Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

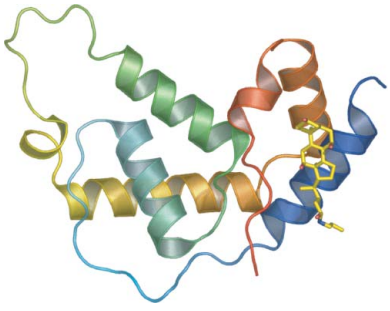
Structure at 1.6 Å resolution of the protein from gene locus At3g22680 from *Arabidopsis thaliana*

Simon T. M. Allard, Craig A. Bingman, Kenneth A. Johnson, Gary E. Wesenberg, Eduard Bitto, Won Bae Jeon and George N. Phillips Jr

Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, USA

Received 20 May 2005
Accepted 22 June 2005
Online 30 June 2005

PDB Reference: At3g22680 gene product, 1vk5, r1vk5sf.



The gene product of At3g22680 from *Arabidopsis thaliana* codes for a protein of unknown function. The crystal structure of the At3g22680 gene product was determined by multiple-wavelength anomalous diffraction and refined to an R factor of 16.0% ($R_{\rm free}$ = 18.4%) at 1.60 Å resolution. The refined structure shows one monomer in the asymmetric unit, with one molecule of the non-denaturing detergent CHAPS {3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate} tightly bound. Protein At3g22680 shows no structural homology to any other known proteins and represents a new fold in protein conformation space.

1. Introduction

At3g22680 was selected for structure determination by our group at the Center for Eukaryotic Structural Genomics (CESG) because it was predicted to have little or no structural homology to any protein in the PDB. The At3g22680 gene of *Arabidopsis thaliana* encodes a protein with a molecular weight of 18.0 kDa (residues 1–157) and a calculated isoelectric point of 5.11. In this paper, we report the three-dimensional structure of the At3g22680 protein at 1.60 Å determined by multiple-wavelength anomalous diffraction (MAD; Hendrickson, 1991). The final refined At3g22680 structure shows one molecule of CHAPS tightly bound to one monomer of At3g22680 in the asymmetric unit. The structure was determined under the National Institutes of Health NIGMS Protein Structure Initiative.

2. Materials and methods

The gene encoding the At3g22680 protein was cloned and a selenomethionine-labeled protein was purified following the standard CESG pipeline protocol for cloning (Thao *et al.*, 2004), protein expression (Sreenath *et al.*, 2005), protein purification (Jeon *et al.*, 2005) and overall information management (Zolnai *et al.*, 2003). Briefly, the target was cloned from mRNA from the *A. thaliana* ecotype Columbia T87 cell line. The gene sequence was verified by DNA sequencing and then cloned into a pQE80-derived expression vector, yielding an N-terminal His_6-tag fusion protein. The expression vector was then transformed into Rosetta cells (Novagen, Madison, WI, USA) and the cell culture grown in polyethylene terephthalate beverage bottles using Terrific Broth medium. Selenomethionyl protein was expressed in a B834(DE3) pLacI+RARE expression host using PASM-5052 autoinducing medium (Studier, 2005) for 25 h at 298 K in 500 ml batches in 2 l bottles. Both native and selenomethionyl protein were purified by IMAC chromatography and concentrated to 10 mg ml^{-1} before being stored in a buffer containing 100 m*M* NaCl, 0.3 m*M* tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and 10 m*M* HEPES pH 7.0 at 193 K. The target sequence and extent of selenomethionine incorporation was confirmed by ESI and MALDI mass spectrometry. The oligomeric state of the At3g22680 gene product was analyzed by analytical gel-filtration chromatography. For this purpose, a Superdex 200 10/300 column was used mounted on an AKTA FPLC chromatographic system.

