PROTEINS: Structure, Function, and Bioinformatics 67:1185–1188 (2007)

STRUCTURE NOTE

NMR Structure of Hypothetical Protein TA0938 From *Thermoplasma acidophilum*

Daniel Monleón,[1,2] Adelinda Yee,[3] Cheryl Arrowsmith,[3] and Bernardo Celda[1,4]*
[1]*Department of Physical Chemistry, University of Valencia, Burjassot 46100, Valencia, Spain*
[2]*Research Unit, Hospital Clinico Universitario, University of Valencia, Valencia 46010, Spain*
[3]*Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, ON, Canada, M5G2M9*
[4]*CIBER Bioengineering, Biomedicine and Nanomedicine, ISCIII, Spain*

Introduction. Structural proteomics is an emerging scientific field aiming to obtain one or more representative 3D structures for every structural domain family in nature by application of high throughput structure determination techniques. These structures, and the corresponding protein production vectors and resonance assignments, will provide a valuable resource for structural and functional studies of the thousands of proteins and their homologues that are targeted by the international structural proteomics efforts.[1] *T. acidophilum* is a thermoacidophilic archaeon[2] that inhabits a hot and highly acidic environment in which few organisms are viable. The genome of *T. acidophilum* is one of the smallest among free-living organisms.[3] TA0938 is a 110-residue conserved hypothetical protein from *T. acidophilum* with unknown function. From BLAST[4,5] database search, TA0938 closest homologues are hypothetical proteins in *Sulfolobus solfataricus* (Q97VD3_SULSO, ~50% sequential identity over 90% of total length) and in *Sulfolobus tokodaii* (Q96XA7_SULTO, ~36% sequential identity over 86% of total length), both with unknown structure and function. Here, we present the solution structure of TA0938 determined by NMR spectroscopy. On the basis of this solution three-dimensional structure and the amino acid sequence analysis, TA0938 seems to be an uncharacterized protein with a novel fold and a putative Zn-binding motif. These results may be the bases for posterior studies on structure-function relationship in proteins with similar fold.

Materials and methods. A recombinant protein consisting of the full sequence of TA0938 (110 amino acids) was expressed in *E. coli* BL21-Gold (DE3) cells containing the pET-15b expression vector (Novagen). Cells were grown at 37°C to an OD_{600} of 1.0–1.2 and induced with 1 m*M* IPTG for overnight at 15°C. The protein was purified to homogeneity using metal affinity chromatography. The purified protein contained the complete sequence of TA0938 plus additional N-terminal histidine tag (MGSSHHHHHHSS GLVPRGSH). U-^{15}N and U-^{13}C, ^{15}N samples were produced in 2X M9 media supplemented with Zinc sulphate, biotin, ^{15}N ammonium chloride, and ^{13}C glucose. ^{15}N-labeled or ^{13}Cl^{15}N-labeled protein solution was prepared in 20 m*M* Tris (pH = 6.7), 100 m*M* NaCl, 1 m*M* DTT, 0.01% NaN3, 1 m*M* benzamidine, 95% H_20/5% D_2O. The concentration of the purified protein in the NMR samples ranged between 0.8 and 1.5 m*M*. Molecular mass and purity of the ^{13}C, ^{15}N protein in the NMR samples were confirmed by mass spectroscopy.

NMR spectra for backbone resonance assignments were recorded at 25°C on a Bruker AVANCE DRX 500 MHz spectrometer equipped with pulse-field gradient triple-resonance probes. Additional NMR spectra for side-chains resonance assignments were recorded at 25°C on a Bruker AVANCE DMX 800 MHz spectrometer equipped with pulse-field gradient triple-resonance probes. Linear prediction to double number of points was used in the ^{1}H, ^{13}C, and ^{15}N indirect dimensions to improve digital resolution. Spectra were processed using the XwinNMR 3.1 software package. No significant differences were detected between ^{15}N-HSQC recorded over the protein with and without Zn. Sparky 3.91[6] and home made shell and perl scripts were used for semi automated peak picking and peak lists filtering.[7] Resonance assignments of TA0938 were obtained mainly by the combination of manual and automatic techniques as specified elsewhere[8] (BMRB entry 6812). Over-

Grant sponsor: Ministerio de Educación y Ciencia, Spain; Grant numbers: SAF 2004-06297, SAF 2004- 20971-E; Grant sponsor: Ministerio de Sanidad; Grant number: ISCIII 2003-G03/185; Grant sponsor: Ontario Research and Development Challenge Fund; Grant sponsor: Programa Ramon y Cajal (to D.M.).

