STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN 2053-230X

Crystal structure of the YajQ-family protein XC_3703 from *Xanthomonas campestris* pv. *campestris*

Zhixin Zhao,* Zhen Wu and Jun Zhang

College of Biology, Pharmacy and Food Engineering, Shangluo University, Shangluo 726000, People's Republic of China.

*Correspondence e-mail: xzhaoslx@yahoo.com

Received 20 June 2016

Accepted 11 August 2016

Edited by T. C. Terwilliger, Los Alamos National Laboratory, USA

Keywords: XC_3703; YajQ; c-di-GMP; receptor.**PDB reference:** XC_3703, 5b7w**Supporting information:** this article has supporting information at journals.iucr.org/f

As an important bacterial second messenger, bis-(3',5')-cyclic diguanylate (cyclic di-GMP or c-di-GMP) has been implicated in numerous biological activities, including biofilm formation, motility, survival and virulence. These processes are manipulated by the binding of c-di-GMP to its receptors. XC_3703 from the plant pathogen *Xanthomonas campestris* pv. *campestris*, which belongs to the YajQ family of proteins, has recently been identified as a potential c-di-GMP receptor. XC_3703, together with XC_2801, functions as a transcription factor activating virulence-related genes, which can be reversed by the binding of c-di-GMP to XC_3703. However, the structural basis of how c-di-GMP regulates XC_3703 remains elusive. In this study, the structure of XC_3703 was determined to 2.1 Å resolution using the molecular-replacement method. The structure of XC_3703 consists of two domains adopting the same topology, which is similar to that of the RNA-recognition motif (RRM). Arg65, which is conserved among the c-di-GMP-binding subfamily of the YajQ family of proteins, together with Phe80 in domain II, forms a putative c-di-GMP binding site.

1. Introduction

Cyclic dinucleotides, including c-di-GMP, c-di-AMP and cyclic GMP-AMP, are second messengers that are found in bacteria (Hengge *et al.*, 2016; Davies *et al.*, 2012) and metazoa (Sun *et al.*, 2013; Wu *et al.*, 2013). In bacteria, c-di-GMP seems to be more versatile than the others, and it has been shown to be implicated in various processes such as biofilm formation, adhesion, resistance, virulence, alginate production and cell-cycle control (Hengge *et al.*, 2016; Ryan, 2013; Römling *et al.*, 2013). The concentration of c-di-GMP in the cell is regulated by three types of protein domains involved in its synthesis and degradation. C-di-GMP is synthesized by GGDEF domains from two molecules of GTP, whereas its degradation is catalyzed by EAL and HD-GYP domains, leading to the formation of linear pGpG and two AMP molecules at different rates (Paul *et al.*, 2004; Ryan *et al.*, 2006; Christen *et al.*, 2005; Schirmer & Jenal, 2009; Krasteva *et al.*, 2012). The GGDEF, EAL and HD-GYP domains are named after the conserved amino-acid residues in the corresponding motifs; GGDEF belongs to the diguanylate cyclase (DGC) family, while EAL and HD-GYP are members of the phosphodiesterases (PDEs). Intriguingly, those domains are always fused with other signals or cue-sensing domains that regulate the c-di-

© 2016 International Union of Crystallography

Table 1

Macromolecule-production information.

Source organism	<i>X. campestris</i> pv. <i>campestris</i> 8004
DNA source	Genome
Forward primer†	GGAATCCCATATGCCGTCCTTCGACGTGATTTCC
Reverse primer‡	GTACTCGAGTCAATCGCGAAATTGTGGAAGTGC
Cloning vector	pET-23b
Expression vector	pET-23b
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced	MPSFDVISEVDKHELTNAVDQANRELDTRFDKGV- VEAKFELEDGKVINQSAPSDQVQKQMTDILRA- RLLAGRIDVRCLEFGDVETNLGARGKQVTVKQ- GIEQKQAKQLVAKLKEAKLKVEAQINGDKLRV- TGKKRDDLQDAIAVLKKADFELPLQFDNFRDL- EHHHHH

† The NdeI site is underlined. ‡ The XhoI site is underlined.

