Journal of Biomolecular NMR **30:** 215–218, 2004.

Letter to the Editor: Hypothetical protein At2g24940.1 from *Arabidopsis thaliana* has a cytochrome b5 like fold*

Jikui Song, Dmitriy A. Vinarov, Ejan M. Tyler, Mark N. Shahan, Robert C. Tyler & John L. Markley**

Center for Eukaryotic Structural Genomics, Department of Biochemistry, University of Wisconsin-Madison, WI 53706-1544, U.S.A.

Received 19 April 2004; Accepted 9 June 2004

Key words: cell-free protein production and labeling, steroid receptor, structural genomics

Biological context

Progesterone is believed to exert rapid non-genomic actions through its interaction with membrane associated progesterone receptors (MAPRs) (Bramley, 2003; Li and O'Malley, 2003). BLAST sequence searches (Altieri et al., 1995) for mammalian MAPRs and putative MAPRs from plants have identified that these proteins all contain a cytochrome b5-like ligand-binding domain (Mifsud and Bateman, 2002). Interestingly, unlike cytochrome b5 itself, these MAPRs domains appear not to bind heme and not to be involved in redox reactions. Their distinct biological functions suggest that these steroid receptors adopt the cytochrome b5 domain as a template in order to build their own ligand-binding pockets (Mifsud and Bateman, 2002).

The Center for Eukaryotic Structural Genomics is engaged in determining the three-dimensional structures of novel proteins from eukaryotic gene families. Its target selection algorithm selected *Arabidopsis thaliana* putative protein At2g24940.1 for structure determination. The biochemical function of At2g24940.1 currently is unknown. Its ∼40% sequence identity with mammalian MAPR suggests that At2g24940.1 may act as a steroid binding protein. In addition, its sequence is distantly similar to that of cytochrome b5 (Figure 1). Here we describe the three-dimensional structure of At2g24940.1 as determined by NMR spectroscopy. At present, no structure of a MAPR is available from the Protein Data Bank. Thus, the structure of At2g24940.1 may provide clues to the function of a class of steroid binding proteins in plants.

Methods and results

A wheat germ cell-free protein expression method was used in preparing protein samples for NMR spectroscopy. The peU-(N)-His6 vector used in the translation reaction was a generous gift from Professor Endo at Ehime University, Matsuyama, Japan. The wheat germ extract used for *in vitro* translation reactions was purchased from the CellFree Sciences, Ltd. (Yokohama, Japan). The protocols used for small-scale *in vitro* transcription and translation were based on published procedures (Madin et al., 2000; Sawasaki et al., 2002). The peU-(N)-His6-At2g24940.1 plasmid was constructed and verified by restriction digest. The entire At2g24940.1 gene was sequenced to confirm the absence of any spurious mutations in the coding region. The protein sample used for the NMR measurements corresponded to the full-length At2g24940.1 gene product plus an N-temlinal extension, N_{ext}- = MG(H)6LE-, added in the cloning process (109 amino acid residues in total). Protein expression was carried on a small scale (50 μl reaction) to determine the level of protein production and the solubility of the protein construct. The results showed that N_{ext}-At2g24940.1 was produced at ∼0.1–0.2 μg/μl and that the product was >90% soluble (as determined from comparison of SDS-PAGE from the supernatant and pellet). Isotopically labeled samples of the protein, [^{15}N]-N_{ext}-At2g24940.1 and [^{13}C,^{15}N]-N_{ext}-At2g24940.1, were produced individually in larger scale reactions (8 ml). Each of these runs resulted in approximately 2.5 mg of >95% pure protein. Each of the purified proteins appeared as a single band on SDS-PAGE at the expected molecular mass of 12 kDa.

A single [^{13}C,^{15}N]-labeled protein sample at approximately 0.5 mM protein concentration in 90% H_2O/10%D_2O containing 10 mM KH_2PO_4, 50 mM KCl, 0.02% (w/v) sodium azide at pH 7.0 was used in recording the NMR data needed to determine the

*Structure data have been deposited at PDB (1 TOG) and NMR data at BMRB (bmr 6138).

**To whom correspondence should be addressed, E-mail: markley@nmrfam.wisc.edu

```
At2g24940 MEFTAEQLSQYNGTDESKPIYVAIKGRVFDVTTGKSFY-GSGGDYSMFAGKDASRALGKM
RAT_CYB5  -AVTYYRLEEVAKRNTAEETWMVIHGRVYDITRFLSEHPGGEEVLLEQAGADATESFEDV
PIG_MAPR  RDFTPAELRRFDGVQDPR-ILMAINGKVFDVTKGRKFY-GPEGPYGVFAGRDASRGLATF

At2g24940 SKNEE-----------DVSPSLEGLTEKEINTLNDWETKFEAKYPVV---GRVVS-----
RAT_CYB5  GHSPDAREMLKQYYIGDVHPNDLKPKDGDKDPSKNNSCQSSWAYWIVPIVGAILIGFLYR
PIG_MAPR  CLDKEA--------LKDEYDDLSDLTPAQQETLNDWDSQFTFKYHHV---GKLLKEGEEP
```

Figure 1. Multiple-sequence alignment: amino acid sequences of At2g24940.1, rat outer mitochondrial membrane cytochrome b5 (RAT_CYB5), and porcine membrane associated progesterone receptor (PIG_MAPR). Positions conserved in 2 or 3 of these sequences are colored green or cyan, respectively. All of these sequences, with the exception of At2g24940.1, extend beyond the aligned domains shown.

