Journal of Biomolecular NMR **29:** 387–390, 2004.

Letter to the Editor: Solution structure of a homodimeric hypothetical protein, At5g22580, a structural genomics target from *Arabidopsis thaliana*

Gabriel Cornilescu[a,c,*,**], Claudia C. Cornilescu[a,c,*,**], Qin Zhao[a,c], Ronnie O. Frederick[a,c], Francis C. Peterson[b,c], Sandy Thao[a,c] & John L. Markley[a,c]

[a]*Department of Biochemistry, University of Wisconsin–Madison, Madison, WI 53706-1544;* [b]*Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI 53226;* [c]*Center for Eukaryotic Structural Genomics*

Received 5 December 2003; Accepted 16 January 2004

Key words: At5g22580, CESG, Center for Eukaryotic Structural Genomics, homodimer, residual dipolar couplings

Biological context

The Center for Eukaryotic Structural Genomics (CESG) is a collaborative effort aiming to expand the knowledge of sequence-structure-function relationships by solving a large number of three dimensional protein structures, in a high throughput fashion. CESG's initial target is the genome of the model plant *Arabidopsis thaliana*, whose sequence was completed recently ('The Arabidopsis Initiative', 2000).

Here we present the NMR structure of a homodimeric putative protein with unknown function coded for by gene At5g22580 from *Arabidopsis thaliana* (Figure 1a).

Methods and results

The At5g22580 gene was cloned into pET15b (Novagen, Madison, WI), between the NdeI and BamHI restriction sites. The construct coded for a 19-amino acid N-terminal fusion containing a thrombin-cleavable $(His)_6$ tag and provided for expression under the control of a T7 promoter. The plasmid was transformed into *E. coli* Rosetta(DE3)/pLysS strain (Novagen, Madison, WI). The cells were grown on a minimal medium containing [^{15}N]-ammonium chloride (for ^{15}N-protein) or [^{15}N]-ammonium chloride and [U-^{13}C]-glucose (for the production of ^{15}N;^{13}C-protein) (Q. Zhao et al., *J. Struct. Funct. Genom.*, accepted). The harvested cells were disrupted using BugBuster HT (Novagen), and the protein was purified by immobilized metal affinity chromatography (IMAC) on a Ni-NTA column from Qiagen (Valencia, CA). The N-terminal His-Tag was cleaved with biotinylated thrombin (Novagen).

All isotropic NMR data were acquired at 25 °C on a 280 µl sample at pH 6.5 containing 1.0 mM ^{13}C;^{15}N-protein, 50 mM NaPi, and 0.1 mM NaN_3 in 93% H_2O and 7% ^2H_2O. The sample used for residual dipolar coupling measurements contained 0.3 mM ^{13}C;^{15}N-protein in the same solvent system, but with the addition of 5% $C_{12}E_5$ polyethylene glycol (Fluka)/hexanol mixture (0.96 surfactant/alcohol molar ratio) (Ruckert and Otting, 2000).

[^{1}H-^{15}N] HSQC, HNCO, HNCACB, CBCA(CO)NH, HBHACONH, CCONH, HCCONH, 3D ^{15}N-edited NOESY (t_{mix} = 120 ms) and 3D ^{13}C-edited NOESY (t_{mix} = 150 ms) spectra were acquired on Varian INOVA 600 and 800 MHz, Bruker AVANCE 500 and 750 MHz spectrometers. Data were processed using NMRPipe package (Delaglio et al., 1995).

$^1D_{NH}$ residual dipolar couplings (RDCs) were measured at 35 °C for 105 out of 111 residues from a 2D IPAP ^{15}N-^{1}H HSQC (Ottiger et al., 1998). The magnitude (A_a = 13.2) and rhombicity (η = 0.53) of the alignment tensor were obtained initially from a powder pattern distribution of the $^1D_{NH}$ data (Clore et al., 1998) and subsequently adjusted to values obtained by singular value decomposition fit to the lowest energy monomeric refined structures using the program PALES (Zweckstetter and Bax, 2000). Attempts to measure scalar couplings through hydrogen bonds (Cornilescu et al., 1999b) for this sample failed, as would be consistent with ^{15}N T_2 relaxation times shorter than 80 ms. Subsequent ^{15}N T_2 measurement (data not shown) confirmed this and yielded uniform values around 55 ms for the rigid part of the molecule (residues 6 to 105). These results suggested that the protein might not be monomeric in solution. A dif-

