PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of the protein At3g01520, a eukaryotic universal stress protein-like protein from *Arabidopsis thaliana* in complex with AMP

Do Jin Kim,[1] Eduard Bitto,[2] Craig A. Bingman,[3] Hyun-Jung Kim,[4] Byung Woo Han,[1]* and George N. Phillips, Jr.[3,5]*

[1] Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742, Korea

[2] Department of Chemistry and Biochemistry, Georgian Court University, Lakewood, New Jersey 08701

[3] Department of Biochemistry, Center for Eukaryotic Structural Genomics, University of Wisconsin-Madison, Madison, Wisconsin 53706

[4] Laboratory of Stem Cell and Molecular Pharmacology, College of Pharmacy, Chung-Ang University, Seoul 156-756, Korea

[5] BioSciences at Rice and Department of Chemistry, Rice University, Houston, Texas 77251

ABSTRACT

Members of the universal stress protein (USP) family are conserved in a phylogenetically diverse range of prokaryotes, fungi, protists, and plants and confer abilities to respond to a wide range of environmental stresses. *Arabidopsis thaliana* contains 44 USP domain-containing proteins, and USP domain is found either in a small protein with unknown physiological function or in an N-terminal portion of a multi-domain protein, usually a protein kinase. Here, we report the first crystal structure of a eukaryotic USP-like protein encoded from the gene At3g01520. The crystal structure of the protein At3g01520 was determined by the single-wavelength anomalous dispersion method and refined to an *R* factor of 21.8% (R_{free} = 26.1%) at 2.5 Å resolution. The crystal structure includes three At3g01520 protein dimers with one AMP molecule bound to each protomer, comprising a Rossmann-like α/β overall fold. The bound AMP and conservation of residues in the ATP-binding loop suggest that the protein At3g01520 also belongs to the ATP-binding USP subfamily members.

Proteins 2015; 83:1368–1373.

Key words: universal stress protein; *Arabidopsis thaliana*; At3g01520.

Grant sponsor: National Institute of Health; grant numbers: U01 GM098248, U54 GM074901, and P50 GM64598 (Protein Structure Initiative); Grant sponsor: Ministry of Science, ICT and Future Planning of Korea; grant number: 2011-0030001 and 2013M3A6A4043695 (Tumor Microenvironment Global Core Research Center and Global Frontier Project funded through the National Research Foundation); Grant sponsor: Offices of Biological and Environmental Research; Grant sponsor: Basic Energy Sciences of the US Department of Energy; Grant sponsor: National Center for Research Resources of the National Institutes of Health.
Do Jin Kim and Eduard Bitto contributed equally to this work.

*Correspondence to: Byung Woo Han, Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742, Korea. E-mail: bwhan@snu.ac.kr (or) George N. Phillips, Department of Biochemistry, Center for Eukaryotic Structural Genomics, University of Wisconsin-Madison, WI 53706.
E-mail: georgep@rice.edu
Received 14 January 2015; Revised 20 April 2015; Accepted 24 April 2015
Published online 28 April 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24821

INTRODUCTION

The gene At3g01520 of *Arabidopsis thaliana* encodes 175-residue universal stress protein (USP)-like protein (Pfam accession: PF00582), which is widely found in the genomes of bacteria, as well as fungi, protozoa, and plants.[1] The USP superfamily represents a set of small cytoplasmic proteins whose expressions are affected by a wide range of internal or external stresses.[1,2] Genetic evidence has subsequently shown that USP mediates survival of cells in response to a wide variety of stress states, including nutrient starvation, exposure to heat, acid, heavy metals, oxidative agents, osmotic stress, antibiotics, and uncouplers of oxidative phosphorylation.[3–5]

Structures of many bacterial and viral USP proteins have been solved so far. Interestingly, while the USP domain structure of MJ0577 from *Methanocaldococcus jannaschii* was solved with a bound ATP, *Haemophilus influenzae* UspA lacks both ATP-binding activity and ATP-binding residues.[6,7] This suggests that the USP domain family proteins fall into two major groups depending on their nucleotide binding capabilities. However, despite the knowledge of bacterial USP proteins, the functional diversity of the USPs in other organisms, including various plant species, has not been well defined.[8]

