Journal of Biomolecular NMR (2005) 33: 51–56
DOI 10.1007/s10858-005-1275-5

NMR structure note

Solution structure of MTH0776 from *Methanobacterium thermoautotrophicum*

Godwin Amegbey[a], Paul Stothard[a], Ekaterina Kuznetsova[b], Adelinda Yee[c], Cheryl H. Arrowsmith[c] & David S. Wishart[a,*]

[a]*Departments of Computing Science and Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E8;* [b]*Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada M5G 1L6;* [c]*Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5G 2C4*

Received 16 April 2005; Accepted 11 July 2005

Key words: methanogenesis, NMR, novel fold, protein structure

Biological context

Methanobacterium thermoautotrophicum is a methanogenic archaebacterium that grows optimally at ~65 °C. Originally isolated from sewage ponds in southern Illinois, *M. thermoautotrophicum*is a microorganism that biodegrades waste materials and produces natural gas in the process. Due to its potential economic and environmental importance the *M. thermoautotrophicum* genome was sequenced in the mid 1990's (Smith et al., 1997) in an effort to better understand the underlying biochemistry and microbiology of its unique biodegradation capabilities. The genome (1,751,377 bp) consists of 1855 open reading frames, with roughly 65% of the proteins having assigned functions.

MTH0776 is a 101 amino acid residue protein from *M. thermoautotrophicum* and belongs to a distinct orthologous group (COG 4033) consisting of 11 proteins varying in size from 92 to 124 residues found only in methanogenic archaebacteria. MTH0776 is part of a two-protein operon with MTH0777, which is also specific to methanogenic archaebacteria. The fact that this two-component operon is conserved in all methane metabolizing archaea, strongly suggests that both proteins may have an important role in methanogenesis and that these proteins may act in a similar portion of the pathway, potentially as binding partners. In an effort to better understand MTH0776, its role in methanogenesis and the molecular interactions between MTH0776 and MTH0777, we have undertaken the NMR structure determination of MTH0776. Here we wish to report the NMR solution structure of MTH0776 which exhibits a completely novel α/β fold.

*To whom correspondence should be addressed. E-mail: david.wishart@ualberta.ca

Methods and results

The protein expression, purification and chemical shift assignment of MTH0776 has been reported previously (Amegbey et al., 2004). The assigned ^{15}N HSQC spectrum is shown in Figure 1. NOE assignments and distance restraints for all structure calculation were obtained using 3D ^{15}N-NOESY HSQC, 3D ^{13}C-NOESY HSQC as well as two-dimensional ^{1}H–^{1}H NOESY spectra (Zhang et al., 1994). All NOESY spectra were recorded with mixing times of 80 ms and 110 ms. Assigned NOE restraints were classified into four distance ranges: 1.8–2.7 Å, 1.8–3.5 Å, 1.8–5.0 Å, and 1.8–6.0 Å corresponding to strong, medium, weak and very weak NOE intensities, respectively. NOE peak intensities were measured by volume integration of well resolved peaks. Pseudo-atom corrections were added to the upper distance limits where appropriate (Wuthrich, 1986). A 0.5 Å correction was

