PROTEINS: Structure, Function, and Bioinformatics 60:552–557 (2005)

STRUCTURE NOTE

NMR Structure Determination of the Conserved Hypothetical Protein TM1816 from *Thermotoga maritima*

Linda Columbus,* Wolfgang Peti, Touraj Etezady-Esfarjani, Torsten Herrmann, and Kurt Wüthrich†
Joint Center of Structural Genomics and Department of Molecular Biology, The Scripps Research Institute, La Jolla, California

Introduction. The proteome of *Thermotoga maritima* (TM), a hyperthermophile with an optimal growth temperature of 80°C, is a focus of the high-throughput structure determination pipeline developed by the Joint Center for Structural Genomics (JCSG; www.jcsg.org).[1] *T. maritima* is a eubacterium with a small genome (1877 predicted genes) and is one of the deepest and most slowly evolving lineages in Eubacteria.[2] Of the gene products, 46% are of unknown function[3]; these proteins have been targeted by the JCSG consortium.[4] As part of this initiative, this article presents the NMR structure of the protein TM1816 in solution.

The input and the statistics for the structure calculation of TM1816 are summarized in Table I. The residual DYANA target function value of 1.71 Å^2 and the backbone atom root-mean-square deviation (RMSD) value of 0.35 Å for the best 20 DYANA conformers after energy-minimization [Fig. 1(A) and Table I] are representative of a high-quality structure determination. The structure of TM1816 contains 3 α-helices and a 5-stranded β-sheet. The sequence arrangement of the regular secondary structures, ββββαβαβα [Fig. 1(A and B)], is similar to that of the ribonuclease H superfamily. The β-sheet is comprised of the residues Met 1–Val 7 (β1), Tyr 26–Val 32 (β2), Val 39–Glu 44 (β3), Leu 67–Val 70 (β4), and Lys 87–Ile 89 (β5) with topology 54123. Strand β1 is parallel to β3, β4, and β5, and antiparallel to β2. The helices α1 (residues 56–62) and α2 (residues 75–84) are arranged in parallel on one face of the β-sheet, and α3 (residues 96–105) is approximately perpendicular to α1 and α2 on the opposite face of the β-sheet. The loop connecting β1 and β2 (L1; Figs. 1(B) and 2(A)] is 19 residues long, with several hydrogen bonds and hydrophobic residue contacts that tether it to the rest of the structure.

The crystal structure of TM1816, which includes a 12-amino acid N-terminal tag [Protein Data Bank (PDB) ID: 1O13], was also determined by the JCSG. The NMR and X-ray structures are very similar, with the residues 4–43 and 55–110 superimposable with a pairwise backbone RMSD of 1.1 Å [Fig. 1(C)]. The regions with the largest variations are loop L3 and helix α3. Loop L3 [Figs. 1(B) and 2(A)], which is disordered in the NMR structure [Fig. 1(A)], is well-ordered in the crystal structure. A number of crystal contacts seem to account for these variations between the NMR and X-ray structures. Specifically, intermolecular interactions of loop L1 and strand β3 are found with loop L3 of a symmetry-related molecule, which includes formation of a short intermolecular parallel β-sheet between strand β3 (residues 38–42) and loop L3 (residues 51–54). Additional intermolecular contacts are made between loop L2 and helix α3.

TM1816 belongs to the Cluster of Orthologous Group (COG) 1433, which contains 28 functionally uncharacterized conserved proteins from 13 different organisms, which are predominantly archaeal species. The structures of 2 other COG 1433 members, MTH1175[6] [from *Methanobacterium thermoautotrophicum* (MTH)] and TM1290,[7,8] were previously determined. A comparison of the 3 sequences [Fig. 2(A)] shows that an optimal alignment yields 33% identity between TM1290 and TM1816, and 30% identity between MTH1175 and TM1816. The architectures of the 3 proteins, TM1816, TM1290, and MTH1175, are identical and all regular secondary structures are superimposable, but with local variations in some of the loops. The backbone atoms of TM1816 have a pairwise RMSD of 4.1 Å with respect to MTH1175 (superposition for best fit of the residues 1–9, 11–47, 58–75, and 77–112), and 4.5 Å with respect to TM1290 (residues 2–9, 11–32, 58–70, and 74–111) (N.B., for each protein, the NMR conformer with the smallest RMSD to the mean coordinates of the bundle of 20 best confomers was used for the RMSD calculation). As seen in Figure 2(A), several small gaps in loops L1, L2, and L3 are required for optimal alignment of the 3 sequences. Superpositions of the backbone traces of TM1816 with MTH1175 and TM1290 are illustrated in

