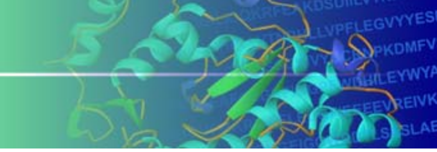


Novel structural results from the Center for Eukaryotic Structural Genomics (CESG)

G. Phillips, Jr., J. Markley, B. Fox, C. Newman, B. Volkman, M. Sussman, D. Aceti, S. Allard, M. Bagguley, C. Bingman, E. Bitto, P. Blommel, E. Brauer, B. Burns, J. Cao, C. Cornilescu, J. Doreleijers, Z. Eggers, J. Ellefson, R. Frederick, H. Geetha, B.W. Han, D. Hruby, W.B. Jeon, J. Kunert, M. Larson, M. Lee, P. Lee, J. Li,C. Loushin Newman, B. Lytle, J. McCombs, Z. Miller, R. Narayana, B. Nelson, A. Olson, F. Peterson, J. Primm, K. Raghunath, B. Ramirez, N. Ravoof, D. Repp, M. Riters, N. Rosenberg, M. Runnels, G. Sabat, I.K. Sam, S. Sarlati, K. Seder, M. Shahan, S. Singh, J. Song, H. Sreenath, E. Steffen, Z. Sun, T. Tiradani, D. Troestler, E. Tyler, R. Tyler, E. Ulrich, D. Vinarov, F. Vojtik, J. Waltner, L. Wang, J. Warrick, G. Wesenberg, M. Westler, R. Wrobel, J. Zhang, Q. Zhao, and Z. Zolnai.

University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, http://www.uwstructuralgenomics.org

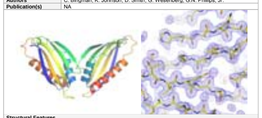
Abstract

The Center for Eukaryotic Structural Genomics (CESG) has deposited approximately 40 structures as of November, 2004. These structures have resulted from a total of 87 which have crystallized and 48 which have been determined to have disperse HSQC spectra, *i.e.,* folded. Most of the proteins showed little or no sequence homology to proteins in the Protein Data Bank at the time of selection or PDB deposition. Some of the structures have been chosen and presented below to illustrate the range of results obtained to date. The structures have come from a combination of X-ray crystallography and NMR, and from E. coli-based protein production and our cell-free based system. Some structures were requested by researchers outside of CESG, but most were selected by our internal target selection algorithm. For more information about the operation of the CESG pipeline please see other posters about bioinformatics, cloning and expression, protein production, NMR, and crystallomics/crystallography. Supported by NIH Grant P50 GM64598.

No sequence similarity; new fold

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	05358	
Organism	*Arabidopsis thaliana*	
Gene Designator	At1g24000.1	
PDB Entry	1VJH	Deposition: 20-Feb-2004
Function	Unknown	
Produced From	*E. coli* B834(DE3)	
Structure by X-ray	Resolution: 2.1 Å, Se-MAD	R-value (R-free): 18.6% (23.9%)
	No. of Residues: 122 aa, (13,758x2)	Subunits/ASU: 2
Data Collected At	Advanced Photon Source, BioCARS 14-BM-D	
Authors	C. Bingman, K. Johnson, D. Smith, G. Wesenberg, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

At1g24000 was annotated at the time of selection as a member of the Bet_v_1 allergin family, but at the time of selection and deposition, there were no significant sequence matches to anything in the PDB. The protein has a βαββαββα, with the five beta strands forming an antiparallel beta sheet. The structure was solved by a 3-wavelength SeMet MAD experiment. There are two copies of the protein in the asymmetric unit, probably indicating a dimer as the biological unit. The crystal structure reveals extra density that may be a low molecular weight polyethylene glycol, pointing to a binding site for low molecular weight hydrophobic ligands. At1g24000 has a high degree of sequence similarity to five other hypothetical proteins in *Arabidopsis*. A DALI search reveals only structural similarity to itself, so At1g24000 probably represents a new fold. In addition to the gene complex in *Arabidopsis*, At1g24000 also displays significant sequence similarity to genes from tobacco, peach, grape, and sugar beet.

