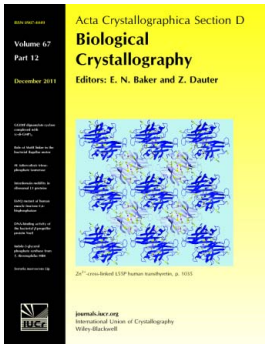
Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449
Editors: **E. N. Baker and Z. Dauter**

The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)$_2$ at the active site

Chao-Yu Yang, Ko-Hsin Chin, Mary Lay-Cheng Chuah, Zhao-Xun Liang, Andrew H.-J. Wang and Shan-Ho Chou

Acta Cryst. (2011). D**67**, 997–1008



Acta Crystallographica Section D: Biological Crystallography welcomes the submission of papers covering any aspect of structural biology, with a particular emphasis on the structures of biological macromolecules and the methods used to determine them. Reports on new protein structures are particularly encouraged, as are structure–function papers that could include crystallographic binding studies, or structural analysis of mutants or other modified forms of a known protein structure. The key criterion is that such papers should present new insights into biology, chemistry or structure. Papers on crystallographic methods should be oriented towards biological crystallography, and may include new approaches to any aspect of structure determination or analysis. Papers on the crystallization of biological molecules will be accepted providing that these focus on new methods or other features that are of general importance or applicability.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)$_2$ at the active site

Chao-Yu Yang,[a] Ko-Hsin Chin,[b] Mary Lay-Cheng Chuah,[c] Zhao-Xun Liang,[c] Andrew H.-J. Wang[d] and Shan-Ho Chou[a,b,e]*

[a]Institute of Biochemistry, National Chung Hsing University, Taichung 40227, Taiwan, [b]Agricultural Biotechnology Center, National Chung Hsing University, Taichung 40227, Taiwan, [c]School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore, [d]Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei, Taiwan, and [e]Graduate Institute of Basic Medical Science, China Medical University, Taichung 40402, Taiwan

Correspondence e-mail: shchou@nchu.edu.tw

Received 26 July 2011
Accepted 30 September 2011

PDB Reference:
XCC4471^{GGDEF}–c-di-GMP complex, 3qyy.

Cyclic diguanosine monophosphate (c-di-GMP) is a key signalling molecule involved in regulating many important biological functions in bacteria. The synthesis of c-di-GMP is catalyzed by the GGDEF-domain-containing diguanylate cyclase (DGC), the activity of which is regulated by the binding of product at the allosteric inhibitory (I) site. However, a significant number of GGDEF domains lack the R*xx*D motif characteristic of the allosteric I site. Here, the structure of XCC4471^{GGDEF}, the GGDEF domain of a DGC from *Xanthomonas campestris*, in complex with c-di-GMP has been solved. Unexpectedly, the structure of the complex revealed a GGDEF-domain dimer cross-linked by two molecules of c-di-GMP at the strongly conserved active sites. In the complex (c-di-GMP)$_2$ adopts a novel partially intercalated form, with the peripheral guanine bases bound to the guanine-binding pockets and the two central bases stacked upon each other. Alteration of the residues involved in specific binding to c-di-GMP led to dramatically reduced K_d values between XCC4471^{GGDEF} and c-di-GMP. In addition, these key residues are strongly conserved among the many thousands of GGDEF-domain sequences identified to date. These results indicate a new product-bound form for GGDEF-domain-containing proteins obtained *via* (c-di-GMP)$_2$ binding at the active site. This novel XCC4471^{GGDEF}–c-di-GMP complex structure may serve as a general model for the design of lead compounds to block the DGC activity of GGDEF-domain-containing proteins in *X. campestris* or other microorganisms that contain multiple GGDEF-domain proteins.

1. Introduction

Cyclic diguanosine monophosphate (c-di-GMP) is a key bacterial secondary messenger that was first described as an allosteric activator of cellulose synthase in *Gluconacetobacter xylinus* (Ross *et al.*, 1987), but is now known to regulate a diverse range of important cellular functions, such as virulence, biofilm formation, cellular morphology and motility, in most eubacteria (Paul *et al.*, 2004; Römling *et al.*, 2005; Jenal & Malone, 2006). c-di-GMP is synthesized from two molecules of GTP by diguanylate cyclase (DGC) containing the GGDEF domain (Chan *et al.*, 2004; Wassmann *et al.*, 2007; De *et al.*, 2008) and is degraded by phosphodiesterase (PDE) containing the EAL domain (Christen *et al.*, 2005; Schmidt *et al.*, 2005) or the HD-GYP domain (Ryan *et al.*, 2006). c-di-GMP has been found to bind to a plethora of effector proteins, including Clp (Leduc & Roberts, 2009; Chin *et al.*, 2010; Tao *et al.*, 2010), PelD (Lee *et al.*, 2007), PleD (Chan *et al.*, 2004; Wassmann *et al.*, 2007), WspR (De *et al.*, 2008), FleQ (Hickman & Harwood, 2008) and PilZ domain-containing proteins (Benach *et al.*,

Table 1
Data-collection and structure-refinement statistics for the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex.

Values in parentheses are for the outermost shell.

	Inflection	High remote	Native
Beamline	NSRRC BL13B1	NSRRC BL13B1	NSRRC BL13C1
Wavelength (Å)	0.97934	0.96415	0.97622
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters (Å, °)	$a = b = 87.54$, $c = 87.88$, $\alpha = \beta = \gamma = 90$	$a = b = 87.54$, $c = 87.88$, $\alpha = \beta = \gamma = 90$	$a = b = 87.28$, $c = 87.86$, $\alpha = \beta = \gamma = 90$
Resolution range (Å)	30–2.50 (2.59–2.50)	30–2.50 (2.59–2.50)	30–1.90 (1.97–1.90)
Total observations	94216 (9544)	93885 (9433)	276945 (11763)
Unique observations	12266 (1193)	12270 (1194)	11378 (1131)
Multiplicity	7.7 (8.0)	7.7 (7.9)	10.1 (10.4)
Completeness (%)	99.7 (100)	99.7 (100)	99.9 (100)
R_{merge}† (%)	11.2 (34.4)	9.5 (29.9)	6.7 (43.9)
$\langle I/\sigma(I)\rangle$	16.8 (6.6)	20.6 (7.5)	33.5 (6.0)
Refinement statistics			
R_{cryst}‡/R_{free}§ (%)	20.36/22.86		
Average B factors (Å^2)			
Protein residues	26.01		
c-di-GMP	15.38		
Mg^{2+} ions	15.25		
Waters	33.70		
No. of atoms			
Protein residues	2445		
c-di-GMP	92		
Mg^{2+} ions	2		
Di(hydroxyethyl) ether	3 × 7		
Waters	186		
Ramachandran plot¶, residues in (%)			
Most favourable regions	99.3		
Additionally allowed regions	0.7		
R.m.s. deviation from ideal geometry			
Bond lengths (Å)	0.005054		
Bond angles (°)	1.15906		

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure-factor amplitudes, respectively. § R_{free} is the same as R_{cryst} but for 5.0% of the total reflections chosen at random and omitted from refinement. ¶ The percentages of residues located in the most favourable and additionally allowed regions were calculated *via* the *MolProbity* program using the default parameters (Chen *et al.*, 2010). No outliers were found.

2007), as well as RNA molecules acting as riboswitches (Sudarsan *et al.*, 2008), to regulate a wide variety of downstream gene expression. Owing to the significant roles played by c-di-GMP in bacterial physiology, its cellular concentration needs to be tightly regulated. This is reflected by the presence of abundant proteins involved in c-di-GMP metabolism in the bacterial kingdom (Galperin *et al.*, 2001). For example, the genome of the plant pathogen *Xanthomonas campestris* (*Xcc*) encodes 37 such proteins: 23 with a GGDEF domain, five with an EAL domain, eight with both GGDEF and EAL domains and three with an HD-GYP domain (Ryan *et al.*, 2007, 2010). In addition, many of these proteins also contain additional sensory and signal transduction domains such as PAS, GAF, REC, HAMP or transmembrane (TM) helix domains (Galperin, 2004), suggesting that c-di-GMP turnover activity is responsive to a wide variety of environmental cues, many of which remain essentially unknown. It is imperative to analyze the structures and functions of these components in order to fully understand the underlying mechanism of the c-di-GMP-related regulation network in bacteria.

The biosynthesis of c-di-GMP is catalyzed by GGDEF domains working cooperatively to condense two molecules of GTP into one molecule of c-di-GMP. The GGDEF domains are usually accompanied by other accessory domains that regulate the catalyzed process (Schirmer & Jenal, 2009). These accessory domains are believed to draw two GGDEF domains together to form a competent homodimer to activate c-di-GMP synthesis. To date, the crystal structures and functions of two full-length GGDEF-domain proteins have been reported: those of the two-component response regulators PleD from *Caulobacter crescentus* (Chan *et al.*, 2004; Wassmann *et al.*, 2007) and WspR from *Pseudomonas aeruginosa* (De *et al.*, 2008, 2009). The PleD protein contains a receiver domain (REC), a degenerate receiver domain (REC′) and a GGDEF domain, while WspR contains a REC domain, a stalk domain and a GGDEF domain. Activation of the GGDEF domain in PleD is induced by the phosphorylation of its REC domain, followed by REC-assisted dimer formation and domain rearrangement to allow an efficient encounter of two symmetric catalytic domains (Chan *et al.*, 2004; Wassmann *et al.*, 2007), while activation of the GGDEF domain in WspR seems to adopt a different phosphorylation-independent route (De *et al.*, 2008, 2009); it becomes active through dimerization of its REC and stalk domains. However, inhibition of DGC activity in both the PleD and WspR regulators seems to be achieved by a similar mechanism, *i.e. via* binding of $(c\text{-di-GMP})_2$ to the allosteric I sites comprising an R*xx*D motif and other less conserved secondary inhibition residues (Schirmer & Jenal, 2009). The noncompetitive product inhibition of PleD and DgcA has been shown to have an inhibition constant of $\sim$1 μM, which is approximately in the range of the cellular concentration of c-di-GMP (Chan *et al.*, 2004; Christen *et al.*, 2006). To date, no other mode of DGC inhibition has been suggested. However, it is important to note that a recent computational analysis of 867 prokaryotic genomes revealed that approximately half of over 10 000 GGDEF sequences lack the canonical allosteric inhibition sites (Seshasayee *et al.*, 2010). Whether these significant numbers of GGDEF-domain proteins are also subject to product inhibition and how this is achieved remain largely unknown. However, it has also been reported that the ability of GGDEF domains to convert GTP to c-di-GMP is intrinsic to these domains, yet is strongly affected by neighbouring protein domains or by other interacting proteins (Ryjenkov *et al.*, 2005).