A. thaliana contains 44 USP domain-containing proteins, and USP domain is found either in a small protein with unknown physiological function or in an N-terminal portion of a multidomain protein, usually a protein kinase.[1] Most of the *A. thaliana* USP-like proteins are annotated as "adenine nucleotide α-hydrolase-like superfamily protein", although it has not been reported whether they bind or hydrolyze ATP.[9,10] In this study, we report the crystal structure of a eukaryotic USP-like protein At3g01520 from *A. thaliana* in complex with AMP at 2.5 Å resolution. This work represents the first structure of eukaryotic USP family protein, thus contributing to expand our knowledge on the structural features of the diverse USP protein family members.

MATERIALS AND METHODS

Native and SeMet-labeled At3g01520 proteins were cloned and purified following the standard Center for Eukaryotic Structural Genomics (CESG) pipeline protocol described in detail elsewhere.[11–14] Crystals of the protein At3g01520 were grown at 293 K by the hanging-drop method from 10 mg mL^{-1} protein solution in buffer (50 m*M* NaCl, 3 m*M* NaN_3, 0.3 m*M* TCEP, and 5 m*M* MES-NaOH, pH 7.0) mixed with an equal amount of well solutions containing 18% (v/v) PEG 2K, 5% (v/v) DMSO, and 100 m*M* PIPES, pH 6.5. They belong to space group *P*1, with unit-cell parameters $a = 63.35$, $b = 65.66$, $c = 73.01$ Å, $\alpha = 75.45$, $\beta = 75.04$, $\gamma = 66.11°$. Crystals were cryoprotected by soaking in a cryosolution containing 20% (v/v) PEG 2K and 100 m*M* Tris-HCl, pH 8.0, and a final concentration of 20% (v/v) ethylene glycol. X-ray diffraction data for native and selenomethionine crystals were collected at the 22-ID synchrotron beamlines at the Advanced Photon Source of Argonne National Laboratory.

We have solved the crystal structure of At3g01520 by SAD phasing at 2.50 Å resolution. Twenty selenomethionine sites were located with the program SHELXD.[15] After initial refinement, six At3g01520 protomers were defined in an asymmetric unit. Out of 18 methionine residues in an asymmetric unit, two of the residues showed definable alternate conformations, explaining the two extra selenomethionine sites found with SHELXD. The selenomethionine sites were used to calculate the phases with RESOLVE.[16] The SAD-phased electron density map was of high quality and was readily interpreted by the automatic model building procedure of RESOLVE.[16] The structure was completed using alternate cycles of manual building in Xfit[17] and refinement in REFMAC.[18] Noncrystallographic symmetry constraints, as defined in the standard refinement protocol of Refmac,[18] were applied throughout the refinement. NCS constraints were not released due to low redundancy of the data as the structure was refined in P1 space group. All refinement steps were monitored using an R_{free} value based on 5.3% of the independent reflections. The stereochemical quality of the final model was assessed using PROCHECK[19] and MolProbity.[20] The coordinates and structure factors have been deposited in the Protein Data Bank (http://www.pdb.org) for immediate release with PDB ID code 2GM3. Table I summarizes the data collection, phasing, and model refinement statistics.

Analytical size-exclusion chromatography was performed on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 50 m*M* Bis-Tris at pH 7.0, 150 m*M* NaCl, 0.3 m*M* TCEP at 277 K. The samples were injected using a 0.25 mL loop, and the flow rate was 0.5 mL/min for 36 mL. Six protein standards from Gel Filtration Marker Kit (Sigma-Aldrich) were used to calibrate the column. The protein standards used were thyroglobulin (MW 669,000), apoferritin (MW 443,000), β-amylase (MW 200,000), alcohol dehydrogenase (MW 150,000), bovine serum albumin (MW 66,000), and carbonic anhydrase (MW 29,000). Output data were subsequently exported and analyzed in Excel (Microsoft)

RESULTS AND DISCUSSION

We have determined the crystal structure of the protein At3g01520 in complex with AMP. The final model includes six AMP-bound protomers in a crystallographic asymmetric unit. The R and R_{free} factors were 22.1 and 26.1%,