GMP pathway in a spatial and temporal manner (Römling *et al.*, 2013).

Compared with the vast numbers of DGC, EAL and HD-GPY domains, identified potential c-di-GMP-binding proteins are few, despite their important roles in signal transduction. The diversity of c-di-GMP-binding proteins makes it difficult to find this type of protein by using bioinformatic servers to detect the conserved motifs in the case of c-di-GMP synthetase and hydrolase enzymes. To date, only a few types of c-di-GMP receptors have been fully characterized, such as PilZ-domain, degenerate GGDEF-domain and EAL-domain proteins (Chou & Galperin, 2016; Krasteva *et al.*, 2012). Besides these protein receptors, certain types of RNA (riboswitches) also can function as a receptor of c-di-GMP; riboswitches also have a limited phylogenetic distribution, resulting in the same c-di-GMP signalling conundrum that only a few types of c-di-GMP receptors function downstream from an overwhelming number of DGC enzymes.

Recently, XC_3703 from *Xanthomonas campestris* pv. *campestris*, a member of the YajQ family, was identified as a novel c-di-GMP-binding protein with a relatively high binding affinity ($K_d = 2 \mu\text{M}$; An *et al.*, 2014). The binding of c-di-GMP to XC_3703 results in disruption of the interaction between DNA and its coactivator XC_2801, a transcription factor from the LysR family which controls virulence-related gene expression. Although biochemical data have provided plenty of knowledge about the c-di-GMP-regulated pathway, the detailed molecular mechanism and structural basis remain largely unknown. In order to clarify how c-di-GMP binds to XC_3703 and regulates protein–protein interaction, the overexpression, purification and structural analysis of the XC_3703 protein are reported in this paper.

2. Materials and methods

2.1. Cloning, expression and purification of XC_3703 from *X. campestris* pv. *campestris*

The XC_3703 gene (GenBank ID AAY50743.1) was amplified from *X. campestris* pv. *campestris* genomic DNA using *Pfu* DNA polymerase and inserted into pET-23b vector (Novagen, Merck) via NdeI and XhoI restriction sites, generating a C-terminally His₆-tagged protein facilitating the

Table 2

Crystallization.

Method	Vapour diffusion
Plate type	Hanging drop
Temperature (K)	293
Protein concentration (mg ml ⁻¹)	20
Buffer composition of protein solution	25 mM Tris–HCl pH 8.0, 100 mM NaCl
Composition of reservoir solution	0.1 M bis-tris pH 5.5, 20% PEG 3350
Volume and ratio of drop	2 μl , 1:1
Volume of reservoir (ml)	1

purification of recombinant protein (Table 1). The forward and reverse primers were GATATACATATGCCGTCCTTCGACGTGA and GTGGTGCTCGAGTCAATCGCGGA-AATTGTCG, respectively. The construct was confirmed by DNA sequencing and transformed into *Escherichia coli* BL21 (DE3) cells.

E. coli BL21 (DE3) cells containing the pET-23b-XC_3703 plasmid were grown in Luria broth (LB) medium supplied with $100 \mu\text{g ml}^{-1}$ ampicillin. When the OD₆₀₀ reached 0.6, isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM to induce protein expression, and at the same time the temperature was decreased to 289 K. The cells were collected by centrifugation (20 min, 5000g) about 18 h later.

