PROTEINS: Structure, Function, and Bioinformatics 57:218–220 (2004)

STRUCTURE NOTE

Crystal Structure of the Protein from Gene At3g17210 of *Arabidopsis thaliana*

Craig A. Bingman,[1,3] Kenneth A. Johnson,[1,3] Francis C. Peterson,[2,3] Ronnie O. Frederick,[1,3] Qin Zhao,[1,3] Sandy Thao,[1,3] Brian G. Fox,[1,3] Brian F. Volkman,[2,3] Won Bae Jeon,[1,3] David W. Smith,[1,3] Craig S. Newman,[1,3] Eldon L. Ulrich,[1,3] Adrian Hegeman,[1,3] Michael R. Sussman,[1,3] John L. Markley,[1,3] and George N. Phillips, Jr.[1,3]*

[1]*Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin*
[2]*Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin*
[3]*Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin*

Introduction. The Center for Eukaryotic Structural Genomics (CESG) is dedicated to determining the structures of novel proteins from *Arabidopsis thaliana.* Proteins, or at least apparent open reading frames, have been prioritized for structure solution by nuclear magnetic resonance (NMR) and/or X-ray crystallography. The protein from gene locus At3g17210 was in the second-highest tier of the project list, indicating no significant demerits: no known homologous structure, not targeted by another structural genomics group, no predicted transmembrane segments, no identifiable signal sequence, relatively low Cys content, no proline-rich segments. The target also had several desirable predicted attributes (absence of low-complexity sequences, and high predicted solubility, and prediction to belong to an unknown structural class). The laboratory information management system used to track progress and data from initial selection to coordinate deposition is the Sesame system.[1]

Materials and Methods. The target was cloned from mRNA from the T87 *Arabidopsis thaliana* cell line.[2] The N-terminal tag sequence was GSSHHHHHHSS-GLVPRGSH, with GSH remaining after thrombin cleavage. Selenomethionyl protein was expressed in a B834(DE3) pLacI+RARE expression host, using Studier's PASM-5052 autoinducing medium (Willam Studier, personal communication) for 25 h at 25°C in 500-mL batches in 2-L pop bottles.[3] Ten and nine-tenths (10.9) grams of cells from one liter of culture were lysed by sonication, and the hexahistidine fusion construct was captured on a 5-mL HiTrap Chelating HP column charged with Ni^{2+} (Amersham). The recombinant protein (254 mg) was eluted with a linear 0-500 mM imidazole gradient, desalted into 100 mM NaCl, 10 mM Tris:HCl (pH 8.0), 1 mM Ca^{2+} and cleaved with thrombin (Novagen) overnight at 4°C. The cleaved target was subjected to subtractive IMAC chromatography, and the purified cleaved target was desalted into 100 mM NaCl, 10 mM Na HEPES (pH 7.0), and concentrated to 14 mg/mL. The target identity and extent of selenomethionine incorporation (> 98%) was confirmed by ESI and MALDI mass spectrometry.

Crystallization conditions for the selenomethionyl protein were very similar to those for native protein. Optimum crystals formed from hanging-drop vapor-diffusion droplets with 2 μL protein + 2 μL reservoir solution against a reservoir of 30%(w/v) MEPEG 5 K, 120 mM Na citrate, 100 mM Na HEPES (pH 6.0). Crystals began to form within three days and reached full size by two weeks. As the crystal growth solution was found to be an effective cryosolvent, crystals were mounted in loops and cooled by direct immersion in liquid nitrogen. The space group was $P6_2$, a = b = 55.468, c =57.673. Multi-wavelength anomalous diffraction data around the Se(K) edge were collected at APS/BioCARS/14IDB on a MARCCD detector, using inverse-beam geometry and reduced with HKL2000[4] and CCP4 programs.[5] Optimum wavelengths for the experiment were determined by a fluorescence scan on the crystal (Table I).

Solution of the Se position and phase improvement were performed with SOLVE/RESOLVE.[6,7] A single, ordered Se atom was found, and the final figure of merit was 0.57. RESOLVE auto-traced 97 of 112 residues in the recombinant target.[8] The structure was improved by iterative rounds of refinement (REFMAC5[9] or CNS[10]) against the low-energy remote data set with five percent of the data withheld from refinement, and model-building using XFIT.[11] A tightly bound metal ion was identified in the

Coordinates for the crystal structure have been deposited in the PDB (accession 1Q4R).

Grant sponsor: NIH National Institute for General Medical Sciences; Grant number: P50 GM64598.

*Correspondence to: George N. Phillips, Jr., Center for Eukaryotic Structural Genomics Department of Biochemistry, University of Wisconsin–Madison, Madison, WI 53706
E-mail: phillips@biochem.wisc.edu.