*Correspondence to: Bernardo Celda, Departamento de Química Física, Universitat de Valencia, C/ Dr. Moliner, 50, Burjassot 46100, Valencia, Spain. E-mail: bernardo.celda@uv.es

Received 8 February 2006; Revised 26 May 2006; Accepted 19 June 2006

Published online 21 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21143

all, ~99% of backbone assignable protons and ~96% of total protons were available for our study.

For structure calculation purposes, ^{13}C-NOESY-HSQC (τ_m of 100 and 200 ms) and ^{15}N-NOESY-HSQC (τ_m of 100 ms) spectra were recorded at 800 MHz (for a review on the experiments used, see Cavanagh et al.[9]). NOE cross-peak assignment was obtained using a combination of manual and automatic methods. A preliminary fold was calculated on the basis of manually unambiguously assigned NOEs on the 3D NOESY spectra at 100 ms of mixing time. In this stage, NOE peaks were classified as weak, medium, and strong intensity and upper limit of 5.0, 4.0, and 3.0 Å in distance restraints were accordingly applied. NOE assignments were extended and structure was further refined by spectra/structure iterative semi-automated analysis with the NOAH module in the program DYANA[10] including all spectra. Peak lists of the NOESY spectra were obtained by interactive peak picking using the "restricted peak picking" option of the program SPARKY. Backbone dihedral restraints were derived from the $^{1}H_{\alpha}$ and $^{13}C_{\alpha}$ secondary chemical shifts using TALOS.[11] A summary of the final set of structural restraints used for torsion angle dynamics calculations together with other statistics for the ensemble of the 20 lowest target function values conformers is reported in Table I. The program MOLMOL[12] was used to analyze the 20 energy-minimized conformers with lowest NOE violations and to calculate solvent accessible surface and electrostatic charges distributions. MOLMOL was also used to prepare drawings of the structures.

