PROTEIN STRUCTURE REPORT

Crystal structure of AGR_C_4470p from *Agrobacterium tumefaciens*

SERGEY M. VOROBIEV,¹ HELEN NEELY,¹ JAYARAMAN SEETHARAMAN,¹
LI-CHUNG MA,² RONG XIAO,² THOMAS B. ACTON,² GAETANO T. MONTELLIONE,² AND
LIANG TONG¹

¹Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, New York 10027, USA

²Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Department of Biochemistry, Robert Wood Johnson Medical School, Northeast Structural Genomics Consortium, Piscataway, New Jersey 08854, USA

(RECEIVED November 10, 2006; FINAL REVISION December 7, 2006; ACCEPTED December 12, 2006)

Abstract

We report here the crystal structure at 2.0 Å resolution of the AGR_C_4470p protein from the Gram-negative bacterium *Agrobacterium tumefaciens*. The protein is a tightly associated dimer, each subunit of which bears strong structural homology with the two domains of the heme utilization protein ChuS from *Escherichia coli* and HemS from *Yersinia enterocolitica*. Remarkably, the organization of the AGR_C_4470p dimer is the same as that of the two domains in ChuS and HemS, providing structural evidence that these two proteins evolved by gene duplication. However, the binding site for heme, while conserved in HemS and ChuS, is not conserved in AGR_C_4470p, suggesting that it probably has a different function. This is supported by the presence of two homologs of AGR_C_4470p in *E. coli*, in addition to the ChuS protein.

Keywords: protein structure; structural genomics; heme utilization enzyme; HemS; ChuS; molecular evolution; gene duplication

AGR_C_4470p from *Agrobacterium tumefaciens*, the causative agent of crown gall disease in plants and a widely used tool for tDNA transfer to plant cells, is a hypothetical protein with molecular mass of 20 kD. It shares 38% amino acid identity with Q74XI2_YERPE from *Yersinia pestis*, and a putative heme iron utilization function for AGR_C_4470p was suggested in the SwissProt database.

Recently, two crystal structures of heme utilization proteins were reported: heme oxygenase ChuS from

Escherichia coli (Suits et al. 2005, 2006) and heme transport protein HemS from *Yersinia enterocolitica* (apo- and heme-bound forms) (Schneider et al. 2006). Both proteins are about twice the size of AGR_C_4470p (molecular weight of about 39 kDa). They contain two domains with the same fold. For ChuS, the two domains could be superimposed with a root-mean-square deviation of 2.1 Å, but the sequence identity between them is only 19% (Suits et al. 2005).

Results and Discussion

The structure of AGR_C_4470p contains a central, six-stranded anti-parallel β-sheet that is flanked by α-helices (Fig. 1A). In addition, there is a three-stranded anti-parallel β-sheet on the surface of the monomer, which

Reprint requests to: Liang Tong, Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, USA; e-mail: ltong@columbia.edu; fax: (212) 865-8246.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062663307>.

Figure 1. Structure of AGR_C_4470p. (A). Schematic drawing of the structure of the dimer of AGR_C_4470p from *Agrobacterium tumefaciens*. The two monomers are colored in yellow and magenta. (B). Overlay of the structure of AGR_C_4470p dimer (in yellow and magenta) with that of *E. coli* ChuS (in slate). The arrow points to the long loop connecting the two domains of ChuS. (C). Overlay of the structure of AGR_C_4470p dimer (in yellow and magenta) with that of *Y. enterocolitica* HemS (in green). The heme is shown as a stick model. (D) A close up of the heme binding site in *Y. enterocolitica* HemS (in green) and comparison with the equivalent region in AGR_C_4470p (in magenta). The proximal ligand in HemS does not have a counterpart in AGR_C_4470p.

mediates the formation of a (crystallographic) dimer (Fig. 1A). The third strand of this β -sheet has a highly irregular structure and is placed next to the central β -sheet of the other monomer in the dimer (Fig. 1A). This extends the central β -sheet to nine strands, and the two β -sheets bury 990 Å² of surface area in the dimer. Our light-scattering studies show that this protein is mostly dimeric in solution (unpublished data). The monomer also contains a long helix at the N terminus, which extends away from the rest of the protein (Fig. 1A). The conformation of this helix is stabilized by crystal packing.

