J. Biochem. 2009;146(5)667–674
doi:10.1093/jb/mvp105

NMR Solution Structure of HP0827 (O25501_HELPY) from *Helicobacter pylori*: Model of the Possible RNA-binding Site

Sun-Bok Jang[1], Chao Ma[1], Ji-Yoon Lee[1], Ji-Hun Kim[1], Sung Jean Park[1], Ae-Ran Kwon[2] and Bong-Jin Lee[1,*]

[1]Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, San 56-1, Shillim-Dong, Kwanak-Gu, Seoul 151-742; and [2]Department of Herbal Skin Care, College of Herbal Bio-Industry, Daegu Haany University, 290, Yugok-Dong, Gyeongsan-Si, Gyeongsangbuk-Do 712-715, Korea

Received June 1, 2009; accepted July 6, 2009; published online July 15, 2009

The HP0827 protein is an 82-residue protein identified as a putative ss-DNA-binding protein 12RNP2 Precursor from *Helicobacter pylori*. Here, we have determined 3D structure of HP0827 using Nuclear Magnetic Resonance. It has a ferredoxin-like fold, β1–α1–β2–β3–α2–β4 (α; α-helix and β; β-sheet) and ribonucleoprotein (RNP) motifs which are thought to be important in RNA binding. By using structural homologues search and analyzing electrostatic potential of surface, we could compared HP0827 with other RNA-binding proteins (sex-lethal, T-cell restricted intracellular antigen-1, U1A) to predict RNA-binding sites of HP0827. We could predict that β sheets of HP0827, especially β1 and β3, are primary region for RNA binding. Consequently, similar to other RNA-binding proteins, RNP motifs (Y5, F45, F47), positively charged and hydrophobic regions (K32, R37, K40, K41, K43, R70, R73) are proposed as a putative RNA-binding sites. In addition, differences in amino acids composition of RNP motifs, N, C-terminal residues, loop-region fold and the orientation of α1-helix with other RNA recognition motif proteins could give specific biological functions to HP0827. Finally, the study on natural RNA target is also important to completely understand the biological function of HP0827.

Key words: *Helicobacter pylori*, HP0827, RNP motif, RRM, single-stranded DNA-binding proteins.

Abbreviations: HP, *Helicobacter pylori*; RNP, ribonucleoprotein; RRM, RNA recognition motif; PABP, PolyA-binding protein; OD, optical density; NMR, Nuclear Magnetic Resonance; CD, circular dichroism; CSI, chemical shift index; NOESY, nuclear overhauser enhancement spectroscopy; TROSY, transverse relaxation optimized spectroscopy; PDB, protein data bank; TIA-1, T-cell-restricted intracellular antigen-1; UV, ultraviolet.

Helicobacter pylori (*HP*) is a micro-aerophilic, Gram-negative, spiral-shaped, human pathogenic bacterium which is present in almost half of the world population (*1*). It induces gastric ulcer and is also associated with the development of gastric cancer and mucosa-associated lymphoid tissue lymphoma (*2*). *Helicobacter pylori* colonizes the human gastric mucosa and can persist in the stomach of patients with or without the clinical symptoms for a lifetime (*3*). The identification on the clinical consequences of *HP* infection is assuredly one of the major discoveries within the past 20 years in medicine. It revolutionized pathogenic understanding and therapeutic concept in gastroenterology.

Currently, three complete *HP* genomes including strains 26695, J99 and HPAG1 have been determined (*4–6*). Because of its importance as a human pathogen, the relationship between structure and biological function of *HP* protein may be used for drug discovery and vaccine development.

As part of our structural genomics effort on *HP*, we have determined the 3D structure of single-stranded (ss)-DNA-binding protein 12RNP2 precursor protein HP0827 (O25501_HELPY) from *HP* strain 26 695. The HP0827 gene encodes a protein of 82 amino acids with calculated *pI* value of 8.96 and molecular weight of 9.4 kDa.

In eukaryotes, the RNA recognition motif (RRM) is one of the most abundant protein domains. According to Pfam database (*7*), a total of 6056 RRM motifs have been identified in 3541 different proteins. For a long time, the RRM domain was considered to exclusively belong to the eukaryotic organism. However, it was recently found that the RRM domain is also present in bacteria, viruses and mitochondria (*8*).

HP0827 has one RRM which contains two ribonucleoprotein (RNP) motifs (RNP1 and RNP2). RRMs are found in a variety of RNA-binding proteins including RNPs. RRM protein superfamily is known to play a biologically important role in RNA metabolism, covering from transcription to translation, splicing (U1A) (*9*), pre mRNA binding (hnRNP C) (*10*) or polyadenylate binding (PABP) (*11*). It is possible to contain from one (as in 12RNP1 and 12RNP2) to four RNA-binding domains—RBDs (PABP). Based on a number of RBDs, RRM protein can have different binding aspects to RNA, and varieties of interaction modes between RRM protein and RNA are possible (*12*). RRM usually interacts with ssRNA, but is

*To whom correspondence should be addressed. Tel: +82-2-880-7869, Fax: +82-2-872-3632, E-mail: lbj@nmr.snu.ac.kr

also known to interact with ssDNA like HP0827 (*13*). For example, Rho is a precursor protein which is an essential transcription termination factor in most eubacterial species and contains RNP1 motif. Although the Rho protein binds to either ssDNA or ssRNA with similar affinities, the Rho protein is remarkably activated only when it is bound to ssRNA (*14*). Thus, we could speculate that the ssRNA binding could be important in activation of precursor protein like HP0827.