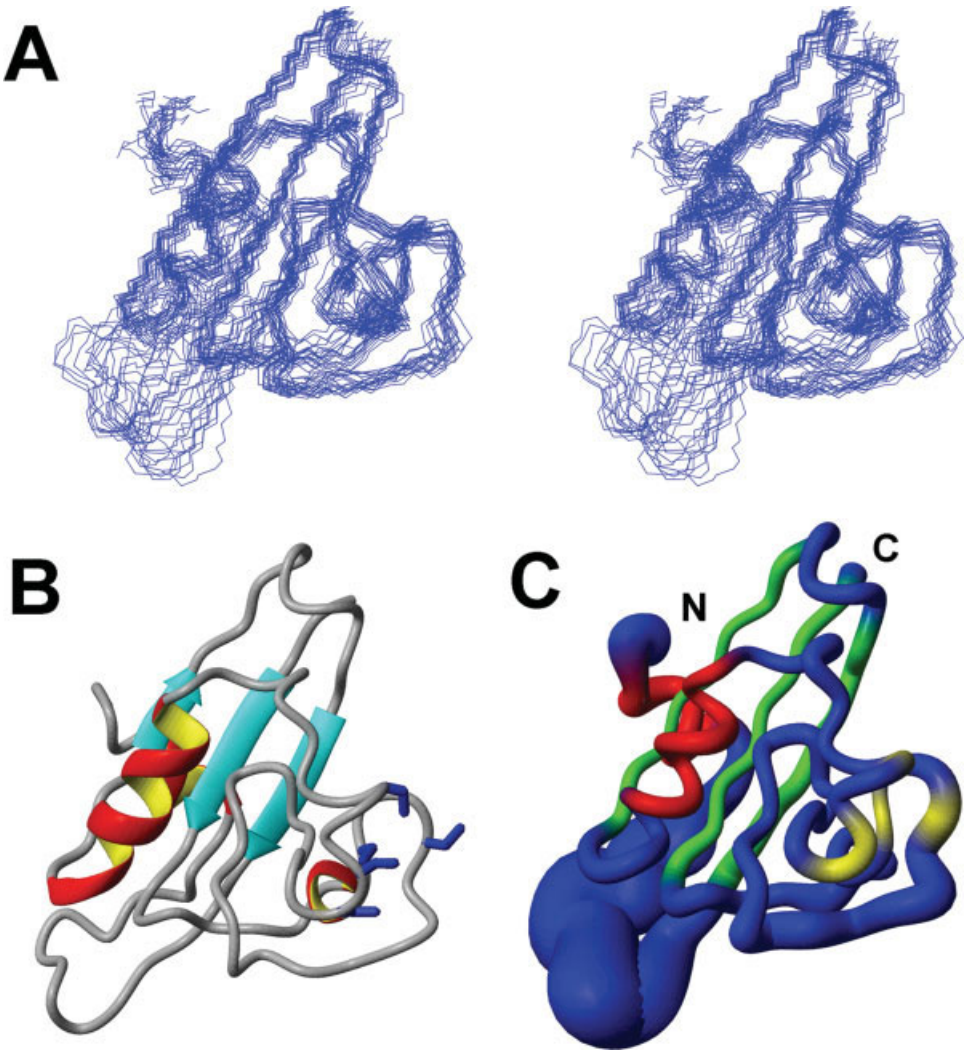

Fig. 1. NMR solution structure of TA0938. **A**. Stereo view of the superposition of the final 20 structures over the average NMR structure. **B**. Ribbon diagram depicting lowest target function NMR structure of TA0938 of *Thermoplasma acidophillum* (PDB accession code 2FQH). The α-helices are shown in red and yellow and β-sheets are shown in cyan. Side-chains of cysteine residues are shown in blue. Chemical shift values for cysteines Cβ in TA0938 were: Cys20, 44.14 ppm; Cys23, 32.28 ppm; Cys42, 36.02 ppm; Cys43, 34.36 ppm; and Cys46 43.10 ppm **C.** Sausage presentation of backbone of TA0938. Thickness of the cylindrical rod is proportional to the mean of the global displacements of the Cα atoms in the 20 DYANA best conformers. The β-strands are shown in green, the α-helices in red and cysteine residues in yellow.

TABLE I. Structural Statistics for the Ensemble of 20 Lowest DYANA Target Function Structures Out of 200 Calculated for TA0938

Distance restraints	
All	2754
Intra residue	926
Sequential ($\|i - j\| = 1$)	692
Medium range ($2 \leq \|i - j\| \leq 5$)	423
Long range ($5 \leq \|i - j\|$)	713
Dihedral angle restraints	
Φ, Ψ	58, 58
Residual DYANA target function	1.2 ± 0.3
Residual NOE distance constraint violations	
> 0.2 Å	11 ± 3
Maximum (Å)	0.24 ± 0.12
Residual van der Waals close contacts violations	
> 0.2 Å	1 ± 1
Maximum (Å)	0.06 ± 0.04
Pair wise r.m.s.d residues 3 to 88 and 101 to 107	
Backbone atoms	0.7 ± 0.2
Heavy atoms	1.4 ± 0.4
Ramachandran plot (%)	
Residues in most favored regions	83
Residues in additional allowed regions	15
Residues in generously allowed regions	2
Residues in disallowed regions	0

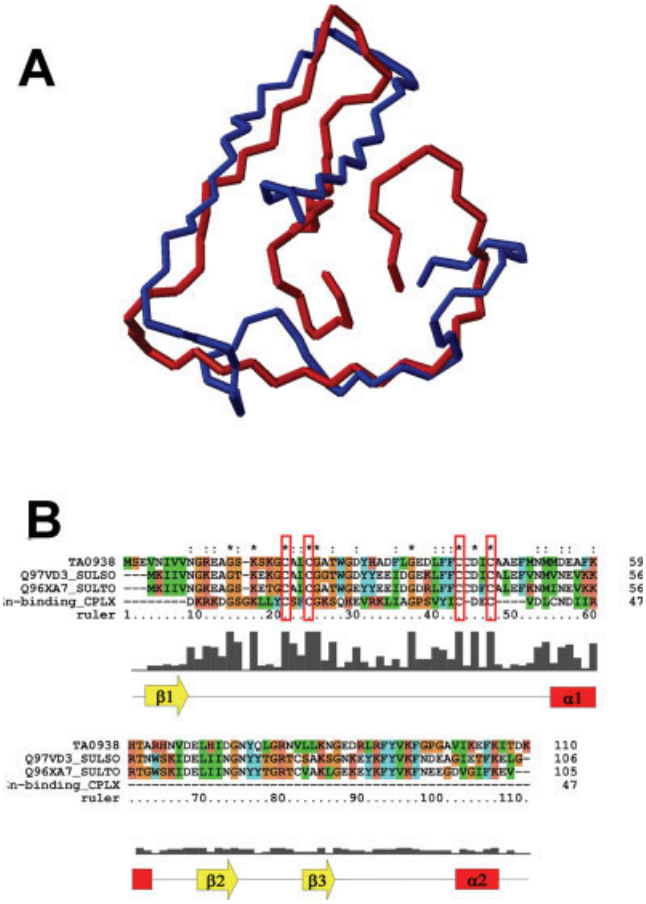

Fig. 2. **A**. Comparison between cysteine-enriched regions of TA0938 (in red) and E. Coli CLPX chaperone Zinc binding domain dimer (PDB code 1OVX, in blue). **B**. Sequence alignment of TA0938, closest homologues hypothetical proteins Q97VD3_SULSO and Q96XA7_SULTO and cysteine-rich segment from *Escherichia coli* CLPX chaperone Zinc binding domain. Secondary structure elements in TA0938 structure have been plotted in the sequence. Cysteines have been marked inside a rectangle.