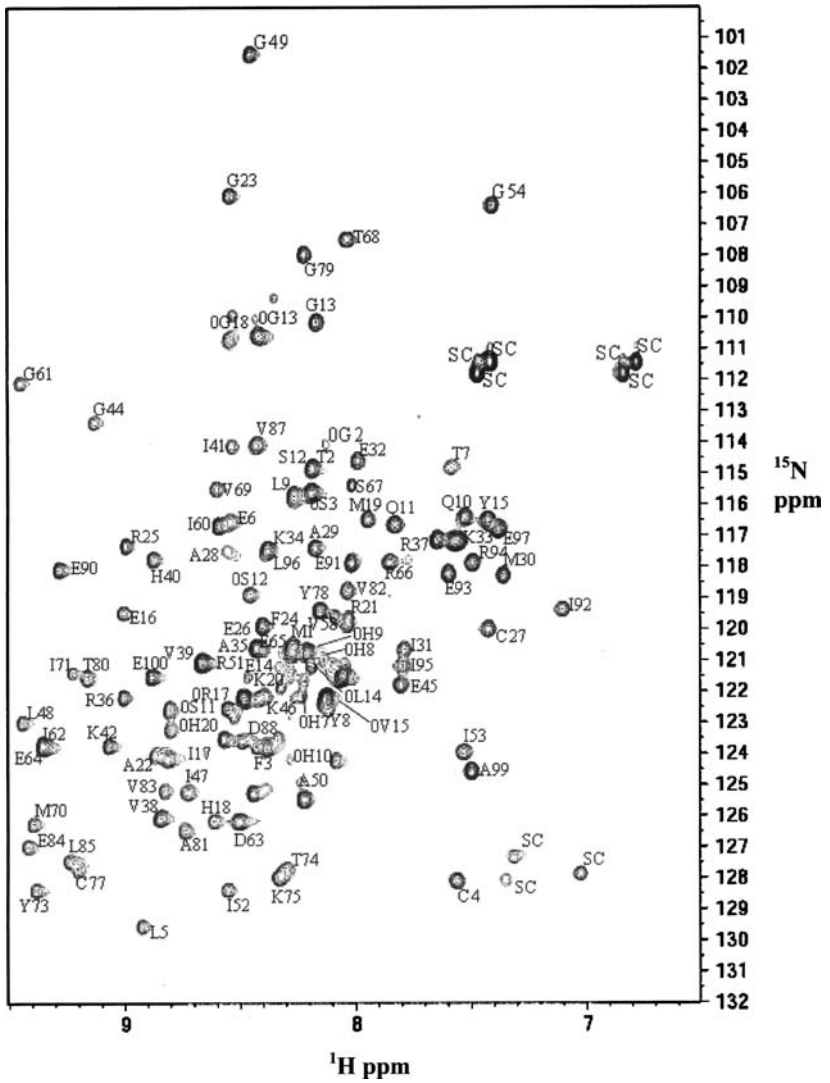

Figure 1. (a) $^{1}H-^{15}N$ HSQC spectrum from 1.5 mM MTH0776 in 700µl of a buffer solution pH 6.8, made up of 50 mM NaH_2PO_4, 300 mM NaCl, 70µl D_2O, 1 mM DSS, 15 mM DTT and 10µl of a 3% solution of sodium azide, collected on a Varian Inova 500 MHz spectrometer, The Oxxx labels indicate residues belonging to the poly-His tag.

applied to the upper bounds for NOEs involving non-stereospecifically assigned methylene protons, 1.0 Å for NOEs involving methyl protons and 2.0 Å for NOEs involving non-stereospecifically assigned protons on opposite sides of aromatic rings. Torsion angle restraints were generated using TALOS (Cornilescu and Delaglio, 1999) based on the observed $^{13}C\alpha$, $^{13}C\beta$, $^{1}H\alpha$, and ^{1}HN chemical shifts. A total of 84 ϕ backbone torsion angle restraints were used in combination with $^{3}J_{HNH\alpha}$ coupling constant constraints obtained from a 3D HNHA experiment. Backbone ϕ angles were assigned an uncertainty of $\pm 10°$ for residues in well-defined helical or beta sheet regions. A total of 80 backbone ψ dihedral angle restraints were derived via TALOS and assigned an uncertainty of $\pm 30°$. Hydrogen bond restraints ($d_{O-HN} = 1.8–2.7$ Å and $d_{O-N} = 2.8–3.7$ Å A for each hydrogen bond) for slowly exchanging amide protons were identified via several methods including a D_2O exchange ^{15}N HSQC experiment, the pattern of sequential and inter-strand NOEs involving ^{1}HN and $C_\alpha H$ protons (Wuthrich, 1986) and from the chemical shift index (Wishart and Sykes, 1994).

