PROTEINS: Structure, Function, and Bioinformatics 65:1055–1057 (2006)

STRUCTURE NOTE

Solution Structure of TA0895, a MoaD Homologue from *Thermoplasma acidophilum*

Jinwon Jung,[1,2] In Young Yeo,[1] Eunmi Hong,[1] Adelinda Yee,[3] Cheryl H. Arrowsmith,[3] and Weontae Lee[1]*
[1]*Department of Biochemistry, College of Science and National Research Laboratory, Yonsei University, Seoul 120-749, Korea*
[2]*Institute of Life Science and Biotechnology, Yonsei University, Seoul 120-749, Korea*
[3]*Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5G 2M9*

Introduction. *Thermoplasma acidophilum* is a thermoacidophilic archaeon that thrives at a temperature of 59°C and a pH of 2, conditions in which few organisms are viable. It is one of the smallest archaeal organisms with plasma membrane.[1] The *T. acidophilum* genome encodes for the 90-residue, hypothetical protein TA0895 (TrEMBL ID: Q9HJR9). BLAST[2] search indicates that TA0895 has 21% sequence identity with the small subunit of *E. coli* molybdopterin synthase (MoaD) and it has many homologues in archea, including *T. volcanium, G. sulfurreducens, P. horikoshii, P. furiosus,* and *T. kodakarensis* [Fig. 1(A)]. Here, we present the solution structure of TA0895 and it suggests evidence that TA0895 protein shares its function with *E. coli* MoaD.

Methods and Materials. TA0895 was cloned into the pET-15b expression vector, and the recombinant protein was expressed in *E. coli* BL21 (DE3) cells (Novagen). This vector encodes for 18 additional amino acids which confer a six-histidine affinity tag (His-tag) at the N-terminal and Gly and Ser residues at the C-terminal. The recombinant TA0895-expressing bacteria were grown on a minimal M9 medium containing $^{15}NH_4Cl$ (to obtain ^{15}N-labeled protein) or ^{13}C-glucose, $^{15}NH_4Cl$ (to obtain ^{15}N, ^{13}C-labeled protein) at 37°C until the optical density reached 0.6 at 600 nm. Protein expression was induced with 1 m*M* IPTG for 5 h at 37°C. The recombinant protein was purified using immobilized-metal affinity chromatography followed by gel filtration chromatography. The concentration of the final NMR sample was about 1.0–1.5 m*M*. The condition of all NMR samples was as follows: 1.5 m*M* ^{13}C-, ^{15}N-, or ^{15}N-labeled protein, 50 m*M* potassium phosphate, and 0.01% NaN_3 in 90% H_2O and 10% D_2O at pH 7.4.

All NMR data was acquired at 298 K using Bruker DRX 500 or Varian INOVA 600 MHz spectrometers. [^1H-^{15}N] HSQC, HNCA, HNCACB, CBCA(CO)NH, HNHA, (H)CC(CO)NH, H(CC)(CO)NH, 3D ^{15}N-edited NOESY (τ_m = 150 ms), and 3D ^{13}C-edited NOESY (τ_m = 150 ms) spectra were acquired for the structure analysis. Acquired FIDs were processed by NMRPipe,[7] and the spectra were analyzed with Sparky.[8] Backbone assignments were performed using AUTOASSIGN[9] and were manually analyzed with Sparky. Backbone dihedral angle restraints were derived from chemical shifts of the backbone atoms using TALOS.[10] Hydrogen bond restraints were determined based on slow exchanging amide protons identified by recording ^1H-, ^{15}N HSQC after dissolving the lyophilized $^{13}C/^{15}N$-labeled protein sample in 100% D_2O. Hydrogen bond acceptors were identified from the ensemble structure calculated without distance restraints of hydrogen bonds.[11] Structure calculations were performed using both manually and automatically assigned NOE restraints with CYANA.[12] The restraints used for the final structure refinement are 1047 distance constraints from NOEs, 84 from hydrogen bonds, and 109 dihedral angle restraints. No distance and angle violation was observed for structure calculations. The final 20 energy-minimized structures with the lowest target function energy were selected for further analysis. Structures were analyzed and validated using the program PROCHECK.[13]

