Journal of Biomolecular NMR, **22:** 291–294, 2002.
KLUWER/ESCOM

Letter to the Editor: Solution structure of the hypothetical protein MTH0637 from *Methanobacterium thermoautotrophicum*

Antonio Pineda-Lucena[a,c,**], Gwan-Su Yi[a,c,**], Xiaoqing Chang[a,c], John R. Cort[b,c], Michael A. Kennedy[b,c], Aled. M. Edwards[a,c] & Cheryl H. Arrowsmith[a,c,*]
[a]*Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, ON, Canada, M5G 2M9;* [b]*Enviromental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, EMSL 2569 K8-98, Richland, WA 99352, USA; and* [c]*Northeast Structural Genomics Consortium*

Received 26 November 2001; Accepted 19 December 2001

Key words: heteronuclear NMR, *Methanobacterium thermoautotrophicum*, structural genomics, structural proteomics

Biological context

After completion of the sequencing phase of the genome projects, attention has turned to determine the structure and function of all the proteins encoded by the newly sequenced genes. Structural proteomics projects are expected to answer basic questions about the conformational space available to proteins, and reveal information about pharmaceutical applications. The number of pharmaceutical targets is expected to increase dramatically in the post-genomic era, and detailed information about the structure of the proteins will facilitate the development of drugs against these potential targets (Brower, 2001).

Here we describe the solution structure of MTH0637 as part of a structural proteomics pilot project on the feasibility of the high-throughput generation of samples from *Methanobacterium thermoautotrophicum* for structural studies (Christendat et al., 2000). MTH0637 is an archaeal protein of unknown function (Smith et al., 1997) that shows significant sequence identity to three other archaeal proteins, also of unknown function (Figure 1). These include AF2072 from *Archaeoglobus fulgidus*, MJ0618 from *Methanococcus jannaschii*, and PAB7122 from *Pyrococcus abyssi*. The structure of MTH0637 reveals a new variation of a common fold. The similarity with a known fold could not be predicted from the protein sequence. Although the biochemical function is unknown, patterns of conserved residues on the protein surface suggest a conserved function for all three sequence homologues.

*To whom correspondence should be addressed. E-mail: carrow@uhnres.utoronto.ca
**These authors contributed equally to this project.

Methods and results

A recombinant protein consisting of the full sequence of MTH0637 (104 amino acids) was expressed in *E. coli* BL21-DE3 cells containing the pET-15b expression vector (Novagen). Cells were grown at 37 °C to an OD_{600} of 0.6 and induced with 1 mM IPTG for 5 hours at 25 °C. The protein was purified to homogeneity using metal affinity chromatography. Subsequently, the N-terminal tag was removed using thrombin and benzamidine-sepharose. The purified protein contained the complete sequence of MTH0637 plus three additional N-terminal residues (Gly-Ser-His) remaining after proteolytic cleavage of the His_6 affinity tag. U-^{15}N and U-^{13}C,^{15}N samples were produced in standard M9 media supplemented with ^{15}N ammonium chloride (1 g l^{-1}) and ^{13}C glucose (2 g l^{-1}). ^{15}N-labeled or ^{13}C/^{15}N-labeled protein solution was prepared in 25 mM sodium phosphate (pH = 6.5), 150 mM NaCl, 1 mM DTT, 95% H_20/5% D_2O. The concentration of the purified protein ranged between 1.0–1.5 mM.

All NMR spectra were recorded at 25 °C on a Varian INOVA 600 MHz spectrometer equipped with pulsed field gradient triple-resonance probes. Linear prediction was used in the ^{13}C and ^{15}N di-

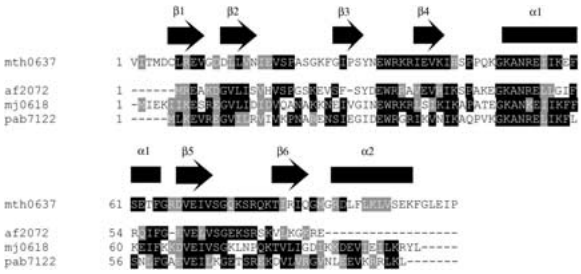

Figure 1. Sequence alignment of MTH0637 with three predicted archaeal proteins, AF2072 (*Archaeoglobus fulgidus*), MJ0618 (*Methanococcus jannaschii*), and PAB7122 (*Pyrococcus abyssi*). Identical and similar residues are highlighted in black and gray, respectively. Black arrows indicate β-strands regions of MTH0637 and black rectangles correspond to α-helical regions.