Structures for MTH0776 were calculated with X-PLOR-NIH 2.9.6 using centre averaging (Schwieters et al., 2003). Note that the structures presented here include a histidine residue at the N-terminus in addition to the native sequence of the protein. Using the chemical shift index (Wishart and Sykes, 1994), regions of secondary structures in the protein sequence were initially identified from the assigned chemical shifts. Only NOE-derived distance constraints involving the Hβ, Hα and HN atoms of these structured regions along with the full set of intra-residue and sequential NOEs were used in the first stage of structure generation. Initially a set of 60 structures was generated using the *ab initio* simulated annealing protocol (*sa._new.inp*) applying standard default parameters. After this initial step, several ambiguous long-range NOE assignments were clarified by analyzing only those structures that exhibited good geometrical and energy parameters. Subsequently the same simulated annealing protocol was repeated with additional and/or corrected NOE data sets to generate a new set of 60 structures. NOE violations were again corrected (usually by extending their upper distance limits) after visually reassessing the corresponding NOE spectral intensities. Typically these problematic NOEs were borderline cases falling between the strong/medium, medium/weak or weak/very weak NOE intensity categories. In the third step, dihedral angles and hydrogen bond restraints were added, and the *sa_new.inp* protocols were run again to generate another 60 structures. These structures were then refined using the *refine.inp* protocol in X-PLOR. The final set of 20 structures was selected on the basis that no inter-proton distance restraint violation could be greater than 0.5 Å, no angle violation could be greater than 5°, no bond-length violation could be greater than 0.01 Å and no bond angle violations from ideality could be greater than 2°. A total of 1844 NOE-derived distance restraints (578 long range, 422 medium range, 698 sequential and 146 intra-residue NOEs), 164 dihedral angle restraints, and 40 hydrogen bond restraints were used to generate the final structural ensemble. The final set of 20 of structures was analyzed using PROCHECK-NMR (Laskowski et al., 1996) and MOLMOL (Koradi et al., 1996). SuperPose (Maiti et al., 2004) was used to visualize all the structures and to calculate RMSD values. The three dimensional coordinates for the final 20 structures have been deposited in the Protein Data Bank (PDB) under accession number 1Z9V. Statistical parameters for the ensemble of 20 calculated structures are presented in Table 1. All structures (Figure 2a and b) exhibit good covalent geometry, and for all 20 structures, more than 99.0% of the main chain (ϕ, ψ) angles fall in the core or allowed regions of the Ramachandran plot (Table 1) as determined using PROCHECK-NMR.

MTH0776 is an alpha/beta protein and is made up of six β-strands with three α-helices forming the secondary structural topological arrangement

Table 1. Structural statistics for the final ensemble of 20 structures for MTH0776

Distance restraints	
All	1844
Intra residue	146
Sequential ($\mid i-j \mid = 1$)	694
Medium range ($2 \leq \mid i-j \mid \leq 5$)	422
Long range ($5 < -i-j\mid$)	578
Dihedral angle restraints	
All	164
Phi (Φ)	84
Psi (ψ)	80
Hydrogen bonds	40
RMSD from ideal geometry	
Bonds (Å)	0.00258 ± 0.00005
Angles (deg)	0.5865 ± 0.0047
Impropers (deg)	0.4525 ± 0.0033
RMSD from experimental data	
NOEs (Å)	0.0461 ± 0.0012
Dihedrals (deg)	0.7216 ± 0.0238
Ramachandran plot analysis (without PMF refinement)	
Most favoured	76.0%
Allowed	21.5%
Generously allowed	2.5%
Disallowed	0.1%
RMSD from the mean structure (Å)	
All residues	
Backbone	0.51 ± 0.12
All heavy atoms	1.49 ± 0.13
Equivalent resolution for stereochemical parameters (Å)	
Percentage residues in A, B, L	2.7
H-bond energy	1.8
Chi-1 pooled	2.4
Chi-2 trans	2.5