Grant Sponsor: National Institutes of Health, Protein Structure Initiative: Grant number: P50 GM62411. Grant sponsor: National Institutes of Health; Grant number: 1F32GM068286 (to L. Columbus). Grant sponsor: Austrian Science Foundation (FWF) E. Schrödinger Fellowship (to W. Peti). Grant sponsor: Max Kade Foundation Fellowship (to W. Peti).

†K. Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at TSRI.

*Correspondence to: Linda Columbus, The Scripps Research Institute, Department of Molecular Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037. E-mail: columbus@scripps.edu

Received 5 November 2004; Accepted 14 December 2004

Published online 3 June 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20465

TABLE I. Input for the Structure Calculation and Characterization of the Energy-Minimized NMR Structures of TM1816

Quantity	Value[a]
NOE upper distance limits	2105
Intraresidual	450
Short-range	564
Medium-range	416
Long-range	675
Dihedral angle constraints	116
Residual target function value, Å^2	1.71 ± 0.21
Residual NOE violations	
Number ≥0.1 Å	22 ± 5 (14–30)
Maximum, Å	0.15 ± 0.06 (0.13–0.39)
Residual dihedral angle violations	
Number ≥2.5°	1 ± 1 (0–1)
Maximum, °	2.6 ± 0.6 (1.5–3.8)
AMBER energies, kcal/mol	
Total	−3862.18 ± 166.42
van der Waals	−352.07 ± 18.65
Electrostatic	−4653.57 ± 166.34
RMSD from ideal geometry	
Bond lengths, Å	0.0078 ± 0.0002
Bond angles, °	1.988 ± 0.036
RMSD to the mean coordinates, Å[b]	
bb (4–43, 55–110)	0.35 ± 0.03 (0.31–0.41)
ha (4–43, 55–110)	0.91 ± 0.08 (0.79–1.01)
Ramachandran plot statistics, %[c]	
Most favored regions	68
Additional allowed regions	25
Generously allowed regions	4
Disallowed regions	3

[a]Except for the top 2 entries, the average value for the 20 energy-minimized conformers with the lowest residual DYANA target function values is reported. The standard deviation is also listed, with the maximum and minimum values given in parentheses.
[b]bb indicates the backbone atoms N, C^{α} and C′; ha stands for all heavy atoms. The numbers in parentheses are the residues for which the RMSD was calculated.
[c]As determined by PROCHECK.[5]

Figure 2(B and C). Between TM1816 and MTH1175, the largest variations occur in the loops L1 and L3, with smaller variations in loops L5 and L6. The major structural differences between TM1816 and TM1290 are in the conformations of loops L1, L2, and L3, predominately due to the varying lengths of the loops, and in the orientation of the helix α2 [Fig. 2(A)]. As seen from the sequence alignment, TM1290 has 1 residue less in loop L1 than TM1816, 3 residues less in loop L2, and 1 additional residue in loop L3. In addition to these local structural variations, the surface charge distributions, particularly on the side of the molecule that contains loops L2 and L4 (i.e., viewed from the bottom in the orientation of the molecule in Fig. 1), are significantly different between the 3 proteins.

