in complex with GMP showing that *PmGH* retained phosphodiesterase activity in the crystal (Fig. 4A). However, pre-soaking of *PmGH* protein crystals in the presence of 100 mM EDTA followed by c-di-GMP revealed well defined difference electron density for a bound c-di-GMP molecule at the active site, displacing the metal bound succinate ions found in the nucleotide free structure. The c-di-GMP was modelled unambiguously with the two guanines in a *cis* conformation so that the molecule presents a V-shaped conformation when bound to *PmGH* (Fig. 4B). This is in contrast to what is observed in EAL domain proteins where c-di-GMP is bound in a more extended conformation (Navarro *et al.*, 2009). Crystal packing presents a less solvent exposed metal binding site for one of the monomers of the *PmGH* dimer and binding of nucleotide is observed only to

monomer B. The structure reveals that only the G-site Fe remains bound in this monomer of *PmGH* in these EDTA treated crystals, whereas the other less accessible monomer subunit still has Fe bound at all three metal sites, but with reduced occupancy (~ 50%) for the M and H sites and shows no difference density for bound nucleotide.

The bound c-di-GMP is buried within the large pocket formed primarily by the HD-GYP domain claws (65% of its accessible surface, 523 Å² buried), with one of the phosphate groups pointing towards the tri-iron centre, while the ribose and guanine bases are stacked against the claws of the chela (Fig. 4B and C). Superposition of the *PmGH* monomer bound to c-di-GMP, which has only the G-site Fe occupied, onto the tri-iron *PmGH* structure shows the *PmGH* c-di-GMP complex structure to be essentially unchanged on binding of c-di-GMP apart from E185 which is disordered due to the loss of the M-H Fe metal pair as well as better defined density for the residues of loop L10/11 which interact directly with the bound c-di-GMP. This analysis shows the bound c-di-GMP to interact with the middle Fe(III) (M-site) through one of the non-bridging phosphate oxygens of one hydrolysable phosphate group and to the hydroxyl group of Y285, the signature Y residue of the GYP motif, through a hydrogen bonding interaction with a non-bridging oxygen of the other hydrolysable phosphate (Fig. 4B). One of the guanine bases, base-1, forms three hydrogen bonds to *PmGH*, two through the C6 carbonyl oxygen of the guanine with the NH1 guanidinium group of R314 and the NZ amine nitrogen of K317 and a third through the guanine N7 atom and the NH1 guanidinium group of Arg-314 (Fig. 4B). Guanine base-2 on the other hand interacts with *PmGH* only through the C2 amine

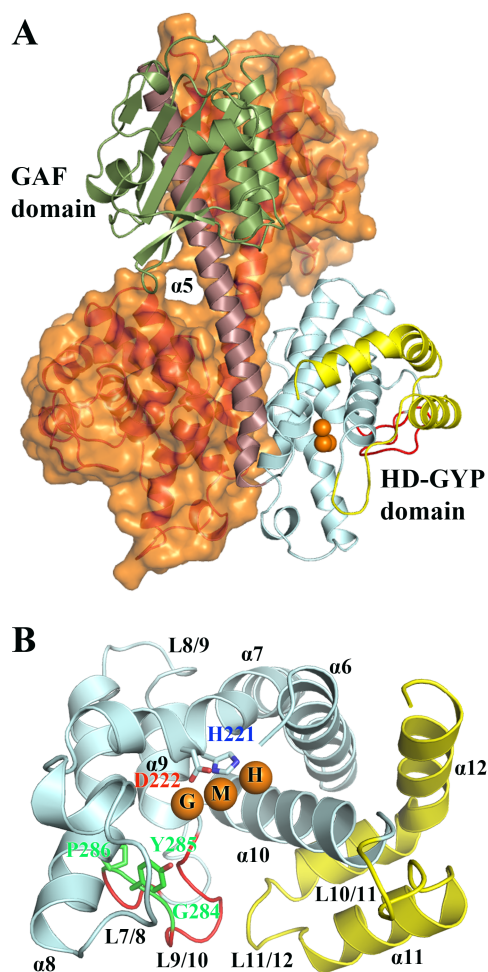


Fig. 2. Structure of *PmGH*.

A. Structure of the *PmGH* homodimer. Molecule A of the dimer is shown in ribbon representation with the GAF domain in green, the long inter-domain dimerization helix $\alpha 5$ in maroon, the core HD domain in cyan, the GYP motif-containing loop in red and the additional surface decorating α -helices which complete the HD-GYP domain in yellow. Molecule B of the dimer is shown in ribbon representation in orange with a semi-transparent surface. The trinuclear iron centre is shown as orange spheres. B. Detailed view of the HD-GYP domain of *PmGH* in ribbon presentation. Colour codes as in (A) with the addition of the HD and GYP motif residues shown in ball and stick. Labelling for α -helices and turns are shown. The central metal iron has been labelled as the middle site (M) and the two flanking metal sites as H and G, to reflect their proximity to the HD and GYP motifs respectively.

group that makes a hydrogen bond to the main-chain carbonyl of K235. Hydrophobic interactions with Y44, A309 and L310 complete the binding interactions of *PmGH* with c-di-GMP (Fig. S2). Comparison of GMP and c-di-GMP binding shows that the guanine base of the GMP molecule superposes with the guanine base-1 of c-di-GMP (Fig. 4D). In the case of the tri-iron *PmGH*-GMP complex, the GMP phosphate moiety bridges the M and H metal sites as

opposed to the inferred sole interaction of one of the hydrolysable phosphate groups with the middle Fe M-site based on the crystal structure of the mononuclear Fe *PmGH*-c-di-GMP complex (Fig. 4A and D).

