STRUCTURE NOTE

Crystal Structure of Gene Locus At3g16990 from *Arabidopsis thaliana*

Paul G. Blommel,[1] **David W. Smith,**[1] **Craig A. Bingman,**[1] **David H. Dyer,**[2] **Ivan Rayment,**[2] **Hazel M. Holden,**[2] **Brian G. Fox,**[1] **and George N. Phillips Jr.**[1]*
[1]*Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin*
[2]*Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin*

Introduction. The Center for Eukaryotic Structural Genomics is dedicated to determining the structures of novel proteins from eukaryotic organisms. Open reading frames are scored using thirteen different categories (i.e. new fold prediction, solubility prediction, small percentage of low complexity sequence, etc.) and then ranked to indicate their suitability for study by nuclear magnetic resonance (NMR) or X-ray crystallography. Gene locus At3g16990 from *Arabidopsis thaliana* was given a suitable score for study, with the only major demerits being a large cysteine residue count, a moderate new fold prediction and a predicted low expression based on gene chip results. Here, we report the crystal structure of the protein from *Arabidopsis thaliana* gene locus At3g16990 as determined by single wavelength anomalous dispersion (SAD) phasing.

Materials and Methods. The protein was synthesized as a Se-Met derivative, as previously described.[1] The Sesame laboratory information management software package[2] was used to collect data from cloning, cell growth and protein purification procedures.

The Se-met labeled protein was crystallized from a protein solution (15 mg/mL protein in 25 mM NaCl and 10 mM Tris pH 7.5) mixed in a 1:1 ratio with the well solution (55 mM sodium acetate pH 4.5, 1.05M ammonium sulfate) in a hanging drop crystallization trial using VDX® trays. The crystals were found to belong to the space group $P4_12_12$ with unit cell constants of $a = b = 62.700$, $c = 287.621$. Phasing was accomplished with the SAD method using data collected at the Se peak with the SOLVE[3]/RESOLVE[4] software package. The phasing effort located 7 (of 8 possible) selenium atom sites in the asymmetric unit, which were input into the RESOLVE program to create an initial traceable map. Further modeling and refinement were completed using TNT[5] and Turbo,[6] while final maps and models were built and refined using consecutive iterations of refinement with Refmac5[7] and model building with Xfit.[8] The final structure consisted of seven alpha helices per molecule and two molecules per asymmetric unit with the two molecules being separate entities. The final refinement model consisted of residues 5–219 (of 221) of chain A and residues 4–219 of chain B, as the first four to five and the final two residues

were too disordered to fit. The final model also contained 253 water molecules and four sulfate groups per asymmetric unit. Table I lists the data and refinement statistics. Coordinates and structure factors were deposited with the Protein Data Bank (entry 1Q4M).

Results and Discussion. Upon final structure determination, the coordinates for At3g16990 were sent to the DALI server[9] to search for proteins with similar three-dimensional structures. Of the solutions returned, four had *Z*-scores above 10. These four structures correspond to heme oxygenase fragment (PDB 1N45, 8% IDE, 3.3 Å RMSD), ribonuclease reductase (PDB 1XSM, 9% IDE, 3.1 Å RMSD), methane monooxygenase hydroxylase (PDB 1MHY-D, 5% IDE, 2.8 Å RMSD), and heme oxygenase (PDB 1J77-A, 9% IDE, 3.4 Å RMSD). Each of these other proteins contains at least one iron atom (in either a heme cofactor or in some other iron binding site). In contrast, At3g16990 shows no evidence of tightly bound iron atoms, as it exhibits no distinct ligand-to-metal charge transfer spectrum typical of iron-sulfur or iron-tyrosinate proteins and no suitable peaks of unidentified intense electron density. Also, in comparing sequences, At3g16690 does contain a heme binding histidine similar to those in 1J77 and 1N45, but unlike both of these structures, the pocket where a heme would bind is filled with a residue side chain (valine in At3g16990 vs. glycine in both 1J77 and 1N45) thereby sterically hindering the potential binding of a heme group. Furthermore, At3g16990 has a pocket (discussed in more detail below along with mystery density) in a region similar to the iron binding pocket of 1XSM, but does not have enough iron binding residues pointing inside the pocket to fully capture an iron ion.

TABLE I. Summary of Data Collection, Crystal Structure and Refinement Parameters

Data Set	Energy (eV)	# Refl	Compl	Redun	R_{sym}
Peak	12.663	45956	83.7	5.2	0.043
Space group			$P4_12_12$		
Unit cell			$a = 62.70$, $b = 62.70$, $c = 287.62$		
Resolution range (Å)			24.92–2.08 (2.14–2.08)		
Completeness			85.4 (48.2)		
Redundancy			5.2 (3.4)		
I/σ			26.6 (5.6)		
R_{sym}			0.048 (0.214)		
R_{cryst}			22.1 (25.2)		
R_{free}			27.8 (30.8)		
Average B factor (Å^2)			32.9		
RMSD bond lengths (Å)			0.018		
RMSD bond angles (°)			1.529		

Numbers in parentheses indicate the highest resolution shell.