Table I
Statistics for Data Collection, Phasing, and Model Refinement

Data collection and phasing[a]	
Space group	*P*1
Cell dimensions	
a, *b*, *c* (Å)	63.35, 65.66, 73.01
α, β, γ (°)	75.45, 75.04, 66.11
Data set	Se λ1 (peak)
X-ray wavelength (Å)	0.9793
Resolution (Å)[b]	50.00–2.50 (2.59–2.50)
Total/unique reflections	133,684/36,006
Completeness (%)	97.2 (89.4)
R_{merge} (%)[c]	5.8 (23.9)
Figure of merit[d] for SAD phasing: 0.51	
Refinement	Data set: Se λ1 (peak)
Resolution (Å)	50.0–2.50
R_{work}[e]/R_{free}[f]	0.221/0.261
No. of protein atoms/mean B-factor (Å^2)	7,084[g]/60.5
No. of ligand atoms (AMP)/mean B-factor (Å^2)	138/47.1
No. of solvent atoms/mean B-factor (Å^2)	13/46.1
Ramachandran plot analysis (for Chain A)	
Most favored regions	143 (97.3%)
Additional allowed regions	4 (2.7%)
Disallowed regions	0 (0%)
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.016
Bond angles (°)	1.50

[a]Data collected at the Sector 22-ID of the Advanced Photon Source.
[b]Numbers in parentheses indicate the highest resolution shell of 20.
[c]$R_{merge} = \Sigma_h \Sigma_i |I(h)_i - <I(h)>| / \Sigma_h \Sigma_i I(h)_i$, where $I(h)$ is the observed intensity of reflection h, and $<I(h)>$ is the average intensity obtained from multiple measurements.
[d]Figure of merit = $<|\Sigma P(\alpha) e^{i\alpha}| / \Sigma P(\alpha)|>$, where α is the phase angle and $P(\alpha)$ is the phase probability distribution.
[e]$R_{work} = \Sigma||F_o| - |F_c|| / \Sigma |F_o|$, where $|F_o|$ is the observed structure factor amplitude and $|F_c|$ is the calculated structure factor amplitude.
[f]R_{free} = *R*-factor based on 5.1% of the data excluded from refinement.
[g]Number of nonhydrogen protein atoms included in refinement.

respectively, for the 20.00 − 2.50 Å data (Table I). The structure is resolved from residue 5 to residue 175. Loops comprising residues 56–65 and 137–145 are missing in all the chains in the asymmetric unit, suggesting that these are highly flexible regions. One At3g01520 protomer comprises a central five-stranded, parallel β-sheet (β4-β5-β1-β2-β3), which is surrounded by five α-helices [Fig. 1(A)]. It is folded into the classic Rossmann-like α/β-fold, resembling the structures of other reported structures of USP family members.[6,7,21–24]

Although it was previously reported that the protein At3g01520 lacks dimer-forming hydrophobic β5 region,[9] the hydrophobic β5 sequence is highly conserved and the protein At3g01520 also exists as a dimer in the crystal [Fig. 1(B,C)]. The protein At3g01520 has buried surface area of 1,017 Å^2 per monomer, which corresponds to ~12% of the monomer surface area (87,875 Å^2). The central β-sheets of the two protomers form an extended β-sheet by β5 strands from each protomer in the dimer structure and this feature has been widely observed in other reported USP structures in a dimeric state [Fig. 1(B)].[6,7,21–24] The dimer interface is mediated through van der Waals packing of a number of hydrophobic residues (Ile38, Phe46, Phe127, Pro157, and Met159) as well as electrostatic interactions formed by side chains of Thr6, Arg40, Asn42, Asp45, and Asp126. Based on the dimeric interface analysis from crystallographic dimer packing and size-exclusion chromatography data [Fig. 1(D)], we propose that the protein At3g01520 forms a dimer for its physiological conditions.