For purification, the protein was subjected to two steps of chromatography: nickel-affinity chromatography and size-exclusion chromatography. The cell pellet was resuspended in ice-cold buffer A (25 mM Tris–HCl pH 8.0, 500 mM NaCl) and lysed by sonication. The lysate was centrifuged at 16 000g for 50 min and the supernatant was loaded onto Ni-chelating Sepharose (GE Healthcare) equilibrated with buffer A. After washing with buffer B (25 mM Tris–HCl pH 8.0, 500 mM NaCl, 15 mM imidazole), the bound protein was eluted using buffer C (25 mM Tris–HCl pH 8.0, 500 mM NaCl, 250 mM imidazole). Finally, the protein was further purified using a Superdex 200 (GE Healthcare) column equilibrated in buffer D (25 mM Tris–HCl pH 8.0, 100 mM NaCl). The purified XC_3703 fractions were collected according to the results of SDS-PAGE. The final protein concentration for crystallization was 20 mg ml^{-1} . All protein-purification steps were performed at 277 K.

2.2. Crystallization

The XC_3703 protein was screened for crystallization with commercial available screening kits using the sitting-drop vapour-diffusion method at 293 K. Several hits were obtained in the initial screen and a promising condition was chosen to optimize the crystal. During optimization, the XC_3703 protein was mixed with an equal volume (1 μl) of reservoir solution and the hanging-drop vapour-diffusion method was used at 293 K. The best condition consisted of 0.1 M bis-tris pH 5.5, 20% (w/v) polyethylene glycol 3500 (Table 2). A rod-shaped crystal appeared overnight and took another day to grow to a suitable size for diffraction.

Table 3

Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

Data collection	
Diffraction source	BL17U, SSRF
Wavelength (Å)	0.9791
Temperature (K)	100
Detector	ADSC Q315r
Crystal-to-detector distance (mm)	250
Rotation range per image (°)	1
Total rotation range (°)	180
Exposure time per image (s)	0.2
Space group	$P2_1$
a, b, c (Å)	51.51, 65.47, 52.82
α, β, γ (°)	90.00, 92.69, 90.00
Mosaicity (°)	0.87–1.26
Resolution range (Å)	50.0–2.10 (2.17–2.10)
Total No. of reflections	72609
No. of unique reflections	20471
Completeness (%)	99.1 (98.4)
Multiplicity	3.5 (3.6)
$\langle I/\sigma(I) \rangle$	11.8 (4.0)
$R_{\text{r.i.m.}}^\dagger$ (%)	4.6 (24.8)
Refinement	
Resolution range (Å)	35.82–2.08 (2.13–2.08)
Completeness (%)	97.28 (74.3)
No. of reflections, working set	18495 (1000)
No. of reflections, test set	1976 (104)
Final R_{cryst}	0.2438 (0.3025)
Final R_{free}	0.2840 (0.3349)
No. of non-H atoms	
Protein	2558
Water	116
Total	2674
R.m.s. deviations	
Bonds (Å)	0.003
Angles (°)	0.582
Average B factors (Å ²)	
Protein	52.69
Water	49.01
Ramachandran plot	
Most favoured (%)	94.03
Allowed (%)	5.97
Disallowed (%)	0

$^\dagger R_{\text{r.i.m.}}$ is defined as $\sum_{hkl} [N(hkl)/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

2.3. Data collection and processing

For X-ray data collection, crystals were immersed in reservoir solution with 15% (v/v) glycerol serving as a cryo-protectant for several seconds and were then flash-cooled in a 100 K nitrogen stream. Native X-ray diffraction data were collected using an ADSC Q315r CCD detector on beamline BL17U1 at Shanghai Synchrotron Radiation Facility (SSRF), with a crystal-to-detector distance of 250 mm. The crystals were rotated through 180° with 1° oscillation and 0.2 s exposure per image. The collected data were indexed, integrated and scaled with *HKL-2000* (Otwinowski & Minor, 1997). Data-collection statistics are listed in Table 3.