Initial crystallization trials were conducted using native protein and the sitting-drop method, with 0.75 µl protein being mixed with an

Table 1
Summary of crystal parameters and data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native	Se peak	Se edge	Se remote
Space group	$P3_121$			
Unit-cell parameters (Å, °)	a = 83.45, b = 83.45, c = 60.58, α = 90.00, β = 90.00, γ = 120.00			
Data-collection and phasing statistics				
Energy (keV)	12.664	12.662	12.660	12.862
Wavelength (Å)	0.97900	0.97916	0.97931	0.96393
Resolution range (Å)	36.13–1.60 (1.66–1.60)	34.48–1.69 (1.75–1.69)	34.52–1.69 (1.75–1.69)	31.18–1.69 (1.75–1.69)
No. of reflections (measured/unique)	350451/32231	341949/27190	341994/27160	342434/27411
Completeness (%)	99.2 (92.2)	97.5 (96.4)	97.2 (94.1)	98.0 (98.1)
$R_{\rm merge}$†	0.037 (0.302)	0.098 (0.295)	0.068 (0.313)	0.072 (0.305)
Redundancy	10.9 (10.4)	12.6 (12.5)	12.6 (12.4)	12.5 (12.1)
Mean $I/\sigma(I)$	38.61 (9.30)	20.55 (5.50)	19.32 (4.60)	19.70 (4.73)
Mean FOM of phasing		0.72		

† $R_{\rm merge} = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection.

Table 2
Summary of refinement and model statistics .

Resolution range (Å)	34.30–1.60
Data set used in refinement	Native
No. of reflections (total/test)	30570/1633
$R_{\rm cryst}$†	0.160
$R_{\rm free}$‡	0.184
R.m.s.d. bonds (Å)	0.019
R.m.s.d. angles (°)	1.646
Average B factor (Å^2)	23.62
No. of CHAPS molecules	1
No. of ethylene glycol molecules	4
No. of sulfate ions	3
No. of water molecules	172
Ramachandran plot, residues in	
Most favorable region (%)	97.2
Additional allowed region (%)	2.8
Generously allowed region (%)	0.0
Disallowed region (%)	0.0

† $R_{\rm cryst} = \sum_h ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_h |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. ‡ $R_{\rm merge}$ was calculated as $R_{\rm cryst}$ using 5.1% of the randomly selected unique reflections that were omitted from structure refinement.

equal volume of mother liquor. Two conditions produced microcrystalline material: (i) 0.8 *M* Li_2SO_4 and 100 m*M* HEPPS pH 8.5 at 293 K and (ii) 2% PEG dimethyl ether 500, 1.0 *M* $(NH_4)_2SO_4$ and 100 m*M* Bis-Tris propane pH 8.5 at 277 K. In an attempt to produce better crystals, a stochastic screen was used (Segelke, 2001) and the protein concentration was reduced to 5 mg ml^{-1}. Well diffracting crystals of both the native and selenomethionine form of At3g22680 were obtained by the hanging-drop method with 2 µl protein being mixed with an equal volume of mother-liquor solution containing 0.73 *M* $(NH_4)_2SO_4$, 40 m*M* $MgSO_4$, 0.41%(*w*/*v*) of the non-denaturing zwitterionic detergent CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate} and 100 m*M* HEPES pH 8.5 at 293 K. Diamond-shaped crystals (100 × 80 × 50 µm) typically appeared in 24 h. Crystals of both the native and selenomethionyl proteins belong to space group $P3_121$, with native crystal unit-cell parameters $a = b = 83.5$, $c = 60.6$ Å and selenomethionyl crystal unit-cell parameters $a = b = 83.6$, $c = 61.1$ Å. Crystals were cryoprotected by soaking in well solution supplemented with increasing concentrations of ethylene glycol up to a final concentration of 35%(*v*/*v*). X-ray diffraction data for native and selenomethionyl crystals were collected at BioCARS 14-BMD and sector 32-IDB beamlines at the Argonne National Laboratory Advanced Photon Source, respectively. All X-ray diffraction data were integrated and scaled using the *HKL*2000 package (Otwinowski & Minor, 1997). The asymmetric unit contains one monomer of At3g22680, which contains five SeMet residues. The program *SOLVE* (Terwilliger & Berendzen, 1999) was used to locate the selenium sites and calculate phases, with all five selenium sites being included for phase calculations. Electron-density modification was subsequently performed with the program *RESOLVE* (Terwilliger, 2000). The automatic tracing procedure in *ARP*/*wARP* (Perrakis *et al.*, 1999) was used to produce a preliminary model which had 96% of all final model residues and side chains placed. The structure was completed using alternate cycles of manual building in *Xfit* (McRee, 1999) and refinement in *REFMAC* (Murshudov *et al.*, 1997). Following two initial rounds of refinement using peak data to 1.69 Å, the model was refined using native data to 1.60 Å. Electron density for the steroid core and the first eight C atoms of the tail of a single CHAPS molecule was clearly visible in the first experimentally phased electron-density map and was included in subsequent refinement cycles. The initial CHAPS model and parameter files used for structure refinement were obtained from the HIC-up database (Kleywegt & Jones, 1998). All refinement steps were monitored using an $R_{\rm free}$ value based on 5.1% of the independent reflections. The stereochemical quality of the final model was assessed using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Lovell *et al.*, 2003). The Ramachandran plot demonstrated that 97.2% of the residues lie in the most favored regions and 2.8% of the residues are in the additionally allowed regions.

