NMR structure of the conserved hypothetical protein TM0979 from Thermotoga maritima

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2005

Impact Factor: 2.63 · DOI: 10.1002/prot.20352 · Source: PubMed

5 AUTHORS, INCLUDING:

Torsten Herrmann
French National Centre for Scientific Resea…
74 PUBLICATIONS **3,265** CITATIONS
SEE PROFILE

Olga Zagnitko
Sanford Burnham Prebys Medical Discover…
22 PUBLICATIONS **4,332** CITATIONS
SEE PROFILE

Kurt Wüthrich
The Scripps Research Institute
742 PUBLICATIONS **75,935** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 59:387–390 (2005)

STRUCTURE NOTE

NMR Structure of the Conserved Hypothetical Protein TM0979 from *Thermotoga maritima*

Wolfgang Peti,[1,*] Torsten Herrmann,[1] Olga Zagnitko,[2] Slawek K. Grzechnik,[2] and Kurt Wüthrich[1]
[1]*The Scripps Research Institute, Department of Molecular Biology and the Joint Center of Structural Genomics, La Jolla, California*
[2]*University of California San Diego, Center for Research in Biological Systems and the Joint Center of Structural Genomics, La Jolla, California*

Introduction. The TM0979 gene of *Thermotoga maritima*[1] encodes a 87-residue conserved hypothetical protein with a molecular weight of 9876 Da and a calculated isoelectric point of 4.7. There is no functional assignment available for this gene. In this manuscript the NMR structure of TM0979 is described. In the context of the Joint Center of Structural Genomics (JCSG)[2,3] structural proteomics initiative, TM0979 was identified as a potential structural proteomics target[3] on the basis of a high-quality one-dimensional (1D) ^{1}H NMR spectrum. Single crystals were obtained, but the resolution was too low to use X-ray diffraction for the structure determination. The structure determination of TM0979 was therefore pursued using NMR spectroscopy. Based on the resulting three-dimensional (3D) structure and the amino acid sequence, we propose that TM0979 is structurally related to the DsrH protein of *Chromatium vinosum*, which has been implicated in intracellular sulfur oxidation and/or siroheme assembly.[4]

The 3D structure of TM0979 (Fig. 1) was determined using standard NMR methods for the sequence-specific backbone and side chain assignments.[5,6] NOESY peak picking, NOESY peak assignment and 3D structure calculation were performed automatically, using the ATNOS/CANDID/DYANA software package.[7–9] Characteristic parameters for the 3D NMR structure determination of TM0979 are summarized in Table I, and a visual display of the molecular structure is afforded by Figure 1.

The TM0979 structure contains in the sequence order α1–β1–α2–β2–α3–(3_{10})–β3–α4 four α-helices, a 3_{10}-helix and a three-stranded parallel β-sheet, which is located in the center of the protein and surrounded by the five helices. Heteronuclear ^{15}N relaxation data show that TM0979 is a rigid protein with intermolecular fluctuations limited to the sub-ns time scale. Only a surface-exposed loop centered on residue E40 and immediately following helix α2, shows increased mobility. Additionally, the C-terminal pentapeptide segment shows increased mobility, as indicated by reduced ^{15}N{^{1}H}-NOE values (data not shown).