Here, we report the structure of the GGDEF domain of an *X. campestris* diguanylate cyclase (XCC4471^{GGDEF}) in complex with c-di-GMP. The structure of the complex indicates that two molecules of c-di-GMP reside in the GGDEF active site. This is quite unanticipated, as usually only one molecule of c-di-GMP is expected to be located at the active site. Each c-di-GMP molecule is also found to adopt an unusual partially intercalated conformation. This c-di-GMP binding mode is different to that adopted by c-di-GMP binding at the allosteric inhibitory sites, which adopts a self-intercalated stack comprising four guanine bases. Such a novel GGDEF–c-di-GMP complex structure implies the existence of an unusual product-bound mode for DGCs.

2. Materials and methods

2.1. Reagents

c-di-GMP was produced by an enzymatic method using an altered thermophilic DGC enzyme as described previously (Rao *et al.*, 2009).

2.2. Cloning, expression and purification of XCC4471

The coding region corresponding to XCC4471^{GGDEF} was PCR-amplified directly from the plant pathogen *Xcc* strain 17 using the forward primer 5′-TACTTCCAATCCAATGCTATGCGCGATCTCAAGCGGCA and the reverse primer 5′-TTATCCACTTCCAATGTCAGCCATCGCGCACCAGCCGCC to form the fragment of the required length. The PCR fragment had the correct size on agarose gel electrophoresis, which was further confirmed by DNA sequencing. A ligation-independent cloning (LIC) approach was used to obtain the desired constructs (Wu *et al.*, 2005). The final construct codes for an N-terminal His_6 tag, a 17-amino-acid linker and the XCC4471^{GGDEF} target under the control of a T7 promoter. Overexpression of the His_6-tagged target protein was induced by the addition of 0.5 m*M* IPTG at 293 K for 20 h. The target protein was purified by immobilized metal-affinity chromatography (IMAC) on a nickel column (Sigma). The His_6 tag and linker were cleaved from XCC4471^{GGDEF} by tobacco etch virus (TEV) protease at 277 K for 16 h. For crystallization, XCC4471^{GGDEF} protein was further purified on a Superdex 75 column (ÄKTA, Pharmacia). The final target protein exhibited a purity of greater than 99% and contained only an extra tripeptide (SNA) at the N-terminal end. SeMet-labelled XCC4471^{GGDEF} was prepared in a similar way and was produced using *E. coli* strain BL21 (DE3) as the host in the absence of methionine but with ample amounts of SeMet (100 mg l^{-1}) in M9 medium. Induction was conducted at 293 K for 24 h by the addition of 0.5 m*M* IPTG. Purification of the SeMet-labelled XCC4471^{GGDEF} protein was carried out using the same protocols as those established for the native protein.

2.3. Construction of XCC4471$^{GGDEF(128–291)}$ single point mutants

XCC4471^{GGDEF} single point mutants were prepared using the QuikChange site-directed mutagenesis kit (Stratagene; Vandeyar *et al.*, 1988) and the resulting sequences were confirmed by DNA sequencing. The primers 5′-CTTACCGAAGCCCTGAATGCCCGCGGTTGCGAGCAGGC, 5′-TCGTGCTGTTCGTGCTGGCCATGGACAACCTCAAGCC, 5′-GACAACCTCAAGCCGATCGTCGACCGCTTCGGCCACCT, 5′-TCGGCCTACGGCTGGCTGCGTGCGCAGGACTGGATCGG, 5′-GCGCAGGACTGGATCGGCGCCTGGGGCGCCTGGGGCGGCGACGAATT, 5′-GACTGGATCGACTGGATCGGCCGCTGGGCCGCCGACGAATTCCTGATCGGC, 5′-TCGGCCGCTGGGGCGGCGCCGAATTCGCCGAATTCCTGATCGGCGT and their corresponding complementary primers were used to introduce R146A, D173A, N181V, G205R, R212A, G214A/G215A and R216A substitutions, respectively. The mutated proteins were expressed and purified as described above.

2.4. Enzymatic assay of wild-type XCC4471$^{GGDEF(69–301)}$

The DGC activity of XCC4471 was assayed using an EnzChek pyrophosphate assay kit from Molecular Probes Inc. (catalogue No. E-6645). This assay is based on the method originally described by Webb (1992) by measuring the amount of P_i which is generated from PP_i released from the DGC reaction using the enzyme inorganic pyrophosphatase. All kinetic measurements were carried out at 298 K under the assay conditions 50 m*M* Tris pH 7.5, 1 m*M* $MgCl_2$, 0.5 m*M* GTP. Product inhibition was carried out in the presence of 2 and 5 µ*M* c-di-GMP at 0.1, 0.2, 0.3 and 0.5 m*M* GTP substrate. The K_i values were calculated using the equation $K_i = K_m[i]/(K_m^{app} - K_m)$, where 'app' refers to the new or 'apparent' value of K_m in the presence of inhibitor.

2.5. Biophysical measurements of wild-type and mutant XCC4471^{GGDEF} proteins

Samples of both wild-type and mutant XCC4471^{GGDEF} proteins for ITC were dialyzed against the assay buffer (40 m*M* NaCl, 10 m*M* Tris pH 8.0, 0.5 m*M* $MgCl_2$) overnight. The protein samples were first diluted to 30 µ*M* with the assay buffer before loading into the ITC cell. The c-di-GMP was diluted in the same way to 0.5 m*M* before loading into the syringe. 2 µl c-di-GMP solution was then injected into the cell at 3 min intervals. ITC experiments on wild-type and mutant XCC4471^{GGDEF} proteins were carried out at 298 K and the data were fitted using the commercial *Origin* 7.0 program to obtain the ΔH and K_d values.

The analytical ultracentrifugation sedimentation-velocity experiments were performed using a Beckman model XL-A analytical ultracentrifuge (Beckman Coulter). Protein samples and reference solution were separately loaded into a conventional double-sector quartz cell and mounted in a Beckman An-50 Ti rotor. Experiments were conducted at 293 K with a rotor speed of 42 000 rev min^{-1} (130 000*g*). The absorbance of the sample at 280 nm was monitored in a continuous mode, with a time interval of 480 s and a step size of 0.002 cm. Multiple scans at different time points were fitted to a continuous size distribution using the program *SEDFIT* (v.9.4c; Schuck *et al.*, 2002). The differences between the experimental data and the fitted curves were superimposed to

generate the 'residuals' figures (Figs. 6*e* and 6*f*). The small difference values and consistent errors of the residuals indicate that the fit is very acceptable. A continuous distribution of sedimentation coefficients $c(s)$, corresponding to the data from the moving-boundary panels, was also generated.

2.6. Crystallization of the XCC4471^{GGDEF}–c-di-GMP complex

For crystallization, the native protein was concentrated to 5 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl using an Amicon Ultra-10 (Millipore). Screening for crystallization conditions was performed using the sitting-drop vapour-diffusion method in 96-well plates (Hampton Research) at 277 K by mixing 0.3 µl protein solution with 0.3 µl reagent solution in the presence of 0.3 m*M* c-di-GMP and 0.5 m*M* $MgCl_2$. Initial screens, including Clear Strategy Screen I, Structure Screens I and II (Molecular Dimensions), a systematic PEG–pH screen and PEG/Ion screen (Hampton Research), were performed using a Phoenix RE crystallization workstation (Rigaku). Rod-like crystals appeared in one week from a reservoir solution comprising 0.1 *M* CHES pH 9.5, 40% PEG 600. Crystals suitable for diffraction experiments were grown by mixing 1.5 µl protein solution with 1.5 µl reagent solution at 277 K and reached dimensions of 0.01 × 0.01 × 0.05 mm after one week. SeMet-labelled XCC4471^{GGDEF}–c-di-GMP was crystallized in the same way.