*These authors have contributed equally.
**To whom correspondence should be addressed. E-mails: gabrielc@nmrfam.wisc.edu; cclaudia@nmrfam.wisc.edu

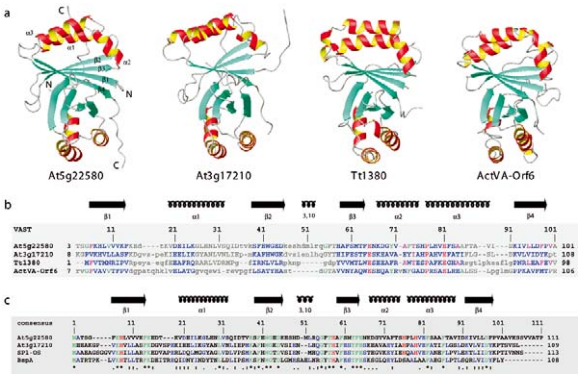

Figure 1. (a) Comparison of ribbon diagrams for At5g22580, its sequence homolog At3g17210 (39%), and two other proteins with similar structures but low sequence identities, Tt13180 (14%) and ActVA-Orf6 (15%). (b) The VAST domain alignment of the above structures. (c) Multiple alignment of the sequences of At5g22580 and At3g17210 (both from *Arabidopsis thaliana*), SP1-OS (stress response, SP1-like protein from *Oryza sativa*), and BspA (boiling stable protein A from *Populus tremula*).

fusion constant measurement (0.12×10^{-5} cm^2 s^{-1}) indicated a molecular size larger than monomeric but smaller than dimeric. A few NOEs proved impossible to assign as monomeric contacts but could be rationalized as intersubunit contacts (Figure 2a). Thus the structure was refined as a symmetrical homodimer. We compared the line widths of the [^{1}H-^{15}N] HSQC cross peaks for several increasing dilution samples up to 0.1 mM and noticed no significant changes, concluding that the monomer-dimer equilibrium doesn't shift to the monomeric form at lower NMR concentrations.

Automatic resonance assignments for 90% of the backbone were obtained using the program Garant (Bartels et al., 1997). Because inspection showed that most of side chain assignments provided by Garant were incorrect, the side chain resonances were assigned manually by means of PIPP/STAPP (Garrett et al., 1991) software.

The TALOS program (Cornilescu et al., 1999a) was used to provide 85 pairs of ϕ/ψ backbone torsion angle restraints and to clearly identify the secondary structural elements (confirmed by local NOEs, Figure 3c). The mobility of residues 1 to 5 and 104-111 (suggested by the T_2 values) was confirmed by their backbone chemical shifts appearing close to their random coil values in the TALOS graphical output.

Hydrogen bond restraints were inferred initially for α-helices and only later for β-strands, when the level of structural refinement allowed their unambiguous alignment within the β-sheet. Two distance restraints of 1.9 Å and 2.9 Å per involved pair of residues were

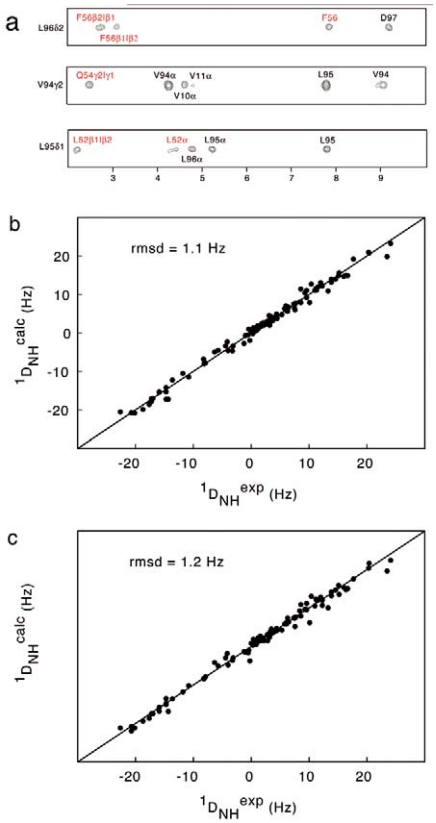

Figure 2. NMR results for the At5g22580 homodimer. (a) Strips from a 3D 13C NOESY spectrum recorded at 750 MHz illustrating intersubunit (red) and intrasubunit (black) NOEs. (b) Correlation between 105 measured $^1D_{NH}$ residual dipolar couplings and those calculated from the lowest energy monomeric structure. (c) Similar correlation between the measured residual dipolar couplings and those calculated from the lowest energy dimeric structure.

