PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structures of the apo and GDP-bound forms of a cupin-like protein BbDUF985 from *Branchiostoma belcheri tsingtauense*

Yang Du,[1] Yong-Xing He,[1] Saren Gaowa,[2] Xuan Zhang,[1] Yuxing Chen,[1] Shi-Cui Zhang,[2] and Cong-Zhao Zhou[1]*

[1] Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China

[2] College of Marine Life Sciences, Ocean University of China, Qingdao, Shandong 266003, People's Republic of China

Key words: BbDUF985; crystal structure; cupin_5 family; induced fit; fluorescence spectrometry.

INTRODUCTION

Cupins are ubiquitous proteins existing in all three kingdoms of life and share a conserved β-barrel fold with a characteristic pocket located at the center of the β-barrel.[1,2] Despite sharing the highly conserved topology, cupins show remarkable variations in their sequences, architecture of domains (comprising either one or two cupin domains), quaternary assembly, and the nature of bound metal ion as well.[3] In the Pfam database (http://pfam.sanger.ac.uk/), cupins are classified into 35 protein families, with greatly diversified functions such as isomerases, epimerase, dioxygenase, nonenzymatic storage proteins, and it is suggested that cupin is one of the most functionally diverse protein superfamily.[2,4] In recent years, large-scale sequencing of genomes and cDNAs has elicited a great number of hypothetical cupins whose biological functions need to be established.

The cephalochordate amphioxus is a modern survivor of an ancient chordate lineage and acts as a living fossil between invertebrates and vertebrates. With the completion of the draft genome of Florida amphioxus *Branchiostoma floridae* in 2008, it was revealed that they are not only the most primitive chordates but also the genomic features appear to have a great deal in common with vertebrates as well.[5] BbDUF985 is a hypothetical protein of 172 residues from *Branchiostoma belcheri tsingtauense* belonging to the cupin_5 family (Pfam entry: PF06172) of the cupin superfamily.[4] Interestingly, although BbDUF985 is a protein from eukaryote, sequence homology analysis indicates that all its closest homologs are prokaryotic and more distantly kindred proteins from higher species.

Previously, we solved the structure of YML079w (sharing 33% sequence identity with BbDUF985) from *Saccharomyces cerevisiae* in the guanosine triphosphate (GTP)-bound form,[6] whereas the apo form structure is absent and further biochemical study was not undertook then. Here, we present the crystal structures of BbDUF985 in both the apo and guanosine diphosphate (GDP)-bound forms, which are the first pair of structures (with and without the ligand) from the cupin_5 family. Moreover, using fluorescence spectrometry, we determined the dissociation constant (K_D) of GDP and investigated the spectra of BbDUF985 with several proposed ligands binding to the proteins in cupin superfamily. Combined with sequence analysis, we speculate that BbDUF985 and its homologs in the cupin_5 family might be involved in the nucleotide transport or metabolism.

Grant sponsor: Ministry of Science and Technology of China; Grant number: 2008AA092600; Grant sponsor: National Natural Science Foundation of China; Grant number: 30870490

*Correspondence to: Cong-Zhao Zhou, Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China.
E-mail: zcz@ustc.edu.cn

Received 31 March 2010; Revised 7 May 2010; Accepted 12 May 2010
Published online 20 May 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22771