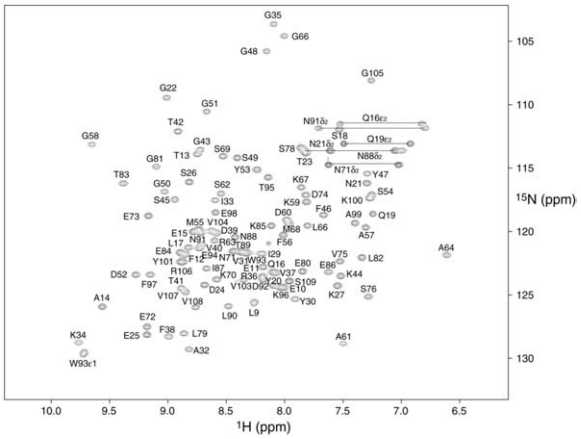

Figure 2. ^{1}H-^{15}N HSQC spectrum (500 MHz, cryogenic probe, 25 °C) of the hypothetical protein At2g24940.1 with peak assignments.

structure. The sample contained ∼1 mM DSS as the internal chemical shift reference.

All NMR spectra were recorded at 25 °C on a 500 MHz Bruker DMX spectrometer equipped with a triple-resonance CryoProbe^{TM} at the National Magnetic Resonance Facility at Madison (NMRFAM). The spectra were processed and analyzed, respectively, with the NMRPipe (Delaglio et al., 1995) and Sparky (http://www.cgLucsf.edu/home/sparky) software packages. NMR data sets collected for backbone assignments were 2D [^{1}H, ^{15}N]-HSQC, 3D HNCACB, 3D HN(CO)CACB, and 3D HNCO. NMR data sets collected for side chain assignments were 2D [^{1}H, ^{13}C]-HSQC, 3D H(CCO)NH, 3D C(CO)NH, and 3D HCCH-TOCSY. Both backbone resonance assignment and side chain assignment of At2g24940.1 were achieved in a semi-automated manner by using the GARANT (Bartels et al., 1996) software program. After the resonance assignment by GARANT and subsequent manual intervention, the backbone assignments (Figure 2) were 99% complete, and the sidechain assignments were ∼90% complete, excluding the added N-terminal expression tag. Assignments of proton resonances from aromatic side chains were based on NOEs between aromatic protons and the $^{1}H^{\beta 2/\beta 3}$ observed in the 3D ^{13}C-edited NOESY-HSQC spectrum. The side chain amide groups of Asn and GIn were assigned on the basis of NOEs between $^{1}H^{\delta(\varepsilon)}$ and the $^{1}H^{\beta 2/\beta 3}$ observed in the 3D ^{15}N-edited NOESY HSQC spectrum. The time-domain NMR data and chemical shift assignments have been deposited in the BioMagResBank database under BMRB accession number 6138.

Structural calculations were performed using the torsion angle dynamics program Cyana (Güntert et al., 1997). Distance constraints were obtained from 3D ^{15}N-edited NOESY-HSQC (τ_{mix} = 125 ms) and 3D ^{13}C-edited NOESY-HSQC (τ_{mix} = 125 ms). Backbone ϕ and ψ constraints and hydrogen bond constraints were generated from assigned chemical shifts by using the programs TALOS (Cornilescu et al., 1999) and CSI (Wishart and Sykes, 1994). Initial NOE assignments and an initial set of structural models were generated by the automated CANDID iterative refinement module of Cyana (Herrmann et al., 2002). Additional NOE assignments were then added, and erroneous ones corrected, through examination of NMR spectra prior to recalculation of structures by Cyana. A total of 2573 NOE distance restraints, 104 backbone ϕ and ψ constraints, and 28 hydrogen bond constraints were used for final calculation of the structure ensemble of At2g24940.1. On the basis of their lowest target function, 20 structures were chosen for further refinement using XPLOR (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 At2g24940.1 (Figure 3) were validated by Procheck-NMR (Laskowski et al., 1996). The statistics for the family of low-target function structures after XPLOR refinement are listed in Table 1. The three-dimensional coordinates for these models have been deposited in the Protein Data Bank (PDB) under the accession number 1TOG.