Mutational analysis of the role of key residues

Alanine substitutions of metal ligands in *PmGH* (E185A, H189A, H221A, D222A, H250A, H276A, H277A and D305A) essentially abolished the phosphodiesterase activity in all cases except for E185 and D305, where

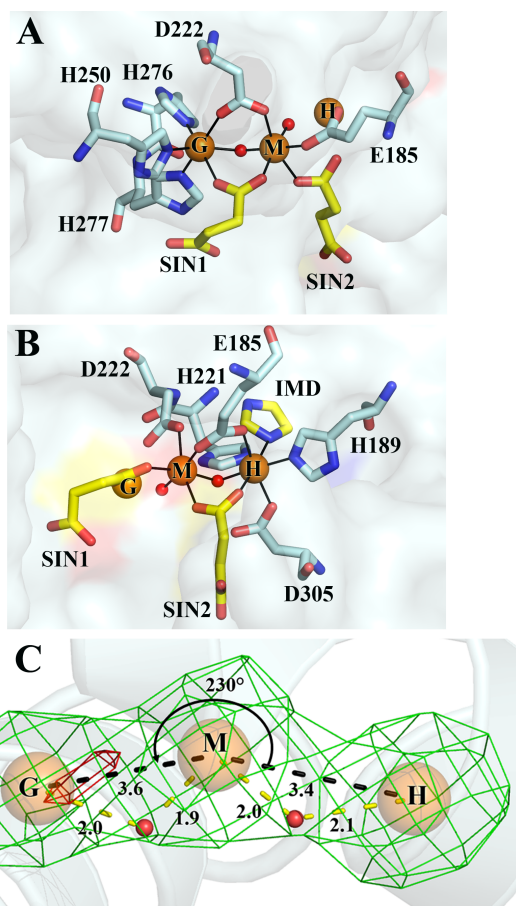


Fig. 3. The tri-iron metal centre of *PmGH*. Detailed views of the tri-iron centre showing the first co-ordination sphere for the G-M metal pair (A) and for the M-H metal pair (B) with the co-ordination of the H and G sites not shown in (A) and (B) respectively for clarity. Protein metal interactions are highlighted as black lines. Fe atoms are shown as orange spheres. Protein side-chain metal ligands are in stick mode, coloured by atom type with carbon in pelican, while the carbon atoms of the metal ligands from the crystallization buffer, 2 succinates SIN-1,2 and an imidazole ion (IMD), are in yellow. (C) Fe-specific difference DANOMAP (Than *et al.*, 2005) in green and Mn anomalous difference map in red, both contoured at $0.043 \text{ e}\text{\AA}^{-3}$. The angle subtended by the tri-iron centre is shown. Black dashed lines depict Fe-Fe distances, while yellow dashed lines indicate bond distances for the pair of μ -hydroxides. Bond distances are in Ångstroms.

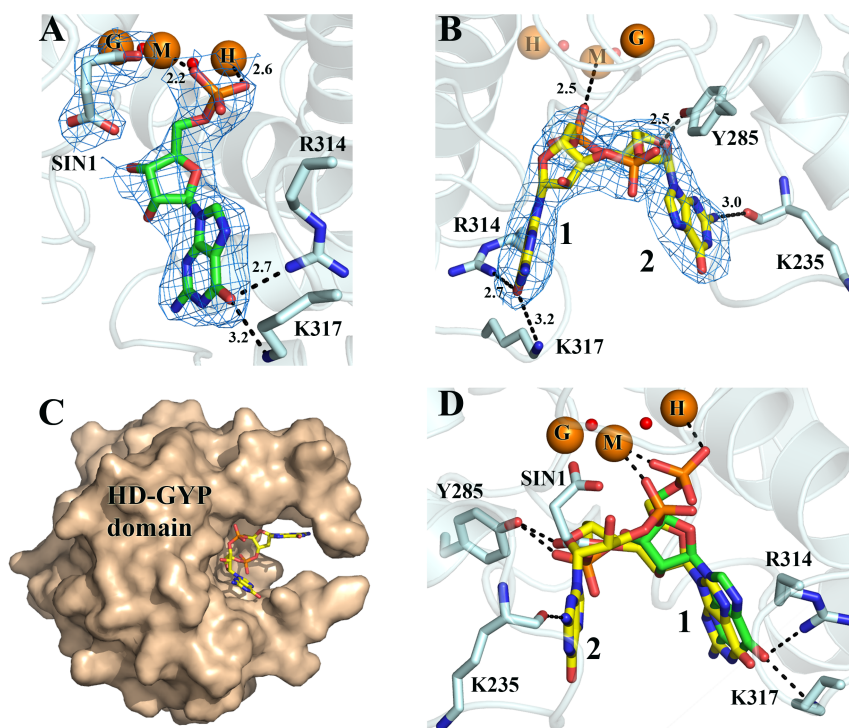


Fig. 4. Substrate binding by *PmGH*.

A. View of GMP shown in stick mode and coloured by atom type bound to *PmGH*. Bonding interactions are represented by dashed lines with distances in Angstroms. Difference electron density map for GMP is contoured at 2σ .

B. View of cyclic di-GMP bound to a metal depleted subunit of *PmGH*. Electron density for cyclic di-GMP is from a bias-removed omit map contoured at 2σ . Fe atoms occupying the middle (M) and HD (H) sites are shown as semi transparent spheres as they are not present in the subunit (due to chelation by EDTA), and are taken from superposition of the equivalent metals from the high resolution structure of *PmGH*.

C. Surface representation of the *PmGH* HDGYP domain monomer subunit showing the binding cavity for cyclic-di-GMP, which is represented in stick mode and coloured by atom type.