2.4. Structure determination and refinement

The crystal diffracted to 2.1 Å resolution and belonged to space group $P2_1$, with unit-cell parameters $a = 51.51$, $b = 65.47$, $c = 52.82$ Å, $\alpha = 90.00$, $\beta = 92.69$, $\gamma = 90.00$ °. Data analysis using *phenix.xtriage* (Adams *et al.*, 2010) indicated that there were

two copies of the XC_3703 protein in the asymmetric unit, with a corresponding Matthews coefficient V_M of 2.456 Å³ Da⁻¹ and a solvent content of 50.0%. The structure of XC_3703 was solved by molecular replacement using another YajQ-family protein (HI1034; PDB entry 1in0; Teplyakov *et al.*, 2003) from *Haemophilus influenzae* as the search model in *Phaser* (McCoy *et al.*, 2007). Model building was carried out using *AutoBuild* in the *PHENIX* software package (Terwilliger *et al.*, 2008). For refinement of the model, TLS refinement was performed with the whole molecule as one TLS group, and R_{work} and R_{free} could be reduced to 24.38 and 28.40%, respectively. The refinement statistics are shown

Figure 1

The purification and overall structure of XC_3703. (a) Gel-filtration chromatography of the XC_3703 protein. This experiment was performed on a Superdex 200 10/300 column (GE Healthcare). The horizontal and vertical axes correspond to elution volume and ultraviolet absorbance ($\lambda = 280$ nm), respectively. The red curve indicates the elution of bovine pancreatic trypsin (24 kDa). The black curve indicates the elution of XC_3703 (19.2 kDa). The blue curve indicates the elution of hen egg-white lysozyme (14.4 kDa). (b) SDS-PAGE analysis of the eluted peak fraction. Left lane, molecular-weight marker (labelled in kDa); right lane, XC_3703 protein. (c) Two molecules of XC_3703 in one asymmetric unit are shown as a cartoon with chain A in cyan and chain B in green.

in Table 3. All molecular graphics were generated using PyMOL (<http://www.pymol.org>).

3. Results and discussion

3.1. Crystallization of XC_3703

The XC_3703 gene was first inserted into the pET-28b vector, generating an N-terminally His₆-tagged protein, which was purified to homogeneity for crystal screening. Crystals grew overnight and data were collected from this type of crystal. However, the crystals only diffracted to 7–8 Å resolution; thus, it was inferred that the long and disordered N-terminal sequence from the vector (about 20 residues) may spoil the quality of the crystal, so the N-terminal His tag was removed by thrombin protease during purification. However, no hits were obtained after rescreening of either apo XC_3703 or the c-di-GMP–XC_3703 complex. This result suggests that the N-terminal His tag could facilitate crystallization by

providing extra crystal contacts, although this type of crystal diffracted poorly. Finally, C-terminally tagged protein was expressed using the pET-23b vector and purified. The protein showed good behaviour and a monomeric state on size-exclusion chromatography (SEC), which was consistent with a previous report (Saveanu *et al.*, 2002; Fig. 1a). After gel filtration, the protein appeared to have a purity of about 97% and a molecular weight about 19 kDa as detected by SDS–PAGE (Fig. 1b). Finally, an apo-form crystal was obtained and optimized to a suitable size for a diffraction experiment.

3.2. The structure of XC_3703 shows a tandem RNA-recognition motif (RRM)-like architecture

There are two XC_3703 molecules in the asymmetric unit, and superposition of these two molecules gives an r.m.s. deviation of only 1.25 Å for 161 aligned residues (Fig. 1c). Both chains contain all 161 amino-acid residues. The structure of XC_3703 consists of two domains (domain I and domain II)

Figure 2

Domain organization of XC_3703. (a) Schematic representation of XC_3703. The green boxes indicate domain I and the cyan box indicates domain II. (b) The structure of XC_3703 is shown as a cartoon. Domain I is depicted in green and domain II is depicted in cyan. Secondary-structure elements referred to in the text are labelled. The residues in the central β -sheet are shown as ball-and-stick models. (c) A canonical RRM domain from human RBM7 is shown as a cartoon in yellow (PDB entry 5iqq; Sofos *et al.*, 2016). The secondary-structure elements are labelled according to the canonical topology. The RNP1 and RNP2 sequences are shown as ball-and-stick models in slate and orange, respectively. (d) Superposition of domain II of XC_3703 and human RBM7 shown as a cartoon. Domain II of XC_3703 is coloured cyan and human RBM7 is coloured yellow.