Table 1. Statistics for the 20 conformers of At2g24940.1 that represent its solution structure

Distance constraints	
Long [(i−j)>5]	537
Medium [1<(i−j)≤5]	687
Sequential [(i−j)=1]	619
Intraresidue [i=j]	768
Dihedral angle constraints (ϕ and ψ)	107
Hydrogen bond constraints	28
Average pairwise RMSD to the mean structure (Å)	
Residues 11–109	
Backbone (C^{α}, C′, N, O)	0.72 ± 0.11
Heavy atoms	1.10 ± 0.11
Regular secondary structure elements	
Backbone (C^{α}, C′, N, O)	0.50 ± 0.11
Heavy atoms	0.89 ± 0.13
Target function ($Å^2$)	0.37 ± 0.12
Violations of distance constraints	
Mean number of violations >0.1 Å	4 ± 2
Sum of violations Å	2.4 ± 0.4
Maximum violation Å	0.18 ± 0.04
Violations of dihedral angle constraints	
Mean number of violations > 2°	1 ± 1
Sum of violations (°)	11.4 ± 3.1
Maximum violation (°)	1.80 ± 0.64
Violations of Van der Waals distances	
Mean number of violations > 0.1 Å	0 ± 0
Sum of violations (Å)	2.1 ± 0.5
Maximum violation (Å)	0.13 ± 0.04
Ramachandran statistics (% of all residues)	
Most favored	84.3
Additionally allowed	14.0
Generously allowed	0.8
Disallowed	0.8

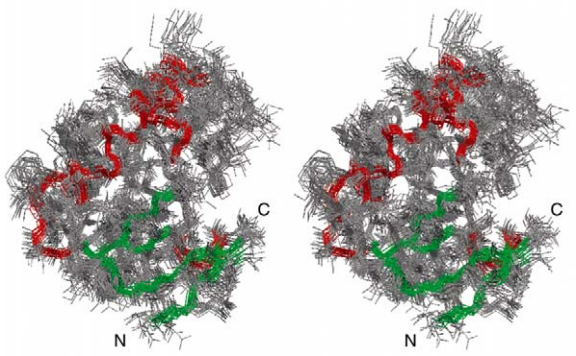

Figure 3. Stereoscopic view of the aligned 20 conformers that represent the structure of At2g24940.1. For clarity, the disordered N-terminal residues (1–9) are not shown. The N- and C-termini of the displayed portion of the protein are labeled. The colors for backbone represent: (red) helices, (green) β-sheets, and (gray) other regions. The superpositions were made using the program MOLMOL (Koradi et al., 1996).

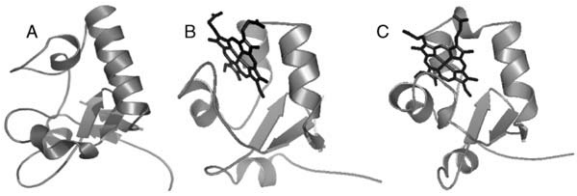

Figure 4. Ribbon diagrams of (A) the NMR structure of At2g24940.1, (B) the X-ray structure of rat outer mitochondrial membrane cytochrome b5 (PDB ID 1B5M), and (C) the X-ray structure of N-terminal domain (residues 3–84) of the chicken sulfite oxidase (1SOX). The iron-bound hemes are shown in black. The ribbon cartoons were generated with PYMOL (http://www.pymol.org).

Discussion

In an effort to establish a procedure for high-throughput NMR structure determinations, the Center for Eukaryotic Structural Genomics has implemented the use of the automatic resonance assignment program GARANT and the automatic NOE assignment program CANDID in CYANA. In the case of At2g24940.1, GARANT provided 75% of the backbone assignments on the basis of initial peak lists generated with the Sparky software. The backbone and sidechain resonance assignments were then completed by hand. With these data as input, the CANDID module of Cyana generated a consistent tertiary fold on the first pass, which served as a good starting point for further resonance assignments and structure refinement. The final ensemble of conformers determined with Cyana; had a backbone r.m.s.d. of 0.72 Å and an average target function of 0.37 $Å^2$, in good agreement with the experimental NMR data. However, since the solution NMR structures often suffer from poor Ramachandran statistics, we used X-PLOR for a final refinement step in which physical force fields and environment solvents were added. Structural validation results from Procheck-NMR indicated that the percentage of residues (residues 11–109) falling into the most favored regions in Ramachandran plot increased from 72.0% to 84.3% as a result of this final refinement.

