Journal of Biomolecular NMR, **21:** 63–66, 2001.
KLUWER/ESCOM

Letter to the Editor: NMR-based structure of the conserved protein MTH865 from the Archaeon *Methanobacterium thermoautotrophicum*

Gregory M. Lee[a], Aled M. Edwards[b,c], Cheryl H. Arrowsmith[b] & Lawrence P. McIntosh[a,*]
[a]*Department of Biochemistry and Molecular Biology, Department of Chemistry, Biotechnology Laboratory, and the Protein Engineering Network of Centres of Excellence, University of British Columbia, Vancouver, BC, Canada V6T 1Z3;* [b]*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; and* [c]*Banting and Best Department of Medical Research, C.H. Best Institute, University of Toronto, Toronto, ON, Canada M5G 1L6*

Received 29 May 2001; Accepted 11 July 2001

Key words: helical bundle, heteronuclear NMR, *Methanobacterium thermoautotrophicum*, structural proteomics

Biological context

As part of a structural proteomics effort, the solution structure of MTH865 from *Methanobacterium thermoautotrophicum* ΔH (henceforth MTH) has been determined using NMR spectroscopy. This protein was one of many targets surveyed in an initiative aimed at establishing the feasibility of a high-throughput NMR analysis of all non-membrane proteins encoded by the MTH genome. Since the inception of this program in 1999, 638 of a possible 1871 open reading frames have been cloned and expressed, leading to the determination of 7 X-ray crystal structures and 13 NMR-derived ensembles of unique MTH proteins (Christendat et al., 2000; unpublished data).

MTH865 is an archaeal protein of unknown function (Smith et al., 1997). A BLAST (Altschul et al., 1990) search indicates that it is conserved, showing significant sequence homology to three predicted archaeal proteins, also of unknown function (Figure 1). These include Vng074c from *Halobacterium* sp. *NRC-1*, MJ0738 from *Methanococcus jannaschii*, and a fragment of AQ2056 from *Aquifex aeolicus*. As with MTH, these archaea are classified as extremophiles: *Halobacterium* sp. *NRC-1* is an extreme halophile, *M. jannaschii* is a thermophile that grows at extreme pressures, and *A. aeolicus* is a hyperthermophile. In addition, there is limited sequence homology to small segments of two bacterial ABC transporter proteins, StrW from *Streptomyces glaucescens* and histidine permease, HisP, from *Salmonella typhirmurium*. To ascertain whether clues to the function of MTH865 could be obtained through structural features or relationships not apparent at the level of sequence, a detailed analysis of the protein was performed using NMR spectroscopy.

*To whom correspondence should be addressed. E-mail: mcintosh@otter.biochem.ubc.ca

Methods and results

Expression, initial purification by Ni^{2+}-affinity chromatography, and proteolytic cleavage of isotopically labeled MTH865 was as previously described (Mackereth et al., 2000). Additional purification was achieved by anion exchange HPLC (Poros 20HQ resin), using 50 mM Tris-HCl buffer (pH 7.5) and a 0 to 1 M KCl gradient. Prior to the NMR data collection, the protein was dialyzed and concentrated in 50 mM phosphate buffer (pH 6.5) containing 50 mM NaCl, 5 mM 1,4-dithiolthreitol, and 3 mM sodium azide. Electrospray ionization mass spectrometry revealed a mass of 8675 Da for the unlabeled protein, consistent with that of full length MTH865, plus three additional N-terminal residues (Gly-Ser-His) remaining after proteolytic cleavage of the His_6 affinity tag.

Samples of ca. 0.5 mM uniformly ^{15}N or $^{15}N/^{13}C$-labeled protein in the buffer described above, with either 10% or 99% D_2O for signal lock, were used for the NMR studies. Most spectra were recorded at 30 °C on a 500 MHz Varian Unity spectrometer equipped with a triple resonance z-gradient probe. Spectra were processed and analyzed with FELIX 2000 (Molecular Simulations, Inc., San Diego, CA). Backbone and side-chain spectral assignments, including those of stereospecific methyl and β-methylene protons, were obtained as previously described (Johnson et al., 1996; Mackereth et al., 2000).