In this study, we have determined the solution structure of HP0827 containing a conserved structure pattern of RRM motif that consists of four β-strands and two α-helices arranged in an α/β sandwich by using NMR. To predict the putative RNA-binding sites of the HP0827, we used a structural comparison method with the structures of the sex-lethal, T-cell restricted intracellular antigen-1 (Tia-1) and U1A which were known as RNA-binding proteins before. Consequently, we could propose the putative RNA-binding sites of HP0827, and this knowledge will be helpful to find the exact RNA-binding sites and patterns of HP0827 in future studies.

MATERIALS AND METHODS

Materials—A genomic DNA from *HP* (ATCC 700392) was obtained ATCC (Manassas, VA, USA). Resins for Ni^{2+} affinity column (Chelating Sepharose Fast Flow resin) and Superdex^{TM} column were purchased from Amersham Pharmacia biotech.

Sequence Alignments—The sequences of HP0827, sex-lethal, TIA-1 and U1A were aligned with ClustalW2 (*15*) and Jalview (*16*) program available on the website (http://www.jalview.org/).

Protein Sample Preparation—The predicted ORF of HP0827 was amplified from *HP* 26 695 genomic DNA using standard PCR methods and ligated into a pET-21a (+) expression vector (Novagen). A pET-21a (+) expression plasmid containing both HP0827 and a C-terminal hexahistidine tag that facilitates protein purification was transformed into *Escherichia coli* strain BL21 (DE3) host cells for large-scale protein production. The accuracy of the cloning was confirmed by DNA sequencing. The cells were grown at 37°C in M9 media that contains $^{15}NH_4Cl$ and $^{13}C_6$-glucose (Cambridge Isotope Laboratories, Inc, Andover, MA, USA) as nitrogen and carbon sources, respectively. When the OD_{600} of cells reached at 0.5, 0.25 mM IPTG was added. After further cell growth for 4 h at 37°C, the cells were harvested. The cell pellets were resuspended in 20 mM sodium phosphate (pH 7.9), 0.5 M NaCl buffer and lysed by sonication at 4°C. The recombinant HP0827 protein was purified by applying the supernatant from the cell lysate onto a Ni^{2+}–NTA column (Qiagen; 3 ml of resin per litre of cell culture) previously equilibrated with the binding buffer. The Fractions containing protein were concentrated to about 1 ml and applied to superdex-75 (Pharmacia) column that had been equilibrated with the final buffer (20 mM MES, pH 6.5, 100 mM NaCl, 1 mM DTT and 0.1 mM PMSF). The NMR sample was ~1 mM ^{15}N- and $^{15}N/^{13}C$-labelled protein prepared by 90% H_2O/10% D_2O buffer solution.

CD—Circular dichroism (CD) spectra were recorded using 2 mm path length cuvette in a J-715 spectropolarimeter (Jasco, Japan) equipped with a peltier temperature control system (Model PTC-348WI). The protein concentration was in the range of 5–10 μM. Samples were scanned three times with a bandwidth of 1 nm and a response time of 2 s in a wavelength range between 190 and 250 nm at a rate of 50 nm/min. To examine the thermodynamic properties, the temperature scanning was conducted with the rate of 1°C/min in the temperature range of 20–80°C at 222 nm. By monitoring changes in secondary structure content, apparent transition temperature (T_{app}) values were estimated by fitting the data using the five-parameter sigmoid function from curve-fitting program Sigmaplot, followed by determining the inflection point by numerical differentiation of the curves with Origin 8.

Gel Filtration Chromatography—Gel filtration chromatography analysis was performed to calculate the molecular mass of the native protein using non-denatured protein molecular weight marker kit (Amersham Biosciences, UK). HPLC system (Hitachi, Japan) using a BioSep-SEC-S 3000 column (Phenomex) was equilibrated and eluted with 20 mM MES, pH 6.5, 100 mM NaCl, 1 mM DTT and 0.1 mM PMSF buffer.

Oligomeric state determination by gel filtration are carried out by comparing an elution volume parameter, such as K_d of the protein of interest, with the values obtained for several known calibration standards. Equation 1 provides the expression for the K_d; in this equation, V_o is the column void volume, V_e is the sample elution volume, and V_c is the geometric column volume. The molecular weight of an unknown protein can be determined from the calibration curve (plot of K_d versus the logarithm of molecular weight) once its K_d value is calculated from its measured elution volume. The sample was eluted with 1 ml/min of flow rate since elution volume could be same with elution time. The LMW Gel Filtration Calibration Kit contains ribonuclease A (13.7 kDa), chymotrypsinogen A (25.0 kDa), ovalbumin (43.0 kDa), albumin (67.0 kDa) and blue dextran 2000 (2000 kDa). Blue dextran 200 and tyrosine are used to know void volume and geometric volume of gel filtration column, respectively.

$$K_d = \frac{V_e - V_o}{V_c - V_o} \quad (1)$$

NMR Spectroscopy—All NMR spectra were recorded at 308 K on Bruker AVANCE 500, 600 and 900 (equipped with a cryoprobe) spectrometers. The backbone and side chain assignments were performed by using 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HBHA(CO)NH, ^{15}N-TOCSY-HSQC, C(CO)NH-TOCSY spectra. Aromatic ring resonances were assigned by using 3D ^{15}N-NOESY-HSQC (mixing time 100 ms), ^{13}C-NOESY-HSQC (mixing time 100 ms). Chemical shifts were referenced to DSS externally. NMR spectra were processed using the program NMRPipe/nmrDraw (*17*) and analysed with the program NMRView.

