PROTEINS: Structure, Function, and Bioinformatics 54:794–796 (2004)

STRUCTURE NOTE

NMR Structure of the Hypothetical Protein AQ-1857 Encoded by the Y157 Gene From *Aquifex aeolicus* Reveals a Novel Protein Fold

Duanxiang Xu,[1,4] Gaohua Liu,[1,4] Rong Xiao,[2,4] Tom Acton,[2,4] Sharon Goldsmith-Fischman,[3,4] Barry Honig,[3,4] Gaetano T. Montelione,[2,4] and Thomas Szyperski[1,4]*

[1]*Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, New York*
[2]*Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey*
[3]*Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York*
[4]*Northeast Structural Genomics Consortium*

Introduction. The open reading frame of the hypothetical protein AQ-1857 (SwissProt/TrEMBL ID Y157_AQUAE) encoded in the genome of *Aquifex aeolicus* was selected[1] as a target for the Northeast Structural Genomics Consortium (NESGC; http://www.nesg.org). Here we report the high-quality NMR solution structure of AQ-1857 (NESG target QR6). The 116-residue protein AQ-1857 belongs to the HesB/YADR/YFHF family[2,3] which includes proteins involved in nitrogen fixation[4] and Fe-S cluster assembly,[5,6] and contains the PROSITE[7] consensus pattern for this family:

F-X-[LIVMFY]-X-N-[PG]-[NSKQ]-
X(4)-C-X-C-[GS]-X-S-F.

Materials and Methods. Uniformly (*U*) ^{13}C, ^{15}N-labeled AQ-1857 was cloned, expressed and purified following standard protocols. Briefly, the full length gene (Y157_AQUAE) from *Aquifex aeolicus* was cloned into a pET21d (Novagen) derivative, yielding the plasmid pQR6-21. The resulting construct contains eight nonnative residues at the C-terminus (LEHHHHHH) that facilitate protein purification. *Escherichia coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pQR6-21, and cultured in MJ minimal medium containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. *U*-^{13}C,^{15}N AQ-1857 was purified using a two-step protocol consisting of Ni-NTA affinity (Qiagen) and gel filtration (HiLoad 26/60 Superdex 75, Amersham Biosciences) chromatography. The final yield of purified *U*-^{13}C, ^{15}N AQ-1857 (> 97% homogeneous by SDS-PAGE; 14.4 kDa by MALDI-TOF mass spectrometry) was about 10 mg/L. In addition, a sample which was *U*-^{15}N and 5% biosynthetically directed fractionally ^{13}C-labeled was generated for stereospecific assignment of isopropyl methyl groups.[8] Two samples of 5%^{13}C,*U*-^{15}N and *U*-^{13}C,^{15}N AQ-1857 were prepared at concentrations of 1.0 mM in 95% H_2O/5% 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 data were collected at 20°C on Varian INOVA 600 and 750 spectrometers. The spectra were processed and analyzed using the programs NMRPipe[9] and XEASY,[10] respectively. Resonance assignments were obtained as described[11] using a suite of reduced-dimensionality NMR experiments, including 3D HNNCAHA, $H^{\alpha\beta}C^{\alpha\beta}$(CO)NHN, HCCH-COSY, and 2D HBCB(CGC-D)HD. These data were complemented by conventional[12] HNNCACB and HC(C)H TOCSY experiments. Assignments were obtained for 93% of the backbone and $^{13}C^{\beta}$, and for 91% of the side chain chemical shifts. Stereospecific assignments were obtained for 44% of the β-methylene groups exhibiting non-degenerate proton chemical shifts, and for all Val and Leu isopropyl moieties. The chemical shifts were deposited in the BioMagResBank (accession code: 5683). Upper distance limit constraints for structure calculations were obtained from 3D ^{15}N-and ^{13}C-resolved [^1H,^1H]-NOESY[12] (Table I). In addition, $^3J_{HN\alpha}$ scalar couplings measured in 3D HNNHA[12] yielded ϕ-angle constraints, and backbone dihedral angle constraints were derived from chemical shifts as described[13] for residues located in regular secondary structure elements (Table I). Structure calculations were performed using the program DYANA.[14]

Grant sponsor: National Institutes of Health; Grant number: P50 GM2413-10; Grant sponsor: the National Science Foundation; Grant numbers: MCB 00075773, DBI-9904841; Grant sponsor: the Center for Computational Research.

*Correspondence to: Thomas Szyperski, Department of Chemistry, University at Buffalo, The State University at New York, Buffalo, New York 14260. E-mail: szypersk@chem.buffalo.edu

Received 18 August 2003; Accepted 21 August 2003

Results and Discussion. Statistics for the structure determination (Table I) show that a high-quality NMR structure was obtained (Fig. 1). AQ-1857 (PDB ID: 1NWB) contains seven β-strands A to F and two α-helices. A(↓), F(↓) and G(↑) form a 3-stranded, and D(↓), E(↑), B(↑) and C(↓) form a 4-stranded sheet. The two sheets form a "sandwich" being rotated by ~45 degrees relative to each other (Fig. 2). The segment 40–45 and the C-terminal tail 102–116 are flexibly disordered in solution.