Structures that are most similar to AGR_C_4470p include the two domains in the structures of ChuS and HemS, identified with the program Dali (Holm and Sander 1993). AGR_C_4470p shares only ~15% overall amino acid sequence identity with these domains, and most of the conserved residues are in the hydrophobic core of the structures. More remarkably, the two monomers of the AGR_C_4470p dimer are arranged in the same way as the two domains in the structures of ChuS (Fig. 1B) and HemS (Fig. 1C). This provides direct structural evidence that ChuS and HemS have evolved by gene duplication, and AGR_C_4470p is an example of

the ancestral single-domain protein. The C terminus of one monomer and the N terminus of the other monomer are far apart in the AGR_C_4470p dimer (Fig. 1A), and both ChuS and HemS contain long linkers between their two domains (Fig. 1B,C).

In the structure of HemS in complex with heme, the heme is tightly clamped between His196 and Phe199 on the proximal side and Arg102, Phe246, and Leu94 on the distal side (Fig. 1D; Schneider et al. 2006). The binding site is also lined by several hydrophobic residues, including Met244, Val253, and Ile255 (Fig. 1D). All of these residues are conserved in ChuS, and ChuS can also bind heme in this pocket (Suits et al. 2006). This is consistent with their possible roles in heme utilization (Suits et al. 2005; Schneider et al. 2006).

In contrast, this heme binding site is not conserved in AGR_C_4470p. The proximal ligand in HemS, His196, does not have a counterpart in AGR_C_4470p (Fig. 1D), and the α -helix embracing the proximal side of the heme is partly disordered and is positioned away from the potential heme-binding pocket in AGR_C_4470p (Fig. 1A). In addition, Leu94, Arg102, and Phe199 have no equivalents in the structure of AGR_C_4470p. On

the other hand, the hydrophobic lining of the pocket is partly conserved: Met244 and Val253 of HemS are equivalent to Leu86 and Val95, although Ile255 of HemS is equivalent to Glu97 in AGR_C_4470p (Fig. 1D). Based on these structural and sequence analyses, it is unlikely that AGR_C_4470p is involved in heme utilization.

AGR_C_4470p belongs to Pfam DUF1008, which includes 38 proteins of unknown function. A sequence alignment of five members of this family is shown in Figure 2. Interestingly, *Escherichia coli* contains two homologs of AGR_C_4470p (one of which is shown in Fig. 2), in addition to ChuS. Residues that line the pocket identified in the structure of HemS are generally conserved among these homologs (Fig. 2), suggesting that this pocket may also have an important role in the functions of these proteins, possibly in the binding of another ligand.

Materials and methods

Expression and purification of *Agrobacterium tumefaciens* hypothetical protein AGR_C_4470p (Northeast Structural Genomics Consortium [NESG] target Atr92) was carried out as a part of the established high-throughput protein production pipeline (Acton et al. 2005). Briefly, the AGR_C_4470p gene was cloned into the pET21 expression vector (Novagen) with a C-terminal hexa-histidine sequence (LEHHHHHH). Selenomethionyl proteins were expressed in *Escherichia coli* BL21(DE3) + Magic, purified using Ni-NTA affinity column (Qiagen) and preparative gel filtration (Superdex 75, GE Healthcare) in a buffer containing 10 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT. The protein's homogeneity was

verified by SDS-PAGE, and the mass was validated by MALDI-TOF mass spectrometry. Purified AGR_C_4470p was concentrated to 12 mg/mL, flash frozen, and stored at -80°C until crystallization.

Crystallization screening was performed using the hanging-drop vapor diffusion method at 18°C . After optimization, AGR_C_4470p crystals useful for structure determination were grown over a reservoir solution containing 50 mM MES (pH 6.0), 50 mM magnesium sulfate, 20% (v/v) PEG400, and Al's oil placed above the reservoir solution to slow water diffusion (Chayen 1997). The crystals were soaked in cryoprotectant containing 100 mM MES (pH 6.0), 100 mM magnesium sulfate, 40% (v/v) PEG400, and 10% (v/v) ethylene glycol, and frozen in liquid propane for data collection at 100 K.

A selenomethionyl MAD data set was collected at beamline X4A at the National Synchrotron Light Source. The diffraction data were processed with the HKL2000 package (Otwinowski and Minor 1997). The crystal belongs to space group *I*222, with cell parameters of $a = 66.2$, $b = 72.3$, and $c = 104.7$ Å. There is one molecule in the crystallographic asymmetric unit.

The programs SHELXE/D (Schneider and Sheldrick 2002) and SOLVE (Terwilliger 2003) were used to locate four selenium sites and to calculate phases to 2.4 Å resolution. Solvent-flattening calculations and partial model building were performed using RESOLVE (Terwilliger 2003), which located 140 residues (72%) in the automated mode. The rest of the model was built manually by using COOT (Emsley and Cowtan 2004) and was refined against 2.0 Å data with the program CNS (Brunger et al. 1998). Refinement statistics are presented in Table 1. The quality of the model was inspected by the programs PROCHECK (Laskowski et al. 1993) and MAGE (Word et al. 1999). The figures were created using the program PyMOL (DeLano 2002). The atomic coordinates and structure factors for AGR_C_4470p have been deposited in the Protein Data Bank, with the accession code 2HQV.

