Journal of Biomolecular NMR **30:** 107–108, 2004.

Letter to the Editor: ^{1}H, ^{13}C and ^{15}N assignments for the *Archaeglobus fulgidis* protein AF2095

Robert Powers[a,*], Thomas B. Acton[b], Yiwen Chiang[b], P.K. Rajan[b], John R. Cort[c], Michael A. Kennedy[c], Jinfeng Liu[d], LiChung Ma[b], Burkhard Rost[d] & Gaetano T. Montelione[b,e]

[a]*Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, U.S.A.;* [b]*Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, U.S.A.;* [c]*Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, DC 99352, U.S.A.;* [d]*Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, U.S.A.;* [e]*Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854, U.S.A.*

Received 18 March 2004; Accepted 4 May 2004

Key words: Archaeglobus fulgidis, NMR, protein AF2095, resonance assignments, secondary structure

Biological context

Structural genomics is providing a means to determine the molecular and cellular function for the vast amount of proteins in the Human proteome that lack any explicit experimental information by characterizing the complete range of protein folds (Montelione, 2001). The Northeast Structural Genomics Consortium (NESG; www.nesg.org) is a pilot project funded by the National Institutes of Health Protein Structure Initiative, focusing on proteins from eukaryotic model organisms including humans. The thermophillic archaea *Archaeglobus fulgidis* AF2095 protein is an example of a protein of unknown biological function targeted for structural analysis by NESG. AF2095 belongs to the Pfam family PF01981 – UPF0099, protein domain family of unknown function that has been found in yeast, archaebacteria and eubacteria. AF2095 has been assigned to NESG Cluster ID:17431, a set of fourteen proteins with high (>∼30%) sequence identity with human, *Drosophila*, *Caenorhabditis elegans*, *Arabidopsis*, yeast, archaeal and eubacterial origin (Liu, 2004). A total of fifty-six proteins are identified when the analysis is expanded to include all available genomes, where determining the NMR solution structure of AF2095 can be leveraged to infer 3D structural information for these proteins. Here we report the near complete ^{1}H, ^{15}N, and ^{13}C NMR assignments and secondary structure of AF2095. These data provide a basis for determining the solution structure of AF2095, for further investigation of the function of this protein and for providing representative structural and functional information for the protein domain family that includes AF2095.

*To whom correspondence should be addressed. E-mail: rpowers3@unl.edu

Methods and experiments

Uniformly ^{13}C,^{15}N-enriched AF2095 (123 amino acids) was cloned, expressed and purified following standard protocols used in the NESG consortium. Briefly, the full length gene (YK95_ARCFU) from *Archaeglobus fulgidis* was cloned into a pET21d (Novagen) derivative, yielding the plasmid pGR4-21. The resulting AF2095 open reading frame contains eight nonnative residues at the C-terminus (LEHHHHHH) of the protein. *Escherichia coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pGR4-21, and cultured in MJ9 minimal medium (Jansson et al., 1996) containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. Initial growth was carried out at 37 °C until the OD_{600} of the culture reached ∼0.8 units. The incubation temperature was then decreased to 17 °C and protein expression was induced by the addition of IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 1 mM. Following overnight incubation at 17 °C, the cells were harvested by centrifugation and lysed by sonication. *U*-^{13}C,^{15}N AF2095 was purified in a two step protocol consisting of Ni-NTA affinity column (Qiagen) and gel filtration column (HiLoad 26/60 Superdex 75 pg, Amersham Biosciences) chromatography. The final yield of pure $U-^{13}$C,^{15}N AF2095 (> 97% by SDS-PAGE; 13.5 KDa by MALDI-TOF mass spectrometry) was approximately 68 mg/l. Samples of *U*-^{13}C,^{15}N AF2095 for NMR spectroscopy were prepared at a protein concentration of 1.0 mM in 95% H_2O/5%

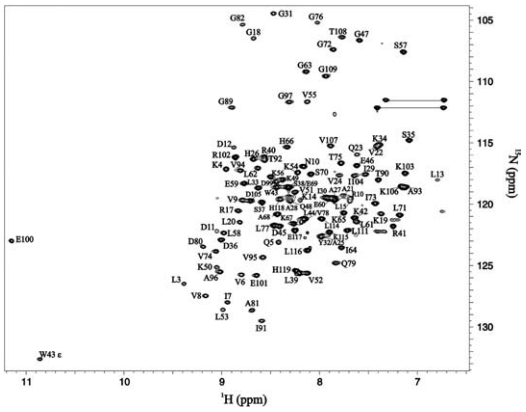

Figure 1. Two-dimensional ^{15}N-HSQC spectrum of AF2095 at 313 K and pH 6.5. The backbone amide and Trp side-chain resonances are assigned. The side-chain amide groups of Asn and Gln are connected by horizontal lines.

