Solution structure of At3g04780.1-des15, an *Arabidopsis thaliana* ortholog of the C-terminal domain of human thioredoxin-like protein

JIKUI SONG, ROBERT C. TYLER, RUSSELL L. WROBEL, RONNIE O. FREDERICK, FRANK C. VOJTEK, WON BAE JEON, MIN S. LEE, AND JOHN L. MARKLEY

Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706-1544, USA

(RECEIVED November 16, 2004; FINAL REVISION December 21, 2004; ACCEPTED December 21, 2004)

Abstract

The structure of At3g04780.1-des15, an *Arabidopsis thaliana* ortholog of the C-terminal domain of human thioredoxin-like protein, was determined by NMR spectroscopy. The structure is dominated by a β -barrel sandwich. A two-stranded anti-parallel β -sheet, which seals off one end of the β -barrel, is flanked by two flexible loops rich in acidic amino acids. Although this fold often provides a ligand binding site, the structure did not reveal an appreciable cavity inside the β -barrel. The three-dimensional structure of At3g04780.1-des15 provides an entry point for understanding its functional role and those of its mammalian homologs.

Keywords: structural genomics; *Arabidopsis thaliana*; NMR; TXNL_HUMAN ortholog

The prevailing method for inferring the molecular role of a putative gene product relies on comparisons with sequences of proteins with known function. However, the current database is insufficient to support the determination of functions for most newly discovered genes. Structural genomics projects aim to alleviate this by expanding knowledge of sequence–structure relationships (Terwilliger 2000). Since the molecular function of a protein frequently is implicated by its three-dimensional structure, structure determinations can provide a shortcut way of linking sequence and function (Mittl and Grutter 2001).

The goals of the Center for Eukaryotic Structural Genomics (CESG) are to develop the technology for producing eukaryotic proteins and for determining their three-dimensional structures. CESG applies this technology in determining structures of novel eukaryotic proteins. The structure described here is the product of gene At3g04780.1 from

Arabidopsis thaliana, which was chosen as a fold-space target. The full-length sequence of At3g04780.1 revealed little sequence identity with any other protein with known structure or function. However, part of this protein shows a 42% sequence identity with the C-terminal domain of the 32-kDa human thioredoxin-like protein (TXNL_HUMAN) (Fig. 1; Lee et al. 1998). The C-terminal domain of the TXNL_HUMAN is rich in acidic residues and has a calculated pI of 4.3, which is also a distinct feature for At3g04780.1 (calculated pI of 4.9). TXNL_HUMAN is expressed in cytoplasm. Although the molecular function of this protein is unknown, its genomic location has suggested that it plays a role in cell apoptosis and cancer (Miranda-Vizuet and Spyrou 2000). The X-ray structure of the N-terminal domain of TXNL_HUMAN has been solved and shown to be a member of thioredoxin family (Jin et al. 2002). However, attempts to crystallize the C-terminal domain of this protein proved unsuccessful. We report here the NMR structure of At3g04780.1-des15 (Fig. 2), which allows one to model the three-dimensional structure of the C-terminal domain of TXNL_HUMAN and provides the foundation for future functional studies of this and other homologous proteins.