The USP family proteins fall into two groups depending on their ATP-binding capabilities.[7,9] In the case of the protein At3g01520, there have been mixed predictions for its ATP-binding capability. According to the overall sequence similarity of At3g01520 to MJ0577 (Protein Data Bank, PDB, ID 1 MJH), which contained naturally acquired ATP in its solved crystal structure, the protein At3g01520 has been classified as a 1 MJH-like plant protein. However, extensive sequence analysis could not reveal conserved ATP-binding residues such as Asp and Ser that have been predicted as ATP-binding residues from the crystal structure of 1 MJH.[9] Interestingly, after model building of all the protein residues in a crystallographic asymmetric unit, the *Fo* − *Fc* omit electron density map clearly showed positive extra electron densities around the similar ATP-binding site of MJ0577 in all six protomers in the asymmetric unit, which could be definitely modeled as an AMP molecule [Figs. 1(A,B) and 2(A)]. Since we did not add any nucleotide supplements in culture media for *E. coli*, this naturally acquired AMP molecule from *E. coli* could be reminiscent of ATP molecule of other ATP-binding USP proteins. Thus, this observation strongly supports that the protein At3g01520 also belongs to the nucleotide-binding USP subfamily members.

Whether the preferred ligand of the protein At3g01520 is ATP or AMP has not been known. However, key residues for the AMP interaction exist on a flexible loop that corresponds to the ATP-binding sites of homologous USP proteins. The bound AMP molecule is buried within a cleft near the surface mainly interacting with a number of main chain nitrogen and carbonyl oxygen atoms [Figs. 1(A) and 2(A)], resembling other AMP/ATP-bound USP structures.[6,21,23,24] The ATP-binding USP subfamily members adopts a similar conformation of ATP-binding loops with a consensus motif G-2x-G-9x-G-(S/T) (x stands for any amino acid).[7,21–24] This ATP-binding consensus sequence motif is also strictly conserved in At3g01520 [Fig. 1(C)]. The phosphate group of the bound AMP is exposed to the solvent, leaving enough room for β- and γ-phosphates for ATP binding. However, unlike other ATP-bound USP structures, the conserved ATP-binding loop region of the protein At3g01520 could not be fully defined in our crystal structure and it may be caused by the absence of β- and γ- phosphate moiety in the case of the bound AMP, resulting in a degree of conformational flexibility. Presumably, the disordered loop of the protein At3g01520 may be able to form an optimum structure in the event