2.7. Data collection and refinement

A crystal was flash-cooled at 100 K under a stream of cold nitrogen. X-ray diffraction data were collected on National

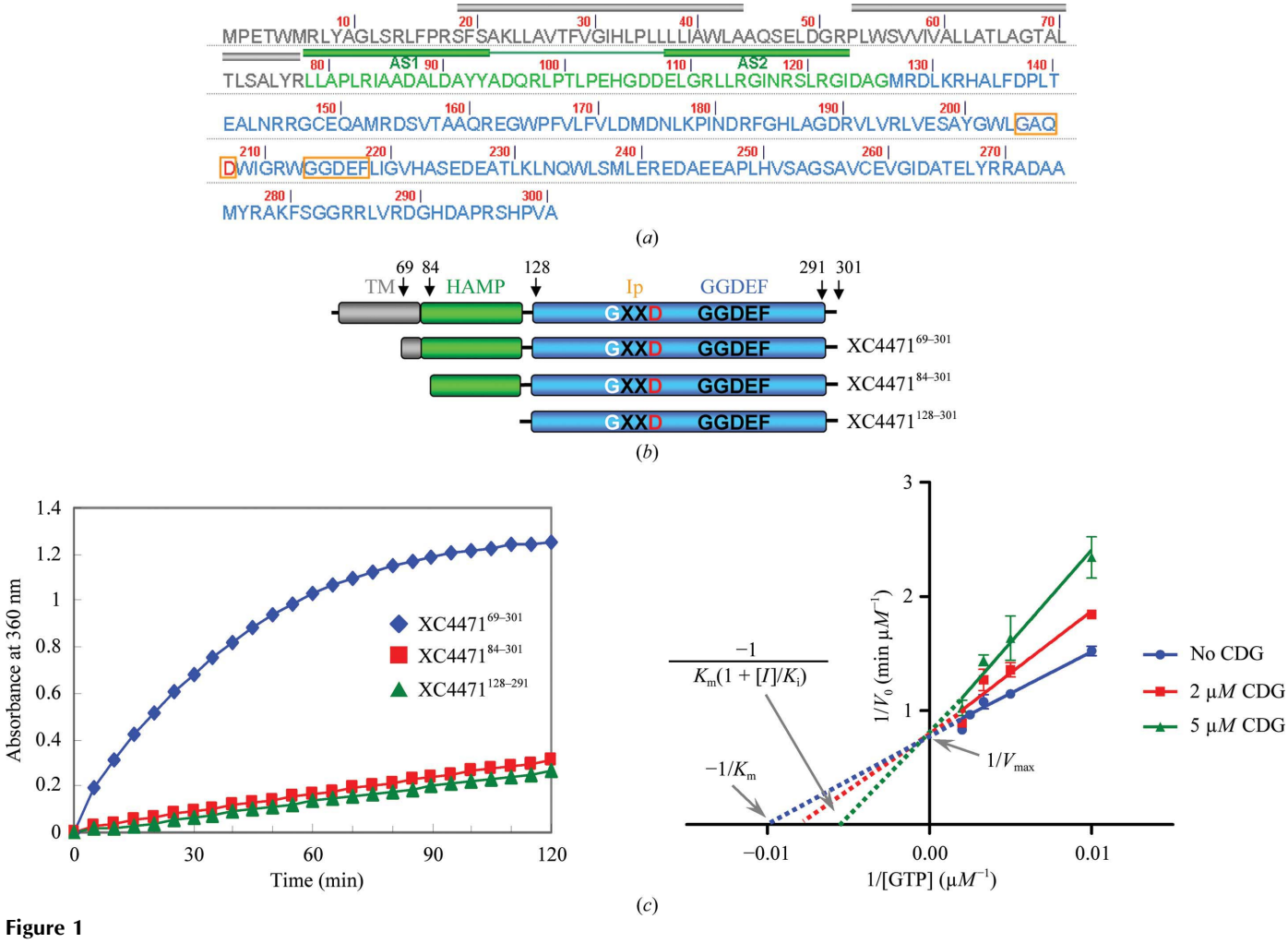


Figure 1
Sequence, domain composition and DGC activity assay of XCC4471. (*a*) The full-length protein sequence of XCC4471. It contains 301 amino-acid residues, with the TM sequence shown in grey letters, the HAMP sequence in green letters and the GGDEF-domain sequence in blue letters. The two predicted TM helices and the two amphiphilic helix sequences (AS) in the HAMP domain are further marked by grey tubes and green tubes, respectively. (*b*) The domain composition of XCC4471. The TM, HAMP and GGDEF domains are coloured similarly as in (*a*). The primary inhibitor site in the GGDEF domain is marked Ip. The starting and ending sequences constructed for kinetic and crystallization studies are indicated by black arrows and residue numbers. (*c*) The DGC activity and inhibition measured using the pyrophosphate assay method. Three different XCC4471 constructs were assayed for their DGC activity and only the construct containing a partial TM sequence gave moderate activity (coloured blue, left figure). The activity of the longest construct was further measured in the presence of 2 and 5 µ*M* c-di-GMP (right figure). The initial velocities *versus* the GTP substrate concentrations are plotted in a Lineweaver–Burk double-reciprocal format (right figure) to measure the values of V_{max} (and k_{cat}), K_m (or apparent K_m) and K_i.

Table 2
Kinetic parameters of XCC4471$^{GGDEF(69–301)}$ at different c-di-GMP concentrations.

The enzyme concentration used in the assay was 2.5 μ*M*. V_{max} and K_i were calculated from the double-reciprocal plot (Fig. 1*c*) and are 1.27 μmol min^{-1} and 6.54 μ*M*, respectively.

c-di-GMP (*M*)	k_{cat} (min^{-1})	K_m (μ*M*)	K_{cat}/K_m (min^{-1} μ*M*$^{-1}$)
0	0.5088 ± 0.0164	105 ± 12	0.00485 ± 0.00057
2		134 ± 21	0.00380 ± 0.00071
5		195 ± 33	0.00261 ± 0.00052

Synchrotron Radiation Research Center (NSRRC) beamlines 13B1 and 13C1, Taiwan. A two-wavelength MAD data set to 2.5 Å resolution was obtained. The data were indexed and integrated using the *HKL*-2000 processing software, giving a data set that is 99.7% complete with an overall R_{merge} of 9.5–11.2% on intensities. The refinement of Se-atom positions, phase calculation and density modification were performed using the program *BnP* (Xu *et al.*, 2005). The model was manually adjusted using the *XtalView*/*Xfit* package (McRee, 1999). *CNS* was then used for refinement to a final R_{cryst} of 20.48% and R_{free} of 23.03% (Brünger *et al.*, 1998). The data-collection and refinement statistics are summarized in Table 1.

3. Results and discussion

3.1. XCC4471$^{69–301}$ is a GGDEF protein containing a HAMP domain and exhibiting DGC activity

GGDEF-domain-containing proteins present in the plant pathogen *Xcc* are usually found to be fused to a variety of signalling domains such as cNMP, HAMP, PAS, GAF *etc.* (Galperin *et al.*, 2001; Galperin, 2004). XCC4471 is a GGDEF-domain-containing protein consisting of 301 amino-acid residues (Fig. 1*a*), with an N-terminal TM domain comprising two predicted TM helices (residues 19–44 and 52–77), a HAMP domain comprising two amphipathic helical sequences AS1 (residues 79–94) and AS2 (residues 109–124), a connector and a GGDEF domain (residues 128–301). The sequences of the HAMP domains from three GGDEF proteins (XCC4471, XCC1263 and XCC2449) from xanthomonads and one from *Vibrio parahaemolyticus* (Vp0117) are found to adopt tandem heptad repeats (boxed in red in Supplementary Fig. S1[1], with two well conserved residues Pro82 and Glu109 in the AS1 and AS2 helices, respectively). Such an N-terminal HAMP domain is presumably connected to the membrane to receive environmental signals to regulate the cytoplasm DGC activity of the downstream GGDEF domain. Since full-length XCC4471 is a membrane protein containing two long TM helices, a series of shorter truncation alternatives were constructed to prevent the potential solubility problem and to allow the measurement of DGC activity (Fig. 1*b*). After extensive trials, we found that XCC4471 becomes soluble after removing the first 69 residues from the N-terminal end. Consequently, three truncations (residues 69–301, 84–301 and 128–291; Fig. 1*b*) were assayed for DGC activity *via* the pyrophosphate method (Webb, 1992). Intriguingly, we found that the shorter proteins (84–301 and 128–291) exhibited only residual DGC activity (Fig. 1*c*). Only the long XCC4471$^{69–301}$ protein containing residues from the partial TM helical and HAMP sequence was found to be fully active. This result indicates that the HAMP and TM domains are required to dimerize the GGDEF domains to achieve DGC activity. Kinetic analyses (Fig. 1*c*) allowed the K_m and k_{cat} values of the active XCC4471$^{69–301}$ truncation to be determined (Table 2).

The DGC activity of XCC4471$^{GGDEF(69–301)}$ was found to be reduced in the presence of c-di-GMP, although there is no allosteric inhibitory R*xx*D sequence (Fig. 2*d*). The enzyme kinetics in the presence of two different c-di-GMP concentrations was measured to determine the K_i value of c-di-GMP against DGC in the XCC4471$^{GGDEF(69–301)}$ protein (Fig. 1*d*). A similar V_{max} value of 1.27 μmol min^{-1} was obtained in the presence of different c-di-GMP concentrations, which is characteristic of a possible competitive inhibition phenomenon for c-di-GMP against DGC. The K_i for c-di-GMP was determined to be 6.54 μ*M*.