In addition to a 3D ^{15}N-NOESY-HSQC (τ_m = 150 ms) recorded at 500 MHz, a simultaneous ^{15}N- and ^{13}C-NOESY-HSQC (τ_m = 120 ms; Pascal et al., 1994) was acquired on an 800 MHz Varian INOVA

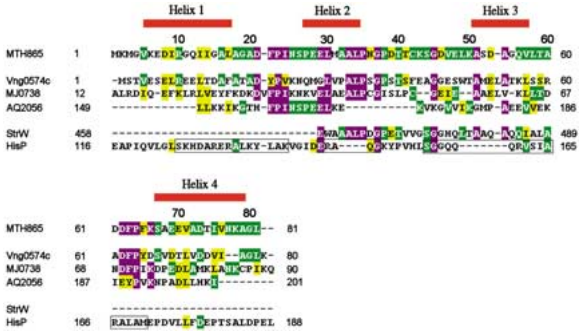

Figure 1. Sequence alignment of MTH865 (GenBank accession AAB85363) with three predicted archaeal proteins, Vng0574c (*Halobacterium* sp. *NRC-1*, accession AAG19089), MJ0738 (*Methanococcus jannaschii*, accession AAB98733) and AQ2056 (*Aquifex aeolicus*, accession AAC07806), all of unknown function. Limited similarity to the ABC transporter proteins StrW (*Streptomyces glaucescens*, accession CAA61417) and HisP (*Salmonella typhirmurium*, PDB accession 1B0U) is also detected. Regions of high sequence homology amongst the archaeal genomes are highlighted in purple. Identical and similar residues to MTH865 are highlighted in green and yellow, respectively. Red bars indicate helical regions of MTH865, while those in the crystal structure of HisP are boxed within its sequence. The figure was created using GENEDOC (Nicholas and Nicholas, 1997) and by visual inspection.

spectrometer, housed at the National High Field Nuclear Magnetic Resonance Centre (University of Alberta, Edmonton, AB, Canada). A constant time $^{13}C/^{13}C/^{1}H$-methyl-methyl NOESY ($\tau_m = 140$ ms; Zwahlen et al., 1998) was acquired on a 600 MHz Varian INOVA spectrometer, housed at the University of Toronto. Peak intensities from the NOESY spectra yielded a final set of 1492 unambiguous and 544 ambiguous distance restraints using 17 iterations (two complete cycles) of ARIA/CNS version 1.0 (Brünger et al., 1998; Nilges and O'Donoghue, 1998). Fifty pairs of backbone dihedral restraints were derived from $^{1}H\alpha$ and $^{13}C\alpha$ secondary chemical shifts using TALOS (Cornilescu et al., 1999), in concert with $^{3}J_{HN-H\alpha}$ coupling constants measured from a 2D HMQC-J spectrum. Twenty-six pairs of hydrogen bond restraints were applied during all 17 ARIA iterations to residues located in regions of helical secondary structure whose backbone $^{1}H^{N}$ were protected for at least 30 min after transfer from H_2O to D_2O buffer. The final 50-structure ensemble was calculated with CNS, starting from randomized coordinates and using r^{-6} sum-averaged ARIA-derived unambiguous and ambiguous distance restraints (Table 1).