3. Results and discussion

The structure of At3g22680 has been solved to 1.60 Å. Data collection, refinement and model statistics are summarized in Tables 1 and 2. The final model contains one monomer in the asymmetric unit, with 77% of all possible residues and side chains built in. No electron density was observed for residues 1–35 and the carboxy-terminal residue Lys157. Retrospective SDS–PAGE analysis of the freezer stocks of At3g22680 indicated that it had undergone partial proteolysis during desalting and final concentration. This may explain the lack of electron density for the amino-terminus of the protein in the crystal structure. The final model also contains one tightly bound CHAPS molecule, four ethylene glycol molecules, three sulfate ions and 172 waters. The At3g22680 monomer belongs to the all-α class of proteins and consists of two 3_{10}-helices and five α-helices ranging from nine to 17 amino acids in size (see Fig. 1). In an attempt to classify the fold of At3g22680, a structural homology search was conducted using both the *DALI* and *VAST* servers (Holm & Sander, 1993; Madej *et al.*, 1995). The *DALI* server identified four very weak potential structural homologs (with $Z > 2.0$). The top score was from an RNA polymerase σ factor from *Streptomyces coelicolor*, with Z = 2.3, an r.m.s.d. of 3.0 Å over 52 aligned C^α residues and 4%

sequence identity (PDB code 1h3l; Li *et al.*, 2002). The *VAST* server identified only one potential structural neighbor, a *Rhodococcus species* haloalkane dehalogenase with a *VAST* score of 4.8, an r.m.s.d. of 2.8 Å over only 30 aligned residues and 13.3% identity (PDB code 1cqw; Newman *et al.*, 1999). Analysis of the superposition of these structures onto At3g22680 revealed that these potential hits represent nothing more than circumstantial alignments of scattered α-helices. The fold of At3g22680 is therefore unique and has not been previously observed in other protein structures.

Sequence analysis of At3g22680 by the profile–profile fold-recognition tool *FFAS*03 (Jaroszewski *et al.*, 2000) confirmed that no previous PDB entry with a family relationship to At3g22680 existed in the RCSB Protein Data Bank at the time of manuscript submission. To identify any biochemically important structural features of At3g22680, we also analyzed patterns of evolutionary conservation within its primary sequence. Analysis by *PSI-BLAST* (Altschul *et al.*, 1990) revealed a rice protein (*Oryza sativano*; primary accession No. Q67IW5) of unknown function that showed 61% identity to At3g22680 over 131 residues. The same rice protein was also identified by a *UNIPROT-BLAST* search (Bairoch *et al.*, 2005), giving an *E* value of 2×10^{-42}; all other matches had an *E* value > 2.0 and were judged to be insignificant. The 61% sequence identity between the *A. thaliana* and *O. sativano* proteins included all the At3g22680 CHAPS-binding residues discussed below. The high level of sequence identity between these two proteins and the fact that they are found only in *A. thaliana* and *O. sativano* suggests that At3g22680 may represent a plant-specific fold.