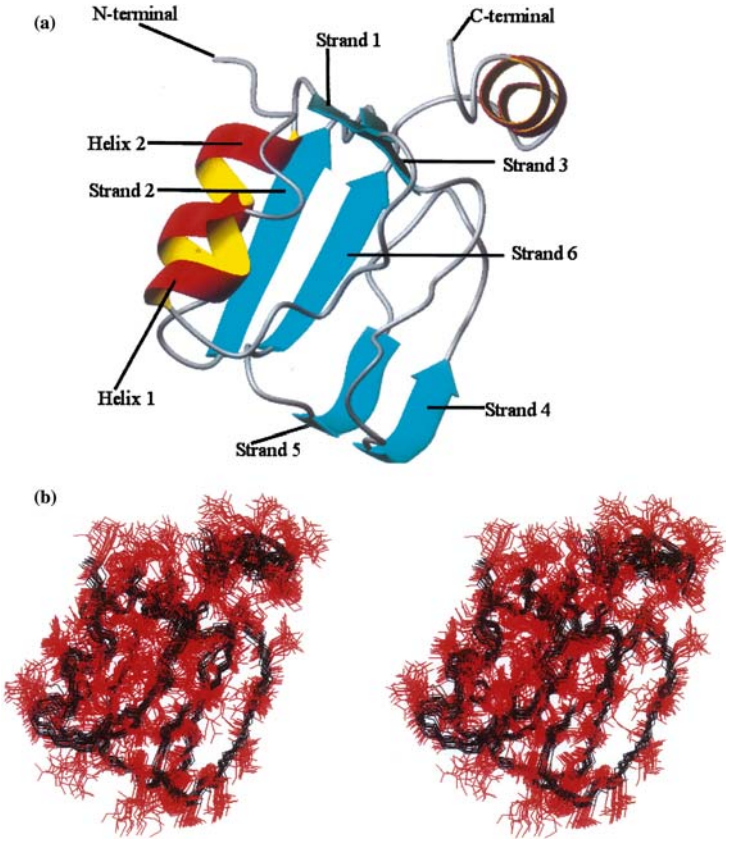

Figure 2. (a) Ribbon diagram of MTH0776 showing the overall fold and the unusual topology of its β-strands and helices. (b) Stereo view of the NMR structural ensemble of the 20 structures of MTH0776 including all heavy atoms. Backbone atoms are coloured black and the side chain atoms are coloured red.

of α_1-β_1-α_2-β_2-β_3-β_4-β_5 -β_6-α_3 (Figure 1a). The β-strands include residues 21–24 (β_1), 37–43 (β_2), 47–52 (β_3), 60–65 (β_4), 69–74 (β_4) and 79–85 (β_6). The first helix (α_1) is the shortest and runs from residues 9–12. The second helix (α_2) runs from residue 27–33 while the third helix (α_3) is the longest and runs from residue 90–99. The structure of MTH0776 is more or less globular in shape with strands β_1 and β_3 being antiparallel and on one side while four strands, (β_2 and β_6 (parallel) and β_4 and β_5 (antiparallel) are on the opposite side. A representative MTH0776 structure was searched against the CE (Shindyalov and Bourne, 1998), DALI (Holm and Sander, 1996) and SCOP (Lo Conte et al., 2002) databases. The results obtained (a best Z score of 1.4 and RMSD of 4.37 Å) indicate no statistically significant match is possible. Therefore it appears that MTH0776 (and other members of COG 4033) exhibit a completely novel fold. Unfortunately, the novel fold makes it difficult to infer any function for MTH0776 on the basis of the NMR structure reported here.

Sequence analysis of MTH0776 was done with BLAST (Altschul et al., 1990), CLUSTALX (Thompson et al., 1997) and CGView (Stothard and Wishart, 2005), A BLAST search for proteins related to MTH0776 yielded hits which were found only in Archaea. All of these proteins were identified as uncharacterised or hypothetical proteins, belonging to the clustered orthologous group (COG) 4033. A multiple sequence alignment of these proteins was performed with CLUSTALX. Gene position and gene order analysis for MTH0776 and MTH0777 was done visually using genome maps generated by CGView for a total of 11 different methanogenic organisms. Our results indicate that these two proteins are always encoded by adjacent genes, although their order does change.