D. Superposition of the structures of *PmGH* bound to cyclic di-GMP and GMP. Both nucleotides are shown in stick mode.

although there is a marked reduction in their ability to hydrolyse c-di-GMP, hydrolysis of c-di-GMP is still detected (Fig. 1B). Mutations in the GYP motif and conserved residues implicated in c-di-GMP recognition (G284A, Y285A, P286A, I294A, R314A and K317A) did not however result in a substantial decrease in catalytic activity (Fig. 1C). Alanine mutation of other conserved residues near the metal centre (D183, D308 and K225) had a similar impact on activity as for metal ligand residues. D308 is located at the C-terminal of helix α 10 forming hydrogen bonds with R192 at the N-terminal of α 5 and the metal ligand H189 highlighting a structural role in stabilizing the metal centre and the HD fold, while K225 contributes to stabilization of the tri-iron centre through hydrogen bonds with the metal ligands E185, D222 and also D183 (Fig. 1B).

Discussion

HD-GYP domain-containing proteins are a large family of the HD superfamily of metal-dependent phosphohydrolases (Aravind and Koonin, 1998). The predicted role of

the HD-GYP domain as a PDE active against c-di-GMP (Galperin *et al.*, 2001) was first demonstrated for RpfG from *X. campestris* (Ryan *et al.*, 2006; Ryan, 2013) with further examples characterized from other bacteria such as *Pseudomonas aeruginosa* (Ryan *et al.*, 2009; Stelitano *et al.*, 2013), *Borrelia burgdorferi* (Sultan *et al.*, 2011), *Thermotoga maritima* (Plate and Marletta, 2012) and *Vibrio cholerae* (Miner *et al.*, 2013). The crystal structure of a HD-GYP protein from *P. marina* (*PmGH*) and its complex with c-di-GMP provides the first structure of an active HD-GYP domain protein that reveals a trinuclear Fe centre. The mode of binding of the cyclic nucleotide differs significantly from that observed for the more extensively characterized EAL domain containing PDEs (Schirmer and Jenal, 2009) while the binding site provides adequate room to allow both hydrolysable phosphates to interact in turn with the metal centre to complete hydrolysis of the c-di-GMP to GMP. The mode of c-di-GMP binding to the tri-iron site extends the diversity of structure and function of bacterial c-di-GMP phosphodiesterases.

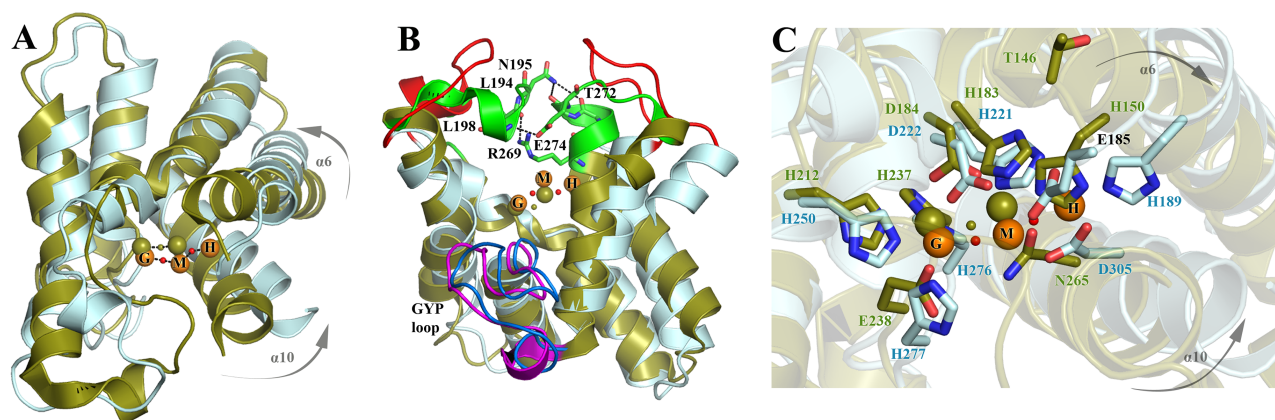


Fig. 5. Comparison of the structures of the HD-GYP domain from *PmGH* and the unconventional HD-GYP domain of Bd1817.

A. *PmGH* and Bd1817 PDE domains are shown in ribbon representation and coloured cyan and olive green respectively. The *PmGH* metal centre is as for Fig. 1, while the Bd1817 binuclear Fe centre with a bridging hydroxide ion is shown as green spheres. The shift in orientation in helices $\alpha 6$ and $\alpha 10$ in *PmGH* when compared with Bd1817 which allows the protein to accommodate the trinuclear metal centre is highlighted.

B. Comparison of the HD-GYP domain nucleotide-binding pocket highlighting the opened and closed conformations observed for *PmGH* and Bd1817 in red and green respectively. Hydrogen bonds proposed to hold the Bd1817 in this close conformation are shown as black dashed lines. The double headed black arrow depicts the distance in Angstroms between the L7/8 and L10/11 loops in *PmGH* for which the active site is in an open conformation.

C. Secondary structure superposition using protein fragments ranging from first to last residue involved in metal co-ordination of *PmGH* and Bd1817 showing only the protein metal ligands.