The 3 proteins are structurally similar to the core domain (residues 99–232) of NafY[9] from *Azotobacter vinelandii*, a 243-residue protein that binds the FeMo-cofactor in order to facilitate the insertion of the cofactor into the apo-dinitrogenase. The NafY core domain binds the FeMo-cofactor, but the full length protein is required to bind apo-dinitroginase.[9] MTH1175, TM1290, and TM1816 each have about 15% sequence identity with the core domain of NafY; however, only TM1816 contains a histidyl residue corresponding to the histidine that was shown to be crucial for FeMo-cofactor binding in NafY[10] [Fig. 3(A)]. Figure 3(B) shows all the identical residues between TM1816 and NafY mapped onto the structure of TM1816. A large portion of these residues are clustered near His 20, which is homologous to the histidine residue implicated in the cofactor binding function of NafY (H121). The conserved hydrophobic residues Phe21, Gly22, Val70, Ile73, and Gly74 [colored yellow in Fig. 3(B)] are structurally adjacent to His 20 and form a conserved motif about the putative cofactor binding site. The hydrophobic patch and the HFG motif [Fig. 3(A)] form a very similar electrostatic surface potential around the critical histidine in both NafY and TM1816, which is qualitatively different from the corresponding conserved basic surface areas in TM1290 and MTH1175 [Fig. 3(C)], and implies that TM1816 contains a local surface structure that is similar to the NafY active site. However, *T. maritima* does not have a nitrogen fixation pathway, and a nitrogenase has not been identified; therefore, TM1816 cannot perform the same role that NafY does in *A. vinelandii*. It is possible that TM1816 binds a metal cofactor similar to NafY (e.g., FeMo, FeW, or FeNi) but, as suggested by the lack of an N-terminal domain, is required for a different pathway.

Materials and Methods. *Protein cloning, expression, and purification:* The gene of TM1816 (SwissProt ID: Q9X2D6) was subcloned into a pET-25b(+) plasmid (Novagen). For expression, the vector was transformed into a BL21-CodonPlus (DE3)-RIL *Escherichia coli* strain (Stratagene). Cell cultures were grown at 37°C to an OD ≈ 0.6 and protein expression was induced with 1 m*M* isopropyl-thio-D-galactoside (IPTG). Unlabeled samples were grown in Luria–Bertani (LB) medium. For uniformly-labeled TM1816, cells were grown in M9 minimal media supplemented with $^{15}NH_4Cl$ (1 g/L) and/or [^{13}C]-D-glucose (4 g/L) for obtaining ^{15}N- or ^{15}N/^{13}C-labeled TM1816.

The cells were harvested 3 h after induction by centrifugation at 5000 *g* for 15 min. The cell pellet was resuspended in lysis buffer [50 m*M* Tris, 1 m*M* ethylene diaminetetraacetic acid (EDTA), protease inhibitor tablets (Roche)] and sonicated during 3 × 1 min on ice. The lysed cells were centrifuged at 4°C for 20 min at 12,000 *g*. The supernatant was filtered, with a cutoff at 0.2 μm and loaded onto a HiTRAP Q FF column (Amersham Biosciences) equilibrated with lysis buffer. TM1816 was eluted with a linear gradient from 0 to 0.5 *M* NaCl. The protein fractions were pooled, heated to 75°C and shaken at 750 rpm for 15 min. The precipitate was pelleted by centrifugation at 20,000 *g* for 20 min. The supernatant was concentrated in an Amicon Ultra concentrator (Millipore), and the purity of the protein was evaluated using mass spectrometry and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE).