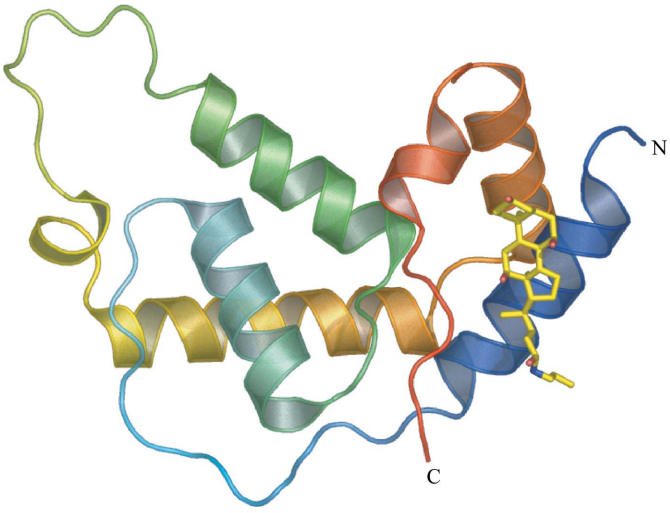


Figure 1
A ribbon diagram of the crystal structure of At3g22680 color coded from the amino-terminus (blue) to the carboxy-terminus (red). CHAPS is depicted in yellow stick format with the elements color coded as follows: oxygen, red; nitrogen, blue; carbon, yellow. The figure was generated using *PyMol* (DeLano, 2002).

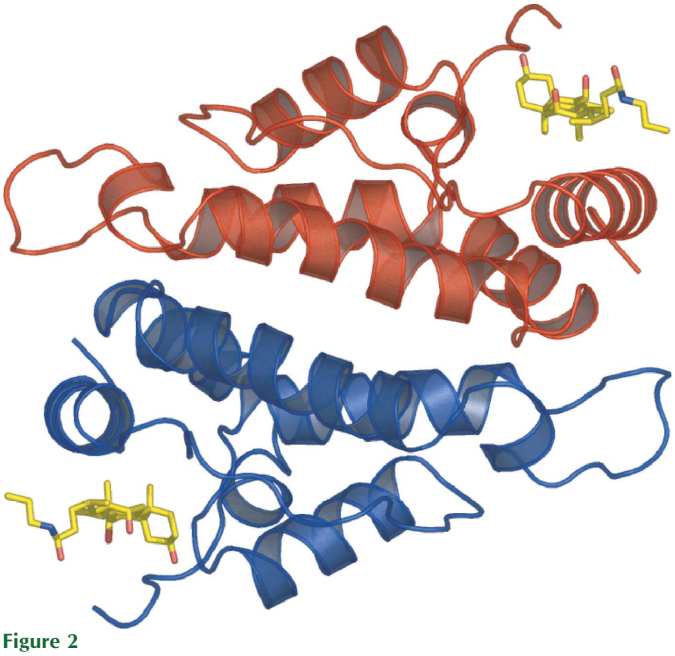


Figure 2
A ribbon representation of the At3g22680 homodimer. The two monomers are related by a crystallographic twofold axis and are shown in different colors (red and blue). The bound CHAPS is shown in the same format as in Fig. 1. The figure was generated using *PyMol* (DeLano, 2002).

The steroid core of the CHAPS molecule is bound to a hydrophobic pocket created by the amino-terminal and carboxy-terminal parts of the At3g22680 monomer. The hydrophobic side of the CHAPS steroid core interacts with the At3g22680 surface by hydrophobic contacts with residues Ala42, Met44, Tyr45 and Tyr48 of the amino-terminal α-helix (residues Leu39–Lys50) and carboxy-terminal residues Trp140, Tyr146, Ile150 and Ile153. The binding of the CHAPS molecule within this hydrophobic pocket suggests that these residues may interact with an unidentified natural ligand.

Analytical gel filtration indicates that the gene product of At3g22680 is likely to be a homodimer *in vitro*. Analysis of crystal packing shows that two At3g22680 monomers related by a crystallographic twofold axis interact *via* a four-helix bundle, utilizing helix residues Leu85–Ser96 and Ile111–Thr130 from each symmetry-related monomer (see Fig. 2). The protein–protein interface area involved in dimer formation is 1152 Å^2, with 71% of the surface area in the buried dimer surface interface being non-polar.

Additional examination of At3g22680 crystal packing reveals that CHAPS is involved in stabilizing crystal-packing contacts between dimers along the crystallographic 3_1 screw axis. Between each crystallographically related 3_1 dimer is a cavity containing one molecule of tightly bound CHAPS. The cavity is made up from the hydrophobic residues discussed above and residues Ile62, Pro63, Phe64, Thr65, Ile72, Ser73, Met74, Gln76 and Leu77 from the monomer related by the crystallographic 3_1 screw axis. The presence of CHAPS at these crystal-packing interfaces is likely to contribute to the good diffraction quality of the crystals (Gall *et al.*, 2003; Itou *et al.*, 2002; Klaholz & Moras, 2000).

