4216–4228 *Nucleic Acids Research, 2012, Vol. 40, No. 9* *Published online 12 January 2012*
doi:10.1093/nar/gkr1305

Structural and biochemical characterization of HP0315 from *Helicobacter pylori* as a VapD protein with an endoribonuclease activity

Ae-Ran Kwon[1], Ji-Hun Kim[2], Sung Jean Park[3], Ki-Young Lee[2], Yu-Hong Min[1], Hookang Im[2], Ingyun Lee[2], Kyu-Yeon Lee[2] and Bong-Jin Lee[2,*]

[1]Department of Herbal Skin Care, College of Herbal Bio-Industry, Daegu Haany University, Gyeongsan 712-715, [2]Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742 and [3]College of Pharmacy, Gachon University of Medicine and Science, Incheon, 406-799, Korea

Received August 29, 2011; Revised December 20, 2011; Accepted December 21, 2011

ABSTRACT

VapD-like virulence-associated proteins have been found in many organisms, but little is known about this protein family including the 3D structure of these proteins. Recently, a relationship between the Cas2 family of ribonucleases associated with the CRISPR system of microbial immunity and VapD was suggested. Here, we show for the first time the structure of a member of the VapD family and present a relationship of VapD with Cas2 family and toxin–antitoxin (TA) systems. The crystal structure of HP0315 from *Helicobacter pylori* was solved at a resolution of 2.8 Å. The structure of HP0315, which has a modified ferredoxin-like fold, is very similar to that of the Cas2 family. Like Cas2 proteins, HP0315 shows endoribonuclease activity. HP0315-cleaved mRNA, mainly before A and G nucleotides preferentially, which means that HP0315 has purine-specific endoribonuclease activity. Mutagenesis studies of HP0315 revealed that D7, L13, S43 and D76 residues are important for RNase activity, in contrast, to the Cas2 family. HP0315 is arranged as an operon with HP0316, which was found to be an antitoxin-related protein. However, HP0315 is not a component of the TA system. Thus, HP0315 may be an evolutionary intermediate which does not belong to either the Cas2 family or TA system.

INTRODUCTION

According to the definition of virulence, virulence genes can be divided into true virulence genes, virulence-associated genes and virulence life-style genes (1). Virulence-associated protein, a product of the *vap* gene in various organisms, may be insufficient but requisite for virulence. Generally, *vap* genes are known as factors or enzyme-producing factors that regulate the expression of true virulence genes or activate virulence factors by translational modification, processing of secretion or that are required for the activity of true virulence factors. Several *vap* genes (*vapA, B, C, D, H* and *I*) are known to exist in various organisms (2–8) but how the products of *vap* genes are related to virulence remains unclear. Furthermore, the precise origin of *vap* sequence is not clear. Previous studies suggested that the *vap* regions may originate from phages via integration events (9). However, the *vap* regions also carry genes similar to plasmid-encoded sequences from other species (5, 9, 10).

VapD, a virulence-associated protein, was identified in *Dichelobacter nodosus* by Katz *et al*. (5). The exact biological role of the VapD protein has not yet been established, but several suggestions have been made. For example, VapD in *Haemophilus influenzae* is known as a toxin (11). In *Rhodococcus equi*, VapA and VapD are reported to be related to acid tolerance (12). In *Actinobacillus actinomycetemcomitans,* the product of the *orf2* gene is known to have high sequence similarity (78.9% amino acid identity) with the *vapD* of *D. nodosus*. That protein is related to plasmid maintenance and incompatibility (13). *vapD* has sequence similarity with the *vapBC* region and functionally has some connection with the *vapBC* operon (6). The *vapBC* operon encodes a variant of the bacterial Toxin–antitoxin (TA). VapBC members are found from Gram-negative, Gram-positive bacteria and Archaea, while the target of VapC as a toxin remains unknown. It has been shown that a toxin of the TA system regulates replication through DNA gyrase (14,15) or translation through mRNA

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

The authors wish it to be known that, in their opinion, the first two authors should be regarded as the joint First Authors

cleavage (16–18). The function of the TA system is still debated, but it is considered as an excellent control mechanism of gene expression able to withstand nutritional stress (19).