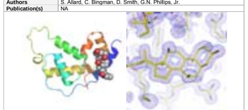
Percent Identity with Nearest PDB Structure at Time Solved	none better than E=1
Pfam Cluster	B_68483
Protonet Cluster Size : Structures in PDB	78 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

New protein fold

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	13974	
Organism	*Arabidopsis thaliana*	
Gene Designator	At3g22680.1	
PDB Entry	1VK5	Deposition: 06-May-2004
Function	Unknown	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 1.60	R-value (R-free): 16.1% (18.4%)
	No. of Residues: 157 (17,924)	Subunits/Molecule: 1
Data Collected At	Advanced Photon Source BioCARS 14-BM-D 14-Feb-2004, COM-CAT 32-IDB 27-Feb-2004	
Authors	S. Allard, C. Bingman, D. Smith, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

At3g22680 appears to represent a new protein fold. The structure was solved by Se-Met MAD. At the time of selection and deposition, there was no significant sequence similarity to structures in the PDB. DALI searches revealed no significant structural similarity to any known protein fold. The quality of crystals of this protein was dramatically improved by including the detergent CHAPS. The above left figure shows CHAPS bound to At3g22680.

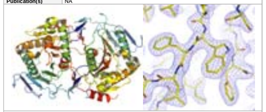
Percent Identity with Nearest PDB Structure at Time Solved	none better than E=1
Pfam Cluster	Pfam-B_96548
Protonet Cluster Size : Structures in PDB	108 : 1

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

New enzyme; new annotation

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	23169	
Organism	*Arabidopsis thaliana*	
Gene Designator	At5g18200.1	
PDB Entry	1VKV	Deposition: 24-Jun-2004
Function	ADP-glucose + phosphate -> glucose-1-P + ADP	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 2.3 Å	R-value (R-free): 22.0% (27.7%)
	No. of Residues: 351 (39,005)	Subunits/Molecule:
Data Collected At	APS: BioCARS 14-ID-B 11-Feb-2004, COM-CAT 32-ID 04-Apr-2004	
Authors	E. Bitto, S. Allard, D. Smith, G. Wesenberg, K. Johnson, C. Bingman, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

The gene locus (At5g18200.1) was chosen for fold-space investigation, since it had only limited amino acid sequence identity to any known protein. The closest relative whose structure was known was a bacterial galactose uridinylyltransferase (GalT). On the basis of this marginal sequence identity (~22%), the At5g18200.1 gene product was annotated in databases as having the same activity as the bacterial protein. The crystal structure of At5g18200.1 confirmed structural similarity to the bacterial GalT enzyme. However, collaboration with Prof. Perry Frey revealed that the enzyme did NOT have appreciable activity for this reaction. This observation underscores the need for actual experimental determination of biological function. Clearly the clues provided by homology in sequence or structure are valuable, but they are just that-clues. Further studies in the Frey laboratory revealed that the enzyme catalyzes a new reaction not previously appreciated in starch metabolism. This finding could have profound effects on our understanding of metabolic pools of ADP-glucose, the first committed step in starch degradation (Frey et. al., to be published), potentially improving nutritional or industrial production of starch in plants.

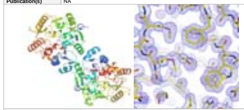
Percent Identity with Nearest PDB Structure at Time Solved	22% of 298 aa (1HXP)
Pfam Cluster	B-7903, B_18006
Protonet Cluster Size : Structures in PDB	89 : 1

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

External request; low homology

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	24674	
Organism	*Arabidopsis thaliana*	
Gene Designator	At5g08170.1	
PDB Entry	1VKP	Deposition: 15-Jun-2004
Function	Agmatine iminohydrolase (EC:3.5.3.12)	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 1.53 Å	R-value (R-free): 14.8% (17.4%)
	No. of Residues: 383 (43,053)	Subunits/Molecule: 2
Data Collected At	Advanced Photon Source COM-CAT 32-ID-B 4-Apr-2004	
Authors	D. Smith, S. Allard, E. Bitto, C. Bingman, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

This target was nominated by David Meinke of Oklahoma State University, and was the first externally nominated target solved by X-ray crystallography. It was selected based on an interesting phenotype in *Arabidopsis* seedlings, identified by genetic screens. Subsequent biochemical characterization by Piotrowski and coworkers established At5g08170 as agmatine iminohydrolase, the final uncharacterized step in polyamine biosynthesis in plants. At5g08170 represents the first structure of an agmatine deiminase, and the first representative of the PPAD family. It is a member of the same superfamily as aminotransferase, unifies the PPAD and known arginine deiminase family, will allow the modeling of a large number of proteins and will facilitate the elucidation of the mechanism of this enzyme. References: 1) Janowitz, T. Kneifel, H. and Piotrowski, M. Identification and characterization of plant agmatine iminohydrolase, the last missing link in polyamine biosynthesis of plants. (2003) *FEBS Let.* 544:258-61.