Relaxation data, measured and analyzed as previously described (Farrow et al., 1994), indicated that MTH865 is a compact monomeric protein (τ_m = 4.7 ns). The major structural feature of MTH865 is a bundle formed by three well-defined α-helices (H1, Val5–Leu17; H2, Ser26–Leu34; H4, Ser67–Ala79) and an irregular helix (H3, Ala51–Val57) (Figure 2A). H1 is orthogonally aligned antiparallel to H4, and H2 antiparallel to H3. Of these, only H3 and H4 are distinctly amphiphilic. A hairpin (Cys42–Leu49) centered on a turn (Ser44–Val47) is present between H2 and H3. This hairpin may be composed of short antiparallel β-strands (residues 42–44 and 46–49) based upon $^{13}C\alpha$, $^{13}C'$, and $^{1}H\alpha$ secondary chemical shifts and $^{3}J_{HN-H\alpha}$ values >9 Hz. Several inter-strand NOE interactions were also observed, including that between Lys43 $^{1}H\alpha$ and Glu48 $^{1}H\alpha$. Furthermore, the amide protons of Val47 and Leu49 were moderately protected from hydrogen–deuterium exchange, possibly resulting from hydrogen bonds involving Cys42 C′/Leu49 H^N and Ser44 C′/Val47 H^N (not used as restraints). However, as evident by high RMS deviations within the structural ensemble (Figure 2A), the precision by which the structure of this region could be defined was limited due to a low number of observable NOEs. This may be explained, in part, by H_2O-exchange, leading to pronounced broadening of the ^{15}N-HSQC peaks of several residues in or near the hairpin. In addition, ^{15}N relaxation measure-

Table 1. NMR restraints and structural statistics of the ensemble calculated for MTH865

Summary of restraints		
NOEs: unambiguous (ambiguous)		
ARIA restraints		
Intraresidue		546 (159)
Sequential		321 (104)
Medium range ($1 < \|i-j\| < 5$)		291 (164)
Long range ($\|i-j\| \geq 5$)		334 (117)
Total		1492 (544)
Dihedral angles		
ϕ,ψ		50, 50
Hydrogen bonds		26 × 2
Deviation from restraints		
NOE restraints (Å)		0.02 ± 0.00
Dihedral restraints (°)		0.54 ± 0.05
Residues in allowed regions of Ramachandran plot[a]		96.0%
Deviation from idealized geometry		
Bonds (Å)		0.01 ± 0.00
Angles (°)		0.53 ± 0.01
Improper angles (°)		0.86 ± 0.02
Dihedral angles (°)		15.55 ± 0.42
Mean energies (kcal mol^{-1})		
E_{vdw}		123.66 ± 7.73
E_{bonds}		15.22 ± 1.00
E_{angle}		92.71 ± 3.98
E_{impr}		18.89 ± 1.57
E_{NOE}		62.75 ± 3.08
E_{cdih}		3.64 ± 0.65
RMSD from average structure (Å)	Residues 5–81	Helical regions
Backbone (N, Cα, C′)	0.48 ± 0.11	0.32 ± 0.08
All heavy atoms	0.97 ± 0.13	0.91 ± 0.22

[a]Calculated from PROCHECK_NMR (Laskowski et al., 1996), excluding Gly, Pro, and the N-terminus.

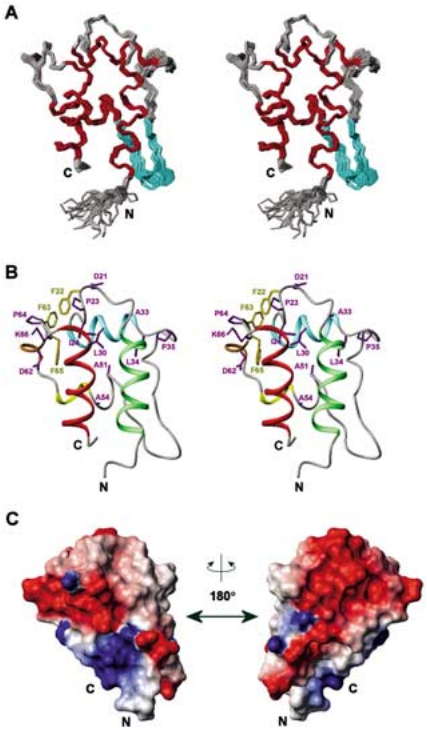

Figure 2. (A) A stereoimage of the MTH865 NMR-derived structure ensemble reveals a bundle of four helices. The top 25 of 50 structures are superimposed over the helical regions (red), with a mean backbone RMSD of 0.32 ± 0.08 Å versus the average structure. The hairpin, corresponding to Cys42-Leu49, is also highlighted (cyan). (B) A stereoimage of the energy minimized average MTH865 structure depicts the helical bundle (H1 green, H2 cyan, H3 yellow, H4 red). The most highly conserved residues amongst the four archaeal genomes (purple), including the three phenylalanines (yellow), help stabilize the hydrophobic core. (C) A space-filling model of MTH865 depicting the surface charge. The figure on the left is displayed in the same orientation as A and B. Regions of negative and positive charge are shown in red and blue, respectively, while those of exposed hydrophobic residues are white or lightly colored. These figures were generated using MOLMOL (Koradi et al., 1996).