3.2. Crystal structure of the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex

Several cocrystallizations of different XCC4471 truncations with c-di-GMP were attempted, but only the shortest XCC4471$^{GGDEF(128–291)}$ truncation could be cocrystallized with c-di-GMP in the presence of Mg^{2+}. The selenomethionine (SeMet) substituted derivative of XCC4471$^{GGDEF(128–291)}$ was further prepared in order to solve the phase problem by the use of the multiple anomalous diffraction (MAD) approach. Needle-like crystals of approximate dimensions 0.02 × 0.02 × 0.05 mm were obtained under similar conditions to the native truncation. The overall fold of the XCC4471$^{GGDEF(128–291)}$ domain is similar to those published previously, comprising a central five-stranded β-sheet (β2β3β1β4β5) surrounded by five α-helices (Fig. 2*b*; Chan *et al.*, 2004; De *et al.*, 2008; Navarro *et al.*, 2009). Except for the α1 helix and several flexible loops, the XCC4471^{GGDEF} structure (plotted in red) superimposes very well on those of PleD (PDB entry 1w25; Chan *et al.*, 2004), WspR (PDB entry 3i5c; De *et al.*, 2009) and MqR89a (PDB entry 3ign; Northeast Structural Genomics Consortium, unpublished work), with root-mean-square deviation values (r.m.s.d.s) of 1.63 Å for 107 C$^{\alpha}$ atoms, 1.63 Å for 103 C$^{\alpha}$ atoms and 1.49 Å for 117 C$^{\alpha}$ atoms, respectively (Fig. 2*c*). However, the XCC4471$^{GGDEF(128–291)}$ sequence lacks several crucial Arg residues found in the primary allosteric (Ip) and secondary allosteric (Is) regions of other GGDEF domains (Fig. 2*a*). The electron density is clear for all residues in the monomer, except for a few residues located in the α4–β4 loop (Glu243–Glu246), which are displayed in grey text in Fig. 2(*a*). In addition, the torsion angles of all residues are found to be located within favourable regions of the Ramachandran plot (99.3% in the most favourable region and 0.7% in the additionally favourable region as calculated by the

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: RR5003). Services for accessing this material are described at the back of the journal.

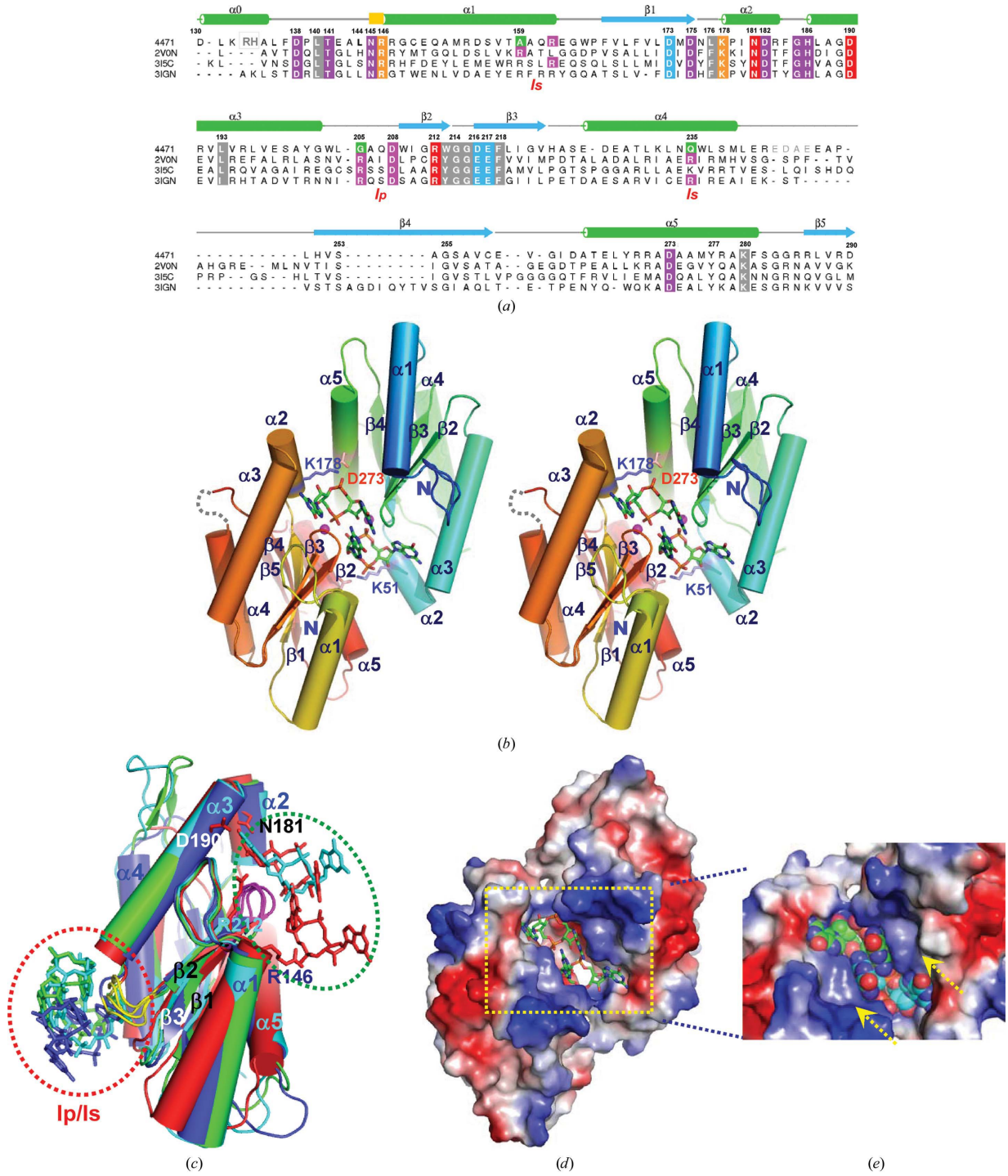


Figure 2
Overall sequence and structural characteristics of XCC4471$^{GGDEF(128–291)}$. (*a*) Sequence alignment of GGDEF domains from XCC4471, PleD (PDB entry 2v0n), WspR (PDB entry 3i5c) and MqR89a (PDB entry 3ign). The residues are numbered according to the sequence of XCC4471, with the secondary-structural elements annotated above the sequence. Many highly conserved residues are observed and are highlighted in different colours for specific recognition: red for c-di-GMP guanine-base recognition, orange for c-di-GMP phosphate diester O-atom recognition, grey for hydrophobic interactions for accommodating the guanine base and purple for indirect recognition. Primary (Ip) and secondary (Is) allosteric inhibitory-site residues are annotated and highlighted in pink. (*b*) Stereo picture of the XCC4471$^{GGDEF(128–291)}$–c-di-GMP dimer. One structure is drawn in warm colours, while the other is drawn in cold colours. The two binding c-di-GMP molecules are drawn in stick representation, together with two Mg^{2+} ions coloured magenta. The invisible E^{243}DAE segment is shown by a dotted line. (*c*) Superimposition of GGDEF-domain structures of XCC4471 (red), PleD (cyan), WspR (blue) and MqR89a (green). The $(c\text{-di-GMP})_2$ molecules binding at the inhibitory site (circled in red) and $(c\text{-di-GMP})_2$ or monomeric c-di-GMP binding at the active site (circled in green) are drawn in stick representation. Some residues involved in specific recognition of the XCC4471^{GGDEF} domain are also drawn and annotated. (*d*) Dimeric XCC4471$^{GGDEF(128–291)}$–c-di-GMP drawn as an electrostatic surface. Two c-di-GMP molecules drawn in stick representation are found embedded in the highly positively charged interface cavity. (*e*) Enlarged region from (*d*) showing two c-di-GMP drawn in van der Waals representation. Two layers of Gly backbone plane atoms are indicated by dotted yellow arrows.

MolProbity program using default parameters; Chen *et al.*, 2010). The asymmetric unit of ligand-bound XCC4471$^{GGDEF(128–291)}$ reveals a dimer connected through interactions between the two opposing antiparallel α-helices (α2 and α5) from each monomer (Fig. 2*b*). Two metal ions (finally interpreted as Mg^{2+}) were also found to be situated in the dimeric interface.

After successfully fitting two GGDEF sequences to the electron-density map in the asymmetric unit, a substantial extra patch of electron density between the two opposing GGDEF domains was clearly revealed. This electron-density patch was subsequently confirmed to be caused by the presence of two molecules of c-di-GMP. When (c-di-GMP)$_2$ was included in calculations, the R_{cryst} and R_{free} values steadily declined to 20.48% and 23.03%, respectively, after iterative calculations. The XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex structure was finally refined to a resolution of 1.90 Å against data collected from the native protein (Table 1). The crystals belonged to space group $P4_32_12$, with a Matthews coefficient and solvent content of 2.32 Å^3 Da^{-1} and 47.12%, respectively. The two molecules of c-di-GMP are found to be situated in the active sites between the two opposing GGDEF domains (Figs. 2*b*, 2*d* and 2*e*) and interact very well with the strongly conserved GGDEF signature motif and with many other conserved residues in >10 000 GGDEF-domain proteins (Fig. 3*a*).

3.3. Specific XCC4471$^{GGDEF(128–291)}$–c-di-GMP binding involves many highly conserved residues

To date, the structures of two GGDEF-domain-containing proteins with (c-di-GMP)$_2$ binding at the primary (Ip) and the secondary inhibition (Is) sites have been reported (Chan *et al.*, 2004; Wassmann *et al.*, 2007; De *et al.*, 2008, 2009; Navarro *et al.*, 2009). The conserved motif at the main Ip site comprises the R^{359}*xx*D residues (numbered according to the PleD sequence). An extra interacting Arg390 residue is also present in PleD, together with several other Arg residues, including Arg313 in the PleD^{GGDEF} domain, Arg198 and Arg242 in the WspR^{GGDEF} domain and Arg148 and Arg178 in the PleD$^{REC'}$ domain at the secondary inhibitory (Is) sites. These extra Arg residues appear to improve the c-di-GMP binding at the allosteric inhibition site (Schirmer & Jenal, 2009; Fig. 2*a*). However, these Arg residues are not generally conserved. In fact, the most conserved Arg and Asp residues in the primary R^{359}*xx*D inhibition site of the GGDEF domain in the PleD and WspR proteins are present only in approximately half of the >10 000 GGDEF-domain sequences, with the other Arg residues in the secondary inhibition site being even less conserved and appearing in only a small portion of the GGDEF sequences identified to date (Seshasayee *et al.*, 2010). However, a significant number of other residues are highly conserved in the identified GGDEF-domain sequences (Fig. 3*a*), including the Asn181, His186 and Asp190 residues involved in specific guanine-base recognition and the Asp173 and Asp216 residues involved in phosphate O-atom recognition (Schirmer & Jenal, 2009). The unique interactions not reported previously are briefly described below and depicted in Fig. 3.