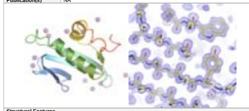
Percent Identity with Nearest PDB Structure at Time Solved	none higher than E=1
Pfam Cluster	PAD_porph
Protonet Cluster Size : Structures in PDB	261 : 1

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Solved by NaBr soak

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	06042	
Organism	*Arabidopsis thaliana*	
Gene Designator	At1g77540.1	
PDB Entry	1XMT	Deposition: 4-Oct-2004
Function	Putative acetyltransferase.	
Produced From	*E. coli* B834(DE3)	
Structure by X-ray	Resolution: 1.15 Å, Br-MAD	R-value (R-free): 14.9% (17.0%)
	No. of Residues: 103	Subunits/Molecule: 1
Data Collected At	Advanced Photon Source, COM-CAT 32-ID	
Authors	E. Bitto, C. Bingman, G. Wesenberg, S. Allard, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

The structure of At1g75540 was solved by a MAD experiment performed with bromide as the anomalous scatterer. This represents the first successful application of Br phasing at CESG. At1g75540 likely functions as acetyl transferase. Its primary structure shows distant homology to proteins with this activity, cataloged within Gcn5-related N-acetyltransferase superfamily. Often, this small domain occurs in multi-domain proteins where it provides substrate acetylation activity. The most similar homolog in the PDB shows 32% sequence identity of 71 aligned residues (1R57, NMR), several others range in 9-13% sequence identity (1BOB, 1NSL, 1KZF, 1YGH, 1GHE, 1QZY). Several of the fold-homologs were co-crystallized with an acetyl coenzyme A, the most common biochemical donor of acetyl groups. The binding site for this cofactor in At1g75540 can thus be inferred to reside in the groove delineated by central helix of the domain (green), several loop residues following the helix and smaller perpendicularly oriented helix (yellow). Some of the residues 43-51 are likely to be involved in the binding of acetyl CoA moiety. This protein aligns to two Pfam-B domains, Pfam-B_34288 over residues 2-29, and Pfam-B_2135 over residues 30-79. The small N-terminal domain seems to be plant-specific.

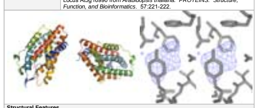
Percent Identity with Nearest PDB Structure at Time Solved	32% over 71 aa (1R57)
Pfam Cluster	B_34288, B_2135
Protonet Cluster Size : Structures in PDB	22 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Mystery density in structure?

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	12240	
Organism	*Arabidopsis thaliana*	
Gene Designator	At3g16990.1	
PDB Entry	1Q4M	Deposition: 4-Aug-2003
Function	Putative seed maturation protein-related	
Produced From	*E. coli* BL21 Rosetta	
Structure by X-ray	Resolution: 2.08 Å Se-SAD	R-value (R-free): 22.4% (27.8%)
	No. of Residues: 220 (24,939x2)	Subunits/ASU: 2
Data Collected At	Advanced Photon Source BioCARS 14-ID-B 26-June-2003	
Authors	G.N. Phillips, Jr., C.A. Bingman, K.A. Johnson, D.W. Smith	
Publication(s)	P.G. Blommel, D.W. Smith, C.A. Bingman, D.H. Dyer, I. Rayment, H. Holden, B.G. Fox, and G.N. Phillips, Jr (2004) Crystal Structure of Gene Locus At3g16990 from *Arabidopsis thaliana. PROTEINS: Structure, Function, and Bioinformatics.* 57:221-222.	



Structural Features

At3g16990 is assigned as a TENA_THI-4 domain protein. It has a sequence similarity of 22% over 213 aa against a subsequently solved structural genomics target from NESG, 1RTW. The protein was considered to be a fold-space candidate at the time of selection, and is one of two TEN_THI-4 domain proteins in the PDB. Electron density maps reveal unassigned density in each monomer (above.) It is thought that this density may represent a pyrimidine metabolite, or MP5, based on homology 1RTW.

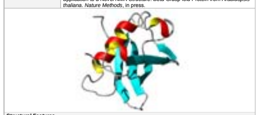
Percent Identity with Nearest PDB Structure at Time Solved	22% over 213 aa (1RTW)
Pfam Cluster	TENA_THI-4
Protonet Cluster Size : Structures in PDB	485 : 2

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

First cell-free structure solved by CESG

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	15176	
Organism	*Arabidopsis thaliana*	
Gene Designator	At3g01050.1	
PDB Entry	1SE9	Deposition: 16-Feb-2004
BMRB Entry	6128	Deposition: 3-Mar-2004
Function	Unknown	
Produced From	Cell-Free (wheat germ extract)	
Structure by NMR	Restraints/Residue: 10.7	Subunits/Molecule: 1
	No. of Residues: 126	Molecular Weight: 12.8 kDa
	Backbone RMSD(7–93): 1.14 Å	All Heavy Atoms RMSD(7–93): 1.60 Å
Data Collected At	Medical College of Wisconsin, Milwaukee, WI	
Authors	B.F. Volkman, B.L. Lytle, F.C. Peterson	
Publication(s)	D.A. Vinarov, B.L. Lytle, F.C. Peterson, E.J. Tyler, B.F. Volkman, J.L. Markley: Protocol for Eukaryotic Cell-Free Structural Proteomics Illustrated through its application to a Novel NMR Structure: Beta-Grasp fold Protein from *Arabidopsis thaliana. Nature Methods*, in press.	