For the NMR studies, TM1816 was dialyzed into 20 m*M*

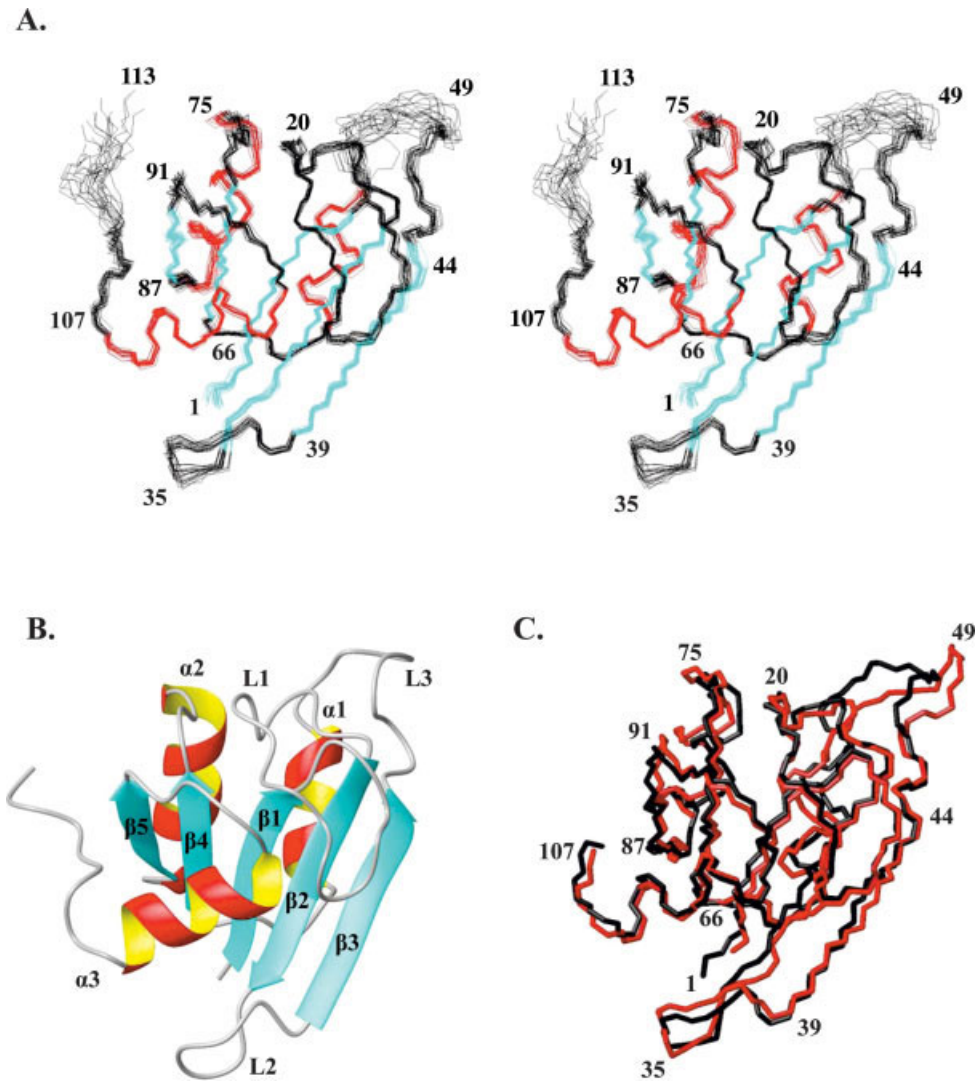

Fig. 1. NMR solution structure of TM1816. Only residues 1–113 of the 124-residue construct are displayed, since the C-terminal undecapeptide segment is disordered. (**A**) Stereoview of backbone traces of the 20 best DYANA conformers after energy minimization, which are superimposed for best fit of the backbone atoms N, C^{α}, and C′ of residues 1–43 and 55–110. Color code: black for loops and other nonregular secondary structures, red for α-helices, and cyan for β-strands. For clarity, select residues are labeled. (**B**) Ribbon diagram of the TM1816 conformer from (A) with the smallest RMSD to the mean coordinates, with the regular secondary structures labeled. Same viewing angle as in (A). (**C**) Comparison of the structures of TM1816 determined by crystallography (red trace) and by NMR in solution [black trace; same conformer as in (B)]. The superposition is for the best fit of the backbone atoms N, C^{α}, and C′ of residues 1–43 and 55–110. Same viewing angle as in (A) and (B). Select residues are labeled, and only residues 1–107 are displayed.

phosphate buffer at pH 6.0 with 4 washes through a concentrator, and 5% D_2O was added. The final concentrations of the protein samples were 2.6 m*M* for unlabeled TM1816, 4.0 m*M* for ^{15}N-labeled TM1816, and 2.1 m*M* for $^{15}N/^{13}C$-labeled TM1816.