Structural and chemical relationships of the trinuclear Fe centre

Although all HD domains share key design features, a striking diversity of catalytic centres have now been identified, containing no metal, mono- or binuclear metal centres, and here a trinuclear metal binding site. In *PmGH* the tri-iron site presents a structure which is distinct from the more extensively studied oxo-centred equilateral triangle trinuclear iron complexes, although the first protein structure with such an oxo-centred iron cluster was only determined relatively recently (Hogbom and Nordlund, 2004). Here, the tri-iron centre of *PmGH* has striking similarities to a series of linear tri-iron complexes that were obtained while designing mimetics for studying primarily carboxylate-bridged bimetallic centres in metalloproteins (Rardin *et al.*, 1992; Kitajima *et al.*, 1993). However in *PmGH* the tri-iron centre has an extended V shape (Fig. 3C). The G and M metal sites of the tri-iron centre align closely with the diiron metal centre of Bd1817, an inactive HD-GYP protein (Lovering *et al.*, 2011) (Fig. 5A) and that of other HD domain containing binuclear Fe sites such as myo-inositol oxygenase (Brown *et al.*, 2006). Comparison of the structure of *PmGH* to Bd1817 shows the overall fold of the HD-GYP domain is maintained but reveals a slight re-orientation of the first and fifth α -helices of the 5-helix HD core domain to accommodate the tri-iron centre (Fig. 5A and C). *PmGH* presents an accessible and far larger binding cavity than is the case for Bd1817 where

the active site is capped by a 'lid' formed by loops corresponding to L7/8 and L10/11 in *PmGH* (Fig. 5B). The Bd1817 interactions responsible for keeping the lid tightly fixed in place are unique to Bd1817 (Fig. 1D), suggesting that these mutations have evolved to intrinsically fix the lid in a closed conformation to prevent binding of nucleotide keeping the protein enzymatically inactive. Despite the fact that a significant number of enzymatically inactive EAL and HD-GYP domains have been demonstrated to play a role in signal transduction and regulation through c-di-GMP binding and/or protein interaction (Navarro *et al.*, 2009; Ryan *et al.*, 2009), the regulatory role of Bd1817 still remains to be revealed (Lovering *et al.*, 2011).

Our observation of the tri-iron centre in the HD-GYP domain of *PmGH* prompted us to perform a phylogenetic comparison of this domain with all relevant proteins in the National Center for Biotechnology Information (NCBI) database. This analysis showed a distinct separation of the HD-GYP domains into two evolutionary groups (Fig. 6). This bipartition is independent of the type of regulatory and/or sensory domain associated with the HD-GYP domain (Fig. 6). Analysis of the sequence alignments highlighted that only seven out of the eight *PmGH* metal ligand residues were shared between these two groups. The one variable ligand corresponds to E185 in *PmGH* which provides a bidentate carboxylate ligand which bridges the M and H metal sites. In the group containing all eight ligands, which includes *PmGH* and TM0186 from *Thermotoga maritima*, a REC/HD-GYP protein that has been characterized



Fig. 6. Maximum parsimony analysis of various HD-GYP domain sequences. The evolutionary history was inferred using the maximum parsimony (MP) method. Tree #1 out of 17 most parsimonious trees (length = 3871) is shown. The consistency index is 0.257040 (0.255115), the retention index is 0.444680 (0.444680), and the composite index is 0.114300 (0.113445) for all sites and parsimony-informative sites (in parentheses). The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Suzuki *et al.*, 2002) with search level 0 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved 122 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 116 positions in the final dataset. Sequence labels are in the following format: (1) organism name followed by a number if more than one sequence is present from the same organism; (2) a +/- sign indicating Gram + or Gram -; (3) other domains present besides the HD-GYP domain are indicated, followed by a number in case of multiple copies; (4) a 3 letter code for the 3 residues triplet subfamily signature corresponding to positions 185–187 in *PmGH*. Extra domains are: REC = CheY homologous receiver domain; GAF = present in cyclic di-GMP phosphodiesterase, Adenyl cyclase, FhlA; PAS = present in Periodic circadian protein, Ah receptor nuclear translocator protein, Single-minded protein; HAMP = present in Histidine kinases, Adenyl cyclases, Methyl-accepting proteins and Phosphatases; HD = extra HD domain of unknown function missing the GYP motif; TPR = domain containing the Teratric Peptide Repeat region; GGDEF = diguanylate cyclase containing the GGDEF motif; SBP = bacterial extracellular Solute-Binding Protein; DUF = Domain of Unknown Function. Evolutionary analyses were conducted in Mega5 (Tamara *et al.*, 2011).

as an active c-di-GMP phosphodiesterase (Plate and Marletta, 2012), the equivalent of the metal ligand E185 is invariably found to be conserved or replaced by an aspartate. A threonine and a glycine consistently follow this conserved metal carboxylate ligand (E/D), giving the signature motif E/D-T-G for this subfamily. Conversely the other subfamily primarily presents a tyrosine or phenylalanine (Y/F) in place of the conserved carboxylate ligand of E185. Despite the lack of a unique signature triplet as in the E/D-T-G group (except in some rare cases), there do exist some well-populated sequence clusters within this second Y/F subfamily identifiable by motifs such as Y-T-Y, Y/I-L-L or F-T-F. The Y/F mutation of the carboxylate metal ligand that in the *PmGH* structure bidentately bridges the M and H metal sites would be expected to impact on the formation or stability of the tri-iron centre. Thus, the Y/F subfamily is more likely to contain a binuclear metal centre in contrast to the E/D-T-G subgroup which are predicted to have a trinuclear metal centre like *PmGH*. The boundary between these two subfamilies is not entirely clear-cut however. For example, RpfG from *X. campestris*, despite phylogenetically clustering within the E/D-T-G subgroup, aligns a glycine in place of the E/D residue, as well as a phenylalanine replacing the H-site metal ligand H189 of *PmGH* (Fig. 1D). The loss of two out of the eight metal ligands utilized by *PmGH* to bind the trinuclear Fe centre suggests that RpfG is more likely to possess a binuclear metal ion centre. This phylogenetic analysis, while suggesting substantial diversity within the HD-GYP family of signalling proteins, at the same time identifies a potential quasi-equal distribution of putative bi- and trinuclear metal centres. Additional crystal structures of representative HD-GYP domain proteins will be required to clearly understand the structural evolution of this family of PDEs and the requirement for a bi- or trinuclear metal centre for catalysis.