Structural Features

The most similar structure in the PDB was the ubiquitin-like protein Rub1 (1BT0), the plant homolog of NEDD8, with 27% structure identity over 76 residues. The beta-grasp fold is ~80 amino acid in length, and is composed of a 5-stranded beta sheet that wraps around a single alpha helix. Ub and ubiquitin-like proteins (UBL) generally serve as covalent modifiers of other proteins in the cell, tagging them for proteasomal degradation or altering their functional behavior. Based on sequence analysis, At3g01050.1 probably does not function as a Ub-like covalent modifier, but instead may be prenylated at a putative C-terminal CAAX box motif, thus targeting the protein and its binding partners to a membrane compartment of the cell. This target aligns to Pfam domain of Pfam-B_16300 over residues 1–99.

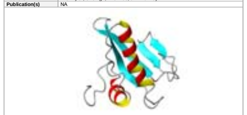
Percent Identity with Nearest PDB Structure at Time Solved	27% over 76 aa (1BT0)
Pfam Cluster	B_16300
Protonet Cluster Size : Structures in PDB	613 : 7

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

CoA substrate verified by NMR

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	06042	
Organism	*Arabidopsis thaliana*	
Gene Designator	At1g77540.1	
PDB Entry	1XO4	Deposition: 5-Oct-2004
BMRB Entry	6338	Deposition: 5-Oct-2004
Function	Putative acetyltransferase	
Produced From	*E. coli* B834(DE3)/pLacIRARE	
Structure by NMR	Restraints/Residue: 19	Subunits/Molecule: 1
	No. of Residues: 103	Molecular Weight: 11.7 kDa
	Backbone RMSD(7–87): 0.48 Å	All Heavy Atoms RMSD(7–87): 0.86 Å
Data Collected At	Nuclear Magnetic Resonance Facility at Madison (NMRFAM)	
Authors	R.C. Tyler, S. Singh, M.S. Lee, J.L. Markley	
Publication(s)	NA	



Structural Features

The most similar structure in the PDB was GCN5-like putative N-acetyltransferase from *Staphylococcus aureus* (1R57) with 32% sequence identity over 71 residues. Several other targets with similar fold were co-crystallized with acetyl coenzyme A or coenzyme A with or without putative substrate. In view of these findings, the titration of ^{15}N-At1g77540.1 with coenzyme A was followed by NMR spectroscopy. The ^{1}H-^{15}N HSQC chemical shift perturbation clearly demonstrated the binding and the location of this ligand.The coenzyme A binding site was mapped to At1g77540.1 resides in the cavity formed by C-terminal residue of β-sheet 4 (β4), loop residues spanning β4 and the N-terminal central helix (α1), N-terminal residues of α1 and residues of smaller perpendicularly oriented helix (α-2) in agreement with the X-ray co-crystal structures. Thus, it is most likely that At1g77540.1 functions as acetyl transferase. This target aligns to two Pfam domains of Pfam-B_34288 over residues 2–29 and Pfam- B_2135 over residues 30–79.

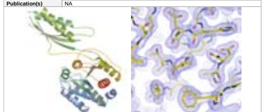
Percent Identity with Nearest PDB Structure at Time Solved	32% over 71 aa (1R57)
Pfam Cluster	B_34288, B_2135
Protonet Cluster Size : Structures in PDB	22 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Plant hormone metabolism; low homology

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	22207	
Organism	*Arabidopsis thaliana*	
Gene Designator	At5g56660.1	
PDB Entry	1XMB	Deposition: 01-Oct-2004
Function	Indole acetic acid amino acid hydrolase (ILL2)	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 2.0 Å SeMet SAD	R-value (R-free): 15.9% (20.4%)
	No. of Residues: 439 (47,811)	Subunits/ASU: 1
Data Collected At	Advanced Photon Source SBC 19-BM	
Authors	E. Bitto, C. Bingman, S. Allard, G. Wesenberg, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features:

At5g56660 (1XMB) represents the first structure of an IAA-amino acid hydrolase. It is involved in the hydrolysis of amide-linked conjugates of auxin (IAA, indole-3-acetic acid). Highest activity was found for IAA-Ala as a substrate, followed by IAA-Ser and IAA-Thr (1). Auxin is one of the most important plant hormones. Auxin levels are regulated by several mechanisms, among them hydrolysis of amide-linked auxin conjugates. BLAST searches reveal over 200 related proteins. There are a large number of closely related IAA-hydrolases (>50% identity) in other plants as rice, wheat and poplar trees. There are also five closely related *Arabidopsis* IAA-aminocid hydrolases that show 40-85% identity in primary sequence (IR3, ILL5, ILL1, ILL6, ILR1). At lower primary structure identities (~30-50%) At5g56660 is similar to a large number of peptidases from microorganisms and eukaryotes. In addition to the known Peptidase_M20 domain (108-421) and M20_dimer domain (213-317), At5g56660 matches Pfam-B_1745, which is found in proteins from Eukaryota, Eubacteria, and Archaea. References: 1) LeClere S *et al.* JBC (2002) 277, 20446-20452.

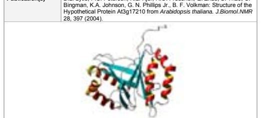
Percent Identity with Nearest PDB Structure at Time Solved	23.8% over 361 aa (1VGY)
Pfam Cluster	Peptidase_M20, B_1745
Protonet Cluster Size : Structures in PDB	189 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Dimer by NMR; low homology

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	13081	
Organism	*Arabidopsis thaliana*	
Gene Designator	At3g17210.1	
PDB Entry	1Q53	Deposition: 6-Aug-2003
BMRB Entry	5843	Deposition: 7-Aug-2003
Function	Unknown	
Produced From	*E. coli* Rosetta(DE3)/pLysS	
Structure by NMR	Restraints/Residue: 18.4	Subunits/Molecule: 2
	No. of Residues: 224	Molecular Weight: 24.4 kDa
	Backbone RMSD: 0.95 Å	All Heavy Atoms RMSD: 1.30 Å
Data Collected At	Medical College of Wisconsin, Milwaukee, WI	
Authors	B.L. Lytle, F.C. Peterson, B.F. Volkman	
Publication(s)	B.L. Lytle, F.C. Peterson, K.L. Kjer, R.O. Frederick, Q. Zhao, S. Thao, C. Bingman, K.A. Johnson, G. N. Phillips Jr., B. F. Volkman: Structure of the Hypothetical Protein At3g17210 from *Arabidopsis thaliana. J.Biomol.NMR* 28, 397 (2004).	



Structural Features

The most similar structure in the PDB to homodimeric (112 residues/monomer) At3g17210.1 shows 35% identity over 108 aligned residues (1RJJ). Additionally, ActVA-Orf6, a bacterial monooxygenase from *Streptomyces coelicolor* (1LQ9) and a protein of unknown function from *Thermus thermophilus* (1IUJ) show structural similarity. Although the two proteins, ActVA-Orf6 and At3g17210, share only 10% sequence identity, their tertiary and quaternary structures are very similar. Because none of the active site residues of ActVA-Orf6 are retained in At3g17210, the latter protein probably has a different function, which remains to be elucidated. This target aligns to Pfam-B domain of Pfam-B_3438 over residues 8–102.

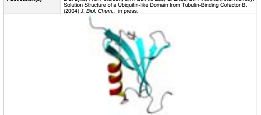
Percent Identity with Nearest PDB Structure at Time Solved	10% over 109 aa (1LQ9)
Pfam Cluster	B_3438
Protonet Cluster Size : Structures in PDB	67 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

C. elegans; SECSG collaboration

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	33910	
Organism	*Caenorhabditis elegans*	
Gene Designator	Ce5O73/F53F4.3	
PDB Entry	1T0Y	Deposition: 13-Apr-2004
BMRB Entry	6176	Deposition: 13-Apr-2004
Function	Putative tubulin binding protein	
Produced From	*E. coli* SG13009[pREP4]	
Structure by NMR	Restraints/Residue: 14.2	Subunits/Molecule: 1
	No. of Residues: 122	Molecular Weight: 13.6 kDa
	Backbone RMSD(2–88): 0.78 Å	All Heavy Atoms RMSD(2–88): 1.26 Å
Data Collected At	Medical College of Wisconsin, Milwaukee, WI	
Authors	B.L. Lytle, F.C. Peterson, S.H. Qui, M. Luo, B.F. Volkman, J.L. Markley	
Publication(s)	B.L. Lytle, F.C. Peterson, S.H. Qui, M. Luo, Q. Zhao, B.F. Volkman, J.L. Markley: Solution Structure of a Ubiquitin-like Domain from Tubulin-Binding Cofactor B. (2004) *J. Biol. Chem.*, in press.	