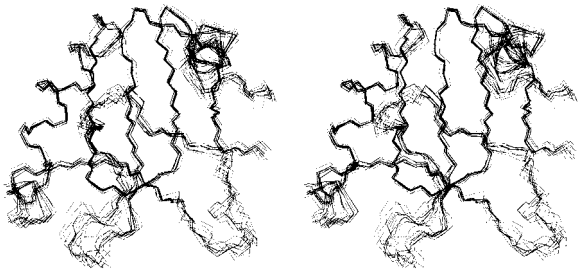

Figure 2. Stereoview of the backbone (N, C^{α}, C') of 10 superimposed NMR-derived structures of MTH0637 of *Methanobacterium thermoautotrophicum* (residues 5-104).

mensions to improve the digital resolution. Spectra were processed using the NMRPipe software package (Delaglio et al., 1995) and analyzed with XEASY (Bartels et al., 1995). SPSCAN (Glaser and Wüthrich) was used to convert nmrPipe formatted spectra into XEASY. The assignments of the ^{1}H, ^{15}N, ^{13}CO, and ^{13}C resonances were based on the following experiments: CBCA(CO)NH, HNCACB, CC(CO)NH-TOCSY, HNCO, HNHA, 3D ^{15}N-edited TOCSY-HSQC, and HCCH-TOCSY (Bax et al., 1994; Kay, 1997). The backbone resonance assignment was achieved mainly by the combined analysis of the HNCACB and CBCA(CO)NH data. The side-chain resonances were identified mainly by the analysis of HCCH-TOCSY. Aromatic ring resonances were assigned based on the analysis of heteronuclear NOESY optimized for the detection of aromatic $^{13}C/^{1}H$ resonances. In the ^{1}H-^{15}N HSQC, 99% backbone amide resonances were assigned. Of the other backbone resonances, 99% have been assigned for C^{α}, 99% for H^{α} and 97% for C'. Moreover, 97% aliphatic side-chains have been assigned for MTH0673.

For structure calculation purposes, a simultaneous ^{15}N- and ^{13}C-NOESY-HSQC (τ_m = 150 ms, Pascal et al., 1994) was acquired. NOE cross peak assignment were obtained by using a combination of manual and automatic procedures. An initial fold of the protein was calculated based on unambiguously assigned NOEs. Later on, the module NOAH in the program DYANA (Güntert et al., 1997) was used. NOAH/DYANA performs automated assignment and distance calibration of NOE intensities, removal of meaningless distance constraints, structure calculation with torsion angle dynamics, and automatic NOE upper distance limit violation analysis. Peak analysis of the NOESY spectra were generated by interactive peak picking with the program XEASY. Backbone dihedral restraints were derived from $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ secondary chemical shifts using TALOS (Cornilescu et al., 1999).

Ninety-eight percent of the manually picked NOE cross peaks were retained for the final structure calculation. They correspond to reliable distance constraints that were unambiguously assigned after multiple cycles of calculations. The program MOLMOL (Koradi et al., 1996) was used to analyze the resulting 10 energy-minimized conformers and to prepare drawings of the structures.