Structural insights into the phosphodiesterase activity of the HD-GYP domain

The structure of the *PmGH*-c-di-GMP complex predicts that the M-site Fe(III) directly interacts with a non-bridging oxygen of one of the scissile phosphate diesters of the c-di-GMP substrate to provide a strong Lewis acid catalyst, while the metal-activated bridging hydroxide ion of the M-H Fe pair is the likely nucleophile for the hydrolysis of the scissile bond as similarly proposed for other metallophosphatases (Williams *et al.*, 1999). However, this structure does not provide a definitive answer to how the O3' leaving group is protonated. The closest conserved residues to the O3' are D183, E185 and K225, with the carboxylate group of E185 situated less than 4 Å away; but the principle role of E185 is expected to be in metal binding. Moreover, the E185A mutant is still capable of hydrolysing c-di-GMP although significantly less effectively than the

native protein while alanine mutation of either D183 or K225 abolishes enzyme activity (Fig. 1A). In the c-di-GMP-*PmGH* complex, the D183/K225 pair is over 5 Å away from the O3' of c-di-GMP. As the structure of the c-di-GMP-*PmGH* complex only contains the G-site Fe (due to EDTA treatment of crystals), binding of the c-di-GMP with an intact trinuclear metal centre is expected to influence the mode of binding of the nucleotide which could place D183 appropriately to act as a general acid for protonation of the O3' leaving group with K225 stabilizing the unprotonated state of D183. Protonation of O3' by bound water has been proposed for EAL PDEs as well as for the HD domain phosphohydrolase YfbR (Zimmerman *et al.*, 2008; Barends *et al.*, 2009; Tchigvintsev *et al.*, 2010) but in the structures presented here no suitably positioned water is observed bound.

The complex of c-di-GMP with *PmGH* shows the nucleotide to bind with the guanine bases in a *cis* conformation which differs to EAL domain proteins, in which c-di-GMP adopts a more extended conformation (Navarro *et al.*, 2009). Similar *cis* conformations for c-di-GMP have been observed in other structures, for example in complexes with riboswitches, the degenerate GGDEF domain protein PeID, the c-di-GMP effector protein domain PilZ (Habazettl *et al.*, 2011; Smith *et al.*, 2011; Li *et al.*, 2012), and with the human STING proteins (Shu *et al.*, 2012), which are stimulators of interferon genes. The *cis*-conformation together with the space provided by the nucleotide binding cavity is proposed to facilitate the sequential hydrolysis of c-di-GMP by HD-GYP PDEs. Based on the structure of the complex of *PmGH* with the final reaction product GMP, which shows the 5'-phosphate of GMP bound to the M-H Fe pair, c-di-GMP hydrolysis may be catalysed by these two sites alone, with the G-site Fe contributing to the hydrolysis rate. This would invoke primarily a two-metal-ion mechanism as seen for other PDEs (Barends *et al.*, 2009; Tchigvintsev *et al.*, 2010). In this case, the two scissile c-di-GMP phosphates would be sequentially hydrolysed by the M-H Fe pair with the phosphate group which initially binds to the hydroxyl group of Y285 being brought into a similar conformation to the phosphate group bound to the M-site Fe(III). After hydrolysis of this M-site bound phosphate group, the hydrolysable phosphate group of 5'-phosphoguananylyl-(3'-5')-guanosine (5'-pGpG), the product of this first hydrolysis step, can be positioned in a similar conformation to the previously hydrolysed phosphate by a rotation around the original twofold axis of the intact c-di-GMP. This would then enable the conversion of the now bound 5'-pGpG to GMP. Equally, it is conceivable that the G-M Fe pair could also contribute to hydrolysis but the capture of the complex between *PmGH* and c-di-GMP due to the presence only of the G-site Fe suggests this Fe may play more of a structural role, although clearly a critical one as single alanine mutations of the conserved protein metal ligands of the G-site

Fe renders the protein inactive. In closing, a direct role in catalysis for all three metal sites in *PmGH* cannot be ruled out as the enzyme has adapted a novel tri-iron centre with the metal ions closely spaced (≤ 3.64 Å) and single point mutations of any of the eight protein metal ligands abolished activity in all cases except for E185 and D305. The impact of these mutations which would disrupt binding of the M-H ion pair maybe artificially mitigated by binding of phosphate or carboxylate ions from the cell as the ligands are positioned on the solvent facing side of the metal centre. The mutations do impair the activity of *PmGH* and thus the predicted distribution of bi- and trinuclear centres in the HD-GYP family may contribute to modulation of PDE activity. Further structural work will be required to decipher unequivocally the roles of the three metal Fe sites which will be aided by a crystal structure of an active binuclear HD-GYP domain together with a global kinetic analysis of both the bi- and tri-nuclear metal containing HD-GYP subfamilies identified here.

