PROTEINS: Structure, Function, and Bioinformatics 61:1114–1116 (2005)

STRUCTURE NOTE

Solution Structure of Conserved Hypothetical Protein HP0894 From *Helicobacter pylori*

Kyung-Doo Han, Sung-Jean Park, Sun-Bok Jang, Woo-Sung Son, and Bong-Jin Lee*

National Laboratory of Membrane Protein Structure (MPS), Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul, Korea

Key words: NMR; structural genomics; unknown protein; ribonuclease; RelE

Introduction. *Helicobacter pylori* is a spiral-shaped Gram-negative bacterium and a human bacterial pathogen that infects approximately half of the world's population.[1] It has unique capacity to adapt in the extreme acidic environment in the stomach and to chronically colonize the epithelium of the stomach.[2,3] It is responsible for diverse gastric diseases such as peptic ulcers, chronic gastritis, mucosa-associated lymphoid tissue lymphoma, and gastric cancer.[4–6]

The genome of *H. pylori* has been fully sequenced for two prototype strains (strain 26695, strain j99). In the chromosome of strain 26695, 1,590 open reading frames (ORFs) were identified. Among the 1,590 ORFs, 499 ORFs have no homologs in other organisms and more have no putative function.[7–9] Determining three-dimensional (3D) structure of these unknown proteins in the paradigm of structural genomics can lead to the inference of the biological function of those and identification of new drug targets.

As part of our structural genomics effort on *H. pylori*, we have determined the solution structure of HP0894. HP0894 (SwissProt/TrEMBL ID O25554) is an 88-residue, conserved hypothetical protein from *H. pylori* strain 26695 with calculated pI value of 8.5 and molecular weight of 10.38 kD.

Materials and Methods. *Protein Expression and Purification.* HP0894 gene was polymerase chain reaction-amplified from the *H. pylori* genomic DNA with specific primers. The amplified DNA fragment was cloned into the plasmid pET21a (Novagen). The recombinant plasmids harboring the target genes were transformed into *Escherichia coli* BL21(DE3) host cells for large-scale protein production. The resulting construct contains eight nonnative residues at the C-terminus (LEHHHHHH) that facilitate protein purification. Uniformly ^{15}N- and ^{15}N/^{13}C-labeled proteins were prepared by growing bacteria in M9 medium using $^{15}NH_4Cl$ and $^{13}C_6$-glucose as nitrogen and carbon sources. The protein was purified using Ni^{2+}-affinity column (Chelating Sepharose Fast Flow resin, Pharmacia). All nuclear magnetic resonance (NMR) samples were dissolved in 90% H_2O/10% D_2O containing ~1 mM ^{15}N- and ^{15}N/^{13}C-labeled protein in 20 mM NaH_2PO_4/Na_2HPO_4 (pH 5.0), 500 mM NaCl, 0.1 mM DTT, and 0.1 mM EDTA.

NMR Spectroscopy. NMR spectra were acquired on a Bruker AVANCE 500 and AVANCE 600 (equipped with a cryoprobe) spectrometer at 303 K. Spectra were processed using NMRPipe/NMRDraw[10] and analyzed using NMRView. Backbone assignments were made from HNCO, HNCA, HN(CO)CA,[11] HNCACB,[12] HN(CO)CACB,[13] and 3D ^{15}N-separated NOESY-HSQC.[14] Aliphatic side-chain assignments were made from 3D ^{15}N-separated TOCSY-HSQC,[15] HCCH-TOCSY, and 3D ^{13}C-separated NOESY-HSQC. Aromatic ring resonances were assigned using 3D ^{13}C-separated NOESY-HSQC. Chemical shifts were referenced to DSS. Slowly exchanging amide protons were monitored by dissolving the protein in D_2O and acquiring a series of ^{15}N-HSQC spectra.

Structure Calculation. Dihedral angle restraints were calculated from chemical shifts using the program TALOS.[16] Upper distance limit restraints were obtained from 3D ^{15}N- and ^{13}C-separated NOESY-HSQC by manual and automatic assignment of NOESY spectra. CANDID module in CYANA 2.0[17] is used to make automatic assignment of NOESY peaks. Hydrogen-bond restrains were based on slowed hydrogen exchange and observation of regular secondary elements from CSI[18] search and NOE patterns. Structure calculations were performed using the program CNS 1.1[19] and standard simulated annealing and torsion angle dynamics. Analyses of final structures were performed using the program PROCHECK-NMR[20] and MOLMOL.[21] The program MOLMOL was used to visualize the structures.

Grant sponsor: Ministry of Health & Welfare, ROK; Grant number: 03-PJ2-PG4-BD02-0001; Grant sponsor: National Research Laboratory Program; Grant number: M1-0203-00-0075; Grant sponsor: 2005 BK21 project for Medicine, Dentistry, and Pharmacy.