Table 3
ITC measurement data for wild-type XCC4471$^{GGDEF(69–301)}$, XCC4471$^{GGDEF(128–291)}$ and mutant proteins with c-di-GMP.

Four kinds of single point mutant proteins were made: (i) for specific c-di-GMP recognition (shown in bold), (ii) for metal-ion interaction (shown in italics), (iii) for alternation back to the R*xx*D motif (underlined) and (iv) for breaking the four-layer base–peptide plane stack (shown in bold italics).

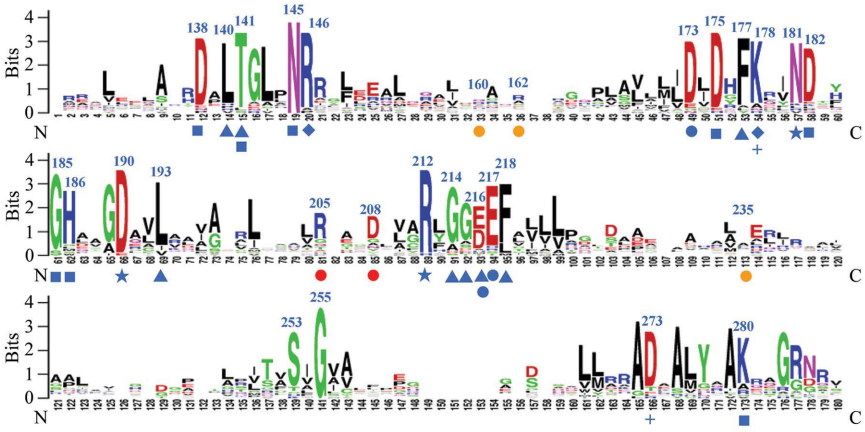
Protein	*N*	K_d (μ*M*)
Wild-type XCC471$^{GGDEF(69–301)}$	1.13 ± 0.0934	6.56
Wild-type XCC471$^{GGDEF(128–291)}$	1.01 ± 0.0133	8.26
N181V	**TW**†	**TW**†
R146A	**TW**†	**TW**†
R212A	**IB**‡	**IB**‡
D173A	*0.003 ± 4.05*§	*197.63*
D216A	*0.948 ± 0.204*	*38.02*
G205R	IB‡	IB‡
G214A	***TW***†	***TW***†
G215A	***TW***†	***TW***†
G214A/G215A	***TW***†	***TW***†

† TW, too weak and not determined. The binding constants of these mutant proteins were too low to be measured using the ITC technique. ‡ IB, inclusion-body formation. These mutant proteins were insoluble and could not be measured by ITC. § The binding constant of this mutant protein reached the measurable limit using the ITC method, with a K_d value of only 0.197 m*M*. Hence, the fitted *N* value is not reliable.

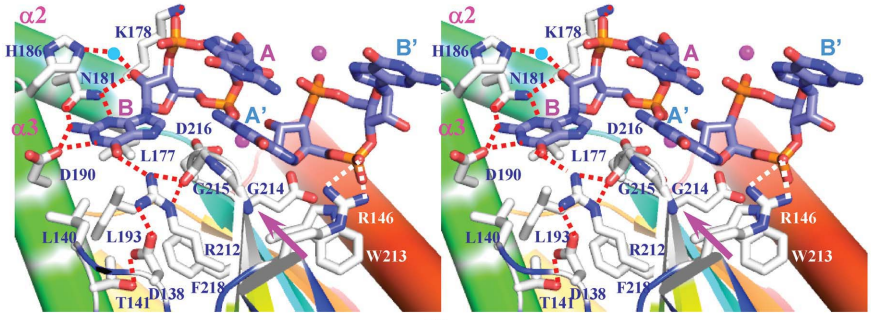
The most unusual structural feature in this complex is the stacking of the backbone-plane atoms of the highly conserved Gly214–Gly215 motif with the guanine base of c-di-GMP. Unlike the self-intercalated form adopted by (c-di-GMP)$_2$ binding at the inhibitory site in other complexes, (c-di-GMP)$_2$ binding at the active site in the XCC4471^{GGDEF}–c-di-GMP complex adopts a partially intercalated form with the peripheral guanine bases bound to the guanine-binding pockets formed by residues from helix α2, helix α3 and loop β2–β3 and the two central bases stacked upon each other (Figs. 3*b* and 3*c*). Owing to this displacement, the space originally occupied by guanine bases B and B′ in the self-intercalated c-di-GMP dimer is now occupied by the Gly214–Gly215 backbone-plane atoms (Gly214 C, Gly214 C′, Gly214 O, Gly215 N and Gly215 C) from the strongly conserved GGDEF motifs (Figs. 3*b*, 3*c* and 3*e*). These five backbone-plane atoms stack well with the C9, C14, N13, C12 and N12 atoms of the c-di-GMP guanine base (atom annotations are shown in Fig. 5*c*), with distances ranging from 3.01 to 3.92 Å. Therefore, the Gly214–Gly215 backbone-plane atoms from each GGDEF signature motif, together with guanine bases A and A′ from c-di-GMP, form a compact four-layer base–peptide plane stack. These novel interactions may explain the need for the highly conserved nature of the Gly214 and Gly215 residues in the GGDEF motif, because alternation of these two residues to any other amino acids will place extra atom(s) at the C$^\alpha$ positions, significantly interfering with the formation of such a compact four-layer base–peptide plane stack. This is indeed confirmed by the greatly reduced binding constants between the GADEF, AGDEF or AADEF single point mutants for c-di-GMP in a series of ITC measurements (see below and Table 3). Interestingly, both glycines are also found to be necessary for DGC activity in the GGDEF-

domain proteins, as mutation of either residue abrogates the activity of this domain (Kirillina *et al.*, 2004).

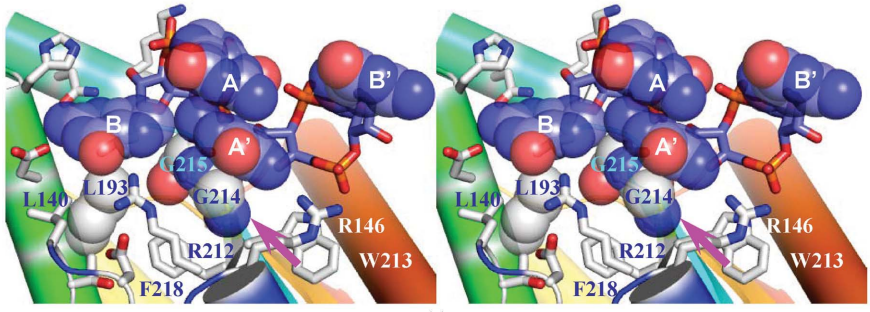
Two Arg residues are found to play crucial roles in stabilizing the c-di-GMP molecule in the active-site pocket. The first is Arg212, which forms two hydrogen bonds to the backbone atoms of Gly215, as well as one hydrogen bond to the c-di-GMP guanine base B O10 atom (Fig. 3*b*). It also forms a salt bridge to the Asp138 side-chain atom, which is then hydrogen bonded to Thr141, forming an extensive hydrogen-bond/salt-bridge network (Fig. 3*b*). Moreover, the Arg212 side chain also stacks well with the ring atoms of Phe218. Arg146 is another well conserved residue that is found to interact with



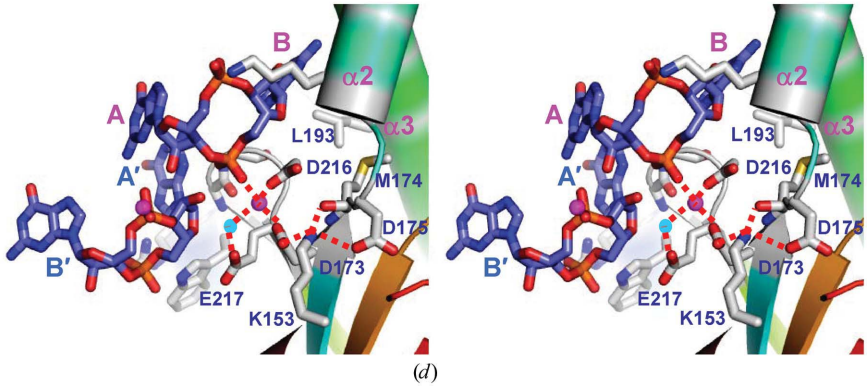
(*a*)



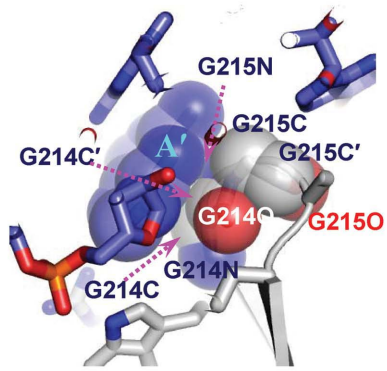
(*b*)



(*c*)



(*d*)



(*e*)

Figure 3
Multiple interactions in the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex. (*a*) Conserved residues of GGDEF domains from 23 *Xcc* genes (Ryan *et al.*, 2007) drawn using the *WebLogo* program (Crooks *et al.*, 2004). The height of each residue reflects the fraction of the corresponding amino acids in that position. Several distinct regions with high sequence conservation spanning residues 138–146, 173–182, 185–193 and 212–218 are observed. Residues in the Ip site are marked by red circles, while those in the Is site are marked by orange circles. Residues involved in c-di-GMP active-site binding are marked by blue stars for specific guanine-base recognition, by blue circles for metal-ion binding, by blue diamonds for c-di-GMP phosphodiester O-atom binding, by blue triangles for hydrophobic cluster residues and by blue squares for residues involved in indirect binding. (*b*) Active-site stereo picture of the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex. The C atoms of c-di-GMP are drawn in light blue and those of residues involved in binding c-di-GMP are drawn in grey. Metal ions are drawn as magenta spheres and water molecules as cyan spheres. Stacking between the Gly214–Gly215 backbone-plane atoms with the guanine bases A′ is indicated by a pink arrow. (*c*) Similar view to (*b*), with the c-di-GMP molecule and the major hydrophobic residue Leu193 drawn in van der Waals representation. (*d*) Stereo picture of the active site of the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex drawn in a different view to (*c*) to reveal the metal ion-binding site. (*e*) The stacking between the Gly214–Gly215 backbone-plane atoms with guanine base A′ drawn in van der Waals representation from a different angle to (*b*). Each specific atom is marked with its name in PDB format.