ments (Supplementary material) indicated conformational mobility on a millisecond timescale (R_{ex} terms) for several amides in the hairpin as well as the adjacent H1. Enhanced nanosecond–picosecond timescale motions are also present for this region, as evident by $^{1}H\{^{15}N\}$ heteronuclear NOE values of ∼0.6 for residues 43 and 45–47 (Supplementary material).

Discussion and conclusions

In the absence of direct biological information, clues to the function of MTH865 are provided by (1) sequence conservation to other archaeal proteins, (2) structural data, including surface features, (3) the genomic organization of MTH, (4) sequence relationship to ABC-transporter membrane proteins, and (5) distant structural relationships with other known proteins involved with nucleotide binding and metabolism.

MTH865 was initially classified as a conserved protein based upon close sequence similarities to Vng0574c from *Halobacterium* sp. *NRC-1* and MJ0738 from *M. jannaschii*. It is likely that all three archaeal proteins have similar structures and functions. Notable in this sequence alignment are two conserved turns containing DFPX moieties (residues 21–24 and 62–65), where X is a hydrophobic residue, which are adjacent in the MTH865 tertiary structural ensemble. The side chains of the Phe and Pro residues are also exposed and form part of the hydrophobic surface of the protein. Another conserved region amongst the archaeal proteins, comprises a $ELX_1X_2ALPX_3GP$ moiety (residues 29–38), where X_1 and X_3 are hydrophobic and polar residues, respectively, and corresponds to H2 and the succeeding turn in MTH865. Although AQ2056 from *A. aeolicus* is a more distant relative, residues 19–30 of MTH865 and 155–166 of AQ2056 share a nearly exact sequence, GXXFPINSPEEL, suggesting that the two proteins share a similar fold, at least in this segment. These and other conserved residues (Figure 2B) help stabilize the hydrophobic core of MTH865.

MTH865 is a compact, monomeric 81-residue protein, suggesting that it is too small to be an enzyme. Therefore, MTH865 could serve as a regulatory or transport molecule or function as a component of a heteromeric macromolecular complex. In agreement with its predicted pI of 4.24, MTH865 is net negatively charged at neutral pH. However, an examination of its surface reveals a distinct dipolar charge distribution (Figure 2C). Four basic residues (Lys2, Lys6, Arg10 and Lys78), located in the interface between H1 and H4, defines a small positively charged patch. By contrast, most of the remainder of the surface of MTH865 is negatively charged. Furthermore, with the exception of Arg10 and Lys66, all the residues conserved amongst the archaeal proteins reside on this negatively charged surface of the MTH865 structure. Two major regions of exposed hydrophobic side chains are also observed in the structural ensemble. The first, which corresponds to the C-terminal ends of, and the loop between, H1 and H2 and the loop preceding H4, contains both DFPX moieties as well as the potentially oxidizable side chain of Cys42. The second patch corresponds to the interface between H1, H3 and H4 closest to the N- and C-termini. These

distinct polar and non-polar regions could serve as interfaces for possible interactions of MTH865 with cellular partners.

A potential clue that MTH865 may function as a subdomain of a nucleotide, nucleoside, or purine binding protein follows from the organization of the MTH genome. Specifically, there is an overlap in the codons that encode the last residue of MTH866, a putative adenine deaminase, and the first residue of MTH865. However, while possibly translated from the same mRNA (Brown et al., 1989), it is unclear what, if any, structural or functional interactions exist between MTH865 and MTH866, particularly since the structure of the latter has yet to be determined.