Figure 2. Sequence alignment of AGR_C_4470p and four homologs. The secondary structure elements are shown above the alignment. Strictly conserved and conservatively substituted residues are highlighted in red, and residues equivalent to those in the binding pocket for heme in HemS are indicated with the green triangles. The sequences shown include AGR_C_4470p from *A. tumefaciens*, ChuX from *E. coli*, ShuX from *Shigella dysenteriae*, and putative heme iron utilization proteins from *Yersinia pestis* and *Vibrio vulnificus*.

Table 1. Summary of crystallographic information

Data collection	
Wavelength (Å) ^a	0.9792
Resolution (Å)	30–2.0
Number of observations	639,506
Redundancy ^b	7.5 (6.5)
Completeness	99.9 (99.9)
<i>I</i> / σ (<i>I</i>)	28.5 (3.4)
<i>R</i> _{merge} (%)	6.7 (54.4)
MAD analysis	
Resolution (Å)	20–2.4
Number of Se sites	4
Figure-of-merit from MAD phasing	0.62
Refinement	
Resolution (Å)	30–2.0
Number of reflections	29,683
Number of protein atoms	1266
Number of water molecules	67
<i>R</i> _{work} (%)	24.1
<i>R</i> _{free} (%) ^c	26.9
RMSD in bond lengths (Å)	0.007
RMSD in bond angles (°)	1.3
Average <i>B</i> factor (Å ²)	42
Ramachandran plot	
Most favored regions (%)	89.0
Additional allowed regions (%)	11.0
Generously allowed regions (%)	0.0
Disallowed regions (%)	0.0

^aOnly the data set at the peak wavelength is shown here.

^bThe numbers in parentheses are for the high resolution shell (2.07–2.00 Å).

^c*R*_{free} is calculated for 4.0% of randomly selected reflections excluded from refinement.

Acknowledgments

We thank Randy Abramowitz and John Schwanof for access to the X4A beamline at NSLS and G. DeTitta of Hauptman Woodward Research Institute for crystallization screening.

This research was supported by grants from the Protein Structure Initiative of the National Institutes of Health (U54 GM074958).

References

- Acton, T.B., Gunsalus, K., Xiao, R., Ma, L., Aramini, J., Baron, M.C., Chiang, Y., Clement, T., Cooper, B., Denissova, N., et al. 2005. Robotic cloning and protein production platform of the Northeast Structural Genomics Consortium. *Methods Enzymol.* **394**: 210–243.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**: 905–921.
- Chayen, N.E. 1997. A novel technique to control the rate of vapor diffusion, giving larger protein crystals. *J. Appl. Crystallogr.* **30**: 198–202.
- DeLano, W.L. 2002. *The PyMol molecular graphics system*. DeLano Scientific, San Carlos, CA.
- Emsley, P. and Cowtan, K.D. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**: 2126–2132.
- Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**: 123–138.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. Procheck—A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–291.
- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Schneider, T.R. and Sheldrick, G.M. 2002. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**: 1772–1779.
- Schneider, S., Sharp, K.H., and Paoli, M. 2006. An induced fit conformational change underlies the binding mechanism of the heme-transport proteobacteria-protein HemS. *J. Biol. Chem.* **281**: 32606–32610.
- Suits, M.D., Pal, G.P., Nakatsu, K., Matte, A., Cygler, M., and Jia, Z. 2005. Identification of an *Escherichia coli* O157:H7 heme oxygenase with tandem functional repeats. *Proc. Natl. Acad. Sci.* **102**: 16955–16960.
- Suits, M.D., Jaffer, N., and Jia, Z. 2006. Structure of the *Escherichia coli* O157:H7 heme oxygenase ChuS in complex with heme and enzymatic inactivation by mutation of the heme coordinating residue His-193. *J. Biol. Chem.* **281**: 36776–36782.
- Terwilliger, T.C. 2003. SOLVE and RESOLVE: Automated structure solution and density modification. *Methods Enzymol.* **374**: 22–37.
- Word, J.M., Lovell, S.C., LaBean, T.H., Taylor, H.C., Zalis, M.E., Presley, B.K., Richardson, J.S., and Richardson, D.C. 1999. Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285**: 1711–1733.