The three-dimensional structure of MTH0637 was determined using a torsion angle dynamics protocol from a total of 1694 NMR-derived constraints. A superposition of 10 low-energy structures is shown in Figure 2, and the structural statistics are given in Table 1. The average global root-mean-square deviation (rmsd) values relative to the mean coordinates are 0.55 $\pm$ 0.33 for the backbone atoms of the residues Asp5-Pro104, and 1.23 $\pm$ 0.23 for all the heavy atoms. Several loops regions within the structure ensemble show variations from the average rmsd found in the ordered regions of the protein. These differences are due to the existence of limited number of observable constraints at those regions (Supplementary material).

The structure of MTH637 has two α-helices (Lys51-Phe64 and Arg88-Lys98) and two antiparallel β-sheets (4 β-strands and 2 β-strands) arranged in a perpendicular manner. The first β-sheet comprises residues Cys6-Val10, Asp13-Asn17, Asp67-Val71 and Thr80-Gln84. The second comprises amino acids Gly28-Ser31 and Ile39-Lys42. Several inter-strand NOE interactions were also observed, and they include contacts between the $^{1}H^{\alpha}$ from Leu7-Val16, Glu9-Leu14, Asp13-Gln84, Asn17-Thr80, Glu19-Gln78, Val68-Ile83, Ile70-Ile81, and Ser72-Lys79.

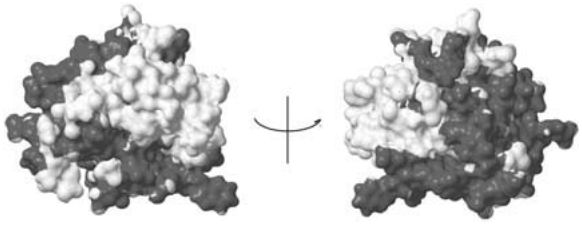

Figure 3. A space-filling model depicting the conserved residues (dark colouring) between MTH0637 and the archaeal proteins. The figure on the left is displayed in the same orientation as in Figure 2.

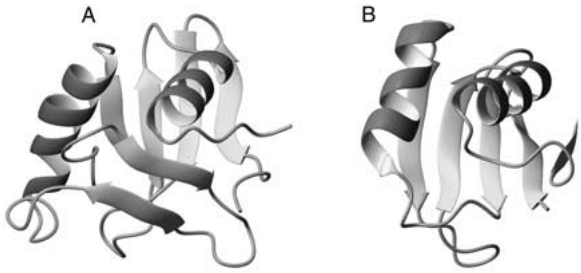

Figure 4. Ribbon diagram depicting (A) the averaged minimized NMR structure of MTH0637 of *Methanobacterium thermoautotrophicum* (residues 5-104) and (B) the human translation initiation factor eIF1.

Table 1. Structural statistics for the ensemble calculated for MTH0637[a]

Distance restraints	
All	1694
Intraresidue	535
Sequential ($\|i-j\| = 1$)	427
Medium range ($2 \leq \|i-j\| \leq 4$)	203
Long range ($\|i-j\| > 4$)	477
Hydrogen bonds	26×2
Dihedral angle restraints	
All	154
ϕ, ψ	76,78
Pairwise r.m.s.d.	
All residues[b]	
Backbone atoms	0.55 ± 0.33
All heavy atoms	1.23 ± 0.23
Ordered regions[c]	
Backbone atoms	0.32 ± 0.09
All heavy atoms	0.76 ± 0.10

[a]Ensemble of the 10 lowest energy structures out of 100 calculated.
[b]Rmsd values for residues 5–104.
[c]Only residues in β-strands and α-helices are included.

Discussion and conclusions

MTH0637 was classified as a conserved protein based on the high sequence homology to MJ0618, AF2072 and PAB7122. In an effort to ascribe functional features to the protein, the location of the most highly conserved portions were maped onto its surface. This analysis revealed that the conserved residues form patches along the three-dimensional structure of the protein (Figure 3). They include hydrophobic and charged residues and, in most cases, are contiguous in the space. The highly conserved residues correspond to the β-strands and the first α-helix of the protein, and are strongly involved in maintaining the core of the protein. The only region of the protein that does not display high sequence similarity to the other proteins is the C-terminal α-helix, which is not involved in many interactions with the rest of the protein.