MATERIALS AND METHODS

Cloning, expression, and purification of BbDUF985

The coding sequence of BbDUF985 was cloned into a pET28a-derived vector. The recombinant protein with a hexa-histidine (6 × His) tag at the N-terminus was overexpressed in *E. coli* Rosetta (DE3) (Novagen, Madison, WI) strain using 2 × YT culture medium (16 g of tryptone, 10 g of yeast extract, and 5 g of NaCl per liter). The cells were grown at 37°C up to an $A_{600\ nm}$ of 0.6. Expression of recombinant BbDUF985 was induced at exponential phase with 0.2 m*M* isopropyl-β-D-thiogalactoside and cell growth continued for another 20 h at 16°C before harvesting. Cells were collected by centrifugation at 4000*g* for 20 min and resuspended in lysis buffer (20 m*M* Tris-Cl, pH 8.0, 150 m*M* NaCl). After 5 min of sonication and centrifugation at 12,000*g* for 30 min, the supernatant containing the soluble target protein was collected and loaded to a Ni-NTA column (GE Healthcare) equilibrated with binding buffer (20 m*M* Tris-Cl, pH 8.0, 150 m*M* NaCl). The target protein was eluted with 200 m*M* imidazole buffer and further loaded onto a Superdex 200 column (Amersham Biosciences) pre-equilibrated with 20 m*M* Tris-Cl, pH 8.0, 150 m*M* NaCl. Fractions containing the target protein were collected and concentrated to 15 mg/mL. The purity of protein was determined on SDS-PAGE and then the protein sample was stored at −80°C.

Crystallization, data collection, and processing

The crystals of BbDUF985 were grown at 289 K with hanging drop vapor-diffusion techniques by mixing 1 μL of the 15 mg/mL protein sample with equal volume of reservoir. Apo-form crystals were obtained in the drop containing 2.0*M* ammonium sulfate, 0.1*M* sodium acetate trihydate pH 4.6, and reached to a maximal size for X-ray diffraction in 1 week. GDP-bound form crystals were obtained by soaking the apo-form crystal to a 10-μL crystallization reservoir containing 2 m*M* GDP molecule for about 30 min and then mounted for X-ray diffraction immediately. The diffraction image of the apo and GDP-bound forms were recorded at 100 K in a liquid nitrogen stream using beamline 3W1A at Beijing Synchrotron Radiation Facility (λ = 1.0000 Å) with MAR 165 mm CCD (MARresearch, Germany) and Rigaku MM007 X-ray generator (λ = 1.5418 Å) with MarRearch 345 image-plate detector (USTC, Hefei, China), respectively. Data were processed with MOSFLM 7.0.4[7] and scaled with SCALA.[8]

Structure solution and refinement

The crystal structures of BbDUF985 were determined by the molecular replacement method with MOLREP[9] using the coordinates of homologous protein from *Shewanella oneidensis* [protein data bank (PDB) code 1yud], which shares 37% sequence identity (156 residues aligned) with BbDUF985 as the search model. The root-mean-square deviation (RMSD) between 1yud and the apo form is 1.5 Å and that between 1yud and the ligand-bound form is 1.4 Å using pairwise Dali server. The initial model was refined by using the maximum likelihood method implemented in REFMAC5[10] as part of CCP4 program suite[11] and rebuilt interactively by using the σ_A-weighted electron density maps with coefficients $2mF_O$-DF_C and mF_O-DF_C in the program COOT.[12] During the later stage, the restrained positional and B-factor refinement was performed using the program phenix.refine[13] and tight non-crystallographic symmetry (NCS) restraints over the two subunits were applied during the refinement. The final models were evaluated with the programs MOLPROBITY[14] and PROCHECK.[15] The final coordinates and structure factors were deposited in the PDB under the accession code of 3LOI and 3LZZ, respectively. The data collection and structure refinement statistics were listed in Table I. All structure figures were prepared with the program PyMOL.[16]