Reprint requests to: John L. Markley, Center for Eukaryotic Structural Genomics, Department of Biochemistry, 433 Babcock Drive, University of Wisconsin-Madison, Madison, WI 53706-1544, USA; e-mail: markley@nmrfam.wisc.edu; fax: (608) 262-3759.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.041246805>.

```

At3g04780.1      1  -SAESASQIPKQGVLDLDFIDWSGVECLNQSSSHSLPNALKQGYREDEGLNLESDADEQL
TXNL_HUMAN      111 PGSNEDTDIPKGYMDLMPFINKAGCECLNESDEHGFNCLRK-----DTTFLESDCEDEQL
At3g04780.1-des15 1  -SAESASQIPKQGVLDLDFIDWSGVECLNQSSSHSLPNALKQGYREDEGLNLESDADEQL
At3g04780.1      60 LIYIPFNQVIKLSFAIKGPEE-EGPKTVKFFSNK-EHMCFSNVNDFPFSDTAEELTEENL
TXNL_HUMAN      166 LITVAFNQPVKLYSMKFQGPDNQGGPKYVKIFINLPRSMDFEAEERSEPTQALELTDDDI
At3g04780.1-des15 60 LIYIPFNQVIKLSFAIKGPEE-EGPKTVKFFSNK-EHMCFSNVNDFPFSDTAEELTEENL
At3g04780.1      118 RGKPVV-LKYVKFQNVRSLTIFIEANQSGSEVTKVQKIALYGSTVETDMKGLKKIEDH-
TXNL_HUMAN      226 REDGIVPLRYVKFQNVNSVTIFVQSNQGEETTRISYFTFTGTPVQATNMNDFKRVVGGK
At3g04780.1-des15 118 RGKPVV-LKYVKFQNVRSLTIFIEANQSGSEVTKVQKIALYGST

```

Figure 1. Alignment of the sequences of *Arabidopsis* At3g04780.1 and human thioredoxin-like protein (TXNL_HUMAN). The conserved and homologous residues are highlighted in green and cyan, respectively. The actual sequence of the protein whose structure was determined (At3g04780.1-des15) is also shown; it was cloned according to an earlier version of the reported sequence of At3g04780.1, which lacked the nucleotides coding for the 15 C-terminal amino acids.

Results and Discussion

The quality of the structure improved markedly after X-PLOR refinement. The final ensemble of 20 conformers determined by CYANA had a backbone r.m.s.d. (residues 9–161) of 0.81 Å and an average target function of 0.45 Å². Following water refinement with X-PLOR, although the backbone r.m.s.d. (residues 9–161) (0.92 Å) did not change significantly, the percentage of residues falling into the most favored regions of the Ramachandran plot increased from 68.6% to 79.1% (Table 1). Figure 3A shows a stereo view of the family of 20 conformers representing the structure of At3g04780.1-des15.

The overall fold of At3g04780.1-des15 is dominated by a β -sandwich jelly roll structure. The β -sandwich is formed

by face-to-face packing of two anti-parallel β -sheets, sheet A composed of four strands (β_1 , residues 14–15; β_5 , 71–78; β_8 , 123–124; and β_{11} , 154–160) and sheet B composed of five strands (β_2 , 26–27; β_4 , 60–70; β_6 , 85–92; β_7 , 109–112; and β_9 , 133–143) (Fig. 3B,C). Internal to the sandwich, the side chains of residues from multiple individual strands pack together to form a large hydrophobic patch. No cavity is present as a potential site for ligand binding. Furthermore, as shown in Figure 3B, the north end of the β -barrel is sealed by a third anti-parallel β -sheet (sheet C), which is composed of two β strands (β_3 , 52–53 and β_{10} , 151–152). Flexible loops flanking sheet C as well as loops encompassing sheets A and B protrude and are exposed to the solvent. The loop residues are largely composed of hydrophilic residues. On the opposite side, the extension and curvature of strands β_4 , β_5 , and β_7 serve to seal the β -barrel. Other secondary structural elements include four short helices (α_1 , residues 16–19; α_2 , 37–40; α_3 , 99–102; and α_4 , 115–118), arranged at discrete regions of the protein.

The VAST (<http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html>) server was used to search the Protein Data Bank for the structural homologs of At3g04780.1-des15.