*Correspondence to: Bong-Jin Lee, National Laboratory of Membrane Protein Structure (MPS), Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, San 56-1, Shillim-Dong, Kwanak-Gu, Seoul 151-742, Korea. E-mail: lbj@nmr.snu.ac.kr

Received 13 April 2005; Accepted 1 July 2005

Published online 17 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20691

TABLE I. Structural Statistics for the 30 Best Conformers of HP0894

Experimental restraints	
NOE restraints	
Short-range ($\|i-j\| \leq 1$)	722
Medium-range ($1 < \|i-j\| < 5$)	202
Long-range ($\|i-j\| \geq 5$)	323
Total	1,247
Hydrogen-bond restraints	10
Dihedral angle restraints	
Φ	54
ψ	56
Root-mean-square deviation to the mean structure (Å) for residues 5–87	
Backbone atoms (N, C^{α}, CO)	0.719 ± 0.19
All heavy atoms	1.19 ± 0.15
Deviation from idealized geometry	
Bonds (Å)	0.00119 ± 0.00003
Angles (°)	0.34083 ± 0.00317
CNS energy (kcal/mol)[a]	
$E_{overall}$	80.43 ± 2.44
E_{bond}	2.13 ± 0.20
E_{angle}	49.25 ± 0.89
$E_{improper}$	2.01 ± 0.31
E_{vdw}	22.90 ± 1.46
E_{noe}	3.95 ± 1.14
E_{cdih}	0.18 ± 0.12
Violations per conformer	
Distance constraints (>0.2 Å)	0
Dihedral angle constraints (>5°)	0
van der Waals (<1.6 Å)	0
Ramachandran plot (%)[b]	
Most favored	76.9
Additionally allowed	21.8
Generously allowed	0.8
Disallowed	0.5

[a]The default parameters and force constants of protein-allhdg.param and anneal.inp in CNS 1.1 were used for calculation.
[b]PROCHECK-NMR was used for calculation.

Results and Discussion. The statistical parameters in Table I showed that a high-quality NMR structure was obtained. HP0894 structure (PDB ID: 1Z8M) has two α-helices, two 3_{10}-helices, and four β-strands (α - α -3_{10}- β -3_{10} - β - β - β). β-Strands form four-stranded anti-parallel β-sheet [Fig. 1(A,B)]. *Sequence Homology.* A PSI-BLAST[22] analysis of the HP0894 sequence against the nonredundant proteins database identified a total of 64 homologs, most of which are annotated as hypothetical or uncharacterized protein. BLAST conserved domain search[23] showed that HP0894 contains a conserved domain of DUF332 (Domain of Unknown Function), which is equivalent to COG 3041 in the National Center for Biotechnology Information Database of Clusters of Orthologous Groups. But, in the Pfam[24] database, HP0894 belongs to plasmid stabilization system protein family (PF05016). Members of this family are involved in plasmid stabilization. But the exact molecular function of these proteins is not known.
Structural Homology. A search for structural homologs with Z score higher than 3.0 using the programs DALI[25] shows HP0894 is structurally similar to Archaeal RelE[26]