Structure Calculation—The NMR solution structure calculation was conducted on the basis of distance restraints and dihedral angle restraints. The restraints of upper distances limit were derived from 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC by manual

and automatic assignments by using CYANA2.0 (*18*). Dihedral angle restraints were calculated from chemical shift using software TALOS (*19*), and overall secondary structure elements were derived from chemical shift index (CSI) (*20*) search and NOE patterns. In addition, hydrogen-bond restraints were collected by using H–D exchange experiment.

CYANA2.0 was used to obtain initial protein structure. The structure of HP0827 was further refined using the software CNS 1.2 (*21*). In our calculations, we have used the simple simulated annealing option with the database-derived distance constraints provided in the same format as NOE distance constraints. Protein structure was calculated by minimizing the violations of the experimental and database-derived constraints and the CNS-built in energy potentials. After a number of refinement cycle, 959 NOE-derived distance restraints (280 long range, 142 medium range, 290 sequential and 247 intraresidue) and 34 dihedral angles were used as final input data. The programs, MOLMOL (*22*) and UCSF-Chimera (*23*) were used to visualize the result of 20 energy-minimized conformers. Analysis on the quality of the final HP0827 structure was accomplished using PROCHECK-NMR (*24*) and AQUA (*25*).

RESULTS

Resonance Assignments—The assignments comprise 92.9% of all ^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα and 81.1% of all side chain atoms. ^{1}H-^{15}N transverse relaxation optimized spectoscopy (TROSY) spectrum of ^{15}N-enriched HP0827 showed very good signal dispersion (Fig. 1). Sequence-specific resonance assignments were performed. The five N-terminal residues have no NOE and are not well defined in the structure. The structure is defined by an average of almost 12 NOEs per residue. The average rmsd is 0.67 ± 0.15 for the backbone atoms. The Ramachandran plot of the 20 energy-minimized conformers indicated 71.4 and 25.0% of the residues located in the most favoured and additional allowed regions, respectively. The NMR structural statistics was summarized in Table 1.

Solution Structure of HP0827—The CD spectrum shows α-helical and β-strand contents of 22–26% and 23–27%, respectively, suggesting that the HP0827 3D structure is well defined in solution. The T_{app} could be determined. The T_{app} value for the HP0827 was 69°C (Fig. 2).

The solution structure of HP0827 [protein data bank (PDB) entry: 2ki2] (Fig. 3) has a ferredoxin-like fold, $\beta1$–$\alpha1$–$\beta2$–$\beta3$–$\alpha2$–$\beta4$ (α; α-helix and β; β-sheet). And, its monomeric state was confirmed by gel filtration chromatography (Fig. 4). The β-strands correspond to residues 2–7 ($\beta1$), 30–33 ($\beta2$), 45–50 ($\beta3$) and 73–76 ($\beta4$). The α-helices correspond to residues 15–26 ($\alpha1$) and 53–61 ($\alpha2$). The four β-strands are arranged in a right-handed twist and form an antiparallel β-sheet that packs against the two α-helices. The Loop3 (*34–44*) is somewhat flexible compared with other regions. Major determinant of RNA-binding specificity might reside in the loop region, since flexibility of the loop could be stabilized upon RNA binding (*14*). Therefore, we could consider this Loop3 region as one of the important regions in the activation of