D_2O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3 at pH 6.5.

All NMR spectra were recorded at 40 °C on Bruker and Varian 500, 600, 750, and 800 MHz NMR systems. Spectra were processed using the NMRPipe software package (Delaglio et al., 1995) and analyzed with PIPP (Garrett et al., 1991) and AutoAssign (Zimmerman et al., 1997). The assignments of the ^{1}H, ^{15}N and ^{13}C resonances were based on the following experiments: CBCA(CO)NH, CBCANH, C(CO)NH, HC(CO)NH, HNCO, HNHA, HNCA, HCCH-COSY and HCCH-TOCSY (Clore and Gronenborn, 1998; Ferentz and Wagner, 2000).

The secondary structure of AF2095 was based on characteristic NOE data involving the HN, Hα and Hβ protons, $^{3}J_{HN\alpha}$ coupling constants, slowly exchanging HN protons and ^{13}Cα and ^{13}Cβ secondary chemical shifts (Wishart and Sykes, 1994; Wüthrich, 1986). The AF2095 NMR structure is composed of three helical regions corresponding to residues 17-45 (α_1); 58-70 and (α_2) and 100-109 (α_3); and a four stranded β-sheet region corresponds to residues 3-11 (β_1); 50-55 (β_2); 75-80 (β_3) and 93-99 (β_4).

Extent of assignments and data deposition

Extensive assignments for the backbone and side-chain resonances of AF2095 were obtained (208/220 HN-N, 109/115 Cα, 104/114 C′, 99/105 Cβ, excluding the LEHHHHHH purification tag) (Figure 1). Initial NMR sample preparations for AF2095 suggested partial aggregation of the protein as evident by relatively low spectral signal-to-noise and a significant number of missing NMR resonance assignments. Optimizing the NMR sample conditions, particularly including data collection at elevated temperature (40 °C) significantly improved the quality of the data enabling the completion of the assignments. This higher temperature is consistent with the extreme thermophillic nature of *A. fulgidis* and is probably closer to the functional conditions for which this protein has evolved. The amino-acid composition of AF2095 (19 L, 14 V, 7 I and 7 A) also complicated the analysis because of the relatively poor dispersion in the methyl region. Compounding this problem is the structural disorder observed for the 12 C-terminal residues, including the hexaHis purification tag (data not shown). Additionally, residues L83 to V86 and K14 to L15 are partially or completely unassigned presumably due to loop mobility since these residues are located between secondary structure elements. Unusually shifted resonance assignments for K49 and P98 were observed presumably due to a close interaction with W43 based on preliminary structural analysis. A table of chemical shifts has been deposited in the BioMagResBank Database (accession code 6058).

Acknowledgements

This work was supported by a grant from the Protein Structure Initiative of the National Institutes of Health (P50 GM62413) and Nebraska Tobacco Settlement Biomedical Research Development Funds. NMR spectra were acquired at both the Center for Advanced Biotechnology and Medicine NMR facility at Rutgers University and the Environmental Molecular Sciences Laboratory (a national scientific user facility sponsored by the U.S. Department of Energy Office of Biological and Environmental Research) located at Pacific Northwest National Laboratory and operated for DOE by Battelle (contract KP130103).

Supporting Information Available: Table containing the ^{1}H, ^{15}N, and ^{13}C resonance assignments for AF 2095.

References

Clore, G.M. and Gronenborn, A.M. (1998) *Trends Biotechnol.*, **16**, 22–34.

Delaglio, F. et al. (1995) *J. Biomol NMR*, **6**, 277–293.

Ferentz, A.E. and Wagner, G. (2000) *Q. Rev. Biophys.*, **33**, 29–65.

Garrett, D.S. et al. (1991) *J. Magn. Reson.*, **95**, 214–220.

Jansson, M. et al. (1996) *J. Biomol. NMR*, **7**, 131–41.

Liu, J. et al. (2004) *Protein Sci.*, in press.

Montelione, G.T. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 13488–13489.

Wishart, D. and Sykes, B. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.

Zimmerman, D.E. et al. (1997) *J. Mol. Biol.*, **269**, 592–610.