The NMR structure of AQ-1857 is the first structure representative of the larger HesB family[2,3] of proteins. No meaningful structural homologues were identified using the programs SKAN,[15] DALI,[16] or CE.[17] This finding strongly supports the notion that AQ-1857 possesses a

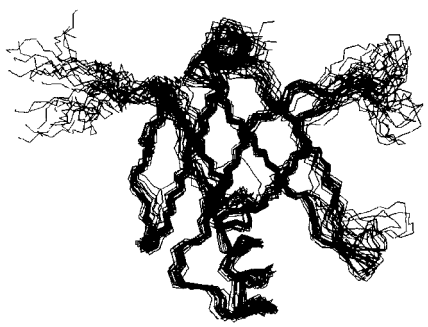

Fig. 1. The 20 DYANA conformers with the lowest residual DYANA target function chosen to represent the NMR solution structure of AQ-1857 are shown after superposition of the backbone heavy atoms N, C^{α} and C′ of the regular secondary structure elements for minimal RMSD.

TABLE I. Statistics of 20 Best DYANA Conformers of AQ-1857[†]

Distance constraints	
All	1328
Intraresidue [$i = j$]	274
Sequential [$(i - j) = 1$]	428
Medium Range [$1 < (i - j) \leq 5$]	242
Long Range [$(i - j) > 5$]	384
Dihedral angle constraints	
ϕ	74
ψ	62
Number of constraints per residue	14.4
Number of long-range constraints per residue	3.8
Average pairwise RMSD (Å) to the mean coordinates	
All residues[a]	
Backbone atoms	1.24 ± 0.24
All heavy atoms	1.77 ± 0.22
Regular secondary structure elements[b]	
Backbone atoms	0.59 ± 0.17
All heavy atoms	1.06 ± 0.16
Distance constraints violations per conformer	
0.1–0.2 Å	1.75
0.2–0.5 Å	1.0
>0.5 Å	0
Dihedral-angle constraint violation per conformer	
0–10°	0.15
>10°	0
Ramachandran Plot	
Residues in most favored regions (%)	81
Residues in additional allowed regions (%)	18
Residues in generously allowed regions (%)	1
Residues in disallowed regions (%)	0

[†]20 conformers with lowest DYANA target function values (0.78 ± 0.16 $Å^2$; range: 0.45–1.04 $Å^2$) out of 100 calculated.
[a]Residues 1–101; the C-terminal segment 102–116 is flexibly disordered in solution.
[b]Residues 14–25 and 77–79 (α-helices), and 10–12, 33–36, 52–53, 63–65, 69–72, 84–89, 94–99 (β-strands).