NMR experiments and data analysis: All NMR data were recorded at 313 K on Bruker Avance600 or DRX800 spectrometers. For the sequence-specific resonance assignments of the polypeptide backbone atoms, the following experiments were recorded: 2D [^{15}N,^{1}H]-HSQC, 3D HNCACB, 3D CBCA(CO)NH, and 3D HNCO.[11] The assigned C^{α} and C^{β} chemical shifts were subsequently used as a starting point to assign the ^{13}C and ^{1}H atoms of the nonaromatic side-chains with 2D [^{13}C,^{1}H]-HSQC and 3D HC(C)H-TOCSY[12] experiments. The ^{1}H spin systems of the aromatic side-chains were assigned by identifying nuclear Overhauser effects (NOEs) between the βCH_2 and the δCH of the aromatic rings, using a 2D [^{1}H,^{1}H]-NOESY (nuclear Overhauser effects spectroscopy) spectrum recorded in D_2O. Further assignments of the aromatic protons were obtained using a [^{1}H,^{1}H]-COSY (correlation spectroscopy) experiment in D_2O.[13] All assignments were done interactively using the program XEASY.[14] Overall, the resonance assignments obtained were complete except for the following: No resonances were assigned for the residues H53, G54, and H118–H124. The backbone amide

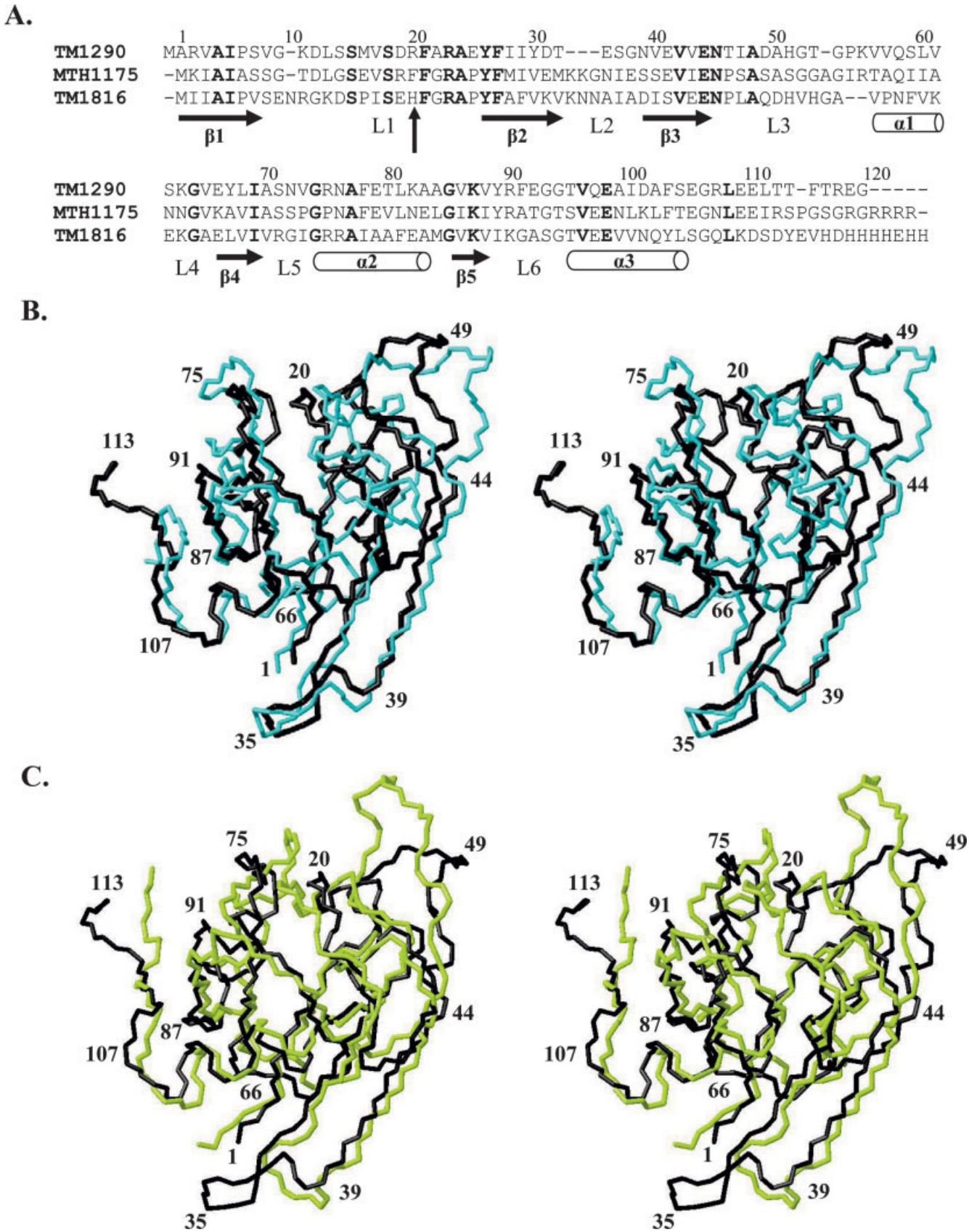

Fig. 2. (**A**) Sequence alignment of TM1816, TM1290, and MTH1175 with the program CLUSTER. The residue numbers above the aligned sequences are based on the TM1816 sequence. Residues that are identical for the 3 sequences are in bold. The regular secondary structures of TM1816 are indicated below the sequence with arrows for the β-strands and cylinders for the α-helices. H20 of TM1816, which is homologous to the histidine residue implicated in FeMo-cofactor binding in NafY from *A. vinelandii*, is identified with a vertical arrow (see Fig. 3). (**B**) Stereoview of a superposition for best fit of residues 1–9, 11–47, 58–75, and 77–112 of the backbone structures of TM1816 (black) and MTH1175 (cyan). Select residues are labeled. (**C**) Same as (B) for TM1816 (black) and TM1290 (green), with superposition for best fit of residues 2–9, 11–32, 58–70, and 74–111.