Structural insights into the multifunctional roles of HD-GYP domains

Recent studies have described a regulatory role for protein-protein interactions involving the HD-GYP domain protein RpfG. Interaction of RpfG with specific GGDEF domain proteins serves to control motility in *X. campestris* (Ryan *et al.*, 2012). Mutational analysis showed that the GYP motif is critical for HD-GYP::GGDEF complex formation but not necessary for the PDE activity of RpfG against c-di-GMP (Ryan *et al.*, 2010). The structure of the *PmGH* HD-GYP reveals that although the GYP containing loop is surface exposed and well ordered, the Y285 of the GYP motif is placed inside the substrate-binding pocket, where it H-bonds to c-di-GMP. Therefore, the structural data presented in this study suggests that if GGDEF domains interact directly with Y285, they need to intercalate with the inner side of the HD-GYP nucleotide-binding pocket. Such a mechanism would clearly prevent c-di-GMP binding and PDE activity, although this has not been observed *in vitro* (Ryan *et al.*, 2012). An intriguing alternative is that RpfG acts as a trigger enzyme for protein complex formation and regulation in a similar fashion to the EAL domain protein YciR of *Escherichia coli* (Lindenberg *et al.*, 2013). In this scenario, RpfG involvement in protein-protein complexes would be determined not only by c-di-GMP binding but also by conformational alterations associated with c-di-GMP degradation, which would be 'reported' via the GYP loop.

With the work described here we now have structures of representatives of all domains involved in c-di-GMP metabolism. However, further structural work is required to gain a fuller understanding of the relative contributions of the amino acids making up the catalytic site of the enzyme and to understand how the HD-GYP domain can utilize

both bi- and trinuclear metal centres to hydrolyse c-di-GMP which we propose exist here. Furthermore, the elucidation of the structures of additional HD-GYP domains from the different classes that we have defined here and of the HD-GYP domain in complex with the GGDEF domain will be necessary to provide a fuller understanding of the regulatory action of this diverse family of signalling proteins.

Experimental procedures

Cloning and protein production

Initial construct design was aided by bioinformatic tools implemented using the OPAL and OPTIC resources as described by Albeck *et al.* (2006) as well as the XTALPRED and HHPRED servers (Söding *et al.*, 2005; Slabinski *et al.*, 2007). The gene encoding for the GAF/HD-GYP protein from *Persephonella marina* EX-H1 (perma_0986; referred to here as *PmGH*) was optimized for structural studies through mutations of its two cysteines to alanines and deletion of the terminal three amino acids which were predicted to be unstructured (Remmert *et al.*, 2012). The construct was synthesized by Genscript inserted in pUC57 and subcloned into pET47b using the STRU-cloning protocol (Bellini *et al.*, 2011) and transformed into *E. coli* BL21 (DE3). BL21 (DE3) cells were grown in LB media and induced with 0.25 mM IPTG; protein overexpression was carried out at 37°C for 1 h. Purification was achieved by Ni²⁺ affinity chromatography using the N-terminal His6 tag followed by tag cleavage using recombinant HRV 3C protease. Protein was concentrated to 10 mg ml⁻¹ for crystallization experiments.

Crystallization, data collection and structure solution

Diffraction quality crystals were obtained using the hanging drop vapour diffusion method from mixing equal volumes of protein with a precipitant solution made up of 0.1 M MES pH 6.5, 0.9 M succinic acid and 2% PEG 2000. *PmGH* crystallized in space group I222 with one dimer in the asymmetric unit. Crystals were cryoprotected with ethylene glycol [25% (v/v)] prior to flash-cooling in liquid nitrogen. The structure was solved using SAD phasing by exploiting the anomalous scattering from the bound iron atoms that were identified by an X-ray fluorescence spectrum prior to the diffraction experiment (Walsh *et al.*, 1999). Native and SAD data at the Fe-K α absorption edge were collected on Diamond Light Source beamlines I02 and I24 respectively. Nucleotide-*PmGH* complexes were obtained by soaking *PmGH* crystals soaked overnight at 277 K with 25 mM c-di-GMP (BioLog/KeraFast) and 30 mM GMP both in the presence and absence of 100 mM EDTA and data evaluated and collected on Diamond beamlines I02, I03, I04 and I04-1. The structure of *PmGH* was determined from a single wavelength anomalous diffraction experiment at the Fe absorption edge determined experimentally on beamline I24 and data extended against a high resolution data set (2.03 Å) collected from a different crystal using beamline I02. All data were processed using XDS within XIA2 (Kabsch, 2010; Winter, 2010) and the Fe sub-

structure was determined using SHELX (Sheldrick, 2010). Building of the structure was aided by automated procedures in Buccaneer (Cowtan, 2006) and manual building was performed with COOT (Emsley *et al.*, 2010). The structure was refined with REFMAC5 (Murshudov *et al.*, 1997) and validated with MolProbity (Chen *et al.*, 2010). Figures of structures were prepared with PyMOL (Schrodinger, 2010).

Alteration of residues in the PmGH HD-GYP domain by mutagenic PCR

Site-directed mutagenesis to introduce the alterations in residues involved in metal binding (E185A, H189A, H221A, D222A, H250A, H276A, H277A, D305A and D305A), substrate binding (Y285A, R314A and K317A) and other residues of interest (D183A, K225A, G284A, P286A, I294A and D308A) was done by using mutagenic PCR in a two-step protocol as previously described (Ryan *et al.*, 2006; 2010). In the first round of PCR, two separate reactions were carried out by using the forward and reverse primers together with one of a pair of primers of complementary sequence carrying the desired alteration and the HD-GYP construct in pET47b as template. (Mutagenic primer sequences will be given upon request.) The products of the first round of PCR were used as templates for a second round of PCR with forward and reverse primers.