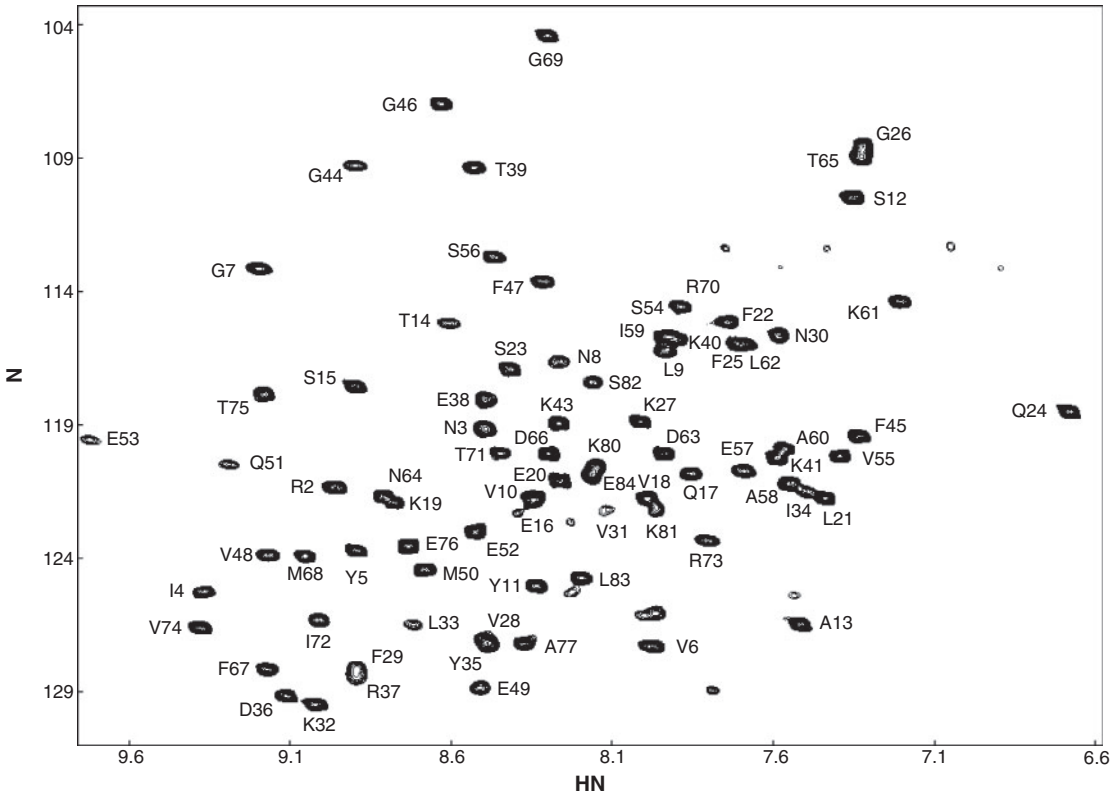

Fig. 1. **Assignments of the ^{1}H-^{15}N-TROSY of the uniformly ^{15}N-labeled HP0827 (pH 6.5, 308 K).**

Table 1. **Structural statistics for the 20 energy-minimized conformers of HP0827.**

Completeness of resonance assignments	
Backbone (%)	96.0
Side chain (%)	81.1
Experimental constraints	
NOE constraints total	959
Intraresidue ($i=j$)	247
Sequential ($\|i-j\|=1$)	290
Medium-range ($1<\|i-j\|<5$)	142
Long-range ($\|i-j\|\geq 5$)	280
Dihedral constraints	
φ	17
ψ	17
RMSD to the mean structure (Å) for residues 1–33, 45–77	
Backbone atoms (N, Ca, CO)	0.67 ± 0.15
All heavy atoms	1.27 ± 0.16
Deviation from idealized geometry	
Bonds (Å)	0.000871 ± 0.0000572
Angles (°)	0.287 ± 0.0032
CNS energy (kcal/mol)[a]	
$E_{overall}$	44.95 ± 1.65
E_{bond}	0.94 ± 0.072
E_{angle}	28.42 ± 0.31
$E_{improper}$	1.44 ± 0.20
E_{vdw}	12.84 ± 0.67
E_{noe}	1.23 ± 0.32
E_{cdih}	0.07 ± 0.04
Violations per conformer	
Distance constraints (>0.1 Å)	0
Dihedral angle constraints (>5°)	0
Van der waals (<1.6 Å)	0
Ramachandran plot (%)[b]	
Ranges: 1–33, 45–77	
Most favoured region	71.4
Additionally allowed region	25.0
Generously allowed region	3.2
Disallowed region	0.4

[a]The default parameters and force constants of protein-allhdg.param, and anneal.inp in CNS 1.2 were used for structure calculation. [b]PROCHECK-NMR was used for calculation.

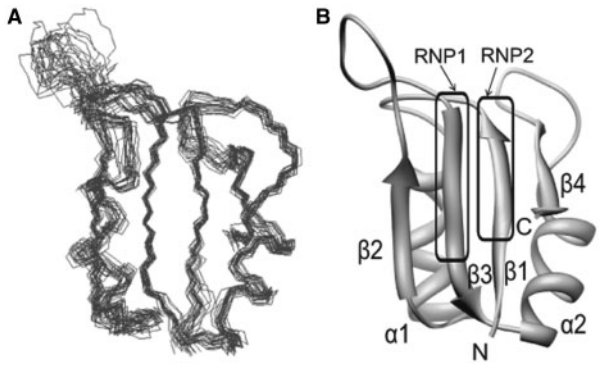

Fig. 3. **Solution NMR structure of HP0827.** (A) The superposition of the final 20 structures over the energy-minimized average structure. (B) Ribbon drawing of the representative conformer of HP0827. Black box represents conserved RNP motif lying side by side.

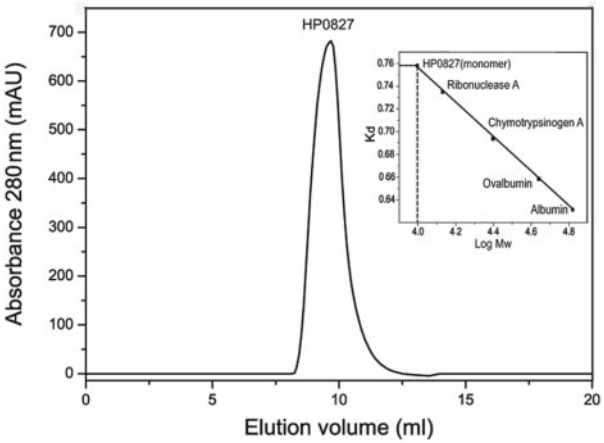

Fig. 4. **Analytical gel filtration profile of HP0827 at pH 6.5 and 25°C on a BioSep-SEC-S 3000 column.** HP0827 elutes with a retention volume of 9.75 ml, corresponding to the mass of the monomer. Inner box shows the calibration on the result of the gel filtration experiment. The dotted line indicates the logarithm of the apparent molecular mass and K_d value of HP0827.