Helicobacter pylori is an important pathogen that infects up to 50% of the human population. It plays an important role in the pathogenesis of peptic ulcer disease, distal gastric lymphoma and even adenocarcinoma (20–22). *VapD* is present in about 60% of *H. pylori* strains (23). *Helicobacter pylori* strain 26695 has only one type of virulence-associated protein, VapD. Two genes in this strain (*hp0315* and *hp0967)* belong to VapD (24). In many prokaryote genomes, there are sequence homologs with the VapD family. According to the Pfam database and BLAST, HP0315 has sequence similarity with the clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein, Cas2. The products of *cas* genes possess functional domains that are typically found in nuclease, helicases, polymerase and polynucleotide-binding proteins (25). Recently, Cas2 proteins were found to be a novel family of endoribonucleases (26).

In this study, we obtained the soluble HP0315 protein through fusion with the immunoglobulin-binding domain of streptococcal *protein* G (GB1) and solved its crystal structure. Here, we present the first structure of a VapD family and its possible function. With this first VapD structure, we can gain a foothold in biological and biochemical investigations of VapD proteins.

EXPERIMENTAL PROCEDURES

Cloning of the HP0315 gene

The *hp0315* gene was PCR amplified from *H. pylori* 26 695 genomic DNA using DNA polymerase. The following primer pair was used: *hp0315*Fwd (5′-CGCGGATCCATGTATGCTTTAGCGTTTGATTTAAAG), *hp0315*Rev (5′-CCGCTCGAGGGATTTCACAATCTCAGTAAAATCG). The PCR product was purified using the AccuPrep®Plasmid Mini Extraction Kit (Bioneer). The purified gene was inserted into the modified pET30a fused with GB1. The recombinant plasmid was transformed into the *Escherichia coli* strain BL21(DE3), and this plasmid was purified and sequenced to confirm the exact gene cloning of HP0315 (Bioneer, Daejeon, Korea). This construct contains N-terminal GB1, which improves the protein stability dramatically. It also contains C-terminal eight nonnative residues (LEHHHHHH) to facilitate protein purification.

Protein expression, purification and site-directed mutagenesis

The cells were grown to an OD_{600} of 0.5 in LB including 100 μg/ml kanamycin after which the protein was overexpressed using 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 4 h at 37°C. The cells were harvested by centrifugation at ∼12 000*g* for 15 min and suspended in 100 ml of lysis buffer consisting of 50 mM Tris–Cl (pH 8.0), 5% (v/v) glycerol, 1 mM β-mercaptoethanol (βME) and 0.1 mM EDTA. Cell lysis was performed by sonication (duration time 1 min, duty time: 40% with 2 min cooling) with a Branson sonifier cell disruptor (Hielscher) in ice. The disrupted cells were then centrifuged at ∼16 000*g* for 1 h. After centrifugation, the supernatant was loaded into a His-tag column equilibrated with the 50 mM Tris–Cl (pH 8.0) buffer supplemented with 500 mM NaCl, 1 mM βME and 0.1 mM EDTA (Buffer A). The pure protein was eluted with an imidazole gradient from 100 to 300 mM imidazole using FPLC (GE Healthcare), and was subsequently bound to a second His-tag column. The pooled protein was dialyzed with Buffer A and concentrated. Finally, the protein was applied to SuperdexTM 75 gel column (GE Healthcare) in 150 mM NaCl, 50 mM Tris (pH 7.0), 0.1 mM EDTA and 1 mM DTT and the pooled protein was concentrated.

To prepare the derivative of GB1_HP0315 substituted by selenomethionine (SeMet), *E. coli* BL21(DE3) cells were grown in 5 ml of M9 medium containing 150 μg/ml kanamycin for 12 h, spun down, resuspended in 50 ml of M9 medium and then grown at 37°C overnight. This culture fluid was inoculated into 2 l of M9 medium. Cells were grown at 37°C, and at an OD_{600} of 0.5, 50 mg/l of selenomethionine, 100 mg/l each of Phe, Thr and Lys and 50 mg/l each of Leu, Ile, Val and Pro were added to block the methionine biosynthesis and to allow selenomethionine substitution (27). IPTG was added 20 min later and the culture was harvested after 4 h. Expression and purification of SeMet-substituted protein were conducted by the same method as described above for the native protein.

Site-directed mutagenesis of GB1_HP0315 was carried out using the QuikChange site-directed mutagenesis kit (Stratagene).