Table 1. Statistics for the 20 energy-minimized conformers of At3g04780.1

Distant constraints	
Long [(i-j) > 5]	918
Medium [1 < (i-j) ≤ 5]	277
Sequential [2 (i-j) = 1]	475
Intraresidue [i = j]	220
Dihedral angle constraints (φ and ψ)	174
Hydrogen bond constraints	68
Target function (Å ²)	0.45 ± 0.09
Average pairwise R.M.S.D. to the mean structure (Å)	
Residue 9–161	
Backbone (C α , C', N, O)	0.92 ± 0.10
Heavy atoms	1.43 ± 0.08
Regular secondary structure elements	
Backbone (C α , C', N, O)	0.58 ± 0.11
Heavy atoms	1.15 ± 0.10
Deviations from idealized covalent geometry	
Bond (Å)	0.018 ± 0.001
Angles (°)	1.184 ± 0.043
Impropers (°)	1.529 ± 0.106
R.M.S.D. from experimental distance restraints (Å)	0.016 ± 0.002
R.M.S.D. from experimental dihedral restraints (Å)	0.579 ± 0.108
Ramachandran statistics (% of all residues)	
Most favored	79.14 ± 2.45
Additionally allowed	16.94 ± 2.21
Generously allowed	2.66 ± 1.17
Disallowed	1.26 ± 0.77

Figure 2. [¹H, ¹⁵N]-HSQC spectrum of At3g04780.1-des15 from *Arabidopsis thaliana* recorded at 600 MHz ¹H resonance frequency. The protein concentration was ~1 mM, the pH was 7.2, and the temperature was 25°C.

Figure 3. Structure of *Arabidopsis* At3g04780.1-des15. (A) Stereoscopic view of the final family of 20 conformers representing the structure. The backbone colors represent (red) α -helix, (green) β -strand, and (dark gray) other regions. The side chains of the protein are shown in light gray. (B) Ribbon diagram. The side chains of the acidic residues surrounding the sheet C are also shown. (C) Schematic of the secondary structure topology. (D) Representation of the surface electrostatic potential, oriented as in C, with positive regions in blue and negative regions in red as calculated by the program MOLMOL (Koradi et al. 1996). (E) Electrostatic surface as in D but rotated by 180° about the vertical axis. For clarity, the disordered N-terminal residues (1–8) are not shown in A,B,D,E.

Among the closest structural neighbors are the *Saccharomyces cerevisiae* anaphase-promoting complex subunit DOC1/Apc10 (PDB accession number 1GQP) with C_{α} r.m.s.d. of 2.9 Å for 127 aligned residues, the *Anguilla anguilla* fucose binding lectin (1K12) with C_{α} r.m.s.d. of 3.0 Å for 129 aligned residues, and the galactose binding domain of *Dactylium dendriodes* galactose oxidase (1GOF)

with C_{α} r.m.s.d. of 3.1 Å for 128 aligned residues (Fig. 4). DOC1/Apc10 is involved in protein–protein interactions, and the other two proteins bind sugars (Ito et al. 1991; Au et al. 2002; Bianchet et al. 2002). Coincidentally, the molecular interactions in these three disparate proteins occur at a region structurally equivalent to sheet C and its two flanking loops of At3g04780.1-des15. This region in Doc1/

Figure 4. Ribbon diagrams (generated with PYMOL, <http://www.pymol.org>) depicting the structures of At3g04780.1-des15 and the three β -sandwich proteins with closest structural similarity: DOC1/APC10, *Saccharomyces cerevisiae* anaphase-promoting complex subunit (PDB accession number 1GQP); fucose binding lectin (shown along with fucose) from *Anguilla anguilla* fucose binding lectin (1K12); galactose oxidase, the galactose binding domain of galactose oxidase (1GOF).

APC10 was proposed to mediate its conserved function in forming the anaphase-promoting complex (APC) (Au et al. 2002); the corresponding regions in fucose binding lectin and galactose oxidase serve as substrate binding pockets (Ito et al. 1991; Bianchet et al. 2002). These observations suggest that this region of At3g04780.1 may play an important role in its biological function. This implication is further reinforced by another finding: Structural analysis of At3g04780.1-des15 reveals that, although its charged residues are spread over the surface, the highest concentration of acidic residues occurs in the proximal region of sheet C (Fig. 3D,E). In particular, the Pro80-Glu81-Glu82-Glu83-Gly84-Pro85 sequence motif (Pro-Asp-Asn-Gly-Gln-Gly-Pro in TXNL_HUMAN), located on the one of the two loops flanking sheet C is a unique feature of At3g04780.1-des15.