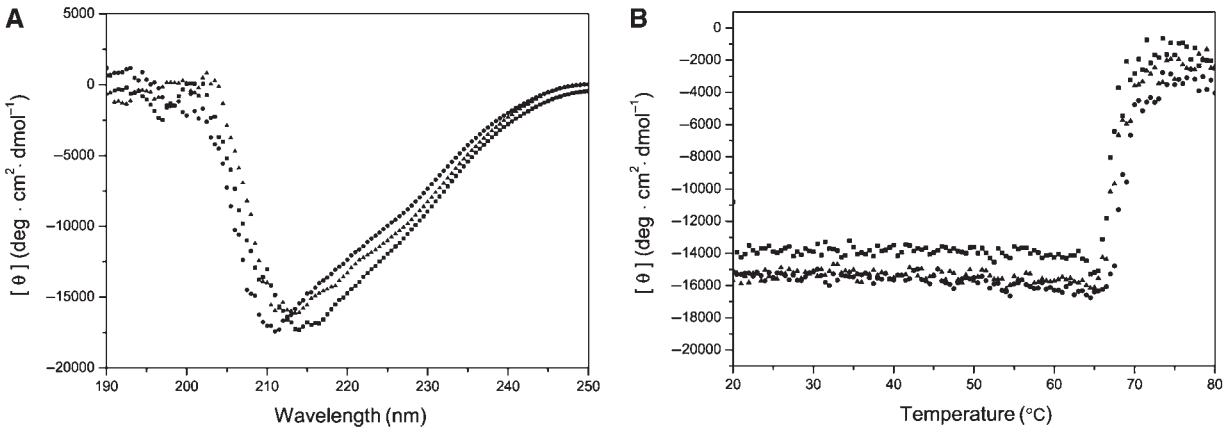

Fig. 2. **CD analysis of HP0827.** (A) Far-UV CD spectra of HP0827 at pH 6.5 (filled circle), pH 5.5 (filled triangle) and pH 4.5 (filled square) are plotted as molar ellipticity versus wavelength; (B) Thermal denaturation profiles of HP0827 at pH 6.5 (filled circle), pH 5.5 (filled triangle) and pH 4.5 (filled square). Changes in molar ellipticity at 222 nm at a scan rate of 1°C/min in the temperature range of 20–80°C were measured.

precursor proteins like HP0827. Although the backbone assignment was almost done in the C-terminal region (78–82), we calculated the structure of HP0827 except C-terminal region (78–82) which is poorly ordered and defined.

Sequence Alignments—A search for structural homologues within the PDB (*26*) using DALI (*27*) server also showed that HP0827 has a structural similarities with proteins belong to RBD superfamily. The superfamily contains the sex-lethal protein (alternative splicing regulator, PDB code: 2sxl-A, *Z*-score: 11.2, RMSD: 2.1, identity: 23%) (*28*), TIA-1 protein (mRNA translation regulator, PDB code: 2rne-A, *Z*-score: 10.8, RMSD: 1.9, identity: 31%) (*29*) and U1A protein (mRNA splicing regulator, PDB code: 1urn-A, *Z*-score: 10.3, RMSD: 2.3, identity: 23%) (*30*). It is notable that all the structural homologues have RNP motifs necessary for RNA binding. And these homologues function as post-transcriptional gene regulators using ssRNA binding and also have a characteristic of ssDNA binding (*28–30*). Although there are some differences between amino-acid compositions in RNP 1 and 2 motifs of these four homologues, the composition is relatively well conserved (Fig. 5A).

DISCUSSION

Comparison of the RNP Motif Regions with other RRM Superfamily Proteins—According to database of gene information (http://cmr.jcvi.org/), HP0827 is identified as a putative ssDNA-binding protein 12RNP2 precursor. In particular, RNP motif is most well known as a crucial region for RNA–protein interaction. We could know that HP0827 contains highly conserved RNP1 and RNP2 sequences using Pfam database (*7*). Generally, this motif shows a high consensus pattern within RRM superfamliy (*31*). RNP1 and RNP2 motifs of HP0827 are lying side-by-side on the adjacent middle antiparallel strands (β3 and β1, respectively) (Fig. 3B). The conserved RNP1 octapeptide sequence is composed of Lys/Arg–Gly–Phe/Tyr–Gly/Ala–Phe/Tyr–Val/Ile/Leu–X–Phe/Tyr, where X could be any amino acid (*32*). In HP0827, RNP1 motif includes Lys–Gly–Phe–Gly–Phe–Val–Glu–Met sequence. Compared with RNP1, RNP2 has less well-conserved hexapeptide sequence, which contains aromatic and aliphatic amino acids, assisting protein binding to RNA. The conserved RNP2 hexapeptide is usually composed of Ile/Val/Leu–Phe/Tyr–Ile/Val/Leu–X–Asn–Leu, where X could be any amino acid (*32*). In HP0827, RNP2 motif includes Ile–Tyr–Val–Gly–Asn–Leu. Generally, the second residue of RNP2 and the third and fifth residues of RNP1 are thought to form an aromatic ring cluster that may play a crucial role in stabilization of protein structure by van der Waals interaction and RNA–protein interaction by aromatic ring stacking and hydrogen bonding (*33*). Like other RNP motifs, HP0827 has conserved three residues (Tyr3 in RNP2 and Phe45, Phe47 in RNP1) and these residues probably have an important role in RNA–HP0827 interaction.