Gel filtration experiment

A gel filtration experiment was conducted to determine the size of the GB1_HP0315. To standardize the column with known molecular weights, the LMW Gel Filtration Calibration Kit (Amersham Pharmacia Biotech) was used. The kit includes ribonuclease A (13.7 kDa), chymotrypsinogen A (25 kDa), ovalbumin (43 kDa), bovine serum albumin (67 kDa) and blue dextran 2000. GB1_HP0315 and molecular weight standards were applied on a BioSep-SEC-S 3000 (Phenomenex) gel filtration column. The molecular weight was calculated as follows:

$K_{av} = (V_e - V_o)/(V_t - V_o)$. The total volume of the resin (V_t) was measured with tyrosine, the void volume (V_o) with blue dextran and V_e with the standards and GB1_HP0315. The standard curve was generated by plotting K_{av} against the logarithm of the molecular weight of the standard proteins. By applying the K_{av} value of GB1_HP0315 to the standard curve, its molecular weight was roughly estimated.

Crystallization and data collection

The SeMet-substituted protein was crystallized by hanging-drop vapor diffusion method at 293 K against a reservoir solution containing 0.2 M $(NH_4)_2SO_4$, 9% (w/v) PEG3350, 5–8% (v/v) glycerol, 100 mM 2-(*N*-morpholino) ethanesulfonic acid [MES (pH 6.0)]. Typical crystals were obtained within 3 days. The crystals grew up to

Table 1. Crystallographic data collection and refinement statistics

Data collection	
Beam line	BL-4 A (PLS)
Wavelength (Å)	0.97951
Resolution range (Å)	30.0–2.80 (2.90–2.80)[a]
Space group	$P3_121$
Unit cell parameters (Å)	$a = 77.81$ $b = 77.81$ $c = 106.31$
Observations (total/unique)	103 334/9568
Completeness (%)	99.8 (100.0)
Redundancy	10.8 (11.0)
R_{sym}[b]	8.3 (53.0)
I/σ	57.4 (5.1)
Phasing	
Phasing power	0.84
Figure of merit	0.44
Refinement	
R_{work}[c] (%)	25.0
R_{free}[c] (%)	28.4
Protein atoms	2044
Water molecules	6
Average B value (Å^2)	54.1
r.m.s.d. from ideal geometry	
Bond length (Å)	0.007
Angles (°)	1.210
Ramachandran analysis (%)	
Most favored region	85.9
Additional allowed region	14.1
Generously allowed region	0.0
Disallowed region	0.0

[a]Numbers in parentheses indicate the statistics for the last resolution shell.
[b]$R_{sym} = \sum|I_{hkl}-<I_{hkl}>|/\sum<I_{hkl}>$, where I_{hkl} = single value of measured intensity of *hkl* reflection, and $<I_{hkl}>$ = mean of all measured value intensity of *hkl* reflection.
[c]$R_{work} = \sum|F_{obs}-F_{calc}|/\sum F_{obs}$, where F_{obs} = observed structure factor amplitude, and F_{calc} = structure factor calculated from model. R_{free} is computed in the same manner as R_{work}, but from a test set containing 10% of data excluded from the refinement calculation.

approximate dimensions of 0.2 mm × 0.2 mm × 0.2 mm. For cryo protection, crystals were transferred to a cryoprotecting solution containing 0.25 M $(NH_4)_2SO_4$, 10% (w/v) PEG3350, 20% (v/v) glycerol and 100 mM MES (pH 6.0) for <2 min in several steps and flash-cooled in a stream of nitrogen gas at 100 K. A set of Se Single Anomalous Scattering (SAS) data was collected at 0.97951 Å using an Area Detector System Corporation Quantum 210 charge-coupled device detector at the beam line 4 A of Pohang Light Source (PLS), Korea. The crystal was rotated through a total of 180° with a 1.0° oscillation range per frame. The data were processed and scaled using the program suit HKL2000 (28). The SeMet-substituted crystal belongs to the space group $P3_121$ with unit cell parameters of $a = 77.81$ Å, $b = 77.81$ Å and $c = 106.31$ Å. The data collection and processing statistics are summarized in Table 1.