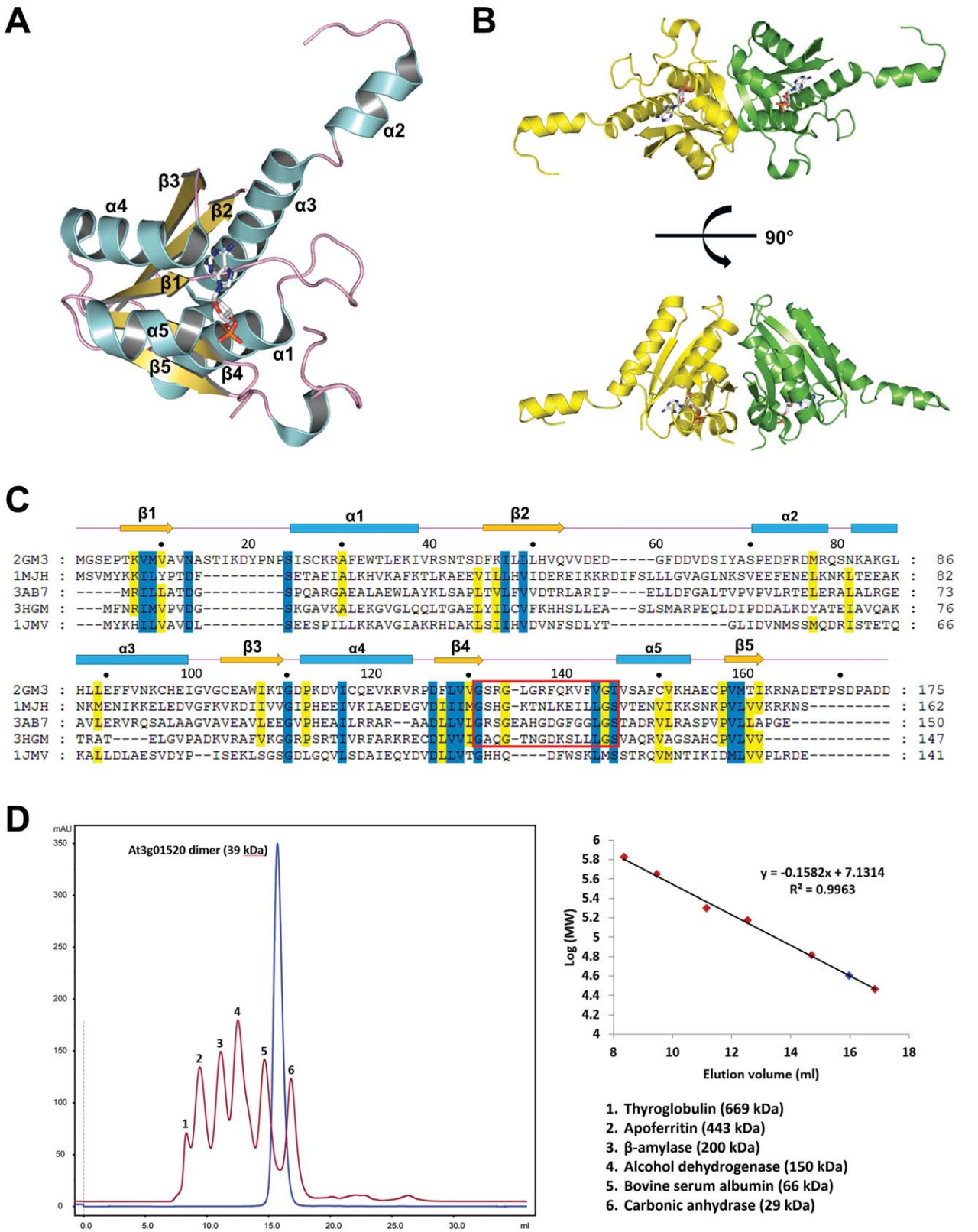

Figure 1

Overall structure of the protein At3g01520. **A**: Monomer structure of the protein At3g01520. α-helices, β-strands, and loops are colored in cyan, yellow, and pink, respectively. The bound AMP is shown in a stick model. **B**: Dimerization of the protein At3g01520. Chain A and B are colored in yellow and green, respectively. The bound AMP is shown in a stick model. **C**: Multiple sequence alignment of USP proteins from *A. thaliana* (PDB ID: 2GM3), *M. jannaschii* (PDB ID: 1 MJH), *T. thermophilus* (PDB ID: 3AB7), *H. elongata* (PDB ID: 3HGM), and *H. influenzae* (PDB ID: 1JMV). The conserved G-2X-G-9X-(S/T) motif is boxed in red. α-helices and β-strands are indicated above the sequences as cyan cylinders and yellow arrows, respectively. **D**: Analytical size-exclusion chromatograms of At3g01520 (blue line) and size markers (magenta line). The position of molecular weight markers is indicated for comparison.