On the level of the amino acid sequence, TM0979 is closely related with the DsrH protein of *Chromatium vinosum*, which is encoded as part of a large operon involved in the sulfur dissimilatory oxidation pathway. Although many genes of this operon have been functionally assigned, the three tandem genes encoding the DsrE, DsrF, and DsrH proteins, which are structurally related to each other, have no attached function as yet. Remarkably, the *dsrEFH* cluster as well as the gene order are conserved in proteobacteria and in some methanogenic archaea such as *E. coli* (genes *yheNML*) and *Methanobacterium thermoautotrophicum* (MTH1357,1358,1360). According to Pott and Dahl,[4] the tentative assignment of these proteins to the siroheme assembly is compromised by their apparent absence in *Archaeoglobus fulgidus*,[10] an organism that contains a dissimilatory siroheme sulfite reductase.[11] However, the presence of related genes in *E. coli*, *H. influenzae*, and methanogenic archaea, which all are organisms that do not contain dissimilatory sulfite reductases but can synthesize assimilatory siroheme sulfite and nitrite reductases, is consistent with the possible involvement of the *dsrEFH*-encoded polypeptides in siroheme-dependent pathways. However, there is no indication of either dissimilatory sulfite reductase or a siroheme biosynthesis pathway in the *T. maritima* genome. Therefore, in general, the functional roles of the DsrEFH proteins, including TM0979 (DsrH), remain to be elucidated. Interestingly, the genomic neighborhood of the *yheNML* cluster in *E.coli* and in other proteobacteria includes a peptidyl-prolyl cis–trans isomerase and several other proteins related to the translation machinery.

TM0979 is the only close homolog of dsrEFH proteins

Grant sponsor: National Institutes of Health, Protein Structure Initiative: Grant number: P50 GM62411; FWF (Austrian Science Fund) E. Schrödinger fellowship J2145

Wolfgang Peti's current address is Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, 70 Ship Street, Providence, RI 02912.

*Correspondence to: Wolfgang Peti, The Scripps Research Institute, MB-44, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail: wolfgangpeti@yahoo.com

Received 27 July 2005; Accepted 23 September 2004

Published online 18 February 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20352

TABLE I. Input for the Structure Calculation and Characterization of the Energy-Minimized Bundle of 20 DYANA Conformers of TM0979

Quantity	Value[a]
NOE upper distance limits	1927
Dihedral angle constraints	422
Residual target function [Å^2]	1.70 ± 0.45
Residual NOE violations	
Number ∃ 0.1 Å	28 ± 4 (20–34)
Maximum [Å]	0.14 ± 0.02 (0.13–0.24)
Residual dihedral angle violations	
Number ∃ 2.5 deg	5 ± 1 (3–7)
Maximum [deg]	4.31 ± 0.56 (3.29–5.39)
Amber energies, [kcal/mol]	
Total	−2925.76 ± 64.69
van der Waals	−197.46 ± 10.30
Electrostatic	−3462.83 ± 65.04
RMSD from ideal geometry	
Bond lengths [Å]	0.0082 ± 0.0002
Bond angles [deg]	2.16 ± 0.05
RMSD to the mean coordinates [Å][b]	
bb (3–85)	0.35 ± 0.05 (0.28–0.52)
ha (3–85)	0.82 ± 0.08 (0.70–1.06)
Ramachandran plot statistics (%)[c]	
Most favored region	61.7
Additional allowed region	31.5
Generously allowed region	5.8
Disallowed region	1.1

[a]Except for the top two entries, the average value for the 20 energy-minimized conformers with the lowest residual DYANA target function values and the standard deviation among them are listed, with the minimum and maximum values given in parentheses.
[b]bb indicates the backbone atoms N, C^α, C′; ha stands for "all heavy atoms." The numbers in parentheses indicate the residues for which the RMSD was calculated.
[c]As determined by PROCHECK.[24]