Structure determination and refinements

The asymmetric unit of GB1_HP0315 crystal contains two molecules. Therefore, the crystal of Se-GB1_HP0315 contains four Se sites (first methionine residue from each GB1 monomer and methionine residue from each HP0315 monomer) in the asymmetric unit. Two selenium sites (of the Se–Met from HP0315) were located with the program SnB (29) using single anomalous scattering (Se peak) data. Two Se–Met sites were used to generate initial phases and initial partial model was built using the program RESOLVE (30). About 50% of the model could be constructed and a more complete model was built with Coot (31). One GB1 part of the two monomers could not be traced. Rigid body refinement followed by torsional simulated annealing and group B factor refinement reduced the R-factors to $R_{work} = 0.283$, $R_{free} = 0.357$. The initial model was subjected to iterative cycles of manual model building with COOT, and crystallographic refinement by CNS (32,33). Solvent molecules became apparent in the later stages of refinement and were added into the model. Further refinement was pursued until no further decrease of the R_{free} was observed ($R_{work} = 0.250$, $R_{free} = 0.284$). The values of R_{work} and R_{free} are a little high because we could not trace one GB1 part of the two monomers. The final model exhibited good stereochemical geometry (Table 1). All stereochemical parameters are within the expected margins and the Ramachandran plot shows 85.9% of residues in the most favored and 14.1% in the additional allowed region, with none in the generously allowed or disallowed region as defined in PROCHECK (34). Refinement statistics are summarized in Table 1. Coordinates and experimental structure factors of the GB1_HP0315 structure have been deposited in the Protein Data Bank with accession code ID 3UI3.

Generation of RNA substrates

HP0315 mRNA substrate was synthesized using the Ambion T7 RNA polymerase MEGAscript high yield transcription kit (Ambion). The long mRNA (308 nt) substrate of *hp0315* was synthesized using *in vitro* transcription reaction. DNA template with a T7 promoter site was prepared by PCR and the following pair of primers was used: Fwd (5′-TAATACGACTCACTATAGGGGAATT GTGAGCGGATAACAATTC), Rev (5′- CTCCGTATT CTTTTTTTAAAATCTC). The short RNA substrates were synthesized by Integrated DNA Technologies (IDT, Coralville, IA, USA).

RNase activity assay

Purified recombinant GB1_HP0315 protein from *H. pylori* (10 μM) was incubated at 37°C with the nucleic acid substrates (1 μg dsDNA (308 nt), 2 μg ssDNA (42 nt) and 1 μg ssRNA (308 nt)) in the presence of 200 mM Tris (pH 7.0) for 60 min. RNase activity assay was carried out at various pH levels (pH 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0 and 11.0). In each reaction, no metal, 2.5 mM Mg^{2+}, Mn^{2+}, Co^{2+} or Ca^{2+} was added, respectively. RNase activity was tested with no monovalent ion, Na^+ from 100 to 500 mM and K^+ from 100 to 400 mM.

A fluorescence experiment proceeded according to the manual of RNaseAlert (IDT). RNaseAlert is a sensitive assay designed for the detection of RNase activity. Basically, a fluorescent-labeled RNase substrate that contains a quencher on one end and a fluorescent molecule on the other emits green fluorescence if it is cleaved by RNase. The fluorescence emitted from cleavage of the

substrate was measured at 520 nm after excitation at 490 nm. GB1_HP0315 (10 μM) was incubated at 37°C with a fluorescence substrate (50 pmol) in the presence of 20 mM Tris (pH 7.0) and 10× RNaseAlertTM buffer for 180 min.

Primer extension inhibition assay

The HP0315 mRNA was used as a template. Each specific DNA primer was synthesized and 5′ labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. Primer (0.8 pmol) was mixed with each template (1 pmol) in a reaction buffer to a total volume of 20 μl. The mixture was heated to 90°C for 5 min and then slowly cooled to room temperature. The reaction mixture was digested by adding various amounts of GB1_HP0315 protein at 37°C for 30 min. A subsequent primer extension reaction was performed by adding 8 μl of 5× buffer [250 mM Tris–HCl (pH 8.3), 375 mM KCl and 15 mM $MgCl_2$], 5 mM DTT, 0.5 mM dNTP and 40 U of Superscript III reverse transcriptase (Invitrogen) for a final reaction volume of 40 μl. cDNA was synthesized at 55°C for 30 min after which the products were subjected to 10% (w/v) denaturing PAGE and analyzed by autoradiography.