Structural Features

This structure is of the N-terminal domain (residues 1–120) of Ce5073, also known as F53F4.3 and tubulin cofactor B (CoB). The most similar structures in the PDB were elongin B, ubiquitin and number of other ubiquitin-like domains. Despite very low sequence similarity (< 20%) to ubiquitin and other beta-grasp proteins, the N-terminal domain of CoB adopts the same fold, consisting of a mixed five-stranded beta sheet that partially enfolds an alpha helix. The function of this domain is unknown, although it has been shown to be essential for the function of cofactor B in promoting correct folding of alpha-tubulin monomers and subsequent assembly into tubulin heterodimers. Since it lacks the C-terminal Gly-Gly sequence involved in conjugation, it is unlikely to function as a ubiquitin-like covalent modifier of other target proteins. Instead, we hypothesize that the domain has separate interaction sites for binding α-tubulin and cofactor E. This target aligns to Pfam domain of Pfam-B_9516 over residues 5–119.

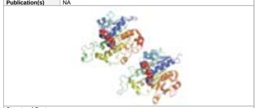
Percent Identity with Nearest PDB Structure at Time Solved	11% over 97 aa (1LM8)
Pfam Cluster	B_9516
Protonet Cluster Size : Structures in PDB	173 : 1

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Tightly bound NADP; low homology

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	23662	
Organism	*Arabidopsis thaliana*	
Gene Designator	At5g02240.1	
PDB Entry	1XQ6	Deposition: 12-Oct-2004
Function	Oxidoreductase (based on bound NADPH)	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 1.8 Å	R-value (R-free): 22.1% (27.8%)
	No. of Residues: 253 (27,059x2)	Subunits/ASU: 2
Data Collected At	Advanced Photon Source Bio-CARS 14-ID-B	
Authors	D. Smith, G. Wesenberg, C. Bingman, S. Allard, E. Bitto, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features:

At5g02240 was a fold space target, with weak similarity to small segments of dehydrogenases in the PDB (22% over 90 aa to 1HD0). All of the weak matches were less significant than E=2. The structure was determined by SeMet MAD, and the protein has a classic dehydrogenase core, with additional N- and C-terminal domains, which match to Pfam-B_27474 and Pfam-B_11732. This structure will enable more confident modeling of the 16 structural neighbors in Protonet, and the 128 proteins with domains in PFam-B_227. DALI searches show good structural overlap to a number of oxidoreductases. The closest match, with a Z score of 24.3, is to biliverdin-IX beta reductase (1HE2). The crystal structure clearly shows that this protein is an oxidoreductase, since it came through protein purification and crystallization with tightly bound NADPH co-factors (see ribbon plot above.)

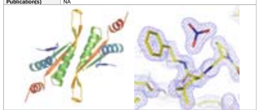
Percent Identity with Nearest PDB Structure at Time Solved	26% over 90 aa 1HD0 (E=2.2)	
Pfam Cluster	Pfam-B_27474	2-116
	Pfam-B_227	117-176
	Pfam-B_11732	203-248
Protonet Cluster Size : Structures in PDB	16 : 0	

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

DNA-binding protein from fold similarity

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Initiative

CESG ORF No.	13081	
Organism	*Arabidopsis thaliana*	
Gene Designator	At2g34160	
PDB Entry	1VM0	Deposition: 24-Aug-2004
Function	Unknown	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 1.80 Å	R-value (R-free): 17.7% (22.1%)
	No. of Residues: 130	Subunits/ASU: 2
Data Collected At	Advanced Photon Source COM-CAT 32-ID-B 04/03/2004	
Authors	S. Allard, E. Bitto, C. Bingman, G. Wesenberg, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

Sequence searches against the PDB indicate that At2g34160 is most similar to the prokaryotic DNA-binding protein Alba (1NH9). The overall level of similarity is significant (E-0.12) over 81 amino acids. Blast searches against Genbank show two different length classes of Alba-like proteins. Alba is of the shorter class. At2g34160 is the only known structural example of the larger class of Alba-like proteins, with the additional C-terminal helix (the final red helix in the ribbon diagram above.) This portion of the protein matches Pfam-B domain 26912. The amino-terminal 18 residues of At2g34160 are disordered in the crystal, and contain a number of positively charged side-chains that may be involved in embracing nucleic acids. There is spatial similarity between the position of bound nitrates in the dimer-dimer interface in this crystal (above) and positions of phosphates in models of Alba-nucleic acid complexes. References: 1) Wardleworth, B. N., Russell, R. J. M., Bell, S. D., Taylor, G. L. and White, M. F. (2002) Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO J*, 21, 4654-4662; 2) Aravind, L., Lakshminarayan, I. M. and Anantharaman, V. (2003) The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biology*, 4, R64.