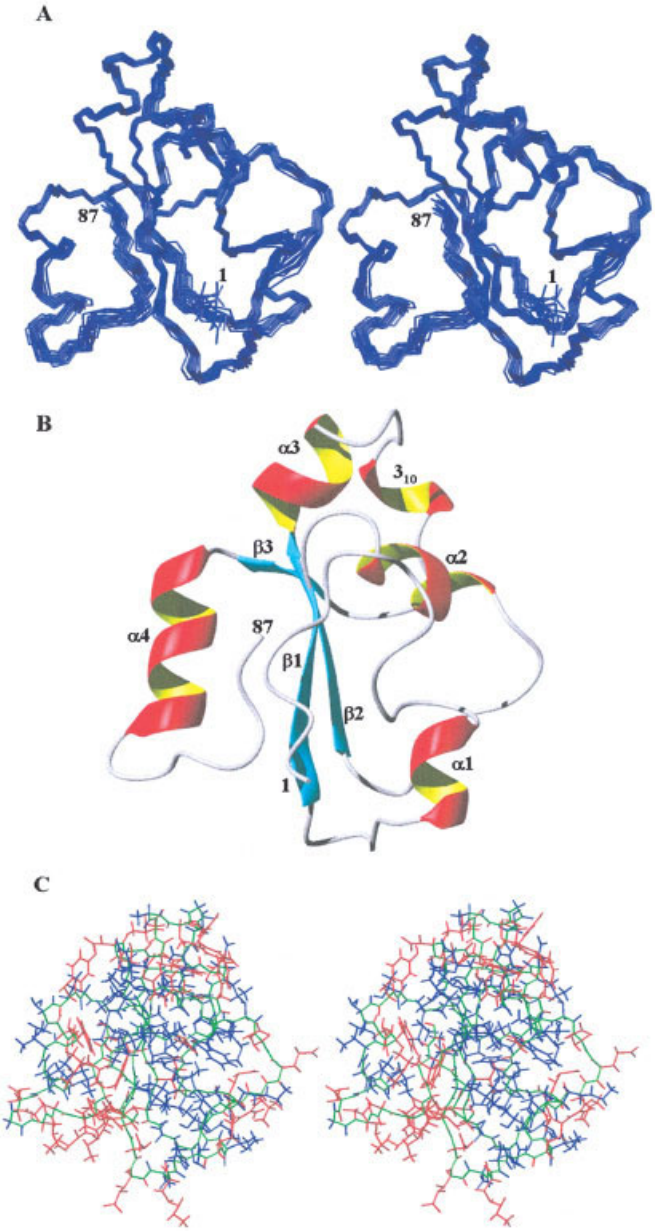

Fig. 1. **A**: Stereo view of the bundle of 20 energy-minimized DYANA conformers representing the NMR structure of TM0979. The superposition is for best fit of the backbone atoms N, C^α and C′ of residues 3–85. The N- and C-terminal residues 1 and 87 are identified. **B**: Ribbon presentation of the closest conformer of TM0979 to the mean coordinates of the bundle of 20 conformers. The regular secondary structure elements are identified. **C**: Closest conformer of TM0979 to the mean coordinates of the bundle of 20 conformers. Color code: green, polypeptide backbone; blue, side chains with an average local displacement smaller than 0.85 Å; red, side chains with an average local displacement equal to or larger than 0.85 Å.

found in *T. maritima*. Its genomic neighborhood consists mostly of hypothetical proteins without any indication of conserved operons. Using the Phydbac web server (http://igs-server.cnrs-mrs.fr/phydbac), we established that another member of the DsrEFH family, TM0981, is located in the immediate genomic neighborhood of TM0979. Close homologs of TM0981 can be found in *E.coli* (YchN) and in several other proteobacteria and archaea. These proteins have largely been annotated as DsrE-like proteins (YchN-type), since they form a family of distant paralogs of DsrE (YheN-type). Although DsrE of the YheN-type is absent in *T. maritima*, it is worth noting that TM0979 (DsrH, YheL homolog) and TM0981 (DsrE of YchN-type) form an operon-like structure.

The structural relationship of TM0979 to the DsrE,F,H family was tested using FFAS[12] and DALI.[13] FFAS showed that the hypothetical protein ychN[14] (PDBid: 1JX7) from *E. coli* was the closest structural neighbor, which is in excellent agreement with the aforementioned analysis. DALI identified the hypothetical protein MTH1491[15] (PDBid: 1L1S) from *Methanobacterium thermoautotrophicum* as the closest structural neighbor. The structures of these two homologous proteins have been determined using X-ray crystallography. Interestingly, MTH1491 and ychN form a trimeric/hexameric structure, respectively, when crystallized, whereas there is no indication that TM0979 forms a multimer in solution under the conditions used during the NMR measurements.