A VAST[10] search of the same model gave a high Z-score of only 3.9 for PDB: 1A32 with a RMSD of 2.6 Å and a sequence identity of 9.6% for 52 residues when compared to the 211 residues of At3g16990. Thus VAST did not return any structurally similar proteins when At3g16990 was queried against the non-redundant PDB database.

A BLAST search[11] on the sequence for At3g16990 produced only one other protein with a high sequence similarity. This protein, pm36 from Glycine max, was annotated to be involved in seed maturation; however, extensive supporting biochemical evidence is not available. Since At3g16990 is also a plant protein, it could also be involved in seed maturation process, but at present there is no evidence to support this conjecture.

The final solution of the structure of At3g16990 contains density that cannot be assigned to any amino acid or crystallization solution component. This density is buried within the protein and does not have any apparent access to solvent. There are three hydrogen bonding donors/acceptors pointing directly toward the density. Two are from the side chains of Asp 47 and Glu 210, while the third acceptor is the carbonyl oxygen from Val 39. The density is stacked between two aromatic side chains, Phe 50 and Tyr 143, suggesting that the molecule may be aromatic. Furthermore, molecular modeling indicates that the unassigned density has a size and shape consistent with a purine or a substituted indole. These molecules have important contributions in signal transduction pathways in plants.[12] Figure 1 shows the location of this density along with the surrounding amino acid side chains.
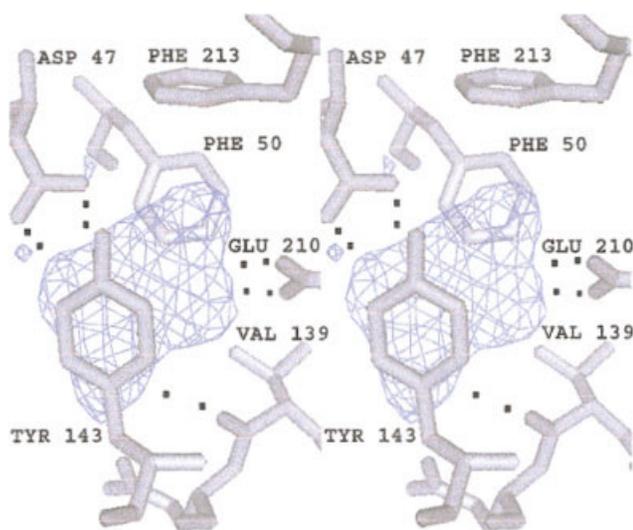
Fig. 1. Stereoview of the unassigned density (rendered at 3σ) within the At3g16990 structure. Potential hydrogen bonding contacts are shown with dotted lines. This image was created using PyMOL.[13]

REFERENCES

1. Smith DW, Johnson KA, Bingman CA, Aceti DJ, Blommel PG, Wrobel RL, Frederick RO, Zhao Q, Sreenath H, Fox BG, Volkman BF, Jeon WB, Newman CS, Ulrich EL, Hegeman AD, Kimball T, Thao S, Sussman MR, Markley JL, Phillips Jr. GN. Crystal structure of At2g03760, a putative steroid sulfotransferase from *Arabidopsis thaliana*. Proteins 2004; In Press.
2. Zolnai Z, Lee JL, Chapman CS, Phillips Jr GN, Rayment I, Ulrich EL, Volkman BF, Markley JL. Project management system for structural and functional proteomics: Sesame. J Struct Func Genomics 2003;4:11–23.
3. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Cryst 1999;D55:849–861.
4. Terwilliger TC. Maximum likelihood density modification. Acta Cryst 2000;D56:965–972.
5. Tronrud DE. The efficient calculation of the normal matrix in least-squares refinement of macromolecular structures. Acta Crystallogr A 1999;55:700–703.
6. Roussel A, Cambillau C. Turbo Frodo. Mountain View California: Silicon Graphics; 1991.
7. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst 1997;D35:240–255.
8. McRee DE. XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. J Struct Biol 1999;125:156–165.
9. Holm L, Sander C. Mapping the protein universe. Science 1997;273:595–602.
10. Gibrat J-F, Madej T, Bryant SH. Surprising similarities in structure comparison. Curr Opin in Struct Bio 1996;6:377–385.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;15:403–410.
12. Ulmasov T, Hagen G, Guilfoyle TJ. Activation and repression by auxin-response factors. Proc Natl Acad Sci 1999;96:5844–5849.
13. DeLano WL. The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific; 2002.