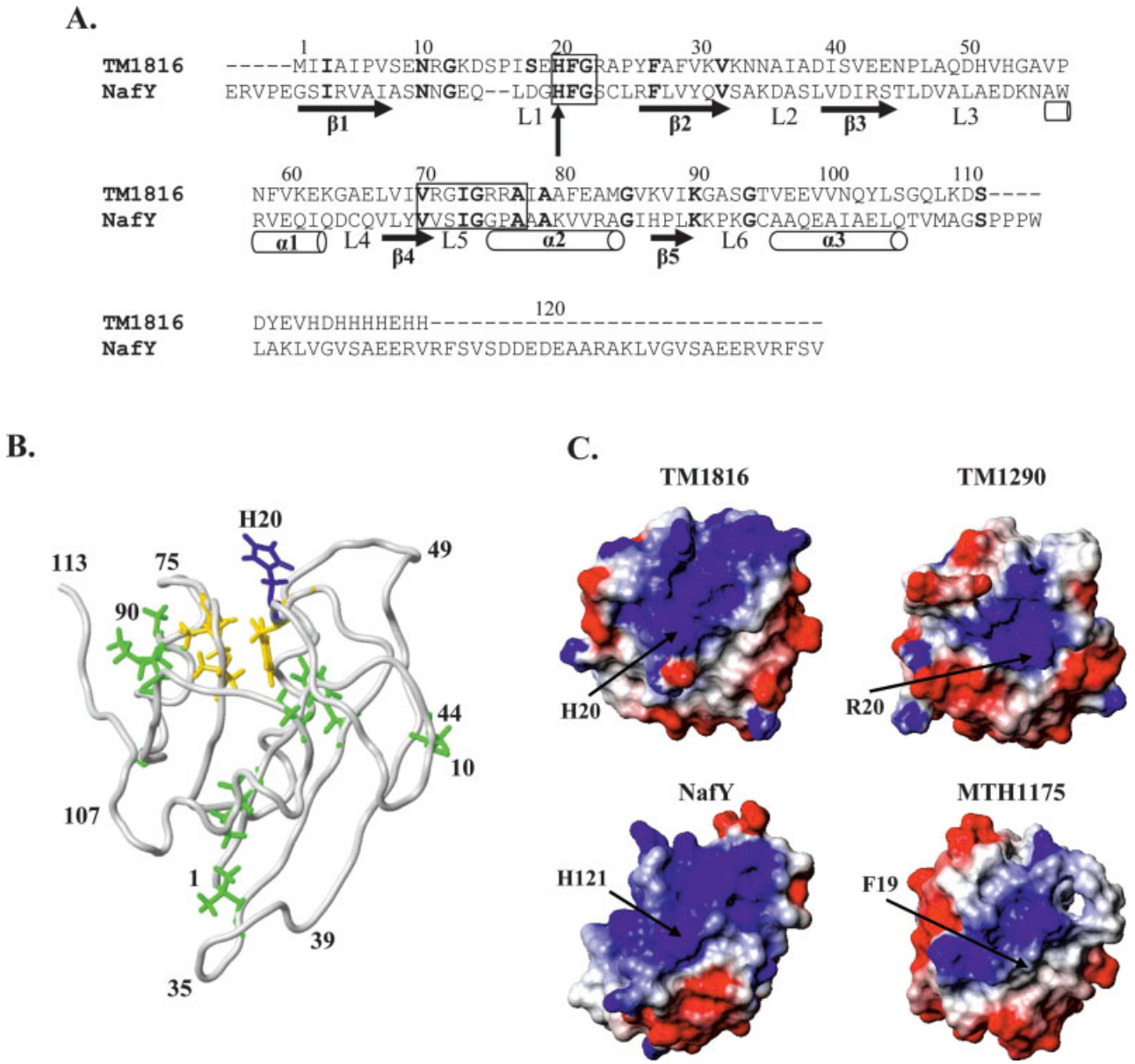

Fig. 3. (**A**) Sequence alignment of TM1816 and the core domain of NafY with the program CLUSTER. The sequence numbers above the alignment are based on TM1816. Residues that are identical in the 2 sequences are in bold. Below the sequence, the regular secondary structures of TM1816 are indicated with arrows for the β-strands and cylinders for the α-helices. H20, which is homologous to the histidine residue implicated in FeMo-cofactor binding in NafY from *A. vinelandii*, is identified with a vertical arrow. The residues in the 2 boxes are located immediately adjacent to His 20 in the 3D structure. (**B**) The polypeptide backbone of TM1816 is shown with the side-chains of the residues that share identity with NafY [boldface residues in (A)]. H20 is colored blue; the structurally adjacent residues F21, G22, V70, I73, G74, and A77 are yellow, and the remaining conserved residues are green. Select residues are labeled. (**C**) Electrostatic surface potential for TM1816 (residues 1–113) and NafY (residues 99–232). The molecules are rotated by 90° about a horizontal axis with respect to (B) in order to expose the conserved histidine residue, which is in the active site of NafY. Color code: blue for positively charged, red for negatively charged, and white for uncharged side-chains. As a reference, the corresponding views for TM1290 and MTH1175 are also shown in (C).

proton and nitrogen resonances were not observed for M1, K13, F21, N34, H51, H53, and R76, but side-chain assignments were nonetheless obtained for M1, K13, F21, N34, and R76. The aromatic protons of H20, H51 and F81, and H^{ϵ} of M1 and M84 were not assigned.