Enzymatic assays on the HD-GYP domain and its variants

The assay buffer and reaction conditions were as described elsewhere (Ryan *et al.*, 2006). Briefly, a standard reaction mixture contained 20 µg of protein, 50 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 10 mM MnCl₂, 0.5 mM EDTA and 50 mM NaCl in a total volume of 600 µl. The assay mixture was warmed to 37°C before the reaction was started by the addition of 27 µl of substrate to give a final concentration of 100 µM. Aliquots of 200 µl were withdrawn to a sterile Eppendorf tube at the indicated time points, and the assay was terminated by placing the tube in a boiling water bath for 3 min. After centrifugation at 15 000 *g* for 2 min, the supernatant was filtered through a 0.22 µm filter before analysis by reverse-phase HPLC on a Hewlett–Packard Model 1090 Series II HPLC system. Samples of 50 µl were injected into a SunFire C-18-T column (150 × 4.6 mm; Waters) and fractionated by using 2% (v/v) acetonitrile/98% Na phosphate buffer (pH 5.8) under isocratic condition at a flow rate of 0.7 ml min⁻¹. Nucleotides were detected at a wavelength of 252 nm.

Accession numbers

The co-ordinates and structure factors have been deposited in the Protein Data Bank, <http://www.pdb.org> PDB ID codes 4MCW (PmGH), 4MDZ (complex with c-di-GMP) and 4ME4 (complex with GMP).

Acknowledgements

The work of the authors has been supported in part by grants awarded by the Wellcome Trust (Project grants WT093314/Z10/Z (M.A.W.) and WT093314MA to J.M.D. and R.P.R. and

WT100204AIA senior fellowship grant to R.P.R.); the Science Foundation of Ireland (SFI 07/IN.1/B955 to J.M.D. and SFI 09/SIRG/B1654 to R.P.R.). We thank Diamond MX village staff and the Diamond Membrane Protein Laboratory for support. Author contributions: M.A.W., R.P.R. and J.M.D. were responsible for the overall project strategy and management. D.B., D.L.C., J.M.D., R.P.R. and M.A.W. designed research; D.B., D.L.C., Y.M., M.B., S.A., J.M.D., R.P.R. and M.A.W. performed research; D.B., D.L.C., J.M.D., R.P.R. and M.A.W. analysed data; D.B., D.L.C., J.M.D., R.P.R. and M.A.W. wrote the paper.

References

- Albeck, S., Alzari, P., Andreini, C., Banci, L., Berry, I.M., Bertini, I., *et al.* (2006) SPINE bioinformatics and data-management aspects of high-throughput structural biology. *Acta Crystallogr D Biol Crystallogr* **62**: 1184–1195.
- Aravind, L., and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci* **23**: 469–472.
- Barends, T.R., Hartmann, E., Griese, J.J., Beittlich, T., Kirienko, N.V., Ryjenkov, D.A., *et al.* (2009) Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* **459**: 1015–1018.
- Bellini, D., Fordham-Skelton, A., and Papiz, M.Z. (2011) STRU-cloning: a fast, inexpensive and efficient cloning procedure applicable to both small scale and structural genomics size cloning. *Mol Biotechnol* **48**: 30–37.
- Boyd, C.D., and O'Toole, G.A. (2012) Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. *Annu Rev Cell Dev Biol* **28**: 439–462.
- Brown, I.D., and Altermatt, D. (1985) Bond-valence parameters obtained from a systematic analysis of the Inorganic Crystal Structure Database. *Acta Crystallogr B* **41**: 244–247.
- Brown, P.M., Caradoc-Davies, T.T., Dickson, J.M., Cooper, G.J., Loomes, K.M., and Baker, E.N. (2006) Crystal structure of a substrate complex of myo-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism. *Proc Natl Acad Sci USA* **103**: 15032–15037.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12–21.
- Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**: 1002–1011.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486–501.
- Galperin, M.Y., Nikolskaya, A.N., and Koonin, E.V. (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol Lett* **203**: 11–21.
- Habazettl, J., Allan, M.G., Jenal, U., and Grzesiek, S. (2011) Solution structure of the PilZ domain protein PA4608 complex with cyclic di-GMP identifies charge clustering as molecular readout. *J Biol Chem* **286**: 14304–14314.
- Hengge, R. (2009) Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* **7**: 263–273.