In conclusion, this work shows that At3g22680 represents a new fold in protein conformational space that is possibly unique to plants.

We acknowledge financial support from NIH National Institute for General Medical Sciences grant P50 GM64598. Use of the Advanced Photon Source and the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research under Contract No. W-31-109-ENG-38. Use of BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources under grant No. RR07707. We acknowledge LS-CAT for time at APS Sector 32 and Joe Brunzelle for facilitating data collection there. Special thanks goes to Ruth M. Saecker for helpful discussions and all the members of the CESG team, including Todd Kimball, John Kunert, Nicholas Dillon, Rachel Schiesher, Juhyung Chin, Megan Riters, Andrew C. Olson, Jason M. Ellefson, Janet E. McCombs, Brendan T. Burns, Blake W. Buchan, Holalkere V. Geetha, Zhaohui Sun, Ip Kei Sam, Eldon L. Ulrich,

Nathan S. Rosenberg, Janelle Warrick, Bryan Ramirez, Zsolt Zolnai, Peter T. Lee, Jianhua Zhang, David J. Aceti, Russell L.Wrobel, Ronnie O. Frederick, Hassan Sreenath, Frank C. Vojtik, Craig S. Newman, John Primm, Michael R. Sussman, Brian G. Fox and John L. Markley.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L. S. (2005). *Nucleic Acids Res.* **33**, D154–D159.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Gall, A. L., Ruff, M. & Moras, D. (2003). *Acta Cryst.* D**59**, 603–606.

Hendrickson, W. A. (1991). *Science*, **254**, 51–58.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Itou, H., Yao, M., Fujita, I., Watanabe, N., Suzuki, M., Nishihira, J. & Tanaka, I. (2002). *J. Mol. Biol.* **316**, 265–276.

Jaroszewski, L., Rychlewski, L. & Godzik, A. (2000). *Protein Sci.* **9**, 1487–1496.

Jeon, W., Aceti, D. J., Bingman, C., Vojtik, F., Olson, A., Ellefson, J., McCombs, J., Sreenath, H., Blommel, P., Seder, K., Buchan, B., Burns, B., Geetha, H., Harms, A., Sabat, G., Sussman, M., Fox, B. & Phillips, G. (2005). In the press.

Klaholz, B. P. & Moras, D. (2000). *Acta Cryst.* D**56**, 933–935.

Kleywegt, G. J. & Jones, T. A. (1998). *Acta Cryst.* D**54**, 1119–1131.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Li, W., Stevenson, C. E., Burton, N., Jakimowicz, P., Paget, M. S., Buttner, M. J., Lawson, D. M. & Kleanthous, C. (2002). *J. Mol. Biol.* **323**, 225–236.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

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

Madej, T., Gibrat, J. F. & Bryant, S. H. (1995). *Proteins*, **23**, 356–369.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Newman, J., Peat, T. S., Richard, R., Kan, L., Swanson, P. E., Affholter, J. A., Holmes, I. H., Schindler, J. F., Unkefer, C. J. & Terwilliger, T. C. (1999). *Biochemistry*, **38**, 16105–16114.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Segelke, B. W. (2001). *J. Cryst. Growth*, **232**, 553–562.

Sreenath, H. K., Bingman, C. A., Buchan, B. W., Seder, K. D., Burns, B. T., Geetha, H. V., Jeon, W. B., Vojtik, F. C., Aceti, D. J., Frederick, R. O., Phillips, G. N. Jr & Fox, B. G. (2005). *Protein Expr. Purif.* **40**, 256–267.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Thao, S., Zhao, Q., Kimball, T., Steffen, E., Blommel, P. G., Riters, M., Newman, C., Fox, B. & Wrobel, R. (2004). *J. Struct. Funct. Genomics*, **5**, 255–265.

Zolnai, Z., Lee, P. T., Li, J., Chapman, M. R., Newman, C. S., Phillips, G. N. Jr, Rayment, I., Ulrich, E. L., Volkman, B. F. & Markley, J. L. (2003). *J. Struct. Funct. Genomics*, **4**, 11–23.