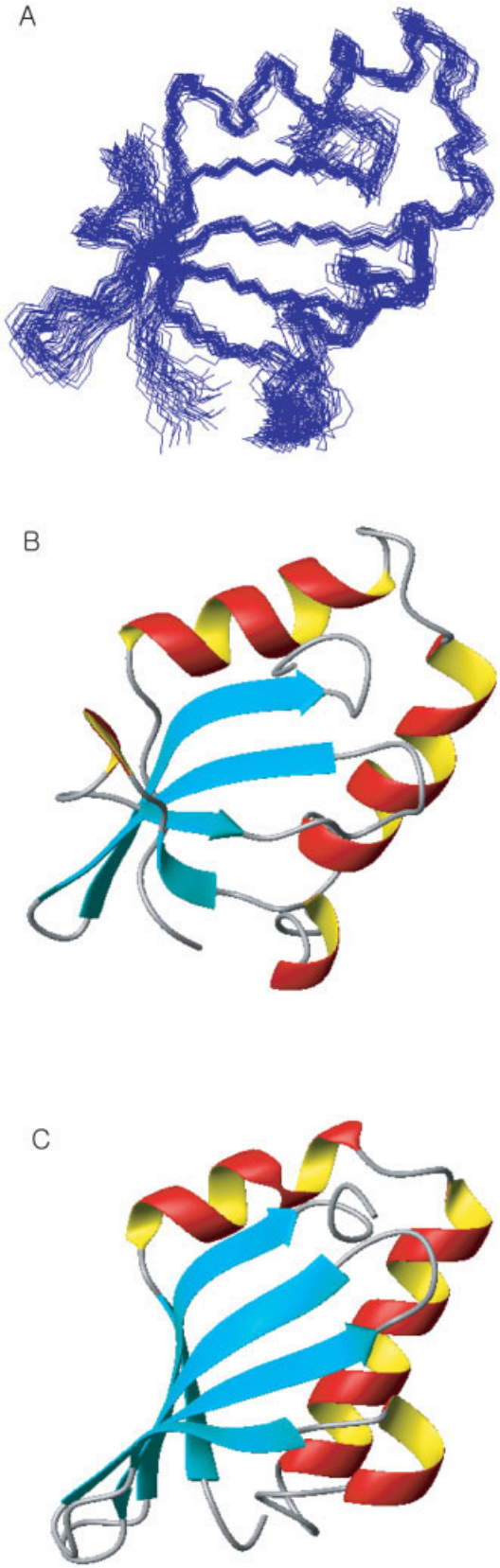

Fig. 1. **A:** The 30 conformers with the lowest energy are shown after superposition of backbone atoms N, C^{α}, and CO of residues 5–83. **B:** Ribbon drawing of the representative conformer of HP0894. **C:** Ribbon drawing of crystal structure of Archaeal RelE with highest structural similarity to HP0894 (PDB ID: 1WMI).

(hypothetical protein Psh013, PDB code 1WMI, Z score= 7.8, RMSD= 2.8) and Ganyloribonuclease (PDB code 1RGE, Z score = 3.3, RMSD = 3.4). These two proteins are both ribonucleases, have the similar number of residues to HP0894 (HP0894: 88 residues, Archaeal RelE: 90, Ganyloribonuclease: 96), share a similar β-sheet topology with HP0894, and have comparable location for two of their helices (Fig. 1C). But they have no detectable sequence homology with HP0894 in PSI-BLAST searches and Blast2 analyses. Archaeal RelE with highest structural similarity to HP0894, whose structure was released recently, is an Archaeal homolog of RelE of *E. coli*, and regarded as a ribosome-dependent ribonuclease.[26] Interestingly, despite no sequential homology with HP0894, Archaeal RelE belongs to Plasmid stabilization system protein family as same as HP0894, according to Pfam database. Considering the result of structural homology search and Pfam database search, therefore, there is the possibility that HP0894 is a ribonuclease.

REFERENCES

1. International Agency for Research on Cancer, World Health Organization. Infection with *Helicobacter pylori.* In: Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7–14 June 1994. IARC Monogr Eval Carcinog Risks Hum 1994;60:177–240.
2. Sachs G, Weeks DL, Melchers K, Scott DR. The gastric biology of *Helicobacter pylori.* Annu Rev Physiol 2003;65:349–369.
3. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1984;1:1311–1315.
4. Rothenbacher D, Brenner H. Burden of *Helicobacter pylori and H. pylori*-related diseases in developed countries: recent developments and future implications. Microbes Infect 2003;5:693–703.
5. Wotherspoon AC, Doglioni C, Diss TC, et al. Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of *Helicobacter pylori.* Lancet 1993;342: 575–577.
6. Peek RM Jr, Blaser MJ. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2002;2:28–37.
7. Tomb JF, White O, Kerlavage AR, et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 1997;388:539–547.
8. Alm RA, Ling LS, Moir DT, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 1999;397:176–180.
9. Popescu A, Karpay A, Israel DA, Peek RM Jr, Krezel AM. *Helicobacter pylori* protein HP0222 belongs to Arc/MetJ family of transcriptional regulators. Proteins 2005;59:303–311.
10. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
11. Kay RE, Ikura M, Tschudin R, Bax A. Three-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 1990;89:496–502.
12. Wittekind M, Mueller L. HNCACB, a high sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the a-carbon and b-carbon resonances in proteins. J Magn Reson Ser 1993;B101:214–217.
13. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J Am Chem Soc 1992;114:6291–6293.
14. Ikura M, Bax A,Clore M, Gronenborn AM. Detection of nuclear Overhauser effects between degenerate amide proton resonances by heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. J Am Chem Soc 1990;112:9020–9021.
15. Gronenborn AM, Bax A, Wingfield PT, Clore M. A powerful method of sequential proton resonance assignment in proteins using relayed ^{15}N-^{1}H multiple quantum coherence spectroscopy. FEBS Lett 1989;243:93–98.
16. Cornilescu G, Delagio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–301.
17. 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.
18. Wishart DS, Sykes BD. Chemical shifts as a tool for structure determination. Methods Enzymol 1994;239:363–392.
19. Brunger AT, Adams PD, Clore GM, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
20. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
21. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
22. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
23. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. Nucleic Acids Res 2004;32:W327–331.
24. Bateman A, Birney E, Cerruti L, et al. The Pfam protein families data base. Nucleic Acids Res 2002;30:276–280.
25. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
26. Takagi H, Kakuta Y, Okada T, Yao M, Tanaka I, Kimura M. Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects. Nat Struct Mol Biol 2005;12(4):327–331.