According to FFAS, TM0979 has no homologs in the *T. maritima* proteome. Models for TM0979 homologs can be accessed at http://www1.jcsg.org/cgi-bin/models/get_mor.pl?key=TM0979. Most models therein describe predicted DsrH-like proteins from different organisms.

Considering the aforementioned situation, the 3D structure of TM0979 from *T. maritima* can be expected to contribute towards establishing the actual function of the

DsrH protein and other members of this family potentially involved in sulfur oxidation pathways in diverse species.

Materials and Methods. *Protein production:* The plasmid pET-25b(+) containing the TM0979 cDNA (TIGR: TM0979; Swissprot:Q9X074) was transformed into *E. coli* strain BL21-CodonPlus (DE3)-IRL (Stratagene). The expression of uniformly $^{13}C/^{15}N$-labeled and ^{15}N-labeled TM0979 was carried out by growing freshly transformed cells in M9 minimal medium containing 4 g/L [$^{13}C_6$]-D-glucose and/or 1 g/L $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively. Cell cultures were grown at 37°C with vigorous shaking to an OD_{600} of 0.6 to 0.7. The expression of TM0979 was induced with 1 mM IPTG. After 3 to 4 h of induction, the cells were harvested by centrifugation, resuspended in extraction buffer (50 mM Tris-HCl at pH 8.0, 1 mM EDTA, 0.1 mg/ml DNase I, 0.1% Triton X-100, Complete Tabs. (Roche)) and lysed by sonication. The cell debris was removed by centrifugation. In the first purification step, ammonium sulfate (40% w/v) was gradually added to the soluble fraction at 4°C for 1 h. After centrifugation the soluble fraction was first desalted by a HiPrep 26/10 desalting column (Pharmacia) and then heated to 75°C for 15 min, whereupon the precipitated protein was removed by centrifugation. Finally, the soluble protein was loaded onto a Hitrap Q FF column (Pharmacia) equilibrated with 50 mM Tris-HCl at pH 7.0, and eluted with a 0–750 mM NaCl gradient. Fractions containing TM0979 were pooled and a second heat purification step was done. The homogeneity of the purified protein was evaluated by SDS-PAGE electrophoresis, and the solution was concentrated to a final volume of 550 μl. The final concentration of TM0979 was between 2 and 10 mM for the different NMR samples.

Data collection: NMR measurements were performed at 313 K on Bruker Avance600 and DRX800 spectrometers using TXI-HCN-z or TXI-HCN-xyz gradient probes. Proton chemical shifts are referenced to internal 3-(trimethylsilyl)-1-propanesulfonic acid, sodium salt (DSS). Using the absolute frequency ratios, the ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS.[16]

Chemical shift assignment and structure calculation: The following spectra were used to achieve the sequence-specific backbone and side-chain assignments of all aliphatic residues: 2D [^1H,^{15}N]-HSQC, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, 3D HBHA(CO)NH, 3D ^{15}N-resolved [^1H,^1H]-TOCSY, 3D HC(C)H-TOCSY.[6] A newly developed aliphatic- aromatic HCCH-TOCSY spectrum has been used for the assignment of the aromatic side chains.[17] Additionally, 2D [^1H,^1H]-NOESY, 2D [^1H,^1H]-TOCSY, and 2D [^1H,^1H]-COSY spectra of a TM0979 sample in D_2O solution after complete H/D exchange of the labile protons have been used to validate the results. The NMR spectra were processed with PROSA[18] or XWINNMR3.5 (Bruker, Billerica, USA), and analyzed with the XEASY software package.[19]

The input for the structure calculation was collected from 3D ^{15}N-resolved [^1H,^1H]-NOESY, 3D ^{13}C-resolved [^1H,^1H]-NOESY (both recorded at 800 MHz with a mixing time of 80 ms) and 2D [^1H,^1H]-NOESY (600 MHz, mixing time 80 ms, D_2O solution) spectra. The amino acid sequence, the aforementioned chemical shift assignment, and the NOESY spectra constituted the input for the automated NOESY peak picking and NOE assignment method of ATNOS/CANDID/DYANA. The 20 conformers from the ATNOS/CANDID cycle 7 with the lowest residual DYANA target function values were energy-minimized in a water shell with the program OPALp,[20,21] using the AMBER[22] force field. The program MOLMOL[23] was used to analyze the resulting 20 energy-minimized conformers and to prepare the drawings of the NMR structure.