used to denote hydrogen bonds for H^N-O and N-O, respectively (Wüthrich, 1986).

2027 approximate interproton distances (including 31 dimeric) were obtained from cross-peak intensities in the two 3D NOESY spectra (Figure 2a). Peak intensities were converted into a continuous distribution of interproton distance restraints, with a uniform 40% distance error applied to take into account spin diffusion in both NOE experiments.

We attempted to solve the structure by automatic NOE assignments using the automatic CANDID iterative protocol of CYANA (Herrmann et al., 2002), but the resulting structures were of poor quality as there was no apriori distinction between intra and inter-monomeric contacts.

At5g22580 adopts a βαβ(3,10)βααβ fold, with the short 3,10 helix (residues 51-53) not being recognized

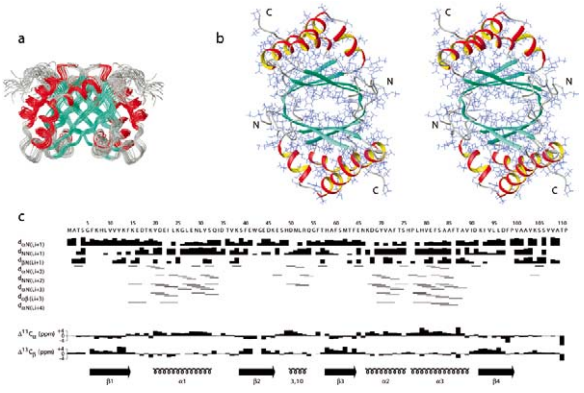

Figure 3. (a) Ensemble of the 20 lowest energy backbone structures of At5g22580 with helices shown in red and β-strands in cyan. (b) Stereo view of the lowest energy structure as a ribbon diagram with side chain heavy atoms shown in blue. (c) Summary of secondary structure elements, local NOE connectivities and C^{α}/C^{β} chemical shifts *versus* the amino acid sequence of At5g22580.

by the MOLMOL (Koradi et al., 1996) criteria in most of the final refined structures (Figure 3a).

All monomeric constraints were symmetrically duplicated for dimer refinement. Structure calculations were carried out using the torsion angle molecular dynamics and the internal variable dynamics module (Schwieters and Clore, 2001) of Xplor-NIH (Schwieters et al., 2003), to ensure preservation of the correct peptide geometry when applying RDC and distance constraints simultaneously. The X-PLOR non-crystallographic symmetry potential term was included to maintain identical structure of the monomeric subunits.

The coordinates of the lowest energy monomeric structure were docked by a rigid body minimization protocol (Clore, 2000). In this approach we readily obtained a dimeric model with the correct relative orientation of its monomeric subunits (confirmed in later refinement stages, vide infra), using 72 (symmetrically duplicated) ambiguous inter-subunit NOE assignments, radius of gyration and RDC constraints.

Discussion and conclusions

The two 3D NOE datasets provided 1996 assigned monomeric NOE constraints. The lack of aromatic side chain assignments and NOE contacts between them was compensated by the long-range orientational information provided by the $^{1}D_{NH}$ RDC constraints. The RDCs were also used for docking the two monomeric subunits into the final symmetric dimer structure.