Results and Discussion. The average root-mean-square deviation (RMSD) values relative to the mean coordinates of 20 representative conformers [Fig. 1(B)] were determined to be 0.47 ± 0.15 Å for the backbone atoms and 0.93 ± 0.08 Å for all heavy atoms, respectively (Table I). The solution structure of TA0895 is composed of four β-strands and three α-helices organized in an

Grant sponsor: NRL program of MOST NRDP; Grant number: M1-0203-00-0020; Grant sponsor: Korean Government (MOEHRD); Grant number: KRF-2004-005-C00129.

Jinwon Jung and In Young Yeo contributed equally to this work.

*Correspondence to: Weontae Lee, Department of Biochemistry, Yonsei University, 134 Shinchon-Dong Seodaemoon-Gu, Seoul 120-749, Korea. E-mail: wlee@spin.yonsei.ac.kr

Received 4 April 2006; Revised 15 June 2006; Accepted 14 July 2006

Published online 3 October 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21171

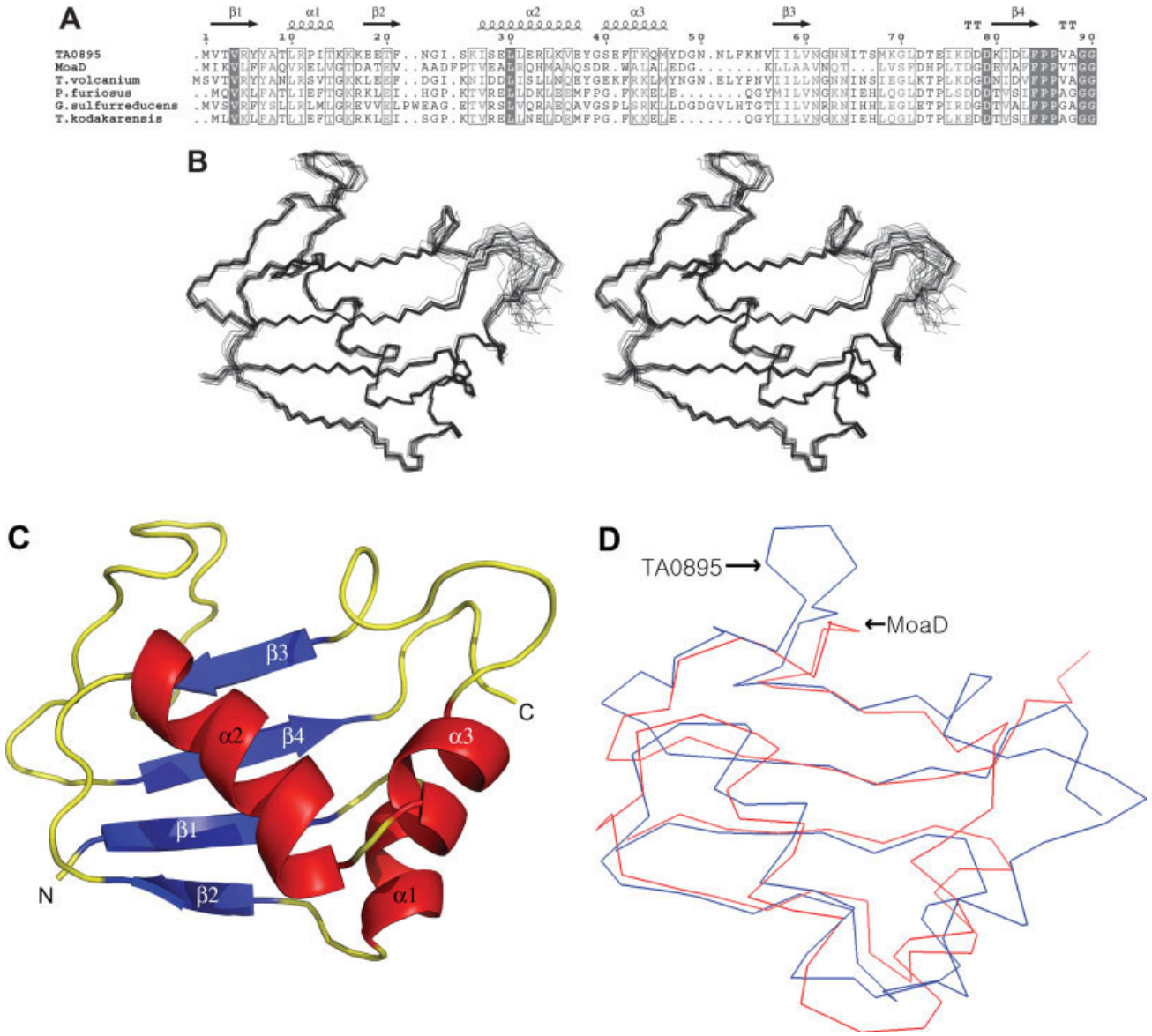

Fig. 1. (**A**) Multiple sequence alignments of TA0895 and its homologues. The alignments were generated by T-COFFEE[3] and decorated using ESPript.[4] Conserved sequences are highlighted in red, and similar sequences are typed in red. The C-terminal region contains the most conserved sequences. There is a 21% sequence identity between TA0895 and MoaD. (**B**) NMR solution structure of TA0895. Shown are the 20 lowest energy conformations from CYANA calculations. A total of 1131 distance restraints including hydrogen bond restraints and 109 angles restraints from TALOS were used for structure generation. The backbone RMSD was calculated as 0.47 Å. (**C**) Ribbon diagram of the TA0895 structure. The β-stands and α-helices are displayed in blue and red, respectively. (**D**) Structural comparison of the Cα trace of TA0895 (blue) and MoaD (red). RMSD value between two structures is 2.39 Å for 74 best-matched Cα pairs. Figures were generated with MOLMOL[5] and PyMOL program.[6]