A 3D structure search using DALI (Holm and Sanders, 1996) showed that there are no other known protein structures that share the MTH0637 fold. However, the MTH0637 fold is related to that of the translation initiation factor eIF1 (Fletcher et al., 1999). Both proteins contain a four-stranded antiparallel β-sheet, and two α-helices. However, MTH0637 contains an additional two-stranded β-sheet inserted between the second β-strand and the first α-helix. This sheet is arranged in a perpendicular manner to the four-stranded β-sheet. Another difference is the extension of the C-terminal α-helix in eIF1 into a fifth β-strand that is arranged in a parallel way to the first β-strand of the β-sheet (Figure 4). An alignment of secondary structure elements does not reveal any common pattern in the residues involved in the maintenance of the respective folds, suggesting that they have evolved separately.

Nevertheless, the similarities between MTH0637 and eIF1 suggest that these two proteins may have similar functions. However, a sequence alignment analysis (Kyrpides and Woese, 1998) showed that there are sequence homologues of eIF1 in *Methanobacterium thermoautotrophicum* (MTH010), and in *Methanococcus jannaschii* (MJ0463). Furthermore, comparison of the sequences indicates that only 3 out of the 14 residues conserved among eIF1 homologs from eukaryotes, archaea and bacteria are present in MTH0637. The charge distribution in both proteins is also very different, and, in the case of MTH0637, does not seem to be compatible with the possibility of an interaction with RNA. Taken together, these results indicate that MTH0637 and its sequence homologues belong to a separate family of proteins involved in a different biological activity.

In summary, MTH0637 represents one of the structurally and functionally unknown proteins in *Methanobacterium thermoautotrophicum.* Here we reported the solution structure of this protein. The structure is composed of two antiparallel and perpendicular β-sheets, and two α-helices. A BLAST search indicated that MTH0637 shares some sequence homology with three other archaeal proteins of unknown structure and function. From the structural point of view, MTH0637 resembles the translation initiation factor eIF1. However, this homology only extends to the structural features, as there exists a sequence homolog of eIF1 in *Methanobacterium thermoautotrophicum* different from MTH0637. Therefore, it is not possible to infer any function for MTH0637. Nevertheless, the structure represents a new fold for an α/β protein. This information will help populate our knowledge base of the set of protein folds available in nature. The chemical shifts have been submitted to the BMRB (accession # 5104), and the structure ensemble has been submitted to the PDB (accession # 1JRM).

Acknowledgements

The authors wish to thank Dr A. Yee and B. Le for their technical assistance. All the spectra were performed at the Environmental Molecular Sciences Laboratory (a national scientific users facility sponsored by the U.S. DOE Office of Biological and Environmental Research) located at Pacific Northwest National Laboratory, operated by Batelle for the DOE. This work was funded by the Ontario Research and Development Challenge Fund, Canadian Institutes of Health Research, and the NIH (P50 GM62413-02).

References

Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **6**, 1–10.

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.*, **239**, 79–105.

Brower, V. (2001) *EMBO Rep.*, **2**, 558–560.

Christendat, D., Yee, A. et al. (2000) *Nat. Struct. Biol.*, **7**, 903–909.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Fletcher, C.M., Pestova, T.V., Hellen, C.U.T. and Wagner, G. (1999) *EMBO J.*, **18**, 2631–2637.

Glaser, R. and Wüthrich, K., http://www.mol.biol.ethz.ch/wuthrich/software/spscan

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.,* **273**, 283–298.

Holm, L. and Sander, C. (1993) *J. Mol. Biol.*, **233**, 5984–6003.

Kay, L.E. (1997) *Biochem. Cell Biol.*, **75**, 1–15.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graphics,* **14**, 51–55.

Kyrpides, N.C. and Woese, C.R. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 224–228.

Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1994) *J. Magn. Reson.*, **B103**, 197–201.

Smith, D.R. et al. (1997) *J. Bacteriol.*, **172**, 7135–7155.