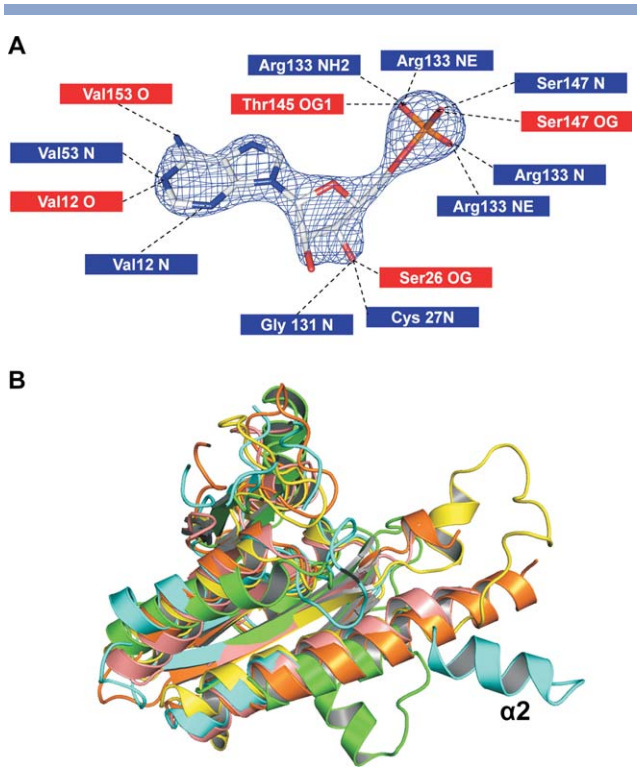

Figure 2

Interaction of the protein At3g01520 with bound AMP. **A**: *Fo-Fc* electron density map of the bound AMP contoured at 3.0 σ. **B**: Aligned monomeric structures of USP proteins. USP proteins from *A. thaliana*, *M. jannaschii*, *T. thermophilus*, *H. elongate*, and *H. influenzae* are colored in cyan, orange, pink, yellow, and green, respectively. The α2 helix of the protein At3g01520 is labeled below.

of interaction with a proper ligand, such as ATP, by reorienting and ordering the flexible loop.

A structural homology search was conducted using the DALI server.[25] Close structural homologs of the protein At3g01520 (DALI *Z* score higher than 15) include a number of established ATP-binding USP subfamily proteins: USP domain of MJ0577 from *M. jannaschii* with *Z* score 15.7, r.m.s.d. distance 3.5 Å [mt]128 aligned C_α residues, and 20% sequence identity (PDB ID 1 MJH[6]); N-terminal USP domain of tandem-type USP TTHA0350 from *Thermus thermophilus* HB8 with *Z* score 15.5, r.m.s.d. distance 2.4 Å over 118 aligned C_α residues, and 20% sequence identity (PDB ID 3AB7[24]); USP TeaD from *Halomonas elongata* with *Z* score 15.3, r.m.s.d. distance 3.9 Å over 120 aligned C_α residues, and 23% identity (PDB ID 3HGM[21]). UspA from *H. influenzae* also shares a high level of structural similarity with *Z* score 15.7, r.m.s.d. distance 3.5 Å over 128 aligned C_α positions, and 20% sequence identity (PDB ID 1JMV[7]), despite its lack of ATP-binding capability and altered ATP-binding loop sequence [Fig. 1(C)]. Interestingly, even though these USP subfamily members share a high degree of structural similarities, the region corresponding to α2 (residues 70-78) of the protein At3g01520 structure shows distinctive conformational diversity [Fig. 2(B)]. It is plausible to speculate that the region is involved in diverse interactions with other cellular components in response to stress signals in various pathways.

In conclusion, we report the first crystal structure of a eukaryotic USP family protein. Its structural similarity with other ATP-binding USP subfamily proteins, as well as the bound AMP and conservation of residues in the ATP-binding loop, suggest that the protein At3g01520 also belongs to the ATP-binding USP subfamily members. The biological function of the protein At3g01520 still remains elusive. However, it has been previously suggested that ATP-binding USP proteins might function as a molecular switch much like the Ras protein family, whose GTP hydrolysis ability is modulated by interaction with a number of regulatory proteins.[22] It is therefore plausible that At3g01520 possibly functions as a molecular switch by its putative ATP-binding and ATP-hydrolyzing properties. Further structure-guided biochemical studies to investigate how potential ATP-hydrolyzing activity of the protein At3g01520 is related to physiological phenotypes will shed light on the role of this uncharacterized USP protein in a wide range of stress-related signaling pathways in *A. thaliana*.

ACKNOWLEDGMENTS

Special thanks go to all the members of CESG team. Data for this study were collected at the 22-ID beamline at the Advanced Photon Source (APS) of Argonne National Laboratory

REFERENCES

1. Kvint K, Nachin L, Diez A, Nystrom T. The bacterial universal stress protein: function and regulation. Curr Opin Microbiol 2003;6:140–145.
2. Nachin L, Nannmark U, Nystrom T. Differential roles of the universal stress proteins of Escherichia coli in oxidative stress resistance, adhesion, and motility. J Bacteriol 2005;187:6265–6272.
3. Nystrom T, Neidhardt FC. Expression and role of the universal stress protein, UspA, of escherichia coli during growth arrest. Mol Microbiol 1994;11:537–544.
4. Nystrom T, Neidhardt FC. Isolation and properties of a mutant of Escherichia coli with an insertional inactivation of the uspA gene, which encodes a universal stress protein. J Bacteriol 1993;175:3949–3956.
5. Nystrom T, Neidhardt FC. Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in escherichia coli. Mol Microbiol 1992;6:3187–3198.
6. Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, Kim SH. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. Proc Natl Acad Sci USA 1998;95:15189–15193.
7. Sousa MC, McKay DB. Structure of the universal stress protein of haemophilus influenzae. Structure 2001;9:1135–1141.