βαβααββ order [Fig. 1(C)]. Most of residues involved in the hydrophobic surface were found between the β-sheets (β1, β3, and β4) and second helix (α2). Residues Leu30, Leu31, Leu34, and Tyr38 of α2 interact with residues on the β-sheets (Val2, Val4, and Tyr6 from β1; Val56, Ile58, and Val 60 from β3; and Ile81 and Leu83 from β4) via hydrophobic contacts.

To find structural homologues, DALI[14] search was performed. Most similar structure from data base search is 1V8C, a structure of MoaD related protein from *Thermus thermophilus* HB8 with Z score of 10.1. The second one is MoaD from *E. coli* (1FMA) with Z score of 7.3. Ubiquitin is also found with Z score of 4.5. The structural similarity between MoaD and ubiquitin was previously reported.[15] BLAST search on PDB reports another structural homologue, which is a homologue protein (1VJK) from *P. furiosus*. Both structures were well superimposed with RMSD of 2.48 Å for Cα.

E. coli MoaD and TA0895 display 21% sequence identity and 52% sequence similarity. If these two proteins share structural homologues, TA0895 might represent a MoaD homologue protein with similar molecular and cel-

TABLE I. Structural Statistics of TA0895

Distant restraints	
All	1047
Intraresidue	271
Sequential ($\|i-j\|=1$)	316
Medium range ($2 \leq \|i-j\| \leq 5$)	178
Long range ($\|i-j\|>5$)	282
Hydrogen bond restraints[a]	84
Dihedral angle restraints	
All	109
Φ	55
Ψ	54
Mean CYANA target function[b]	1.22 ± 0.05 Å^2
RMS deviations from the average coordinate	
Backbone atoms	0.47 ± 0.15 Å
All heavy atoms	0.93 ± 0.08 Å
Ramachandran analysis (%)	
Residues in most favored regions	83.9
Residues in additional allowed regions	13.6
Residues in generously allowed regions	2.5
Residues in disallowed regions	0