Ligand binding assay

Based on principle of protein fluorescence spectroscopy, intrinsic aromatic amino acids fluorescence is strongly influenced by the proximity of other residues, which will change the fluorescence intensity and/or emission wavelength. Thus, this method is a sensitive measurement of the local conformational change caused by aromatic residues.[17] To determine the GDP binding affinity of BbDUF985, intrinsic phenylalanine fluorescence of the protein was monitored by wavelength from 285 to 450 nm using a RF-5301PC spectrofluorophotometer (Shimadzu, Japan). The excitation wavelength was set to 275 nm at a slit width of 3 nm. GDP, deoxythymidine diphosphate (dTDP), glucose-6-phosphate and ectoine were selected to assay fluorescence spectrum, and the identical nonsaturation concentration (200 μ*M*) was adopted. To determine the K_D of GDP, a gradient of GDP concentrations (from a 10-m*M* stock solution) in phosphate buffered saline were added into 1-μ*M* BbDUF985 samples and equilibrated for 10 min before measurement. On GDP binding, the fluorescence intensity of BbDUF985 was decreased, which is proportional to a given ligand concentration. The change of fluorescence intensity could be analyzed using 1:1 binding site model [as illustrated in Fig. 1(A)] adopting the following equation.[18]

$$I = I_0 + [\Delta I \times L/(L + K_D)] \quad (1)$$

where "I" stands for the fluorescence intensity at a given ligand concentration, "I_0" for the fluorescence intensity

Table I
Data Collection and Refinement Statistics

Data collection	Apo form	GDP-bound form
Resolution range (Å)	40.00–2.10	19.94–2.50
overall/outer shell	(2.14–2.10)	(2.64–2.50)
Space group	$I4_122$	$I4_1$
Unit-cell parameters (Å)	$a = b = 91.54$, $c = 173.22$ $\alpha = \beta = \gamma = 90^\circ$	$a = b = 88.86$, $c = 168.17$ $\alpha = \beta = \gamma = 90^\circ$
Wavelength (Å)	1.0000	1.5418
Total reflections	219,274	59,714
Unique reflections	21,913	22,168
Completeness (%)	99.9/98.5	98.6/98.9
overall/outer shell		
$<I/\sigma(I)>$	18.2 (2.2)	15.6 (3.1)
R_{merge}[a] (%)	11.7 (57.7)	5.3 (37.1)
overall/outer shell		
Refinement statistics		
R_{factor}[b] (%)	22.0	21.5
R_{free}[c] (%)	25.0	25.6
Number of protein molecules	1	2
Number of atoms	1474	2839
Rms deviation from target[d]		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	1.078	1.191
Average B factors:	40.6	40.0
Ramachandran plot (%)[e]	99.4/0.6/0	97.2/2.8/0

[a] $R_{merge} = \frac{\Sigma_h \Sigma_i |I_{hi} - \langle I_h \rangle|}{\Sigma_h \Sigma_i I_{hi}}$, where I_{hi} is the ith observation of the reflection h, while I_h is the mean intensity of reflection h.

[b] $R_{factor} = \Sigma \frac{||F_{obs}| - |F_{calc}||}{|F_{obs}|}$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[c] R_{free} was calculated with 5% randomly selected reflections for the Duf apoform and ligand-bound form. No refinement was done on the 5% of randomly selected reflections at any stage.

[d] Root-mean-square deviation of bond lengths or bond angles from ideal geometry.

[e] Percentage of residues in most favoured/additionally allowed/generously allowed/disallowed regions of Ramachandran plot, according to MOLPROBITY.

without ligand, "ΔI" for the change of fluorescence intensity, and "L" for the ligand concentration. Changes in the maximal fluorescence intensity at 337 nm were fit based on the above equation using Origin 7.5 software (Northampton) and the K_D of GDP toward BbDUF985 was determined. All data represent the average of three independent measurements with the standard deviations shown as error bars.

RESULTS AND DISCUSSION

Overall structure

The crystal structure of BbDUF985 in the apo form was refined to the resolution of 2.1 Å. It belongs to the $I4_122$ space group with one molecule in an asymmetric unit (Table I). Most residues are well fitted in the electron density map except for the N-terminal residues Met1-Ser9. The overall structure of BbDUF985 monomer exhibits a typical cupin-like β-barrel topology which comprises an eight-strand β-barrel surrounded by four α-helices [Fig. 1(A)]. Besides, a long loop between β1 and β2 resembles a handle protruding from the core domain. Two monomers are further assembled into a dimer via reciprocal interactions from strands β1, β2, β3, and β8. The dimeric interface is about 1500 Å^2, as shown in Figure 1(A), which involves eight hydrogen bonds and extensive hydrophobic interactions analyzed by PDBsum.[19] In addition, BbDUF985 also exists as a dimer in solution confirmed by gel filtration using Superdex 75 column (GE healthcare), indicating that the protein probably functions as a dimer form (data not shown).