8. Isokpehi RD, Simmons SS, Cohly HH, Ekunwe SI, Begonia GB, Ayensu WK. Identification of drought-responsive universal stress proteins in viridiplantae. Bioinform Biol Insights 2011;5:41–58.
9. Kerk D, Bulgrien J, Smith DW, Gribskov M. Arabidopsis proteins containing similarity to the universal stress protein domain of bacteria. Plant Physiol 2003;131:1209–1219.
10. Verslues PE, Lasky JR, Juenger TE, Liu TW, Kumar MN. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in arabidopsis. Plant Physiol 2014;164:144–159.
11. Jeon WB, Aceti DJ, Bingman CA, Vojtik FC, Olson AC, Ellefson JM, McCombs JE, Sreenath HK, Blommel PG, Seder KD, Burns BT, Geetha HV, Harms AC, Sabat G, Sussman MR, Fox BG, Phillips GN, Jr. High-throughput purification and quality assurance of arabidopsis thaliana proteins for eukaryotic structural genomics. J Struct Funct Genomics 2005;6:143–147.
12. Sreenath HK, Bingman CA, Buchan BW, Seder KD, Burns BT, Geetha HV, Jeon WB, Vojtik FC, Aceti DJ, Frederick RO, Phillips GN Jr., Fox BG. Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. Protein Expr Purif 2005;40:256–267.
13. Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Riters M, Newman CS, Fox BG, Wrobel RL. Results from high-throughput DNA cloning of arabidopsis thaliana target genes using site-specific recombination. J Struct Funct Genomics 2004;5:267–276.
14. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN, Jr., Rayment I, Ulrich EL, Volkman BF, Markley JL. Project management system for structural and functional proteomics: sesame. J Struct Funct Genomics 2003;4:11–23.
15. Uson I, Sheldrick GM. Advances in direct methods for protein crystallography. Curr Opin Struct Biol 1999;9:643–648.
16. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr 2000;56:965–972.
17. McRee DE. XtalView/xfit–a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125:156–165.
18. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255.
19. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. J Mol Biol 1993;231:1049–1067.
20. Lovell SC, Davis IW, Arendall WB, 3rd de Bakker PI, Word JM, Prisant MG, Richardson JS Richardson DC. Structure validation by calpha geometry: phi,psi and cbeta deviation. Proteins 2003; 50: 437–450.
21. Schweikhard ES, Kuhlmann SI, Kunte HJ, Grammann K, Ziegler CM. Structure and function of the universal stress protein TeaD and its role in regulating the ectoine transporter TeaABC of halomonas elongata DSM 2581(T). Biochemistry 2010;49:2194–2204.
22. Drumm JE, Mi K, Bilder P, Sun M, Lim J, Bielefeldt-Ohmann H, Basaraba R, So M, Zhu G, Tufariello JM, Izzo AA, Orme IM, Almo SC, Leyh TS, Chan J. Mycobacterium tuberculosis universal stress protein rv2623 regulates bacillary growth by ATP-binding: requirement for establishing chronic persistent infection. PLoS Pathog 2009;5:e1000460
23. Tkaczuk KL, I AS, Chruszcz M, Evdokimova E Savchenko A, Minor W. Structural and functional insight into the universal stress protein family. Evol Appl 2013;6:434–449.
24. Iino H, Shimizu N, Goto M, Ebihara A, Fukui K, Hirotsu K, Kuramitsu S. Crystal structure of the tandem-type universal stress protein ttha0350 from thermus thermophilus hb8. J Biochem 2011; 150:295–302.
25. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.