At2g24940.1 contains two β-sheets and four helices (Figure 3), arranged in a β-α-β-β-α-α-α-β topology. Residues 30–33, 36–39 and 103–106 form a three-stranded antiparallel β-sheet, while the N- and C- termini of the protein (residues 10–13 and 105–108) join together to form a two-stranded parallel β-sheet. Residues 84–98 form a long α-helix arching

over the β-sheets. Helix 2 (residues 41–47) can be described as a short 3–10 helix (residues 41–43) and an α-helix (residues 44–47) separated by a kink. Signals were not observed for the first seven residues (MGHHHH), but the chemical shifts for His8 are close to random values; thus the N-terminal tag (residues 1–9) of At2g24940.1 appears to be dynamically disordered.

The VAST search software (http://www.ncbi.nlm.gov/StructureNAST/vastsearch.html) was used to compare the structure of At2g24940.1 with those of other proteins in the Protein Data Bank. The most similar structures identified by VAST were the N-terminal domain (residues 3–84) of chicken sulfite oxidase (PDB ID 1SOX) and rat outer mitochondrial membrane cytochrome b5 (PDB ID 1B5M). Figure 4 shows a comparison of these two structures with the mean structure determined for At2g24940.1. Interestingly, both the chicken sulfite oxidase and outer mitochondrial membrane cytochrome b5 are heme binding proteins. The level of sequence identity for the regions aligned with At2g24940.1 is 22% for chicken sulfite oxidase and 17% for rat outer membrane mitochondrial cytochrome b5. Structural features common to these three proteins are notable. In chicken sulfite oxidase and outer membrane mitochondrial cytochrome b5, the hydrophobic heme binding pockets are formed by α-helices lying over the β-sheet. Key residues responsible for the heme binding in cytochrome b5 (Rodriguez-Maranon et al., 1996), His 39 and His 63, are replaced by a tyrosine and a lysine respectively in At2g24940.1. Nevertheless, the structure of At2g24940.1 (Figure 3) contains a crevice at the position analogous to the heme binding pocket of cytochrome b5. This crevice, which is formed by helix 2 (residues 41–47), helix 3 (63–67), helix 4 (84 -98), and the loop between helix 2 and helix 3, may be the site at which ligands bind to At2g24940.1.

Although the function of *Arabidopsis thaliana* At2g24940.1 is not known, its sequence similarity to the steroid binding domain of the mammalian MAPR (Figure 1) and its structure, which indicates the presence of a binding crevice, suggests that it may represent a member of a class of plant steroid binding domains. Interestingly, At2g24940.1 lacks the transmembrane domain that is common to mammalian MAPRs. It is hoped that this structure will inspire the functional characterization of this and other members of this family of plant proteins.

Upon completion of this work, the authors discovered that the RIKEN Structural Genomics (Yokohama, Japan) has solved the NMR structure of the same protein (At2g24940.1) but without the N-terminal extension (PDB accession number 1J03). The structure described here superimposes very closely with 1J03, although the MOLMOL (Koradi et al., 1996) representation of 1J03 does not identify the presence of the two-stranded β-sheet formed by the N- and C-terminal segments (described above), which are supported by chemical shift data.

Acknowledgements

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, OIA9977486, BIR-9214394), and the U.S. Department of Agriculture. We thank Prof Yaeta Endo of Ehime University (Matsuyama, Japan) and CellFree Sciences Ltd. (Yokohama Japan) for advice and encouragement in using the wheat germ cell-free system they have developed, Prof Brian F. Volkman for the X-PLOR scripts used, Dr Jurgen Doreleijers for developing the water refinement protocol, and Dr Min Lee for organizational assistance.

References

Altieri, A.S. et al. (1995) *J. Am. Chem. Soc.*, **117**, 7566–7567.
Bartels, C. et al. (1996) *J. Biomol. NMR*, **7**, 207–213.
Bramley, T. (2003) *Reproduction*, **125**, 3–15.
Brünger, A.T. (1992) *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven.
Cornilescu, G. et al. (1999) *J. Biomol. NMR*, **13**, 289–302.
Delaglio, F. et al. (1995) *J. Biomol. NMR*, **6**, 277–293.
Güntert, P. et al. (1997) *J. Mol. Biol.*, **273**, 283–298.
Herrmann, T. et al. (2002) *J. Mol. Biol.*, **319**, 209–227.
Koradi, R. et al. (1996) *J. Mol. Graph.*, **14**, 51–55.
Laskowski, R.A. et al. (1996) *J. Biomol. NMR*, **8**, 477–486.
Li, X. and O'Malley, B.W. (2003) *J. Biol. Chem.*, **278**, 39261–39264.
Madin, K. et al. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 559–564.
Mifsud, W. and Bateman, A. (2002) *Genome Biol.*, **3**, RESEARCHO068 (Epub).
Rodriguez-Maranon, M.J. et al. (1996) *Biochemistry*, **35**, 16378–16390.
Sawasaki, T. et al. (2002) *FEBS Lett.*, **514**, 102–105.
Schwieters, C.D. et al. (2003) *J. Magn. Reson.*, **160**, 65–73.
Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.