The GDP binding site

Comparative structural analysis using Dali server (http://www.ebi.ac.uk/dali/)[20] revealed a group of structurally similar proteins, such as auxin binding protein (PDB code 1LR5, *Z*-score 11.1), canavalin (PDB code 1CAX, *Z*-score 9.5), sugar phosphate isomerase (PDB code 3I7D, *Z*-score 9.5), dTDP sugar epimerase (PDB code 3EJK, *Z*-score 9.1), and other hypothetical cupins. Combined with the information provided by Pfam database (cupin clan NO. CL0029), we attempted to soak the apo-form crystal with several chemical molecules such as GDP, glucose-6-phosphate, and ectoine. In the end, only GDP can be bound in the crystal structure of BbDUF985. The structure was refined to the resolution of 2.5 Å. It belongs to the $I4_1$ space group with two molecules in an asymmetric unit (Table I). GDP is bound at the center of the β-barrel with the ribose ring and phosphate group moiety exposed to the solvent. The GDP molecule fits well in electron density map ($2mF_O$-DF_C map) contoured at 1.5 σ [Fig. 1(A)], when compared with the YML079w structure in the GTP-bound form, in which only the purine base moiety fits the electron density map.[6]

As shown in Figure 1(B), GDP is stabilized by both hydrophobic interactions and hydrogen bonds. The hydrophobic interaction is mainly contributed by Phe31, Phe140, and Phe142 and five hydrogen bonds are formed involving three water molecules. In detail, N1 of purine base forms H-bond with Wat211 that is further stabilized by Oγ1 of Thr54, while N7 of purine base interacts with Nε2 of His24. Furthermore, O3 of ribose moiety makes H-bond with Wat180 and O1A of α-phosphate group interacts with Wat244 that can fix the phosphate group tail of GDP molecule probably avoiding from wobbling in solvent environment. Besides, N2 of purine base interacts with Oε1 of Asp33 through attractive electrostatic force. To perform a multiple sequence alignment with Multalin[21] and ESPript[22] [Fig. 1(D)], we selected a set of representative homologs in the cupin_5 family. Remarkably, all the residues participating in ligand binding are conserved in the cupin_5 family even between BbDUF985 and relative distant homologs from plant, implying that functional convergence and conservation of GDP binding pattern of these proteins from the cupin_5 family.