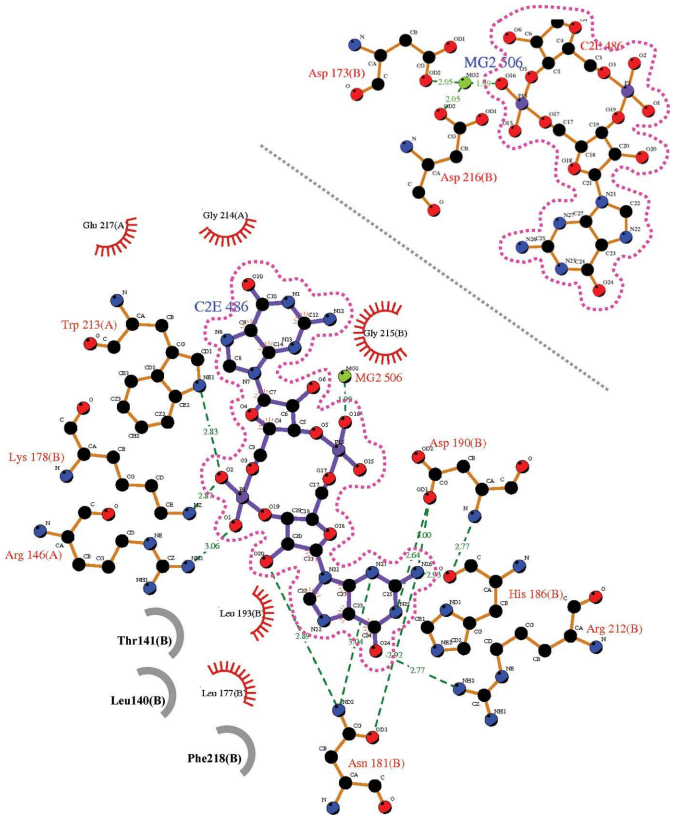


Figure 4
LIGPLOT representation of the GGDEF–c-di-GMP complex revealing bonding atoms and bonding lengths (Wallace *et al.*, 1995). c-di-GMP molecules are circled in pink. Interacting atoms are connected by green dotted lines with bonding lengths indicated. Nonligand residues involved in direct hydrophobic contacts with c-di-GMP are shown as partial red spoked circles, while those indirectly involved are shown as partial grey circles. Mg atoms are shown in green and residues involved in metal binding are shown in the upper right corner.

a nonbonded phosphate O atom of c-di-GMP, which also stacks well with ring atoms of Trp213. Details of the binding atoms and bond lengths between c-di-GMP and the Xc4471 GGDEF domain in a *LIGPLOT* format (Wallace *et al.*, 1995) are summarized in Fig. 4.

3.4. Comparison with reported c-di-GMP structures reveals a novel conformation for (c-di-GMP)$_2$ in the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex

c-di-GMP alone usually forms a self-intercalated dimer stabilized extensively by guanine–guanine base stacking (Egli *et al.*, 1990; Liaw *et al.*, 1990). When bound to proteins, several distinct c-di-GMP conformations have been found, including a self-intercalated dimeric form, a monomeric form with a sandwiched Arg residue and a monomeric open form. A self-intercalated dimeric(c-di-GMP)$_2$ conformation was mainly detected for c-di-GMP binding at the allosteric inhibitory site (R*xx*D) of the GGDEF domain, with four guanine bases intercalating with each other (Chan *et al.*, 2004; De *et al.*, 2008). A monomeric form with a sandwiched Arg residue was discovered for c-di-GMP binding as a monomer at the c-di-GMP switch site (R*xxx*R and D/N*x*S*xx*G) of the PilZ effector protein (Benach *et al.*, 2007). In such a conformation, the space between the two guanine bases is too large to form an efficient base stack. This problem is solved by inserting an Arg residue into the void to form a good cation–π interaction (Benach *et al.*, 2007). A monomeric open form was otherwise mainly detected for c-di-GMP binding at the active site of the EAL domain, with the two guanine bases fully stretched out. Such a conformation contains no guanine–guanine base stacking, which is presumably more suitable for phosphodiester cleavage by EAL domains (Barends *et al.*, 2009; Minasov *et al.*, 2009; Tchigvintsev *et al.*, 2010).

In this manuscript, we report a novel partially intercalated dimeric form for the binding of c-di-GMP at the active site of the XCC4471$^{GGDEF(128–291)}$ domain. This novel dimeric c-di-GMP form is very well revealed in the electron-density map of the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex plotted at a 3σ contour level (Figs. 5*a* and 5*b*). In this mode, two guanine bases A and A′ exhibit partial internucleotide stacking, while the other two guanine bases are completely de-stacked from their respective partner nucleotides (A from B and A′ from B′). This is very different to the self-intercalated dimeric form observed previously for the (c-di-GMP)$_2$ structure present in the inhibitory sites of PleD or WspR, in which guanine bases A and A′ not only stack very well against each other, but also exhibit partial stacking with guanine bases B′ and B, respectively (guanine bases are annotated in cyan text in Fig. 5*d*). We have carefully measured the backbone, five-membered ring and base torsional angles (annotated in Fig. 5*c*) of these two different (c-di-GMP)$_2$ structures in order to detect whether there are torsion-angle differences present in these structures that explain the different structures. There seem to be no major differences between these two sets of torsion angles apart from some minor angle adjustments (data not shown). This indicates that, depending on its binding partner, c-di-GMP is flexible enough to adopt different conformations *via* only minor torsion-angle adjustments.

3.5. Binding strengths of c-di-GMP with GGDEF-domain mutants are consistent with the XCC4471$^{GGDEF(128–291)}$–c-di-GMP cocrystal structure

The strong binding between c-di-GMP and the longer or shorter XCC4471^{GGDEF} domains in solution was confirmed by measuring the K_d of c-di-GMP with wild-type XCC4471$^{GGDEF(69–301)}$ or XCC4471$^{GGDEF(128–291)}$ and with a plethora of shorter single point mutants at several strongly conserved key residues. Two of these measurements *via* the ITC method are shown in Figs. 6(*a*) and 6(*b*), with the overall results summarized in Table 3. c-di-GMP binds to the longer or shorter wild-type XCC4471^{GGDEF} with little difference, exhibiting K_d values of 6.56 and 8.26 μM, respectively, and a 1:1 protein:ligand ratio. Such values are close to the K_i value of 6.54 μM for c-di-GMP as measured in the enzymatic assay (Table 2). The K_d and K_i values are also within the range of physiological concentrations of c-di-GMP in bacteria (Römling *et al.*, 2005; Hickman & Harwood, 2008). Single

point mutants expected to influence (i) specific recognition by c-di-GMP, (ii) the binding of Mg^{2+} and (iii) the formation of a compact four-layer base–peptide plane stack were also generated. For the first type of mutant, mutations of the Asn181 and Arg212 residues involved in recognizing the c-di-GMP guanine base and a mutation of Arg146 involved in recognizing c-di-GMP phosphate were generated and their binding strengths were measured (Fig. 3). N181V and R146A mutations exhibit significantly weaker binding strength to c-di-GMP (too weak to be measured; Table 3). The R212A mutant, however, formed an inclusion body and could not be further analyzed. Single point mutants of the second type expected to influence the magnesium ion binding affinity (D216A and D173A) also led to a reduced affinity for c-di-GMP by approximately fivefold to 25-fold (Table 3). The mutation of Asp173 caused a greater affinity reduction (by 25-fold) than that of Asp216 (by fivefold), perhaps because Asp173 also participates in other interactions, such as those with Met174 in the β1 strand and with Lys178 in the α5 helix (Fig. 3*d*). For the third kind of mutant, several alterations to the highly conserved $G^{214}G^{215}$DEF signature motif were checked. As expected, binding of c-di-GMP in the A^{214}GDEF or GA^{215}DEF single point mutant or in the $A^{214}A^{215}$DEF double point mutant is significantly destabilized and these mutants also exhibit binding constants in a range too weak to be measured (Table 3).