Table 1. Structural statistics

Number of experimental distance restraints[a]	2117	
intraresidue	790	
sequential ($\|i-j\|=1$)	598	
short range ($1<\|i-j\|=5$)	315	
long range ($\|i-j\|>5$)	293	
hydrogen bonds	90	
intermolecular	31	

	R.m.s. deviations			
	⟨Sa⟩[b] monomer	Lowest energy monomer	⟨Sa⟩[b] dimer	Lowest energy dimer
Predicted dihedral restraints[c] (°) (170)	2.90 ± 0.36	2.55	2.77 ± 0.25	2.76
Residual dipolar couplings $^{1}D_{NH}$ (Hz) (105)	1.14 ± 0.09	1.08	1.18 ± 0.07	1.17
Deviations from idealized covalent geometry				
bonds (Å) (1759)	0.00245 ± 0.00003	0.00248	0.00252 ± 0.00005	0.00253
angles (°) (3186)	0.490 ± 0.002	0.489	0.492 ± 0.002	0.489
impropers (°) (919)	0.70 ± 0.04	0.63	0.78 ± 0.06	0.74
Measures of structure quality				
Lennard-Jones energy[d] (kcal mol^{-1})	−362 ± 55	−413	−829 ± 119	−946
Ramachandran analysis[e]				
most favored regions (%)	80.9	82.5	82.2	83.8
allowed regions (%)	18.9	17.5	17.6	16.2
disallowed regions (%)	0.2	0	0.2	0
Coordinate precision[f] Å				
backbone (N, C^{α}, C′)	0.61 ± 0.15		0.89 ± 0.28	
all non-hydrogen atoms	1.25 ± 0.12		1.38 ± 0.23	

The statistics are for the 20 structures (of 100 calculated) with the overall lowest energies.
[a]Per monomeric unit.
[b]⟨SA⟩ represents the 20 lowest-energy structures. For ⟨SA⟩, the values shown are mean ± standard deviation with the number of restraints used to calculate these values shown in parentheses.
[c]Error range used for the ϕ/ψ constraints is twice the standard deviation of the TALOS predicted angles.
[d]The van der Waals energies described by Lennard–Jones potentials were not incorporated into the simulated annealing calculation.
[e]Calculated for residues 6–45 and 58–102.
[f]Average rmsd of the 20 lowest-energy structures from the mean coordinates for residues 6–45 and 58–102.

The initial monomeric refinement resulted in a subset of 20 lowest energy structures (out of 100) with an rmsd from the average of 0.61 Å (Table 1), excepting the disordered termini and a poorly defined loop containing the short 3,10 helix. This loop, connecting strands β2 and β3 (referred throughout as the '46–57 loop'), has a lower number of assigned monomeric NOE contacts but is part of the dimer interface and was expected to be better defined following the dimer refinement.

We initially attempted refining the dimer structure with RDCs by fixing the backbone of one monomer, except for the 46–57 loop and all noninterfacial side chain atoms, and allowing rotational and translational degrees of freedom to the other (with similar rigidization of backbone and noninterfacial side chain atoms). This resulted in a small set of outliers in the correlation between experimental and calculated $^{1}D_{NH}$ RDCs for the lowest energy calculated structure (data not shown). All outliers corresponded to residues within the 46–57 loop. Refinement testing with different sets of noninterfacial backbone residues kept rigid showed slightly different conformations of the 46–57 β-loop involving small rearrangements of 4–5 residues at each

end of the two β-strands connected by the loop. Therefore we decided not to restrict any subset of atoms and used both monomeric and dimeric restraints simultaneously in a full dimeric refinement. Neither the final set of calculated structures nor the subset of lowest energy structures (i.e., 20 out of 100) showed any consistent (i.e. in more than 35% of the calculated structures) NOE violations larger than 0.5 Å as a result of slight rearrangements of backbone and side chain atoms. The dimer interface is stabilized mainly by hydrophobic interactions between the monomeric β-sheets associated in the dimeric β-barrel.

The similar values of the Q factors (Cornilescu et al., 1998) between measured and calculated $^1D_{NH}$ RDCs, both in the initial monomer refinement, $Q = 0.081$, (Figure 2b) and in the final dimeric structures, $Q = 0.088$, (Figure 2c) confirmed that the two mainly rigid monomeric backbone structures are accurately oriented relative to one other in the homodimer.

At5g22580 has a 39% sequence identity and a high structural similarity with another *Arabidopsis* CESG target, At3g17210, with unknown function, whose structure was solved independently by NMR spectroscopy, PDB accession 1Q53 (B.L. Lytle et al. *J. Biomol. NMR*, in press) and X-ray crystallography, PDB accession 1Q4R (G.N. Phillips Jr. et al., manuscript in preparation). The minor structural differences are located primarily in the loop connecting strands β2 and β3 and in the overall shape of the dimeric β-barrel.