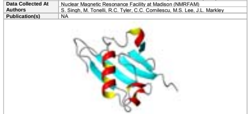
Percent Identity with Nearest PDB Structure at Time Solved	34.6% over 81 aa (1NH9)
Pfam Cluster	Alba, Pfam-B_26912
Protonet Cluster Size : Structures in PDB	10 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

New ubiquitin-like mouse protein

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	34198	
Organism	*Mus musculus*	
Gene Designator	Mm.20273	
PDB Entry	1XO3	Deposition: 5-Oct-2004
BMRB Entry	6337	Deposition: 5-Oct-2004
Function	Unknown	
Produced From	*E. coli* B834(DE3)/pLacIRARE	
Structure by NMR	Restraints/Residue: 23	Subunits/Molecule: 1
	No. of Residues: 101	Molecular Weight: 11.2 kDa
	Backbone RMSD(4–98): 0.28 Å	All Heavy Atoms RMSD(4–98): 0.69 Å
Data Collected At	Nuclear Magnetic Resonance Facility at Madison (NMRFAM)	
Authors	S. Singh, M. Tonelli, R.C. Tyler, C.C. Cornilescu, M.S. Lee, J.L. Markley	
Publication(s)	NA	



Structural Features

The most similar structure in the PDB were the MoaD related protein from *Thermus thermophilus* Hb8 with 13.2% identity over 68 residues (1V8C) and molybdopterin synthase with 23.4% identity over 64 aligned residues. Molybdenum cofactor biosynthesis is an evolutionarily conserved pathway over eubactera, archaea, and eukaryotes. In humans, genetic deficiencies of this enzyme lead to a severe and usually fatal disease. Serious neurological symptoms include attenuated growth of the brain and untreatable seizures. Molybdopterin synthase is a heterotetramer with large and small subunits known as MoaE and MoaD, respectively. The small subunit, MoaD, is ubiquitin-like, and the insertion of the C-terminus of this subunit into a large subunit forms the active site. The active form of MoaD is thiocarboxylated at the invariant C-terminal Gly residue that serves as the sulfur donor in the biosynthetic pathway. As expected, similar to MoaD, the structure of Mm.20273 differs from that of ubiquitin (Ub) largely in the regions outside of the well defined Ub signature secondary structure fold. This target aligns to Pfam domains of Pfam-B_45334 over residues 2–63 and Pfam-B_14998 over residues 64–101.

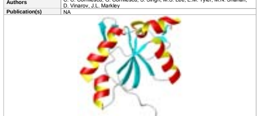
Percent Identity with Nearest PDB Structure at Time Solved	13.2 % over 68 aa (1V8C)
Pfam Cluster	B_45334, B_14998
Protonet Cluster Size : Structures in PDB	11 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Unknown function; low homology

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	20862	
Organism	*Arabidopsis thaliana*	
Gene Designator	At5g66040.1	
PDB Entry	1TQ1	Deposition: 16-Jun-2004
BMRB Entry	6240	Deposition: 15-Jun-2004
Function	Unknown	
Produced From	Cell-Free (wheat germ extract)	
Structure by NMR	Restraints/Residue: 16.5	Subunits/Molecule: 1
	No. of Residues: 129	Molecular Weight: kDa
	Backbone RMSD: 1.05 Å (16–56, 85–128)	All Heavy Atoms RMSD: 1.59 Å (16–56, 85–128)
Data Collected At	Nuclear Magnetic Resonance Facility at Madison (NMRFAM)	
Authors	C. C. Cornilescu, G. Cornilescu, S. Singh, M.S. Lee, E.M. Tyler, M.N. Shahan, D. Vinarov, J.L. Markley	
Publication(s)	NA	



Structural Features

The most similar structure in the PDB was a prototype sulfortransferase for the single-domain Rhodanese Homology Superfamily GlpE from *Escherichia coli* (1GMX) with 22 % sequence identity over 77 residues. The rhodanese homology domain has an alpha beta fold domain duplicated internally in the rhodanese protein. The enzymatically active cysteine containing domain can be found in CDC25 class of phosphatases, sulfide dehydrogenases and stress protein such as senescence specific protein 1 in plants, PspE and GlpE in bacteria and cyanide and arsenate resistance proteins. Additionally, this domain can be found as a single copy in other proteins such as phosphatases and ubiquitin C-terminal hydrolases. Although At5g66040.1 falls within the Rhodanese Superfamily, the true biological function of At5g66040.1 currently is unknown. This target aligns to domains within Pfam trusted matches for rhodanese over residues 13–114.