3.6. c-di-GMP stabilizes the XCC4471$^{GGDEF(128–291)}$ dimer

Analytical ultracentrifugation (AUC) was applied to examine the effect of c-di-GMP on the oligomerization state of the shorter and longer XCC4471^{GGDEF} domains. They were studied at 15 µ*M* in the absence and presence of c-di-GMP (at a 1:1 protein: ligand ratio). The data for the shorter domain show that there is only one moving boundary for the apo and c-di-GMP-bound species (Figs. 6*c* and 6*d*), with the c-di-GMP binding form moving faster than the apo form. The profiles at different time points were fitted to a continuous size distribution with the Lamm equation using the *SEDFIT* program (http://www.analyticalultracentrifugation.com; Schuck *et al.*, 2002). Molecular weights of 19 and 39 kDa were obtained after fitting the data with the Lamm equation, with no other species being detected. An AUC study for the oligomerization state of the longer XCC4471$^{GGDEF(69–301)}$ domain was also carried out in a similar way. A molecular weight of 50 kDa was obtained for the longer domain either in the presence or absence of c-di-GMP (data not shown). These data clearly indicate that the shorter XCC4471$^{GGDEF(128–291)}$ domain adopts a monomer in the

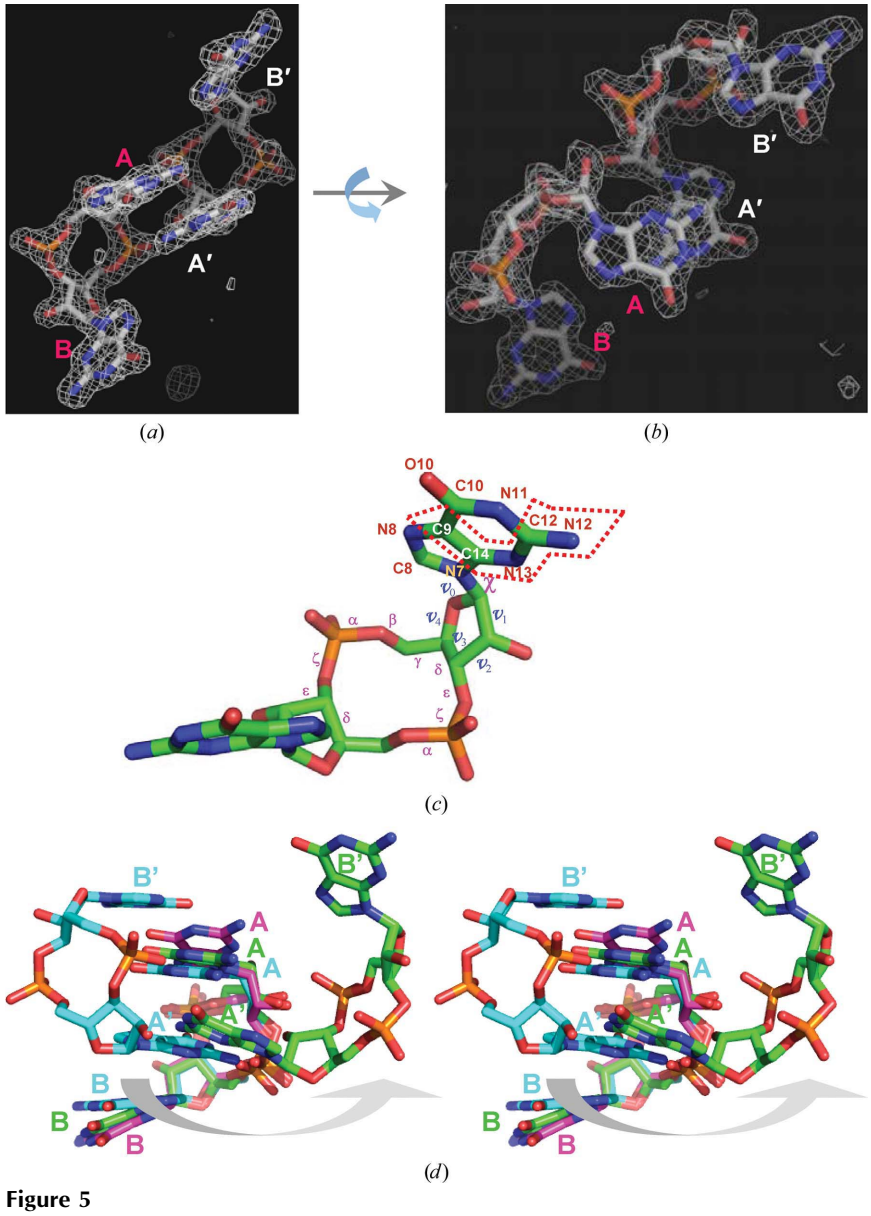


Figure 5
Electron-density map and conformations of $(c\text{-di-GMP})_2$. (*a*) Well visible electron-density map of $(c\text{-di-GMP})_2$ in the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex drawn at a 3σ contour level. This map was obtained using the following procedure. After phases had been determined using the Se-MAD approach, the molecular model was generated, which was repetitively refined with the native data for a higher resolution structure. Following refinement, a large patch of extra electron density was observed. Since the Xc4471 protein was cocrystallized with c-di-GMP, the extra density was believed to be caused by the c-di-GMP ligand. Therefore, c-di-GMP coordinates were included for the final round of data refinement. Near the end of structural determination, an $F_o - F_c$ map was generated to demonstrate that most model error had been eliminated. An OMIT map generated without the ligand coordinates was then created to reveal the final c-di-GMP structure. (*b*) A similar map shown at a different angle. (*c*) Torsion-angle and guanine-base atom annotations in PDB format. (*d*) Superimposed figure of c-di-GMP molecules from the XCC4471$^{GGDEF(128–291)}$–c-di-GMP complex (C atoms in green), the PleD–c-di-GMP complex at the inhibitory site (C atoms in cyan) and the PleD–c-di-GMP complex at the active site (C atoms in magenta).

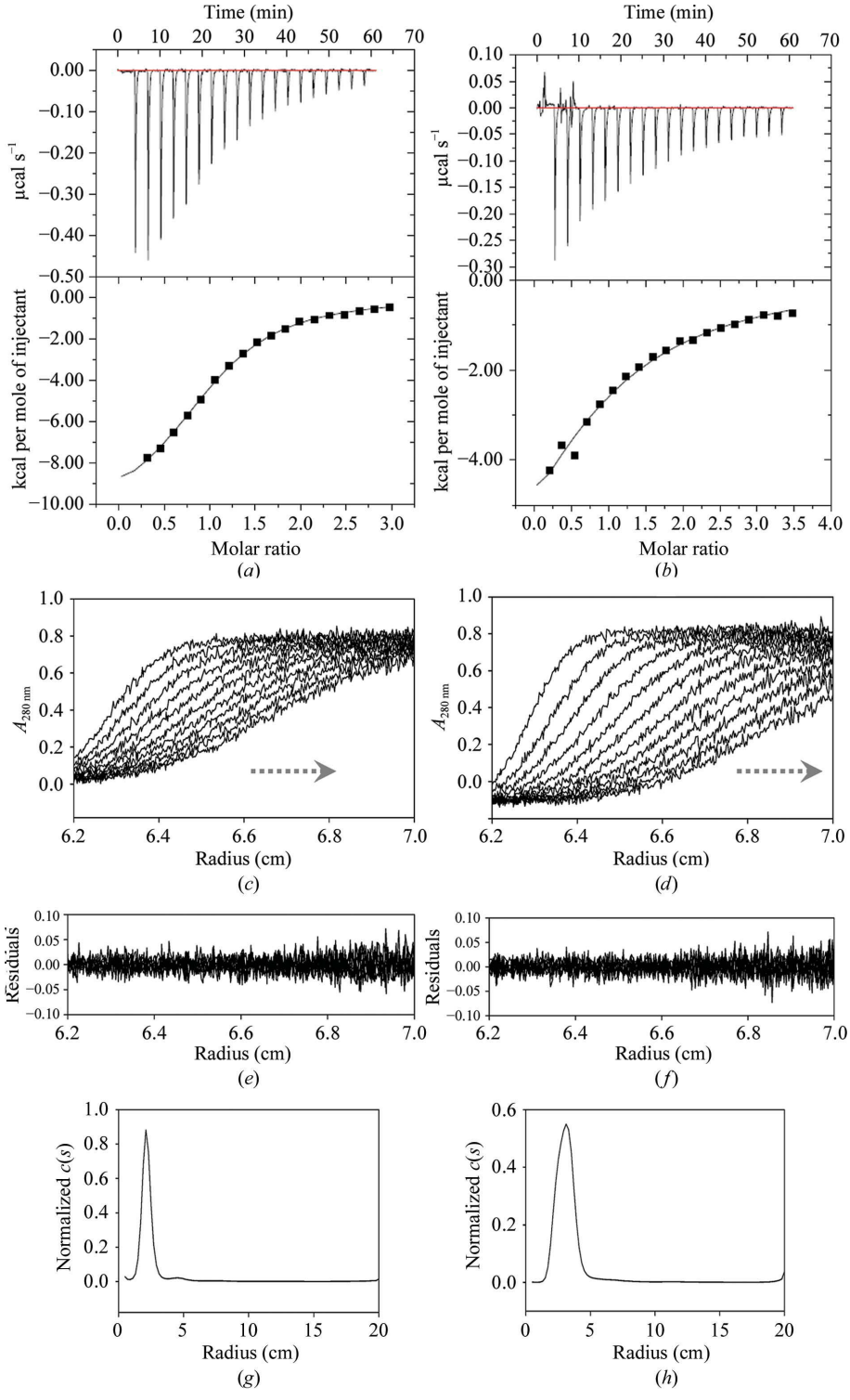


Figure 6
Biophysical measurements of XCC4471 and its single point mutants. Only two representative ITC figures are shown. (*a*) shows the heat released and the fitted curve for wild-type 4471$^{GGDEF(128–291)}$ protein, while (*b*) shows those for the D216A mutant. The sedimentation-velocity profiles of XCC4471$^{GGDEF(128–291)}$ (15 µ*M*) in the absence (*c*) or presence (*d*) of a 1:1 ratio of c-di-GMP were obtained at different times of up to 8 h sedimentation at 42 000 rev min^{-1} (130 000*g*) at 293 K (every second trace is shown) from the absorbance at 280 nm. Superposition of the difference ('residuals') between the experimental and fitted curves in the absence or presence of 1:1 protein:c-di-GMP is shown in (*e*) and (*f*), respectively. Normalized *c*(*s*) distribution analyses were carried out using the *SEDFIT* program, with the resulting curves shown in (*g*) (without c-di-GMP) and (*h*) (with c-di-GMP).

absence of c-di-GMP but readily forms a dimer in the presence of c-di-GMP, while the longer XCC4471$^{GGDEF(128–291)}$ domain adopts a dimer even in the absence of c-di-GMP, consistent with the enzymatic assay data showing that the longer XCC4471$^{GGDEF(128–291)}$ domain is active as a DGC but the shorter domain sequence is less active (Fig. 1*c*).