The structural homologues obtained from DALI search have RNP motifs necessary for RNA binding. Moreover, it was already shown that the RNP motifs of sex-lethal, TIA-1 and U1A interact with RNA (*28–30*). Although there are some differences in amino-acid composition in RNP 1 and 2 motifs between these four homologues, the composition is relatively well conserved. Most of all, coexisting aromatic residues (Tyr, Phe) within RNP motif region are anticipated to have crucial role in RNA binding by forming aromatic ring cluster in β1 and β3 strands. Like the other structural homologues, HP0827 also has conserved Ile4 and Leu9 aliphatic amino acid that is proposed to be involved in the hydrophobic core for hydrophobic interaction with RNA (*8*). However, it is remarkable that these three structural homologues show a little different RNA–protein interaction mode. The position and type of amino acids and RNA-binding pattern are somewhat different among homologues. In RNP2 region, HP0827, TIA-1 and U1A protein have consensus aromatic residues (Tyr5, Phe100 and Tyr13, respectively) at the second position, which is considered to contribute RNA binding. But sex-lethal protein has Y131 at the fifth position in RNP2. This difference in position gives other RNA-binding mode and specificity to other RRM proteins like sex-lethal protein. In RNP1 region, these four homologues have aromatic residues at the third, fifth and eighth position, which are conserved residues for RNA binding. But HP0827 lacks the consensus aromatic residues at the eighth position. As seen in Fig. 5, the β1, β3-sheets in RRM are the central surface region for RNA recognition.

Comparison of the Putative RNA-binding Regions with other RRM Superfamily Proteins—Besides these conserved regions, we could speculate that other contacts could reinforce RNA recognition by analyzing surface charge map (Fig. 6). Positively charged residues might make this RNA–protein binding stronger by electrostatic interaction between negatively charged phosphate backbone of RNA and positively charged residues.

The solvent-accessible surface of the β2, β4 sheets and Loop5 (α2–β4) has highly positively charged basic residues that could assist RNA binding. These kinds of amino acids are also present in sex-lethal protein (R155, K194, R195, K197) (*28*), TIA-1 protein (R127, R169) (*29*) and U1A (K80, R83, K88) (*30*). In HP0827, positively charged K32, R70 and R73 located at β2, β4 sheets and Loop5 are anticipated to be involved in RNA binding. These amino acids are also exposed to solvent (Fig. 6). Therefore, these amino-acid composition and geometry probably can make HP0827 directly contact with RNA more easily.

Roles of Loop3 Region in RNA Binding—The Loop1 (β1–α1), Loop3 (β2–β3) and Loop5 (α2–β4) regions are reported to play important roles in RNA recognition (*34*). In contrast to the Loop1 (L1) and Loop5 (L5), the conformation and length of the Loop3 (L3) are diverse among RRMs (*35*). Therefore, this region could give RRMs many RNA recognition modes. In TIA-1, there is a longer β-sheet region than other homologues which is composed of G135, K136, S137 and K138 which has important role in RNA recognition (*29*). In the case of sex-lethal protein, it is known that aromatic cluster composed of Y160, Y164 and Y166 plays a crucial role in RNA recognition (*28*). In the case of U1A, RNA loop fits into the groove of the Loop3 (β2–β3) (*30*).