Results and Discussion. The 20 lowest target function structures of TA0938 are well-converged, as shown in Figure 1. The structure of TA0938 has two clearly different parts: a region containing all the regular secondary structure elements and a bundle of loops, which contain all cysteines in the protein. The part of the protein with regular secondary structure is formed by a central β-sheet flanked by two α-helices (residues 54–62 and 101–106) at each face of the sheet. In this part of the protein, a very flexible loop is observed between residues 90 and 100. Significant broadening of peaks belonging to this region seems to support the flexibility suggested by the partial disorder observed in this loop. The molecular topology of the protein can be described as βαββα. In this topology, the bundle of loops with the cysteines is located between β1 and α1. The central β-sheet is composed by three strands (residues 3–8, 68–73, and 82–85) with a β2([darrow]), β1(↑), and β3(↑) pattern. The bundle of loops where the cysteines are located (residues 16–46) shows some degree of sequence similarity to reported Zn-binding motifs. A strong NOE between Hβ of cysteines Cys20 and Cys46 suggests that these cysteines may be in disulfide bond distance range. The chemical shifts for the Cβ of these cysteines seem to support the partial formation of this disulfide bond.

A homology three-dimensional structure search using DALI[13] within the Protein Data Bank showed that TA0938 shares no meaningful structural similarity with protein structures reported to date. The highest z score is below the threshold of being significant (i.e. 0.8 versus a threshold of 2.0). These results imply that TA0938 would be an uncharacterized protein with a novel fold. Nevertheless, the conformation detected for the region where the cysteines are located showed some structural similarity to the E. Coli CLPX chaperone Zinc-binding domain [PDB code 1OVX, Figure 2(A)]. Sequence identity between this TA0938 segment and the Zn-binding motif of CLPX was 30%. In addition, Cβ chemical shift values for Cys23, Cys 43, and Cys44 in this loop seemed consistent with Zn-binding cysteine ^{13}C average chemical shift values.[14]

In summary, we present the solution structure of TA0938, a functionally unknown protein in *Thermoplasma acidophillum*. According to this structure, TA0938 would be an uncharacterized protein with a novel fold with a cysteine-rich region, which resembles some Zn-binding motifs. Structures ensemble has been deposited into the Protein Data Bank[15] (PDB accession code 2FQH).

Acknowledgments. The authors thank the SCSIE of the University of Valencia for providing access to the NMR facility and high performance computing facilities. The NMR Facility of the Parc Cientific the Barcelona is also gratefully acknowledged for providing access to the 800 MHz NMR spectrometer. We also thank Bruker España S.A. for technical and economic support as part of

the agreement with the University of Valencia for the development of new techniques in biomolecular NMR.

REFERENCES

1. Yee A, Pardee K, Christendat D, Savchenko A, Edwards AM, Arrowsmith CH. Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. Acc Chem Res 2003;36:183–189.
2. Darland G, Brock TD, Samsonoff W, Conti SF. A thermophilic acidophilic mycoplasm isolated from a coal refure pile. Science 1970;170:1216–1418.
3. 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.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
5. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
6. Goddard TD, Kneller DG. SPARKY 3. 1999. San Francisco: University of California;
7. Monleon D, Colson K, Moseley HNB, Anklin C, Oswald R, Szyperski T, Montelione GT. Rapid analysis of protein backbone resonance assignments using cryogenic probes, a distributed Linux-based computing architecture, and an integrated set of spectral analysis tools. J Funct Struct Genomics 2002;2:93–101.
8. Monleon D, Nune S, Esteve V, Yee A, Arrowsmith C, Celda B. Resonance assignments for the hypothetical protein TA0938 from *Termoplasma acidophilum*. J Biomol NMR 2006;36(5): 36.
9. Cavanagh J, Fairbrother WJ, Palmer AGI, Skelton NJ. Protein NMR spectroscopy: principles and practice. 1995. New York: Academic Press;
10. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
11. 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.
12. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
13. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
14. Kornhaber GJ, Snyder D, Moseley HNB, Montelione GT. Identification of zinc-ligated cysteine residues based on 13Ca and 13Cb chemical shift data. J Biomol NMR 2006;34:259–269.
15. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein data Bank. Acta Crystallogr Sect D 2002;58:899–907.