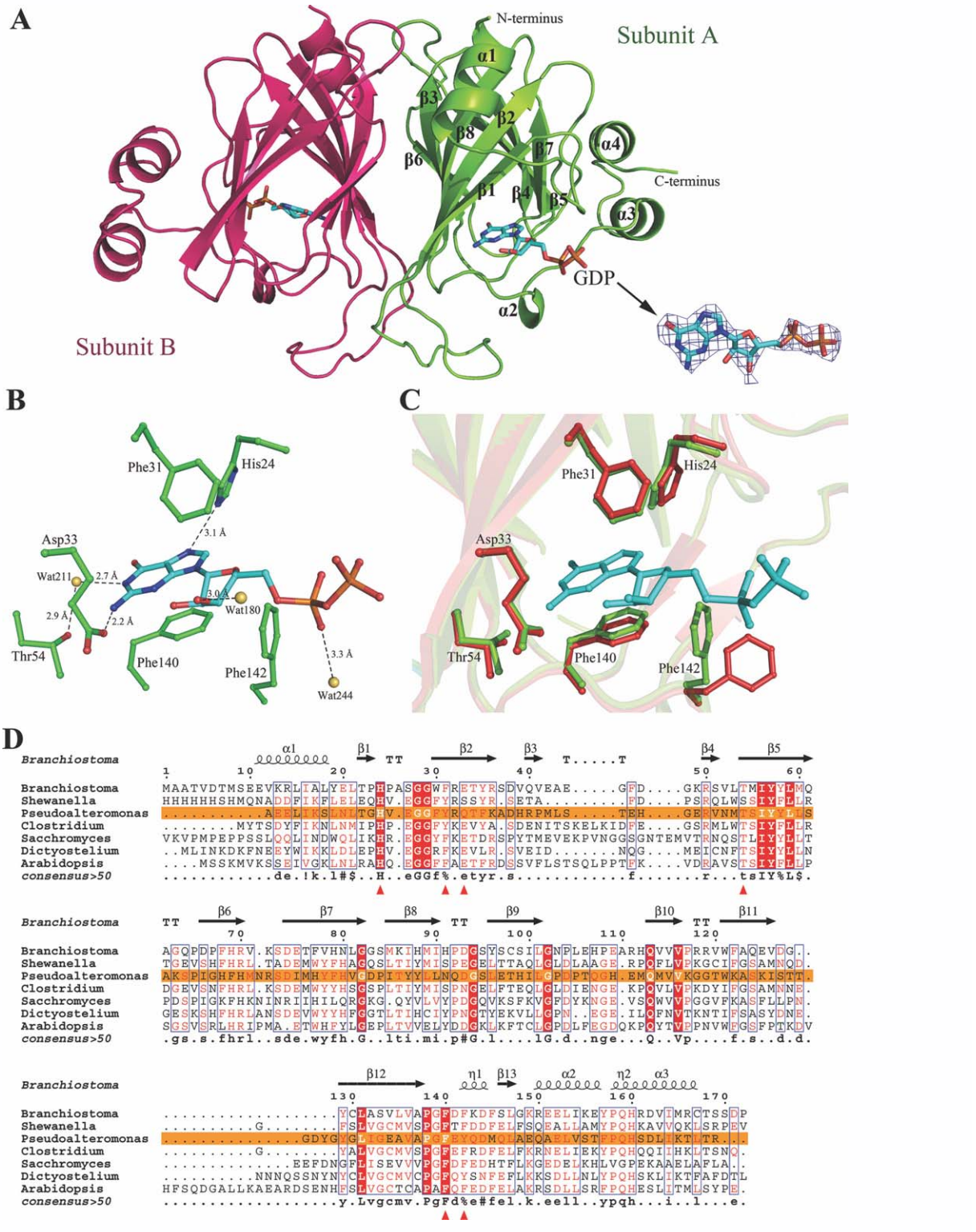

Figure 1

Crystal structures and sequence alignment of BbDUF985. **A**: Schematic representation of BbDUF985 homodimer comprising of subunits A and B with one GDP molecule binding to each subunit. All β-strands and helices are labeled sequentially in subunit A. The black arrow denotes the $2mF_O$-DF_C map of GDP that is contoured at 1.5 σ. **B**: Interactions between BbDUF985 and GDP. **C**: Structural superposition of the apo and GDP-bound forms, which are colored in red and chartreuse, respectively. **D**: Multiple sequence alignments of BbDUF985 (the first row) against the homologs from the prokaryotes *Shewanella oneidensis*, *Clostridium phytofermentans*, and *Pseudoalteromonas atlantica*, fungus *Saccharomyces cerevisiae*, protozoan *Dictyostelium discoideum*, plant *Arabidopsis thaliana*, respectively. The cupin_5 domain of NUP from *Pseudoalteromonas atlantica* are indicated in orange background and the conserved residues involving in GDP binding are marked by red up-triangle arrow.