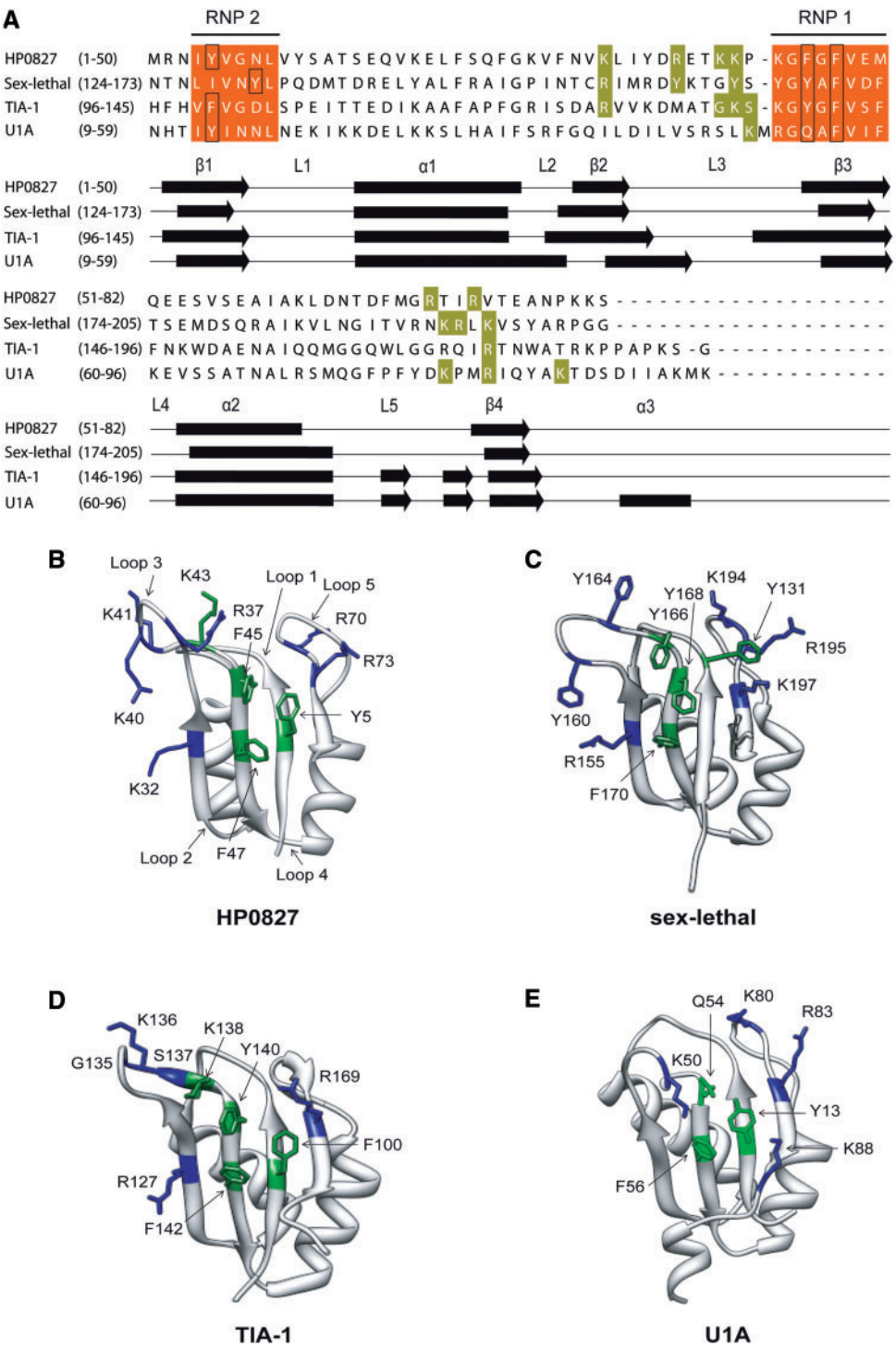

Fig. 5. **Sequence alignment and comparison of structural homologues of HP0827.** (A) The conserved RNP1 and RNP2 sequences are displayed in orange colour. The positively charged and aromatic amino acids outside RNP motif are displayed in green colour. The aromatic amino acids are highlighted by black boxes that are thought to be important for primary RNA binding within RNP motif. The secondary structure elements are indicated as black bars (α-helix and β-sheet) and black lines (coil region). Sequence alignment was generated by using ClustalW2 (*15*) and Jalview (*16*). (B) Structure of HP0827: surface-exposed and positively charged residues (Y5, K32, R37, K40, K41, K43, F45, F47, R70 and R73) shown by stick are putative RNA-binding sites. (C) Structure of sex-lethal protein: Y10, R34, Y39, Y43, Y45, Y47, F49, K73, K74 and K76 are putative RNA-binding sites. (D) Structure of TIA-1 protein: F100, R127, G135, K136, S137, K138, Y140, F142 and R169 are putative RNA-binding sites. (E) Structure of U1A protein: Y13, K50, Q54, F56, K80, R83 and K88 are putative RNA-binding sites. The residues coloured by green and blue are putative amino acids related to RNA binding within and outside RNP motif, respectively.

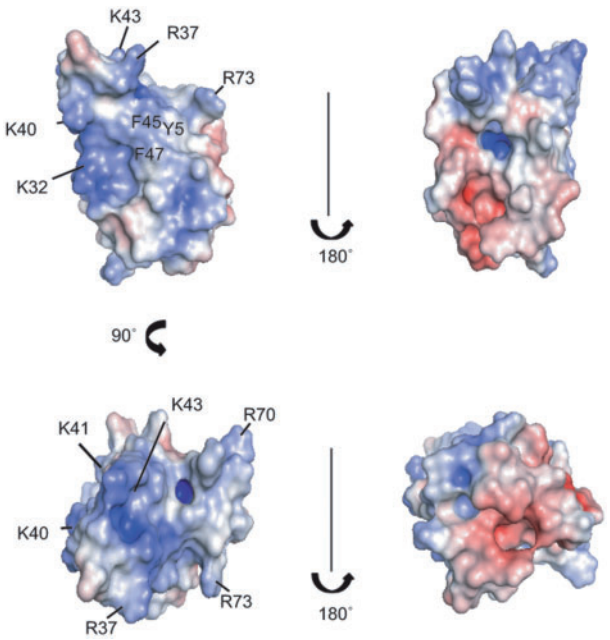

Fig. 6. **Electrostatic potential surface diagrams of HP0827 oriented to show the predicted RNA-binding regions.** The extreme ranges of red (negative) and blue (positive) represent electrostatic potentials of <−5 and >+5 k_bT, where k_b is the Boltzmann constant and T is the temperature. The figure was calculated with APBS showing the accessible surface (*36*).