The conserved proximity of MTH0777 and MTH0776 among all methanogenic archaea sug-

gested that these two proteins may be co-expressed and that they may bind or interact with one another. To check this hypothesis, purified samples of ^{15}N-labelled MTH0776 and unlabelled MTH0777 were prepared (Amegbey et al., 2004). Excess MTH0777 was then added to a 1 mM solution of MTH0776 and dialysed into a buffer consisting of 50 mM NaH_2PO_4, 300 mM NaCl and 15 mM DTT. The protein mixture was then prepared for a ^{15}N-HSQC experiment and the results compared to that of a pure ^{15}N-labelled MTH0776 protein sample in the identical buffer. The preliminary NMR evidence suggests that the two proteins may interact with one another (see supplementary material). ^{15}N T_2 measurements were also done on the purified ^{15}N-labelled MTH0776 protein as well as the ^{15}N-labelled MTH0776/unlabelled MTH0777 complex. The average T_2 relaxation time decreased from 120 ms for the MTH0776 monomer to 65 ms for the putative MTH0776/MTH0777 complex.

In addition to these structural, sequential and spectroscopic analyses, samples of purified MTH0776 (alone) and MTH0776/MTH0777 (in complex), were tested for catalytic activity using a battery of enzyme assays developed at the University of Toronto. These included phosphatase, esterase/lipase, protease, dehydrogenase, oxidase and phosphodiesterase/nuclease activity. Our results indicated that neither MTH0776 nor the putative MTH0776/MTH0777 complex had any of the activities tested.

Discussion and conclusions

The structure of MTH0776 was solved as part of a structural proteomics project initiated by the Ontario Cancer Institute in 1999. The primary mandate of this project was to investigate previously unknown, unidentified or unclassified proteins from a thermophilic archeon (*M. thermoautotrophicum*). One of the key objectives of this project has been to identify novel or never-before-seen protein folds. Given the unique ecological niche occupied by this organism and the high percentage of unknown or unclassifiable sequences in its genome, there was a general expectation that many of these 'unknown' proteins would yield very novel protein structures. Interestingly, of the ~40 *M. thermoautotrophicum* structures solved to date, less than 10% of the structures exhibit truly novel folds (Yee et al., 2003). MTH0776, therefore, is one of the few proteins from *M. thermoautotrophicum* that exhibited a never-before-seen fold, The discovery of novel protein folds is actually becoming increasingly rare, Statistics collected from the PDB and analyses performed on fold classes have shown a steady decline in the percentage of novel folds deposited into the PDB (http://scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-l.67). On average fewer than 3% of newly deposited proteins have distinct or previously unidentified protein folds.

The identification of novel or unique folds among proteins is important, particularly given the intended role of structural proteomics in addressing the protein folding problem. It is quite apparent that the protein fold universe is finite, with perhaps fewer than 2000 unique or distinct folds existing among all living organisms (Govindarajan et al., 1999). To date we have found perhaps 1000 of these folds and so any extension of the 'fold space' provides theoreticians with important new information about what kinds of topologies, packing and secondary structure connections are possible or 'allowed'.

Of course the other key objective of any structural proteomics initiative is to use structure to help with the determination of function. Given that the structure of MTH0776 appears to be unique, this largely makes it impossible to determine the protein's function via structural homology. Therefore a number of ancillary investigations were undertaken to determine the possible nature/function of MTH0776 through sequence comparisons, binding assays and enzymatic tests.