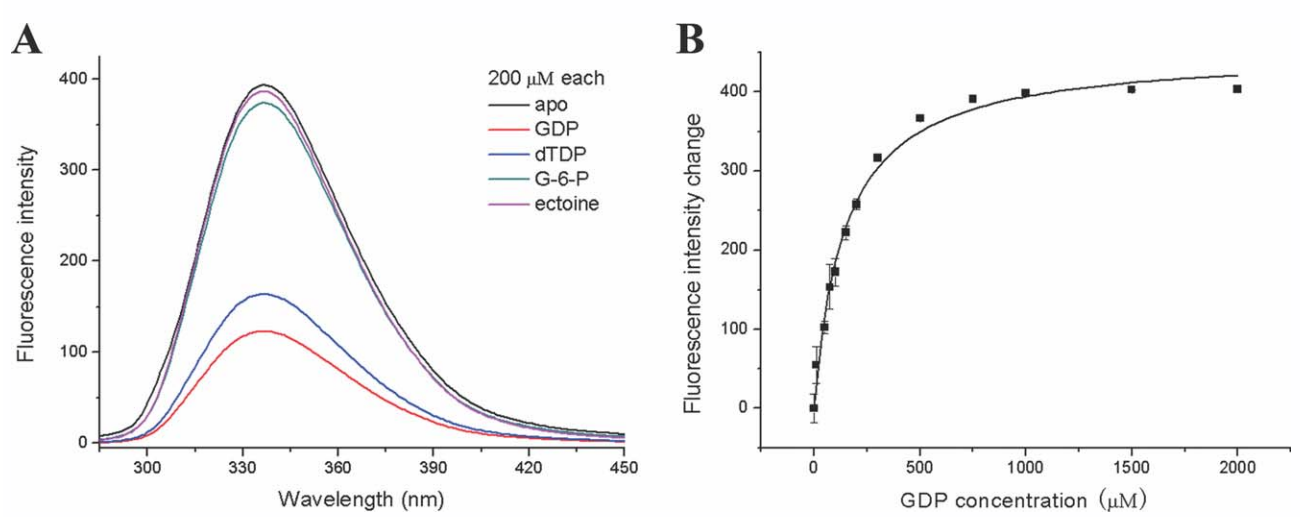

Figure 2

The GDP binding assay. **A**: Fluorescence spectra of BbDUF985 with each selected molecule in contrast to that of the apo form. **B**: The plot of decrease in maximal fluorescence intensity (Y) against the GDP concentration (X) for determination of K_D of GDP.

Although superposition of the apo and GDP-bound forms yields a RMSD of only 0.5 Å, obvious conformational changes are still observed in the pocket on GDP binding, which is mainly contributed by the conserved ligand binding residues [Fig. 1(C)]. As a result of induced fit, residues His24 and Asp33 shift inward to form hydrogen bonds with GDP, and the side chains of Phe31, Phe140, and Phe142 flip toward the purine and ribose rings of GDP to make hydrophobic interactions.

The GDP binding affinity with BbDUF985

The cupin superfamily is a very large group of functionally diverse proteins that are found in all three kingdoms of life: archaea, eubacteria, and eukaryota, and lots of cupin proteins have not been functionally characterized yet.[23] As mentioned above, BbDUF985 belongs to the uncharacterized cupin_5 family of entire 35 cupin families, the residues of which involving in ligand binding are well conserved [Fig. 1(D)], implying that proteins in the cupin_5 family probably share a similar molecular function differing from that of other cupin families.