Validation and deposition: Analysis of the stereo-chemical quality of the models was accomplished using the JCSG Validation Central suite, which integrates seven validation tools: Procheck 3.5.4, SFcheck 4.0, Prove 2.5.1, ERRAT, WASP, DDQ 2.0, and Whatcheck. The Validation Central suite is accessible at http://www.jcsg.org. The ^1H, ^{13}C, and ^{15}N chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the BMRB accession number 6010. The atomic coordinates of the final structure have been deposited in the PDB with the code 1RHX.

Acknowledgments. The JCSG consortium is funded by NIGMS GM062411. W.P. was awarded an Erwin Schrödinger Fellowship (J2145) by the FWF; K.W. is the Cecil H. and Ida M. Green Professor of Structural Biology at TSRI. The use of the high-performance computing facility at The Scripps Research Institute is gratefully acknowledged, and we thank Lukasz Jaroszewski, Robert Schwarzenbacher, and Valerie De Crecy-Lagard for their "bioinformatics" support.

REFERENCES

1. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson LD, Nelson WC, Ketchum KA, and others. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature 1999;399: 323–329.
2. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, and others. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 2002;99:11664–11669.
3. Peti W, Etezady-Esfarjani T, Herrmann T, Klock HE, Lesley SA, Wüthrich K. NMR for structural proteomics of *Thermotoga maritima*: Screening and structure determination. J Struct Funct Genomics 2004;5:205–215.
4. Pott AS, Dahl C. Sirohaem sulfite reductase and other proteins encoded by genes at the dsr locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. Microbiology 1998;144(Pt 7):1881–1894.
5. Wüthrich K. NMR of proteins and nucleic acids. New York: Wiley; 1986.
6. Sattler J, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 1999;34:93–158.
7. 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.
8. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 2002;24: 171–189.
9. 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.
10. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, and others. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature 1997;390:364–370.
11. Dahl C, Kredich NM, Deutzmann R, Truper HG. Dissimilatory sulphite reductase from *Archaeoglobus fulgidus*: physico-chemical properties of the enzyme and cloning, sequencing and analysis of the reductase genes. J Gen Microbiol 1993;139(Pt 8):1817–1828.
12. Jaroszewski L, Rychlewski L, Godzik A. Improving the quality of twilight-zone alignments. Protein Sci 2000;9:1487–1496.
13. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
14. Shin DH, Yokota H, Kim R, Kim SH. Crystal structure of a conserved hypothetical protein from *Escherichia coli*. J Struct Funct Genomics 2002;2:53–66.
15. Christendat D, Saridakis V, Kim Y, Kumar PA, Xu X, Semesi A, Joachimiak A, Arrowsmith CH, Edwards AM. The crystal structure of hypothetical protein MTH1491 from *Methanobacterium thermoautotrophicum*. Protein Sci. 2002;11:1409–1414.
16. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 1995;6:135–140.
17. Peti W, Norcross J, Eldridge G, O'Neil-Johnson M. Biomolecular NMR using a micro-coil NMR probe–new technique for the chemical shift assignment of aromatic side chains in proteins. J Am Chem Soc 2004;126:5873–5878.
18. Güntert P, Dötsch V, Wider G, Wüthrich K. Processing of multi-dimensional NMR data with the new software PROSA. J Biomol NMR 1992;2:619–629.
19. Bartels C, Xia T, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. J Biomol NMR 1995;6:1–10.
20. Luginbühl P, Güntert P, Billeter M, Wüthrich K. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 1996;8:136–146.
21. Koradi R, Billeter M, Güntert P. Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun 2000;124:139–147.
22. Cornell WD, Cieplak P, Bayly CI, Gould lR, Merz J, K.M., Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
23. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
24. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.