forming an elongated conformation (Figs. 2*a* and 2*b*). Each domain contains a four-stranded antiparallel β -sheet packed by two helices from one side. Domain I is composed of the first N-terminal β -strand and residues 101 to the C-terminus, and domain II includes residues 11–100 (Fig. 2*a*). The topology of these two domains is similar, although there is no sequence similarity between them. This topology is reminiscent of the RNA-recognition motif (RRM) α/β sandwich fold with a $\beta\alpha\beta\alpha\beta$ arrangement (Barret *et al.*, 2011; Figs. 2*c* and 2*d*). The RRM motif usually contains a conserved RNP1 or RNP2 sequence featuring aromatic residues (Phe and Tyr), which can stack with the nucleotide bases. However, in the XC_3703 structure no such residues exist on the solvent-exposed face of the central β -strands (Fig. 2*b*). Therefore, XC_3703 seems unlikely to bind RNA in the same manner as classical RRM proteins.

HI1034 from *H. influenzae* was the first protein structure to be solved from the YajQ family (Teplyakov *et al.*, 2003). These two proteins can be superposed with an r.m.s. deviation of 1.50 Å for 159 aligned residues, implying that the structures of the YajQ family are highly conserved. However, there are still differences between the two structures. The last β -strand ($\beta 8$) in XC_3703 is much shorter, and the residues after $\beta 8$ form a bulge conformation (Fig. 3*a*).

3.3. Putative binding site for c-di-GMP in XC_3703

A distinctive feature of XC_3703 is its ability to bind c-di-GMP (An *et al.*, 2014). Besides XC_3703, its homologues

PA4395 from *Pseudomonas aeruginosa* and Smlt4090 from *Stenotrophomonas maltophilia* are also reported to be potential c-di-GMP receptors from the YajQ family. Why the members of this subfamily bind c-di-GMP and others do not remains an open question. We tried to solve the structure of the complex of XC_3703 with c-di-GMP by soaking an apo-form crystal. However, there was no interpretable electron density for c-di-GMP molecules in the data sets. One possible reason is that the apo-form crystal lattice prevented it from binding to c-di-GMP as c-di-GMP induces a large conformation change in the protein. An attempt was made to co-crystallize the protein–ligand complex, but the complex crystals were too tiny for the collection of data. It is known that proteins usually bind to c-di-GMP through specific residues, among which acidic residues (glutamate, aspartic acid) and arginine residues can form hydrogen bonds to the Watson–Crick and Hoogsteen edge of guanine bases, respectively, and tyrosine and phenylalanine residues stack with the guanine bases (Chou & Galperin, 2016). By sequence alignment, Arg65 is conserved among these three c-di-GMP-binding proteins and does not exist in a number of homologues, such as HI1034 from *H. influenzae* and BCK_02545 from *Bacillus cereus* (Figs. 3 and 4), which makes it a candidate for a c-di-GMP-binding residue. Also, the solvent-exposed aromatic residue Phe80 is close to Arg65, further making this region a hotspot for c-di-GMP binding. The electrostatic surface potential of XC_3703 is highly positive in this region, while in HI1034 it is negative, which indicates that XC_3703 and HI1034 have a different binding potential for their interaction

Figure 3

Structural comparison between XC_3703 and its homologue HI1034. (*a*) Superposition of XC_3703 and HI1034. XC_3703 and HI1034 are shown as cartoons in green and yellow, respectively. Secondary-structure elements referred to in the text are labelled. The putative c-di-GMP-binding residues (Arg65 and Phe80, labelled in black) in XC_3703 are shown as ball-and-stick models and the corresponding residues (Ile66 and Ile81, labelled in red) in HI1034 are also shown as ball-and-stick models. (*b*) Electrostatic surface potential of XC_3703 and HI1034 generated by PyMOL. Negative and positive charges are shown in red and blue, respectively. The enlarged views show the putative c-di-GMP-binding site. Domain II from both proteins is shown as a cartoon in green and yellow, respectively, and the important residues are shown as ball-and-stick models.