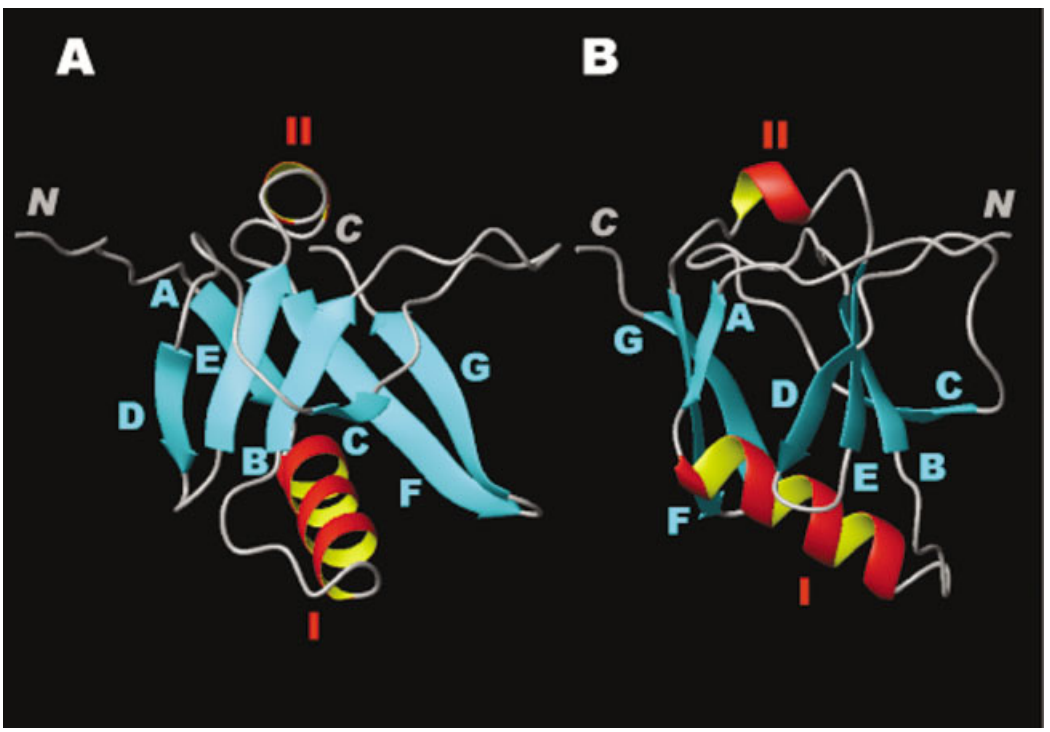

Fig. 2. **A:** The novel fold of AQ-1857: ribbon drawing of the DYANA conformer with the lowest residual target function value. The α-helices I and II are shown in red and yellow, the β-strands A to G are in cyan, other polypeptide segments are in grey, and the N- and C-terminal ends of the protein are indicated as 'N' and 'C'. **B:** Same as in (A), but rotated by 90° about the vertical axis. α-Helix I: residues 14–25, II: 77–79; β-Strand A: 10–12, B: 33–36, C: 52–53, D: 63–65, E: 69–72, F: 84–89, G: 94–99.

hitherto uncharacterized, novel fold (Fig. 2). Interestingly, the PROSITE consensus pattern for the HesB family spans the flexibly disordered C-terminal tail of the protein. In fact, two of the three cysteinyl residues which have been proposed to be involved in iron-sulfur cluster assembly in members of this family[5,6] are located in this tail (not shown in Fig. 1; the third cysteine is located in position 43 in the loop connecting β-strands 2 and 3). It is thus very likely that the flexibly disordered tail is of functional importance, and it is tempting to speculate that this tail adopts an ordered conformation only upon involvement in Fe-S cluster assembly.

Acknowledgements. This work was supported by the National Institutes of Health (P50 GM62413-01), the National Science Foundation (MCB 00075773 to T.S.; DBI-9904841 to B.H), and the *Center for Computational Research* at UB.

REFERENCES

1. Liu, J, Rost, B. Target space for structural genomics revisited. Bioinfomatics 2002;18:922–933.
2. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucl Acids Res 2002;30:276–280.
3. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, Bucher P, Copley RR, Courcelle E, Das U, Durbin R, Falquet L, Fleischmann W, Griffiths-Jones S, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lopez R, Letunic I, Lonsdale D, Silventoinen V, Orchard SE, Pagni M, Peyruc D, Ponting CP, Selengut JD, Servant F, Sigrist CJ, Vaughan R, Zdobnov EM. The InterPro Database, 2003 brings increased coverage and new features. Nucl Acids Res 2003;31:315–318.
4. Huang TC, Lin RF, Chu MK, Chen HM. Organization and expression of nitrogen-fixation genes in the aerobic nitrogen-fixing unicellular cyanobacterium *Synechococcus sp.* Strain RF-1. Microbiology 1999;145:743–753.
5. Ollagnier-de Choudens S, Nachin L, Sanakis Y, Loiseau L, Barras F, Fontecave M. SufA from *Erwinia chrysanthemi.* Characterization of a scaffold protein required for iron-sulfur cluster assembly. J Biol Chem 2003;278:17993–18001.
6. Wollenberg M, Berndt C, Bill E, Schwenn JD, Seidler A. A dimer of the FeS cluster biosynthesis protein IscA from cyanobacteria binds a [2Fe2S] cluster between to protomers and transfers it to [2Fe2S] and [4Fe4S] apo proteins. Eur J Biochem 2003;279:1662–1671.
7. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bionform 2002;3:265–274.
8. Szyperski, T, Neri, D, Leiting, B, Otting, G, Wüthrich, K. Support of ^{1}H NMR assignments in proteins by biosynthetically directed fractional ^{13}C-labeling. J Biomol NMR 1992;2:323–334.
9. Delaglio, F., Grzesiek, S., Vuister, GW, Zhu, G, Pfeifer, J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
10. Bartels C, Xia T, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 1995;6:1–10.
11. Szyperski, T, Yeh, DC, Sukumaran, DK, Moseley, HNB, Montelione, GT. Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. Proc Natl Acad Sci USA 2002;99:8009–8014.
12. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ. Protein NMR spectroscopy. New York: Wiley; 1996.
13. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shifts and sequence homology. J Biomol NMR 1999;13:289–302.
14. Güntert, P, Mummenthaler C, Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
15. Petrey D, Nicholls A, and Honig B. GRASP2: Visualization, surface properties and electrostatics of macromolecular structures. Meth Enzymol 2003;374:494–511.
16. Holm, L, Sander C, Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
17. Shindyalov, IN, Bourne, PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.