The biological function of the human ortholog of At3g04780.1, the C-terminal domain of TXNL_HUMAN, is unknown. Thus far, homologs for TXNL_HUMAN have been identified only in animals and other multicellular eukaryotes (Lee et al. 1998). Implications for the biological function of TXNL_HUMAN come from several experimental observations: TXNL_HUMAN has been shown to copurify with a kinase of the STE20 family (Lee et al. 1998; Miranda-Vizueté and Spyrou 2000), which is proteolytically activated by caspase in apoptosis. Furthermore, the fact that many tumor-related genes are located in the same chromosomal region as TXNL_HUMAN suggests that TXNL_HUMAN may be a tumor-related protein (Miranda-Vizueté and Spyrou 2000). However, no experimental evidence is available to support the implications, and it appears that there are no molecular interactions between TXNL_HUMAN and the STE20 proteins (Lee et al. 1998). The thioredoxin domain in this protein has reducing activity comparable to that of human thioredoxin, although its rate

of reduction by DTT from the oxidized state is much lower (Lee et al. 1998). The three-dimensional structure of the N-terminal domain of TXNL_HUMAN, as determined by X-ray crystallography (Jin et al. 2002), is similar to those of other thioredoxins. However, unlike other thioredoxins, the N-terminal domain of TXNL_HUMAN is monomeric, instead of dimeric, and even more uniquely, a great number of positively charged residues (Lys28, Arg32, Arg39 and His62) occur at its active site (Jin et al. 2002). Dimerization is a key regulatory factor in human thioredoxin (Weichsel et al. 1996). The lack of dimer formation of N-terminal TXNL_HUMAN and the acidic amino acid-rich C-terminal domain have led to the hypothesis that the C-terminal domain of TXNL_HUMAN may function to negatively regulate the activity of the N-terminal thioredoxin domain (Jin et al. 2002). To the best of our knowledge, there are no *Arabidopsis* thioredoxin-like proteins identified at present that exhibit such a feature.

The three-dimensional structure of At3g04780.1-des15 provides an entry point for understanding its functional role. The structure suggests further experiments designed to screen for ligands and/or other proteins involved in the biological role of At3g04780.1 and its mammalian orthologs.

Materials and methods

The protein produced for this study was cloned on the basis of the original sequence published for gene At3g04780.1; a subsequent correction of the ORF lengthened its C terminus by 15 amino acid residues (Fig. 1). The protocols used for gene cloning, cell growth, isotope labeling, and protein purification of target proteins have been described elsewhere (Aceti et al. 2003; Jeon et al. 2004; Tyler et al. 2004). In short, the At3g04780.1-des15 gene was inserted into protein expression vector VP13 (developed at CESC), which produces the protein as a tobacco etch virus (TEV) protease cleavable fusion with maltose binding protein containing an N-terminal (His)₆-affinity tag. The host cells for protein production, *Escherichia coli* B834(DE3)/pLacIRARE, were grown on an autoinduction medium supplemented with ¹³C-glycerol and ¹⁵N-NH₄Cl (Cambridge Isotope Inc.) (Tyler et al. 2004). The fusion protein was initially purified by immobilized metal affinity chromatography (IMAC), and the target protein was subsequently separated from the (His)₆-MBP through TEV protease cleavage followed by a second IMAC purification step (Jeon et al. 2004). The NMR sample for the structure determination contained ~1 mM [U-¹³C, U-¹⁵N]-At3g04780.1-des15, 50 mM KH₂PO₄, 90% H₂O/10% D₂O, and 10 mM DTT at pH 7.2.