Figure 4

Sequence alignment of XC_3703 with its homologues. The secondary structure and sequence numbering are shown above the sequence alignment. The putative c-di-GMP-binding residues are indicated by blue stars.

partners such as c-di-GMP (Fig. 3b). However, further biochemical and structural evidence is needed to confirm this.

4. Conclusion

In this study, we solved the structure of XC_3703, which belongs to the YajQ family. The structure of XC_3703 adopts a tandem RRM fold, although it has lost the conserved RNA-binding residues. Considering that XC_3703 plays an essential role in the virulence of *X. campestris* pv. *campestris*, the structure of XC_3703 provides a promising target for controlling the disease caused by this pathogen of economically important crops.

Acknowledgements

We are grateful to the staff of beamline BL17U1 at Shanghai Synchrotron Radiation Facility (SSRF) for their support during data collection and processing. This project was supported by doctoral research funding (No. 14SKY028) from Shangluo University.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
 An, S.-Q., Caly, D. L., McCarthy, Y., Murdoch, S. L., Ward, J., Febrer, M., Dow, J. M. & Ryan, R. P. (2014). *Plos Pathog* **10**, e1004429.
 Barret, M., Egan, F., Fargier, E., Morrissey, J. P. & O'Gara, F. (2011). *Microbiology*, **157**, 1726–1739.
 Chou, S.-H. & Galperin, M. Y. (2016). *J. Bacteriol.* **198**, 32–46.

- Christen, M., Christen, B., Folcher, M., Schauerte, A. & Jenal, U. (2005). *J. Biol. Chem.* **280**, 30829–30837.
 Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. (2012). *Cell*, **149**, 358–370.
 Hengge, R., Gründling, A., Jenal, U., Ryan, R. & Yildiz, F. (2016). *J. Bacteriol.* **198**, 15–26.
 Krasteva, P. V., Giglio, K. M. & Sondermann, H. (2012). *Protein Sci.* **21**, 929–948.
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
 Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
 Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004). *Genes Dev.* **18**, 715–727.
 Römling, U., Galperin, M. Y. & Gomelsky, M. (2013). *Microbiol. Mol. Biol. Rev.* **77**, 1–52.
 Ryan, R. P. (2013). *Microbiology*, **159**, 1286–1297.
 Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y.-W., Zhang, L.-H., Heeb, S., Camara, M., Williams, P. & Dow, J. M. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 6712–6717.
 Saveanu, C., Miron, S., Borza, T., Craescu, C. T., Labesse, G., Gagy, C., Popescu, A., Schaeffer, F., Namane, A., Laurent-Winter, C., Bârză, O. & Gilles, A.-M. (2002). *Protein Sci.* **11**, 2551–2560.
 Schirmer, T. & Jenal, U. (2009). *Nature Rev. Microbiol.* **7**, 724–735.
 Sofos, N., Winkler, M. B. L. & Brodersen, D. E. (2016). *Acta Cryst.* **F72**, 397–402.
 Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. (2013). *Science*, **339**, 786–791.
 Teplyakov, A., Obmolova, G., Bir, N., Reddy, P., Howard, A. J. & Gilliland, G. L. (2003). *J. Struct. Funct. Genomics*, **4**, 1–9.
 Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* **D64**, 61–69.
 Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C. & Chen, Z. J. (2013). *Science*, **339**, 826–830.