Compared with three homologues, HP0827 has highly disordered Loop3 (β2–β3) loop region. In this region, there are four residues (R37, K40, K41 and K43) that have positive charge. This positively charged region may play an important role in RNA recognition and this disordered region would have a stable conformation when RNA interacts with this disordered region (Fig. 5B). Finally, we compared α-helices of the HP0827 with three structural homologues. The canonical conformation of HP0827 (α1, α2) is similar to those of other proteins. However, the orientation of α1 (*14–25*) to α2 (53–61) helix shows different pattern. Two α-helices of other homologues lie perpendicular to each other, whereas α1 lies parallel to α2 in HP0827 (Fig. 5B). Different orientation of α-helices with other RRMs may give different RNA recognition pattern or function to HP0827. On the basis of these discoveries, the understanding of the molecular recognition between RNA and proteins, and its functional implications would be an important subject in structural biology and biomedical research in the future studies.

PDB ACCESSION CODES

The atomic coordinates for the HP0827 have been deposited in PDB under accession code 2ki2.

ACKNOWLEDGEMENTS

We thank Dr Hookang Im for the critical comments on the manuscript. We also thank Korea Basic Science Institute (KBSI) and National Center for Inter-University Research Facilities (NCIRF) for using their NMR machines.

FUNDING

Ministry of Education, Science and Technology (MEST); New Drug Target Discovery; Grant number, 370C-20070095, Innovative Drug Research Center for Metabolic and Inflammatory Disease; and 2008 BK21 Project for Medicine, Dentistry and Pharmacy.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Cover, T.L. and Blaser, M.J. (1996) Helicobacter pylori infection, a paradigm for chronic mucosal inflammation: pathogenesis and implications for eradication and prevention. *Adv. Intern. Med.* **41**, 85–117
2. Forman, D., Newell, D.G., Fullerton, F., Yarnell, J.W., Stacey, A.R., Wald, N., and Sitas, F. (1991) Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. *BMJ* **302**, 1302–1305
3. Ge, Z. and Taylor, D.E. (1999) Contributions of genome sequencing to understanding the biology of Helicobacter pylori. *Annu. Rev. Microbiol.* **53**, 353–387
4. Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzegerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M., and Venter, J.C. (1997) The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* **388**, 539–547
5. Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F., and Trust, T.J. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* **397**, 176–180
6. Oh, J.D., Kling-Backhed, H., Giannakis, M., Xu, J., Fulton, R.S., Fulton, L.A., Cordum, H.S., Wang, C., Elliott, G., Edwards, J., Mardis, E.R., Engstrand, L.G., and Gordon, J.I. (2006) The complete genome sequence of a chronic atrophic gastritis Helicobacter pylori strain: evolution during disease progression. *Proc. Natl Acad. Sci. USA* **103**, 9999–10004
7. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280
8. Maris, C., Dominguez, C., and Allain, F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131
9. Allain, F.H., Howe, P.W., Neuhaus, D., and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* **16**, 5764–5772

10. Swanson, M.S., Nakagawa, T.Y., LeVan, K., and Dreyfuss, G. (1987) Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol. Cell Biol.* **7**, 1731–1739
11. Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K., and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell Biol.* **6**, 2932–2943
12. Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845
13. Hamimes, S., Arakawa, H., Stasiak, A.Z., and Buerstedde, J.-M. (2005) RDM1, a novel RNA recognition motif (RRM)-containing protein involved in the cell response to cisplatin in vertebrates. *J. Biol. Chem.* **280**, 9225–9235
14. Deborah, M.B., Todd, C.W., Timothy, J.A., John, P.R., and Gordon, S.R. (1998) The NMR structure of the RNA binding domain of *E. coli* rho factor suggests possible RNA-protein interactions. *Nat. Struct. Biol.* **5**, 393–399
15. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680
16. Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004) The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427
17. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293
18. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227
19. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302
20. Wishart, D.S. and Sykes, B.D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4**, 171–180
21. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921
22. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 29–32, 51–55
23. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
24. Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993) Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
25. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.* **8**, 477–486
26. Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., and Zardecki, C. (2002) The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899–907
27. Holm, L. and Sander, C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480
28. Inoue, M., Muto, Y., Sakamoto, H., Kigawa, T., Takio, K., Shimura, Y., and Yokoyama, S. (1997) A characteristic arrangement of aromatic amino acid residues in the solution structure of the amino-terminal RNA-binding domain of Drosophila sex-lethal. *J. Mol. Biol.* **272**, 82–94
29. Kuwasako, K., Takahashi, M., Tochio, N., Abe, C., Tsuda, K., Inoue, M., Terada, T., Shirouzu, M., Kobayashi, N., Kigawa, T., Taguchi, S., Tanaka, A., Hayashizaki, Y., Guntert, P., Muto, Y., and Yokoyama, S. (2008) Solution structure of the second RNA recognition motif (RRM) domain of murine T cell intracellular antigen-1 (TIA-1) and its RNA recognition mode. *Biochemistry* **47**, 6437–6450
30. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438
31. Mattaj, I.W. (1989) A binding consensus: RNA-protein interactions in splicing, snRNPs, and sex. *Cell* **57**, 1–3
32. Bandziulis, R.J., Swanson, M.S., and Dreyfuss, G. (1989) RNA-binding proteins as developmental regulators. *Genes Dev.* **3**, 431–437
33. Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621
34. Birney, E., Kumar, S., and Krainer, A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803–5816
35. Fleming, K., Ghuman, J., Yuan, X., Simpson, P., Szendroi, A., Matthews, S., and Curry, S. (2003) Solution structure and RNA interactions of the RNA recognition motif from eukaryotic translation initiation factor 4B. *Biochemistry* **42**, 8966–8975
36. Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041