[a] Two restraints for each hydrogen bond.
[b] No distance and angle restraint violation are observed during structure refinement.

lular functions of MoaD. NMR structure shows that the overall topology of TA0895 is almost the same as the structure of MoaD. When the structures of these two proteins are superimposed, backbone RMSD is calculated as 2.392 Å for the best-fitted 74 pairs of Cα atoms [Fig. 1(D)]. Additionally, TA0895 and MoaD share their amino acid sequences. Multiple sequence alignments of TA0895 with its homologue proteins suggest four highly conserved residues, showing that the consensus sequence is PPXXGG [Fig. 1(A)]. In MoaD, the carboxyl group of the last glycine residue is considered as a sulfur donor in a thiocarboxylate reaction during molybdopterin synthesis. Its long and flexible C-terminal tail has penetrated deeply into its interacting proteins, such as MoeB(PDB:1JW9_D)[15] or MoaE(PDB:1FMA_D),[16] where it is placed at their active sites. TA0895 also has a similar long and flexible tail which might be essential for interaction with its partner proteins. In addition, many of the hydrophobic residues involved in the interactions between MoaD and MoaE or MoeB are both sequentially and structurally well-conserved. For example, the F7, L52, L59, and F75 residues of *E. coli* MoaD are known to interact with MoaE or MoeB. The corresponding residues in TA0895 (Y7, I57, M68, and F84) have similar chemical properties and they are located in the same positions in the 3D structure. Beside TA0895 belongs to thiaminS (ThiS) family according to Pfam,[17] the structure of ThiS and MoaD is quite different. They have similar topology with ubiquitin, however, MoaD has a helix between first two β-strands and the helix between β2 and β3 is kinked unlike ubiquitin or ThiS. Therefore, it seems to be reasonable to distinguish between MoaD and ThiS by Conserved-Domain Database.[18] All these results lead us to suggest that TA0895 shares some of the same functions as *E. coli* MoaD, representing a MoaD homologue found in *Thermoplasma acidophilum*.

(*Note:* Chemical shifts for TA0895 have been deposited in BioMagResBank (accession code: BMRB-6982) and coordinates for the 20 structures have been deposited into RCSB PDB with accession code 2G1E).

REFERENCES

1. Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. Nature 2000;407:508–513.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
3. Poirot O, Suhre K, Abergel C, O'Toole E, Notredame C. 3DCoffee: a web server for mixing sequences and structures into multiple sequence alignments. Nucleic Acids Res 2004;32:W37–W40.
4. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: multiple sequence alignments in postscript. Bioinformatics 1999;15:305–308.
5. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.
6. DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002.
7. 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.
8. Goddard TD, Kneller DG. *SPARKY* 3. San Francisco, CA: University of California; 1999.
9. Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 1997;269:592–610.
10. Cornilescu F, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–302.
11. Fernandez C, Hilty C, Wider G, Güntert P, Wüthrich K. NMR structure of the integral membrane protein OmpX. J Mol Biol 2004;336:1211–1121.
12. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 2002;319:209–227.
13. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of proteins structures. J Appl Crystallogr 1993;26:283–291.
14. Holm L, Sander C. Alignment of three-dimensional protein structures. Methods Enzymol 1996;266:635–662.
15. Lake MW, Wuebbens MM, Rajagopalan KV, Schindelin H. Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. Nature 2001;414:325–329.
16. Rudolph MJ, Wuebbens MM, Rajagopalan KV, Schindelin H. Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. Nat Struct Biol 2001;8:42–46.
17. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32:D138–D41.
18. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki C, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH. CDD: a conserved domain database for protein classification. Nucleic Acids Res 2005;33:D192–D196.