Disappointingly, a large battery of enzymatic assays covering many common enzymatic functions yielded no useful results. Likewise sequence searches using BLAST, PSI-BLAST and PFAM comparisons yielded little in terms of useful information. Interestingly, the results from our gene order or gene synteny analyses show that both MTH0776 and MTH0777 (or their homologues) were conserved in all methane metabolizing archaebacteria. This level of syntenic conservation is rare and is often a good indication that these two proteins form a two-component operon system.

Given that many proteins in two-component operons often physically bind to each other, we chose to investigate the possible interactions

between MTH0776 and MTH0777 via NMR. The $^{1}H-^{15}N$ HSQC spectrum collected on the putative complex indicates substantial changes in the chemical shifts, intensities and linewidths for a number of amino acid residues that map to a well-defined region on the MTH0776 protein (supplementary material). We also conducted ^{15}N T_2 measurements with both the MTH0776 monomer and the MTH0776/MTH0777 complex, and the results confirm that majority of the residues in MTH0776/MTH0777 complex have shorter T_2's that are characteristic of a molecule of about 30 kD. While more work needs to be done to fully investigate these spectral perturbations, these preliminary data suggests that MTH0776 and MTH0777 bind to each other in solution.

In summary we have presented the solution structure of MTH0776, a functionally unknown protein from *M. thermoautotrophicum*. By all measures this fold is unique and it now allows the structures of at least 11 other related proteins to be modeled using the MTH0776 template. Efforts are ongoing to determine its precise function as well as to further characterise its molecular interaction with MTH0777. Supporting information is available.

Electronic supplementary material is available at http://dx.doi.org/10.1007/s10858-005-1275-5.

Acknowledgements

The authors wish to thank Shan Sundararaj and Trent Bjorndahl for their help with SSASS, an in-house program to aid the sequential assignment and with useful discussions in the structure determination. We would like to thank the Canadian National High Field NMR Centre (NANUC) for their assistance and use of the facilities. Operation of NANUC is funded by the Canadian Institutes of Health Research, the Natural Science and Engineering Research Council of Canada and the University of Alberta, GYA was supported by a Pfizer graduate student scholarship as well as financial support from PENCE for the project. AY is supported by the NIH Protein Structure Initiative grant P50 GM62513-05 to the Northeast Structural Genomics Consortium, the Ontario Research and Development Challenge Fund, and Genome Canada.

References

Altschul, S.E., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.

Amegbey, G., Chang, Z., Stothard, P., Yee, A., Arrowsmith, C. and Wishart, D.S. (2004) *J. Biomol. NMR*, **30**, 459–460.

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

Govindarajan, S., Recabarren, R. and Goldstein, R.A. (1999) *Proteins*, **35**, 408–414.

Holm, L. and Sander, C. (1996) *Science*, **273**, 595–602.

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

Laskowski, R.A., Rullman, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J. (1996) *J. Biomol. NMR*, **8**, 477–486.

Lo Conte, L., Brenner, S.E., Hubbard, T.J.P., Chothia, C. and Murzin, A. (2002) *Nucl. Acid Res.*, **30**, 264–267.

Maiti, R., Van Domselaar, G., Zhang, H. and Wishart, D. (2004) *Nucl. Acids Res.*, **32**, W590–W594.

Schwieters, C.D., Kuszewski, J.J., Tjiandra, N. and Clore, G.M. (2003) *J. Magn. Res.*, **160**, 66–74.

Shindyalov, I.N. and Bourne, P.E. (1998) *Protein Eng.*, **11**, 739–747.

Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Alldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D. and Reeve, J.N. (1997) *J. Bacteriol.*, **179**, 7135–7155.

Stothard, P. and Wishart, D.S. (2005) *Bioinformatics*, **21**, 537–539.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmorgin, F. and Higgins, D.G. (1997) *Nuc. Acids Res.*, **25**, 4876–4882.

Wishart, D.S. and Sykes, B.D. (1994) *Meth. Enzymol.*, **239**, 363–392.

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

Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M. and Arrowsmith, C.H. (2003) *Acc. Chem. Res.*, **36**, 183–189.

Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR*, **4**, 845–858.