All NMR spectra were recorded at the National Magnetic Resonance Facility at Madison (NMRFAM) on a Varian Inova 600 spectrometer equipped with a cryogenic probe. The temperature of the sample was held at 25°C. Data sets collected for resonance assignments and structure determination included 2D [¹H, ¹⁵N] HSQC, 2D [¹H, ¹³C] HSQC, 2D (HB)CB(CGCD)HD, 2D (HB)CB(CGCDCE)HE, 3D HNCACB, HN(CO)CACB, HNCO, HCCONH, CCONH, HCCH-COSY, ¹⁵N edited [¹H, ¹H] NOESY ($\tau_{\text{mix}} = 100$ msec), and ¹³C edited [¹H, ¹H] NOESY ($\tau_{\text{mix}} = 100$ msec).

We used published approaches for NMR spectral processing, data analysis, and resonance assignment (Song et al. 2004). The

GARANT (Bartels et al. 1996) program was used as a semiautomatic approach for determining the backbone and aliphatic side chain resonance assignments. Information from 2D (HB)CB (CGCD)HD and 2D (HB)CB(CGCDCE)HE (Yamazaki et al. 1993) data sets, which correlate the $^{13}\text{C}^\beta$ of an aromatic residue with its $^1\text{H}^{\delta/\epsilon}$, were used to assign the aromatic side chains of Phe, Tyr, and His. The backbone assignments were 99% complete (Fig. 2), and the side chain assignments were ~90% complete. The raw, time-domain, NMR data sets and chemical shift assignments have been deposited in the BioMagResBank database under BMRB accession number 6341.

The TALOS (Cornilescu et al. 1999) software package was used to predict ϕ and ψ torsion angles from the assigned chemical shifts as restraints for structure calculations. The automated CANDID iterative refinement module of the CYANA software package (Herrmann et al. 2002) was used in generating the initial NOE assignments and the initial set of structural models. Additional NOE assignments were then added and erroneous ones corrected through examination of NMR spectra prior to recalculation of structures by CYANA (Güntert et al. 1997). The final structure refinement by CYANA included 76 hydrogen bond constraints, which were generated from NOEs observed as characteristic for the α -helices and β -sheets. The 20 structures with the lowest target function were chosen for further refinement by X-PLOR (Brünger 1992; Schwieters et al. 2003), in which physical force field terms and explicit water solvent molecules were added to the experimental constraints. The final 20 NMR structures of At3g04780.1 were validated by Procheck-NMR (Laskowski et al. 1996), and the statistics for these are listed in Table 1. The coordinates for these structural models have been deposited in the Protein Data Bank (PDB) under the accession number 1XOY.

Acknowledgments

Structure determined under the National Institutes of Health, NIGMS Protein Structure Initiative; coordinates and related data have been deposited at PDB (1XOY) and NMR data at BMRB (bmr 6341).

This research was supported by the NIH Protein Structure Initiative through grant 1 P50 GM64598. NMR data were collected and analyzed in the National Magnetic Resonance Facility at Madison, which is supported by National Institutes of Health grants P41RR02301 (Biomedical Research Technology Program, National Center for Research Resources) and P41GM66326 (National Institute of General Medical Sciences). Equipment in the facility was purchased with funds from the University of Wisconsin, the National Institutes of Health (P41GM66326, P41RR02301, RR02781, RR08438), the National Science Foundation (DMB-8415048, OIA-9977486, BIR-9214394), and the U.S. Department of Agriculture.