4. Conclusions

We have observed a stable 'product-bound' state for a GGDEF domain in complex with two molecules of c-di-GMP. Most conserved residues in the >10 000 GGDEF-domain sequences can be well explained from this complex crystal structure. The conservation of these residues is most likely to be a consequence of their role in substrate binding and catalysis. The (c-di-GMP)$_2$ is found to adopt a novel partially intercalated conformation with two guanine bases forming a novel compact four-layer base–peptide plane stack with the well conserved Gly124–Gly215 backbone atoms in the GGDEF signature motif. The presence of the novel product-bound mode observed for XCC4471 may be characteristic of GGDEF-domain proteins containing a sensor and a transmembrane domain. However, in order to fully understand its mechanism, further studies of full-length XCC4471 crystal structures in different ligand-binding forms are necessary. Since a higher concentration of c-di-GMP is required for biofilm formation, which usually makes bacteria immune to antibiotic treatment (Tamayo *et al.*, 2007), inhibiting the DGC activity of GGDEF domains may be a good way to prevent resistance of these bacteria to antibiotics. However, pathogenic bacteria usually encode many copies of GGDEF domains (Galperin *et al.*, 2001; Galperin, 2004; Ryan *et al.*, 2007), which makes it difficult to discover a universal inhibitor for DGCs in these GGDEF-domain proteins. The XCC4471^{GGDEF}–c-di-GMP complex crystal structure in this report reveals detailed interactions between c-di-GMP and most of the strongly conserved residues in the GGDEF domains, which may serve as a

good model for the design of lead compounds for blocking DGC activity in a cell and for treating biofilm-related chronic diseases.

This work was supported in part by the Ministry of Education, Taiwan, ROC under the ATU plan and by the National Science Council, Taiwan, ROC (grant 97-2113-M005-005-MY3 to S-HC). We appreciate the Structural Genomics Database service provided by the GMBD Bioinformatics Core (http://www.tbi.org.tw), NRPGM, Taiwan. We would also like to thank the Core Facilities for Protein X-ray Crystallography at the Academia Sinica, Taiwan for help in crystal screening, the National Synchrotron Radiation Research Center (NSRRC) in Taiwan and the SPring-8 Synchrotron facility in Japan for assistance in X-ray data collection. The National Synchrotron Radiation Research Center is a user facility supported by the National Science Council, Taiwan, ROC and the Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine, Taiwan, ROC. We also thank Dr Maxwell Dow for a review of and valuable comments on this manuscript.

References

Barends, T. R., Hartmann, E., Griese, J. J., Beitlich, T., Kirienko, N. V., Ryjenkov, D. A., Reinstein, J., Shoeman, R. L., Gomelsky, M. & Schlichting, I. (2009). *Nature (London)*, **459**, 1015–1018.

Benach, J., Swaminathan, S. S., Tamayo, R., Handelman, S. K., Folta-Stogniew, E., Ramos, J. E., Forouhar, F., Neely, H., Seetharaman, J., Camilli, A. & Hunt, J. F. (2007). *EMBO J.* **26**, 5153–5166.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U. & Schirmer, T. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 17084–17089.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Chin, K.-H., Lee, Y. C., Tu, Z.-L., Chen, C.-H., Tseng, Y.-H., Yang, J.-M., Ryan, R. P., McCarthy, Y., Dow, J. M., Wang, A. H.-J. & Chou, S.-H. (2010). *J. Mol. Biol.* **396**, 646–662.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M. & Jenal, U. (2006). *J. Biol. Chem.* **281**, 32015–32024.

Christen, M., Christen, B., Folcher, M., Schauerte, A. & Jenal, U. (2005). *J. Biol. Chem.* **280**, 30829–30837.

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.

De, N., Pirruccello, M., Krasteva, P. V., Bae, N., Raghavan, R. V. & Sondermann, H. (2008). *PLoS Biol.* **6**, e67.

Egli, M., Gessner, R. V., Williams, L. D., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Rich, A. & Frederick, C. A. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 3235–3239.

Galperin, M. Y. (2004). *Environ. Microbiol.* **6**, 552–567.

Galperin, M. Y., Nikolskaya, A. N. & Koonin, E. V. (2001). *FEMS Microbiol. Lett.* **203**, 11–21.

Hickman, J. W. & Harwood, C. S. (2008). *Mol. Microbiol.* **69**, 376–389.

Jenal, U. & Malone, J. (2006). *Annu. Rev. Genet.* **40**, 385–407.

Kirillina, O., Fetherston, J. D., Bobrov, A. G., Abney, J. & Perry, R. D. (2004). *Mol. Microbiol.* **54**, 75–88.

Leduc, J. L. & Roberts, G. P. (2009). *J. Bacteriol.* **191**, 7121–7122.

Lee, V. T., Matewish, J. M., Kessler, J. L., Hyodo, M., Hayakawa, Y. & Lory, S. (2007). *Mol. Microbiol.* **65**, 1474–1484.

Liaw, Y.-C., Gao, Y.-G., Robinson, H., Sheldrick, G. M., Sliedregt, L. A. J. M., van der Marel, G. A., van Boom, J. H. & Wang, A. H.-J. (1990). *FEBS Lett.* **264**, 223–227.

McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.

Minasov, G., Padavattan, S., Shuvalova, L., Brunzelle, J. S., Miller, D. J., Baslé, A., Massa, C., Collart, F. R., Schirmer, T. & Anderson, W. F. (2009). *J. Biol. Chem.* **284**, 13174–13184.

Navarro, M. V., De, N., Bae, N., Wang, Q. & Sondermann, H. (2009). *Structure*, **17**, 1104–1116.

Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004). *Genes Dev.* **18**, 715–727.

Rao, F., Pasunooti, S., Ng, Y., Zhuo, W., Lim, L., Liu, A. W. & Liang, Z.-X. (2009). *Anal. Biochem.* **389**, 138–142.

Römling, U., Gomelsky, M. & Galperin, M. Y. (2005). *Mol. Microbiol.* **57**, 629–639.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de Vroom, E., van der Marel, G. A., van Boom, J. H. & Benziman, M. (1987). *Nature (London)*, **325**, 279–281.

Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y.-W., Zhang, L.-H., Heeb, S., Cámara, M., Williams, P. & Dow, J. M. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 6712–6717.

Ryan, R. P., Fouhy, Y., Lucey, J. F., Jiang, B.-L., He, Y.-Q., Feng, J.-X., Tang, J.-L. & Dow, J. M. (2007). *Mol. Microbiol.* **63**, 429–442.

Ryan, R. P., McCarthy, Y., Andrade, M., Farah, C. S., Armitage, J. P. & Dow, J. M. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 5989–5994.

Ryjenkov, D. A., Tarutina, M., Moskvin, O. V. & Gomelsky, M. (2005). *J. Bacteriol.* **187**, 1792–1798.

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

Schmidt, A. J., Ryjenkov, D. A. & Gomelsky, M. (2005). *J. Bacteriol.* **187**, 4774–4781.

Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J. & Schubert, D. (2002). *Biophys. J.* **82**, 1096–1111.

Seshasayee, A. S., Fraser, G. M. & Luscombe, N. M. (2010). *Nucleic Acids Res.* **38**, 5970–5981.

Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H. & Breaker, R. R. (2008). *Science*, **321**, 411–413.

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

Tao, F., He, Y.-W., Wu, D.-H., Swarup, S. & Zhang, L.-H. (2010). *J. Bacteriol.* **192**, 1020–1029.

Tchigvintsev, A., Xu, X., Singer, A., Chang, C., Brown, G., Proudfoot, M., Cui, H., Flick, R., Anderson, W. F., Joachimiak, A., Galperin, M. Y., Savchenko, A. & Yakunin, A. F. (2010). *J. Mol. Biol.* **402**, 524–538.

Vandeyar, M. A., Weiner, M. P., Hutton, C. J. & Batt, C. A. (1988). *Gene*, **65**, 129–133.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.

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

Webb, M. R. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 4884–4887.

Wu, Y.-Y., Chin, K.-H., Chou, C.-C., Lee, C.-C., Shr, H.-L., Gao, F. P., Lyu, P.-C., Wang, A. H.-J. & Chou, S.-H. (2005). *Acta Cryst.* F**61**, 902–905.

Xu, H., Weeks, C. M. & Hauptman, H. A. (2005). *Acta Cryst.* D**61**, 976–981.