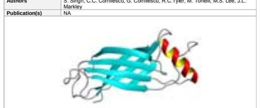
Percent Identity with Nearest PDB Structure at Time Solved	22% over 77 aa (1GMX)
Pfam Cluster	Rhodanese
Protonet Cluster Size : Structures in PDB	53 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

First in class; low homology

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	02361	
Organism	*Arabidopsis thaliana*	
Gene Designator	At1g01470.1	
PDB Entry	1XO8	Deposition: 6-Oct-2004
BMRB Entry	6339	Deposition: 6-Oct-2004
Function	Unknown	
Produced From	*E. coli* B834(DE3)/pLacIRARE	
Structure by NMR	Restraints/Residue:	Subunits/Molecule: 1
	No. of Residues: 151	Molecular Weight: 16.5 kDa
	Backbone RMSD(23–146): 0.34 Å	All Heavy Atoms RMSD(23–146): 0.67 Å
Data Collected At	Nuclear Magnetic Resonance Facility at Madison (NMRFAM)	
Authors	S. Singh, C.C. Cornilescu, G. Cornilescu, R.C.Tyler, M. Tonelli, M.S. Lee, J.L. Markley	
Publication(s)	NA	



Structural Features

The most similar structures in the PDB were human cellular coagulation factor XIII with 6.7% identity over 104 residues (1F13), human Fcari bound to Iga1-Fc with 4.7% over 87 residues (1OWO), and histocompatibility antigen Hla-Dm with 11.9% over 84 residues (1HDM). At1g01470.1 belongs to a class of proteins known as late embryogenesis abundant (LEA) proteins that are expressed at different stages of late embryogenesis in higher plant seed embryos. Different types of these proteins are expressed under stress. The expression of the LEA gene imparts an adaptive mechanism to withstand abiotic stress conditions such as salinity, drought, and high or low temperature. Although extensive studies has been carried out on LEA_2 proteins, this is the first three-dimensional structure solved for this class of protein. The structural similarity of At1g01470.1 to proteins in the mammalian blood coagulation cascade involved in wound healing is intriguing and suggests that the protein may be involved in a plant system that responds to stress of a wounding nature. However tantalizing this conjecture may be, the true biological function of At1g01470.1 currently is unknown. This target aligns to domains within Pfam trusted matches for LEA_2 over residues 2–151.

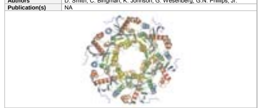
Percent Identity with Nearest PDB Structure at Time Solved	6.7% over 104 aa (1F13)
Pfam Cluster	LEA_2
Protonet Cluster Size : Structures in PDB	24 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

Nucleic acid through the center?

Center for Eukaryotic Structural Genomics (CESG)
Protein Structure Data Summary

CESG ORF No.	22116	
Organism	*Arabidopsis thaliana*	
Gene Designator	At5g06450.1	
PDB Entry	1VK0	Deposition: 12-Apr-2004
Function	Unknown	
Produced From	*E. coli* B834(DE3) p(LacI+RARE)	
Structure by X-ray	Resolution: 2.10	R-value (R-free): 18.3% (23.3%)
	No. of Residues: 206 (23,172x6)	Subunits/Molecule: 6
Data Collected At	Advanced Photon Source BioCARS 14-BM-D 26-Oct-2003	
Authors	D. Smith, C. Bingman, K. Johnson, G. Wesenberg, G.N. Phillips, Jr.	
Publication(s)	NA	



Structural Features

At5g06450 has been assigned to the 35EXOc, a family of proteins that includes the 3'-5' exonuclease proofreading domain present in DNA polymerase I, Werner syndrome helicase, RNase D and other enzymes. The closest 3D match in DALI is to 1XWL, a 3'-5' exonuclease domain, with an RMSD of 3.6 Å over the aligned regions. At5g06450 represents the only structure of a member of this class of proteins with a cyclic quaternary structure. At5g06450 has C_6 rotational symmetry. There is a central cavity of variable dimensions, narrowing to 22 Å width between ordered side chains on the N-terminal face of the hexamer. Disordered loops, invisible in the electron density map, would seem to project into this open annulus, further narrowing the cavity. The disordered loops have the sequence YKYKGS, which is highly positively charged, and contains aromatic groups that could stack with nucleic acid. The overall structure is reminiscent of processivity factors in nucleic acid metabolism and viral exonucleases. The size of the Protonet cluster associated with this protein suggests that it may open up a large number of other proteins to modeling efforts.

Percent Identity with Nearest PDB Structure at Time Solved	none better than E=1
Pfam Cluster	Pfam-B_104596
Protonet Cluster Size : Structures in PDB	61 : 0

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.262.3759; cesginfo@biochem.wisc.edu; http://www.uwstructuralgenomics.org [NIH Grant P50 GM64598].

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (P50 GM 64598).