Received 27 October 2003; Accepted 3 November 2003

Published online 8 July 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20215

TABLE I. Data and Refinement Statistics

Data collection	Peak	Edge	HER	LER
Energy(eV)	12662	12660	12860	12460
Resolution range (Å)		all (20–1.90 Å)		
Reflections				
Measured	75176	74895	70420	78202
Unique	14162	14546	14116	14924
Completeness				
Overall	93.5	92.8	90.4	95.3
Outer shell (1.97–1.90 Å)	52.6	49.0	38.3	63.7
R(merge)				
Overall	4.2	4.2	4.4	4.0
Outer shell	27.4	24.6	24.2	26.8
Redundancy (overall)	5.2	5.2	5.1	5.3
Mean I/sig(I)	23.8	23.9	21.8	25.5
Sigma cutoff	0	0	0	0
Phasing				
Mean FOM after SOLVE		0.54		
Refinement				
F-factor/free R-factor	18.5/23.2			
RMDS bonds (Å)	0.020			
RMDS angles (deg)	1.861			
Average B factor	21.2			
Number of water molecules	109			
Number of metal ions	1			
Ramachandran plot				
Residues in most favorible region (%)			95.7	
Residues in additional allowed region (%)			4.3	
Residues in generously allowed region (%)			0.0	

final rounds of refinement. The final model comprises 103 contiguous ordered amino acid residues, beginning at residue 10 of the recombinant construct and residue 7 of the native protein and continuing to the C-terminus of the protein. The model contains 109 ordered waters, including two associated with the bound metal, which, on the basis of its B-value, occupancy in refinement, is surmised to be a fully occupied Mg^{2+} ion. The final R-overall and R-free were 0.185 and 0.232 for all 7209 observed reflections (94.8% complete) between 20 and 1.9 Å. The R-free value in the 1.95–1.90 Å shell was 0.267. The RMS deviations of bond lengths and angles from ideal values were 0.020 Å and 1.86° (Table I).

Results and Discussion. The crystal structure of At3g17210 protein is best described as a dimer on a two-fold crystallographic axis. The monomers have β-α-β-β-α-α-β secondary structure elements, with antiparallel beta strands, with 2,1,3,4 sheet topology. Strand 2 of one monomer hydrogen-bonds with strand 4 of the other monomer to form an elongated eight-stranded beta barrel. The N- and C-termini emerge from the same face of the dimer, which also contains the magnesium binding site. The fold of the protein and the oligomeric state compare favorably with the structure of this protein as determined also by NMR spectroscopy by CESG.[12]

BLAST sequence searches[13] revealed a high degree of sequence similarity ($P(N) < e^{-10}$) between At3g17210 and the following SWISSPROT entries: Q8LQD2 *Oryza sativa*, Q41049 *Populus balsamifera* pop3 peptide, Q9AR79 *Populus tremula* Boiling stable protein A, Q8L408 *Oryza sativa*, Q9FK81 (At5g22580) *A. thaliana*, Q42482 *Populus* sp. wound responsive mRNA, and O31026 *V. cholera*. Because of the protein's homology to pop3 and the other proteins above, the At3g17210 protein may belong to a class of plant stress response proteins. Although it has an apparent structural similarity to the monooxygenase from the gene actva-orf6 of *Streptomyces coelicolor* (PDB entry 1LQ9; [14]), a structure reported after CESG selected this target, the At3g17210 protein must have a different function, because it does not contain the active-site residues of the monooxygenase.

Acknowledgments. We acknowledge support from the BioCARS beamline at APS/Argonne National Laboratory and members of the CESG team: D. Aceti, P. Blommel, B. Buchan, J. Cao, C. Corenilescu, J. Doreleijers, D. Dyer, H. Geetha, D. Hruby, T. Kimball, B. Ramirez, N. Rosenberg, M. Runnels, K. Seder, J. Shaw, H. Sreenath, J. Song, E. Tyler, D. Vinarov, F. Vojtik, G. Wesenberg, M. Westler, R. Wrobel, J. Zhang, and Z. Zolnai.

REFERENCES

1. 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.
2. Aceti DJ, Blommel PG, Endo Y, Fox BG, Frederick RO, Hegeman AD, Jeon WB, Kimball TL, Lee JM, Newman CS, Peterson FC, Sawasaki T, Seder KD, Sussman MR, Ulrich EL, Wrobel RL, Thao S, Vinarov DA, Volkman BF, Zhao Q. Role of nucleic acid and protein manipulation technologies in high-throughput structural biology efforts. In: Steinbüchel A, editor. Biopolymers. Weinheim: Wiley-VCH; 2003. p 469–496.
3. Sanville Millard C, Stols L, Quartey P, Kim Y, Dementieva I, Donnelly MI. A less laborious approach to the high-throughput production of recombinant proteins in Escherichia coli using 2-liter plastic bottles. Protein Expr Purif 2003;29:311–320.
4. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter Jr., CW, Sweet RM,

editors. Methods in enzymology, Vol. 276, Macromolecular crystallography, Part A. San Diego: Academic Press; 1997. p 307–326.
5. The CCP4 Suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763
6. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55:849–861.
7. Terwilliger TC. Maximum likelihood density modification. Acta Crystallogr D Biol Crystallogr 2000;56:965–972.
8. Terwilliger TC. Automated main-chain model-building by template-matching and iterative fragment extension. Acta Crystallogr D Biol Crystallogr 2002;59:34–44.
9. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255.
10. Brunger AT, et al. Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
11. McRee DE. Xta/View/xfit—A versatile program for manipulating atomic coordinates and electron density. J Struc Biol 1999;125: 156–165.
12. Lytle BL, Peterson FC, Kjer KL, Frederick RO, Zhao Q, Thao S, Bingman CA, Phillips Jr. GN, Volkman BF. Structure of the hypothetical protein At3g17210 from *Arabidopsis thaliana*. J Biomol NMR 2003. Forthcoming.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
14. Sciara G, Kendrew SG, Miele AE, Marsh NG, Federici L, Malatesta F, Schimperna G, Savino C, Vallone B. The structure of Actra-Orfb, a novel type of monooxygenase involved in actinorhodin biosynthesis. EMBO J 2003;22:205–215.