References

- Aceti, D.J., Blommel, P.G., Endo, Y., Fox, B.G., Frederick, R.O., Hegeman, A.D., Jeon, W.B., Kimball, T.L., Lee, J.M., Newman, C.S., et al. 2003. Role of nucleic acid and protein manipulation technologies in high-throughput structural biology efforts. In *Biopolymers* (ed. A. Steinbuechel), pp. 469–496. Wiley-VCH, Weinheim, Germany.
- Au, S.W., Leng, X., Harper, J.W., and Barford, D. 2002. Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the Doc1/Apc10 subunit. *J. Mol. Biol.* **316**: 955–968.
- Bartels, C., Billeter, M., Güntert, P., and Wüthrich, K. 1996. Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *J. Biomol. NMR* **7**: 207–213.
- Bianchet, M.A., Odom, E.W., Vasta, G.R., and Amzel, L.M. 2002. A novel fucose recognition fold involved in innate immunity. *Nat. Struct. Biol.* **9**: 628–634.
- Brünger, A.T. 1992. *X-PLOR version 3.1: A system for X-ray crystallography and NMR*. Yale University Press, New Haven, CT.
- Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**: 289–302.
- Güntert, P., Mumenthaler, C., and Wüthrich, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**: 283–298.
- Herrmann, T., Güntert, P., and Wüthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**: 209–227.
- Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., and Knowles, P.F. 1991. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* **350**: 87–90.
- Jeon, W.B., Aceti, D.J., Bingman, C.A., Vojtik, F.C., Olson, A.C., Ellefson, J.M., McCombs, J.E., Sreenath, H.K., Blommel, P.G., Seder, K.D., et al. 2004. High-throughput purification and quality assurance of *Arabidopsis thaliana* proteins for eukaryotic structural genomics. *J. Struct. Funct. Genomics* (in press).
- Jin, J., Chen, X., Zhou, Y., Bartlam, M., Guo, Q., Liu, Y., Sun, Y., Gao, Y., Ye, S., Li, G., et al. 2002. Crystal structure of the catalytic domain of a human thioredoxin-like protein. *Eur. J. Biochem.* **269**: 2060–2068.
- Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL—A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**: 51–55.
- Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**: 477–486.
- Lee, K.K., Murakawa, M., Takahashi, S., Tsubuki, S., Kawashima, S., Sakamaki, K., and Yonehara, S. 1998. Purification, molecular cloning, and characterization of TRP32, a novel thioredoxin-related mammalian protein of 32 kDa. *J. Biol. Chem.* **273**: 19160–19166.
- Miranda-Vizuet, A. and Spyrou, G. 2000. Genomic structure and chromosomal localization of human thioredoxin-like protein gene (txl). *DNA Seq.* **10**: 419–424.
- Mittl, P.R. and Grutter, M.G. 2001. Structural genomics: Opportunities and challenges. *Curr. Opin. Chem. Biol.* **5**: 402–408.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**: 65–73.
- Song, J., Vinarov, D.A., Tyler, E.M., Shahan, M.N., Tyler, R.C., and Markley, J.L. 2004. Hypothetical protein At2g24940.1 from *Arabidopsis thaliana* has a cytochrome B5-like fold. *J. Biomol. NMR* (in press).
- Terwilliger, T.C. 2000. Structural genomics in North America. *Nat. Struct. Biol.* **7** (Suppl.): 935–939.
- Tyler, R., Sreenath, H.K., Aceti, D.J., Bingman, C.A., Singh, S., Markley, J.L., and Fox, B.G. 2004. Auto-induction medium for the production of [^{15}N] and [^{13}C , ^{15}N]-labeled proteins for NMR screening and structure determination. *Prot. Expr. Purif.* **30**: 215–218.
- Weichsel, A., Gasdaska, J.R., Powis, G., and Montfort, W.R. 1996. Crystal structures of reduced, oxidized, and mutated human thioredoxins: Evidence for a regulatory homodimer. *Structure* **4**: 735–751.
- Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. 1993. 2-Dimensional NMR experiments for correlating $^{13}\text{C}_\beta$ and $^1\text{H}_{\delta/\epsilon}$ chemical-shifts of aromatic residues in ^{13}C -labeled proteins via scalar couplings. *J. Am. Chem. Soc.* **115**: 11054–11055.