Partial sequence homology to segments of two bacterial ABC-transporter membrane proteins, StrW and a histidine permease, HisP, provides further circumstantial evidence of a role for MTH865 in adenine nucleotide binding. Similarities between MTH865 and HisP extend to structure, as HisP also contains a helical bundle, although somewhat out of register with respect to its sequence alignment with MTH865 (Figure 1). This bundle is part of the HisP ATP-binding domain, but does not directly contact the nucleotide. Consistent with this relation, MTH865 does not appear to bind adenine nucleotides, as evident by the lack of any spectral perturbations upon addition of excess quantities of these ligands (data not shown).

Distant structural relationships between MTH865 and other proteins involved with nucleotide binding and metabolism are also observed. A DALI (Holm and Sander, 1993) structural search yielded three such proteins, all containing helical bundles: fragments of elongation factor Ts, an ATP-binding subunit of GroEL, and NusB. However, these three proteins scored low by structural (Z factors $\sim$ 2.0 and RMSD $>$ 2.5 Å) and sequence comparisons ($<$ 20% conservation). This highlights the difficulty in inferring the function of a protein based on a relatively common fold such as a helical bundle.

In summary, we have determined the solution structure of MTH865 via NMR spectroscopy and have shown that it has the architecture of a four-helix bundle. Partial sequence and structural homology suggests that MTH865 may be related to nucleotide-binding proteins and, based upon the genomic organization of MTH, is possibly a structural domain to MTH866, a putative adenine deaminase. Beyond this circumstantial evidence, we find no definitive answers regarding the structure-function relationships of MTH865. This is due to the limited number of known homologs and the absence of any supplementary data on the biological role of MTH865.

The chemical shifts have been submitted to the BMRB (accession # 4996); the structure ensemble has been submitted to the PDB (accession # 1IIO).

Acknowledgements

The authors wish to thank C. Mackereth, S. Gagné, A. Yee, R. Muhandiram, L. Kay, M. Nilges, and J. Linge for their technical assistance. This work was funded in part by a grant from the National Cancer Institute of Canada with funds from the Canadian Cancer Society and the Ontario Research and Development Challenge Fund. C.H.A., A.M.E. and L.P.M. are Canadian Institute of Health Research Scientists, and L.P.M. is an Alexander von Humbolt Fellow. The Protein Engineering Network of Centres of Excellence provided instrument support.

Supplementary material

Figures containing an annotated ^{15}N-HSQC spectrum and ^{15}N relaxation data are available upon request from the authors.

References

Altschul, S.F., Gish, W., Miller, et al. (1990) *J. Mol. Biol.*, **215**, 403–410.

Brown, J.W., Daniels, C.J. and Reeve, J.N. (1989) *CRC Crit. Rev. Microbiol.*, **16**, 287–338.

Brünger, A.T., Adams, P.D., Clore, G.M., et al. (1998) *Acta Crystallogr.*, **D54**, 905–921.

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.

Farrow, N.A., Muhandiram, R., Singer, A.U., et al. (1994) *Biochemistry*, **33**, 5984–6003.

Holm, L. and Sander, C. (1993) *J. Mol. Biol.*, **233**, 123–128.

Johnson, P.E., Joshi, M.D., Tomme, P., et al. (1996) *Biochemistry*, **35**, 14381–14394.

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

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., et al. (1996) *J. Biomol. NMR*, **8**, 477–486.

Mackereth, C.D., Arrowsmith, C.H., Edwards, A.M. and McIntosh, L.P. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 6316–6321.

Nicholas, K.B. and Nicholas Jr., H.B. (1997) *http://www.psc.edu/biomed/genedoc/*

Nilges, M. and O'Donoghue, S. (1998) *Prog. NMR Spectrosc.*, **32**, 107–139.

Pascal, S.M., Muhandiram, D.R., Yamazaki, T., et al. (1994) *J. Magn. Reson.*, **B103**, 197–201.

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

Zwahlen, C., Gardner, K.H., Sarma, et al. (1998) *J. Am. Chem. Soc.*, **120**, 7617–7625.