Given the feature of hydrophobic interaction with GDP contributed by aromatic residues in binding pocket, protein fluorescence spectrometry was applied. Change of fluorescence intensity/wavelength can be used to compare binding affinity of different chemical molecules. Based on the Dali output, four representative compounds (200 μ*M* each) were assayed, listed as GDP, dTDP (nucleotide moiety of substrate for dTDP sugar isomerase), glucose-6-phosphate (substrate for gluocose-6-phosphate isomerase), and ectoine (pyrimidine-like molecule). The fluorescence spectra revealed that GDP and dTDP are able to trigger significant decrease of the fluorescence intensity, whereas gluocose-6-phosphate or ectoine is not [Fig. 2(A)]. In addition, using the spectra with a gradient of GDP concentrations [Fig. 2(B)], we determine the K_D of GDP, which is 146 ± 11 μ*M*, approximately equivalent to the average physiological concentration of GDP (159 ± 51 μ*M*) in most species except for *Homo sapiens*.[24]

Moreover, there are five kinds of domain organization in the cupin_5 family,[4] one of which was defined by a protein of 575 amino acids from *Pseudoalteromonas atlantica*. It comprises both purine nucleoside permease (NUP) domain (N-terminal 53–385 residues) and cupin_5 domain [N-terminal 414–555 residues, as shown in Fig. 1(D)]. This evidence also suggests that the proteins in the cupin_5 family might be related to the nucleotide transport or metabolism. Despite the definite biological function of BbDUF985 remains unclear, however, our structural and biochemical data of BbDUF985 provided hints for further functional characterization of proteins in the cupin_5 family.

ACKNOWLEDGMENTS

The authors thank Beijing Synchrotron Radiation Facility (BSRF) for X-ray data collection.

REFERENCES

1. Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW. Evolution of functional diversity in the cupin superfamily. Trends Biochem Sci 2001;26:740–746.
2. Dunwell JM, Purvis A, Khuri S. Cupins: the most functionally diverse protein superfamily? Phytochemistry 2004;65:7–17.
3. Agarwal G, Rajavel M, Gopal B, Srinivasan N. Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold. PLoS One 2009;4:e5736.

4. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32(Database issue):D138–D141.
5. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. The amphioxus genome and the evolution of the chordate karyotype. Nature 2008;453:1064–1071.
6. Zhou CZ, Meyer P, Quevillon-Cheruel S, De La Sierra-Gallay IL, Collinet B, Graille M, Blondeau K, Francois JM, Leulliot N, Sorel I, Poupon A, Janin J, Van Tilbeurgh H. Crystal structure of the YML079w protein from *Saccharomyces cerevisiae* reveals a new sequence family of the jelly-roll fold. Protein Sci 2005;14:209–215.
7. Powell HR. The Rossmann Fourier autoindexing algorithm in MOSFLM. Acta Crystallogr D Biol Crystallogr 1999;55(Pt 10):1690–1695.
8. Evans PR. Data reduction. In: Proceedings of CCP4 Study Weekend, on Data Collection and Processing. Warrington: Daresbury Laboratory; 1993.
9. Vagin A, Teplyakov A. MOLREP an automated program for molecular replacement. J Appl Crystallogr 1997;30:1022–1025.
10. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53(Pt 3):240–255.
11. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50(Pt 5):760–763.
12. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60(Pt 12 Pt 1):2126–2132.
13. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 2002;58(Pt 11): 1948–1954.
14. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, III, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35(Web Server issue):W375–W383.
15. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
16. DeLano WL. The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.
17. Vivian JT, Callis PR. Mechanisms of tryptophan fluorescence shifts in proteins. Biophys J 2001;80:2093–2109.
18. Smits SH, Hoing M, Lecher J, Jebbar M, Schmitt L, Bremer E. The compatible-solute-binding protein OpuAC from Bacillus subtilis: ligand binding, site-directed mutagenesis, and crystallographic studies. J Bacteriol 2008;190:5663–5671.
19. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM. PDBsum: a web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci 1997;22:488–490.
20. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
21. Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 1988;16:10881–10890.
22. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 2003;31:3320–3323.
23. Khuri S, Bakker FT, Dunwell JM. Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. Mol Biol Evol 2001;18:593–605.
24. Traut TW. Physiological concentrations of purines and pyrimidines. Mol Cell Biochem 1994;140:1–22.