A VAST search (Gibrat et al., 1996) (Figure 1b) revealed a high similarity between the structure of At5g22580 and that of ActVA-Orf6, a bacterial monooxygenase from *Streptomyces coelicolor* (Sciara et al., 2003), which apparently does not require a metal ion or a prosthetic group for catalysis. The conserved active sites residues of ActVA-Orf6 are not conserved in At5g22580, with the exception Tyr72 (aligned to Tyr70). The topology of the active site of ActVA-Orf6 is reproduced in At5g22580, with a cavity created by the β-sheet and the two sequentially connected α-helices. Tyr70 is found in a similar position in the cavities of the two proteins and is surrounded by the side chains of His8, Asp97 (two residues that are conserved among several homologous *Arabidopsis* proteins), Lys67, Thr74 and His79; these residues may play a functional role in interacting with a putative ligand hosted by the deep cavity.

The recently released structure of the Tt1380 protein from *Thermus thermophilus* (Tt1380, PDB accession 1IUJ), is also similar to that of At5g22580. The X-ray coordinates of Tt1380 show a Zn atom in each cavity of the dimer coordinated to a His residue. That histidine is replaced by Ser86 in the structurally aligned At5g22580 sequence; however, residues His8 and His79 in At5g22580 are oriented toward the region occupied by the Zn inside the cleft formed by the β-sheet and the two sequentially connected α-helices.

As shown in Figure 1c, At5g22580 is homologous to At3g17210 from *Arabidopsis thaliana* (39% sequence identity) and to the stress-responsive proteins from *Oryza sativa* (SP1-OS, 35%) and *Populus tremula* (BspA, 30%). These four proteins contain several conserved residues in the regions that correspond to strand β3 and helices α2 and α3, which form a cleft in both *Arabidopsis* proteins. Since this region corresponds to the active site of the bacterial monooxygenase, we speculate that these residues may be part of the active sites of the plant proteins.

The coordinates and NMR structural restraints have been deposited in the PDB (accession 1RJJ), and the NMR assignments and time-domain data have been deposited in the BMRB (accession 6011).

Acknowledgements

The authors thank Marius Clore, Brian Volkman and Jurgen F. Doreleijers for insightful discussions and suggestions. We acknowledge Adrian Hegeman for the mass spectra of At5g22580. Work was supported by NIH CESG Grant P50 GM64598 (J.L. Markley, PI).

References

Bartels, C. et al. (1997) *J. Comput. Chem.*, **18**, 139–149.
Clore, G.M. (2000) *Proc. Natl. Acad. Sci.*, **97**, 9021–9025.
Clore, G.M. et al. (1998) *J. Magn. Reson.*, **133**, 216–221.
Cornilescu, G. et al. (1999a) *J. Biomol. NMR*, **13**, 289–302.
Cornilescu, G. et al. (1999b) *J. Am. Chem. Soc.*, **121**, 2949–2950.
Cornilescu, G. et al. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.
Delaglio, F. et al. (1995) *J. Biomol. NMR*, **6**, 277–293.
Garrett, D.S. et al. (1991) *J. Magn. Reson.*, **95**, 214–220.
Gibrat, J.F. et al. (1996) *Curr. Opin. Struct. Biol.*, **6**, 377–385.
Herrmann, T. et al. (2002) *J. Mol. Biol.*, **319**, 209–227.
Koradi, R. et al. (1996) *J. Mol. Graph.*, **14**, 51–55.
Ottiger, M. et al. (1998) *J. Magn. Reson.*, **131**, 373–378.
Ruckert, M. and Otting, G. (2000) *J. Am. Chem. Soc.*, **122**, 7793–7797.
Schwieters, C.D. and Clore, G.M. (2001) *J. Magn. Reson.*, **152**, 288–302.
Schwieters, C.D. et al. (2003) *J. Magn. Reson.*, **160**, 65–73.
Sciara, G. et al. (2003) *EMBO J.*, **22**, 205–215.
'The Arabidopsis Initiative' (2000) *Nature*, **408**, 796–815.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley Interscience, New York, NY.
Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.