Structure calculation: A 2D homonuclear [^{1}H,^{1}H]-NOESY spectrum recorded in D_2O, a 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectrum recorded in H_2O, and a 3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY spectrum (all with τ_m = 70 ms) were analyzed using the algorithms ATNOS and CANDID[15,16] implemented in the program DYANA.[17] Similar to a previously published protocol,[8] the amino acid sequence of TM1816, the chemical shift lists, and the aforementioned 3 NOESY spectra were used as input. In addition, constraints on the backbone dihedral angles ϕ and ψ derived from the C^{α} chemical shifts[18] were used as input for each cycle. Seven cycles of automated NOE peak identification and assignment were performed in concert

with the DYANA NMR structure calculation. The input parameters for the final structure calculation in cycle 7 are listed in Table I. The 20 conformers (from 100 starting conformers) from cycle 7 with the lowest residual DYANA target function values were energy-minimized in a water shell with the program OPALp.[19,20] The program MOLMOL[21] was used to analyze the 20 energy-minimized conformers and to prepare the molecular models in the figures.

Data deposition: The ^{1}H, ^{13}C, and ^{15}N chemical shifts have been deposited in the BioMagResBank under the BMRB accession number 6198. Atomic coordinates for a bundle of 20 conformers representing the NMR solution structure of TM1816 and the conformer closest to the mean coordinates of this bundle have been deposited in the PDB (http://www.rcsb.org; PDB ID: 1T3V).

Acknowledgments. We thank Dr. Scott Lesley and Heath Klock for the clone of TM1816. We also thank Olga Zagnitko and Slawek K. Grzechnik of the Bioinformatics Core, Joint Center for Structural Genomics, UCSD, for helpful discussion.

REFERENCES

1. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 2002;99:11664–11669.
2. Achenbach-Richter L, Gupta R, Stetter KO, Woese CR. Were the original eubacteria thermophiles? Syst Appl Microbiol 1987;9:34–39.
3. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, Fraser CM. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 1999;399:323–329.
4. 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.
5. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
6. Cort JR, Yee A, Edwards AM, Arrowsmith CH, Kennedy MA. NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from *Methanobacterium thermoautotrophicum*. Struct Funct Genomics 2000;1:15–25.
7. Etezady-Esfarjani T, Peti W, Wüthrich K. NMR assignment of the conserved hypothetical protein TM1290 of *Thermotoga maritima*. J Biomol NMR 2003;25:167–168.
8. Etezady-Esfarjani T, Herrmann, T, Peti W, Klock HE, Lesley SA, Wüthrich K. Letter to the Editor: NMR structure determination of the hypothetical protein TM1290 from *Thermotoga maritima* using automated NOESY analysis. J Biomol NMR 2004;29:403–406.
9. Dyer DH, Rubio LM, Thoden JB Holden HM, Ludden PW. The three-dimensional structure of the core domain of NafY from *Azotobacter vinelandii* determined at 1.8-Å resolution. J Biol Chem 2003;278:32150–32156.
10. Rubio LM, Singer SW, Ludden PW. Purification and characterization of NafY (apodinitrogenase γ subunit) from *Azotobacter vinelandii*. J Biol Chem 2004;279:19739–19746.
11. Bax A, Grzesiek S. Methodological advances in protein NMR. Acc Chem Res 1993;26:131–138.
12. Peti W, Griesinger C, Bermel W. Adiabatic TOCSY for C,C and H,H J-transfer. J Biomol NMR 2000;18:199–205.
13. Wüthrich K. NMR of proteins and nucleic acids. New York: Wiley; 1986.
14. Bartels C, Xia TH, 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.
15. 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.
16. 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.
17. 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.
18. Wishart DS, Sykes BD. The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 1994;4:171–180.
19. 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.
20. Koradi R, Billeter M, Güntert P. Point-centered domain decomposition for parallel molecular dynamic simulation. Comput Phys Commun 2000;124:139–147.
21. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.