- Ho, Y.S., Burden, L.M., and Hurley, J.H. (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J* **19**: 5288–5299.
- Hogbom, M., and Nordlund, P. (2004) A protein carboxylate coordinated oxo-centered tri-nuclear iron complex with possible implications for ferritin mineralization. *FEBS Lett* **567**: 179–182.
- Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**: 125–132.
- Kanacher, T., Schultz, A., Linder, J.U., and Schultz, J.E. (2002) A GAF-domain-regulated adenyl cyclase from *Anabaena* is a self-activating cAMP switch. *EMBO J* **21**: 3672–3680.
- Kitajima, N., Amagai, H., Tamura, N., Ito, M., Morooka, Y., Heerwegh, K., *et al.* (1993) Novel formation of a linear trinuclear ferric complex containing a .mu.-oxo bis(.mu.-acetato) and a .mu.-hydroxo bis(.mu.-acetato) bridge. *Inorg Chem* **32**: 3583–3584.
- Li, Z., Chen, J.H., Hao, Y., and Nair, S.K. (2012) Structures of the PeID cyclic diguanylate effector involved in pellicle formation in *Pseudomonas aeruginosa* PAO1. *J Biol Chem* **287**: 30191–30204.
- Lindenberg, S., Klauck, G., Pesavento, C., Klauck, E., and Hengge, R. (2013) The EAL domain protein YciR acts as a trigger enzyme in a c-di-GMP signalling cascade in *E. coli* biofilm control. *EMBO J* **32**: 2001–2014.
- Loving, A.L., Capeness, M.J., Lambert, C., Hobley, L., and Sockett, R.E. (2011)) The structure of an unconventional HD-GYP protein from *Bdellovibrio* reveals the roles of conserved residues in this class of cyclic-di-GMP phosphodiesterases. *mBio* **2**: e00163-11.
- Martinez, S.E., Wu, A.Y., Glavas, N.A., Tang, X.B., Turley, S., Hol, W.G., and Beavo, J.A. (2002) The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. *Proc Natl Acad Sci USA* **99**: 13260–13265.
- Miner, K.D., Klose, K.E., and Kurtz D.M. Jr (2013) An HD-GYP cyclic di-guanosine monophosphate phosphodiesterase with a non-heme diiron-carboxylate active site. *Biochemistry* **52**: 5329–5331.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240–255.
- Navarro, M.V., De, N., Bae, N., Wang, Q., and Sondermann, H. (2009) Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. *Structure* **17**: 1104–1116.
- Plate, L., and Marletta, M.A. (2012) Nitric oxide modulates bacterial biofilm formation through a multicomponent cyclic-di-GMP signaling network. *Mol Cell* **46**: 449–460.
- Rardin, R.L., Poganiuch, P., Bino, A., Goldberg, D.P., Tolman, W.B., Liu, S., and Lippard, S.J. (1992) Synthesis and characterization of novel trinuclear iron(II) and manganese(II) carboxylate complexes: structural trends in low valent iron and manganese carboxylates. *J Am Chem Soc* **114**: 5240–5249.
- Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**: 173–175.
- Ryan, R.P. (2013) Cyclic di-GMP signalling and the regulation of bacterial virulence. *Microbiology* **159**: 1286–1297.
- Ryan, R.P., Fouhy, Y., Lucey, J.F., Crossman, L.C., Spiro, S., He, Y.W., *et al.* (2006) Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc Natl Acad Sci USA* **103**: 6712–6717.
- Ryan, R.P., Lucey, J., O'Donovan, K., McCarthy, Y., Yang, L., Tolker-Nielsen, T., and Dow, J.M. (2009) HD-GYP domain proteins regulate biofilm formation and virulence in *Pseudomonas aeruginosa*. *Environ Microbiol* **11**: 1126–1136.
- Ryan, R.P., McCarthy, Y., Andrade, M., Farah, C.S., Armitage, J., and Dow, J.M. (2010) Cell-cell signal-dependent dynamic interactions between HD-GYP and GGDEF domain proteins mediate virulence in *Xanthomonas campestris*. *Proc Natl Acad Sci USA* **107**: 5989–5994.
- Ryan, R.P., McCarthy, Y., Kiely, P.A., O'Connor, R., Farah, C.S., Armitage, J.P., and Dow, J.M. (2012) Dynamic complex formation between HD-GYP, GGDEF and PilZ domain proteins regulates motility in *Xanthomonas campestris*. *Mol Microbiol* **86**: 557–567.
- Schirmer, T., and Jenal, U. (2009) Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol* **7**: 724–735.
- Schrodinger, L.L.C. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
- Sheldrick, G. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**: 479–485.
- Shu, C., Yi, G., Watts, T., Kao, C.C., and Li, P. (2012) Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat Struct Mol Biol* **19**: 722–724.
- Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I.A., Lesley, S.A., and Godzik, A. (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**: 3403–3405.
- Smith, K.D., Shanahan, C.A., Moore, E.L., Simon, A.C., and Strobel, S.A. (2011) Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc Natl Acad Sci USA* **108**: 7757–7762.
- Söding, J., Biegert, A., and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**: W244–W248.
- Stelitano, V., Giardina, G., Paiardini, A., Castiglione, N., Cutruzzola, F., and Rinaldo, S. (2013) C-di-GMP hydrolysis by *Pseudomonas aeruginosa* HD-GYP phosphodiesterases: analysis of the reaction mechanism and novel roles for pGpG. *PLoS ONE* **8**: e74920.
- Sultan, S.Z., Pitzer, J.E., Boquoi, T., Hobbs, G., Miller, M.R., and Motaleb, M.A. (2011) Analysis of the HD-GYP domain cyclic dimeric GMP phosphodiesterase reveals a role in motility and the enzootic life cycle of *Borrelia burgdorferi*. *Infect Immun* **79**: 3273–3283.
- Suzuki, Y., Glazko, G.V., and Nei, M. (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc Natl Acad Sci USA* **99**: 16138–16143.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary

- genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.
- Tchigvintsev, A., Xu, X., Singer, A., Chang, C., Brown, G., Proudfoot, M., *et al.* (2010) Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. *J Mol Biol* **402**: 524–538.
- Than, M.E., Henrich, S., Bourenkov, G.P., Bartunik, H.D., Huber, R., and Bode, W. (2005) The endoproteinase furin contains two essential Ca²⁺ ions stabilizing its N-terminus and the unique S1 specificity pocket. *Acta Crystallogr D Biol Crystallogr* **61**: 505–512.
- Walsh, M.A., Dementieva, I., Evans, G., Sanishvili, R., and Joachimiak, A. (1999) Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr D Biol Crystallogr* **55**: 1168–1173.
- Williams, N.H., Lebus, A., and Chin, J. (1999) A structural and functional model of dinuclear metallophosphatases. *J Am Chem Soc* **121**: 3341–3348.
- Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**: 186–190.
- Zimmerman, M.D., Proudfoot, M., Yakunin, A., and Minor, W. (2008) Structural insight into the mechanism of substrate specificity and catalytic activity of an HD-domain phosphohydrolase: the 5′-deoxyribonucleotidase YfbR from *Escherichia coli*. *J Mol Biol* **378**: 215–226.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.