Enzymatic assay against short RNAs

Based on the primer extension products of the pre-digested RNA, the region around the most frequently cleaved site was selected for the short RNA cleavage assay. The short RNA substrates were purchased from IDT and 5′ labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. The labeled RNA substrates (1 pmol) were mixed with 10 μg (0.56 pmol) of GB1_HP0315 protein in a reaction buffer in a total volume of 20 μl. The mixtures were incubated at 37°C for 30 min. The reaction was quenched by the addition of an equal volume of formamide loading buffer [80% (v/v) formamide, 0.025% (v/v) bromophenol blue, 0.025% (v/v) xylene cyanol and 10 mM EDTA (pH 8.0)]. The reaction products were resolved by electrophoresis in 15% (w/v) polyacrylamide and 8 M urea gels using TBE buffer. As nucleotide size markers, an imidazole ladder was prepared by partial RNA cleavage by imidazole. The labeled RNA substrates (1 pmol) were incubated in 2 M imidazole and 0.5 mM EDTA (pH7.0) in a total volume of 20 μl at 90°C for 10 min. An equal volume of formamide loading buffer was then added.

RESULTS

Stability improvement of HP0315

The *hp0315* gene was initially cloned into the pET-21 a(+) vector to produce the C-terminally His-tag fused protein product. This native protein was unstable at a protein concentration that exceeded 20 μM and precipitated easily. GB1 is a relatively small fusion protein among fusion proteins such as glutathione *S*-transferase (GST) and maltose-binding protein. For this reason, GB1 is quite advantageous when used in structural studies using NMR or crystallography, which means that structural studies of protein fused with GB1 can be done even without the cleavage of the fusion tag (35). In our study, solubility of HP0315 was greatly improved by N-terminal GB1 tagging. GB1_HP0315 could be concentrated without any precipitation. To evaluate whether the native structure of GB1_HP0315 was well maintained, the secondary structure of the fusion protein was analyzed. As GB1 is a small protein made up of 23% helical and 39% β sheet secondary structures (PDB ID: 2GB1), it is predicted that GB1 does not have much of an effect on the CD spectra of GB1_HP0315. The prediction was in good agreement with the results of the CD spectra. Two CD spectra (with and without GB1) were similar to each other, indicating that at least the secondary structure of HP0315 was not changed by the fusion protein, GB1 (Supplementary Figure S1).

Dimeric state of HP0315

The molecular weight of GB1_HP0315 was estimated to be 17.4 kDa from the protein sequence. As shown in Supplementary Figure S2, the apparent molecular weight of GB1_HP0315 was determined to be ~41 kDa, which indicated that GB1_HP0315 could be a dimeric protein with a molecular weight of 34.8 kDa. This result means that HP0315 has a dimeric structure, as GB1 itself is a monomeric protein.

Overall structure of HP0315

The dimeric structure of HP0315 fused with a GB1 protein at each N-terminus was determined at a resolution of 2.8 Å. The structure of HP0315 from *H. pylori* consists of 10 secondary structure elements: β1 (residues 1–8), α1 (residues 10–17), α1′ (residues 21–35), β2 (residues 38–41), β3 (residues 44–47), α2 (residues 53–66), α2′ (residues 68–73), β4 (residues 75–87) and α3 (residues 88–93). The monomer has a ferredoxin-like fold. It has the β1-(α1-α1′)-β2-β3-(α2-α2′)-β4-α3 instead of the β-α-β-β-α-β structure of the ferredoxin fold. In the first half of the monomer, α1 is shorter than α1′ and, in the last half of the monomer, α2 is longer than α2′. The α1′ and α2 helices have amphipathic property, and the hydrophobic side of these amphipathic helices faces the inner space. Many hydrophobic residues facing the protein interior form a hydrophobic core for structural stability, which give the structure components stability (Figure 1D and E). In addition, the β strands are stabilized by direct hydrogen bonds between neighboring β-strands.

The dimer of HP0315 is butterfly shaped (Figure 1A). The β4 strand and the α3 helix associate with the adjacent monomer, forming a dimerization interface. Direct hydrogen bonds between β4 from one monomer and β4 from the other monomer contribute to dimerization. The area of the calculated dimerization interface is 2900 Å^2. This area represents 29.3% of the total solvent-accessible surface area of a dimer. HP0315 forms a dimer with a root-mean-square difference (r.m.s.d.) of 0.362 Å between two monomers (Figure 1B), with small differences in the loop between β2 and β3 and a C-terminal region between the monomers.

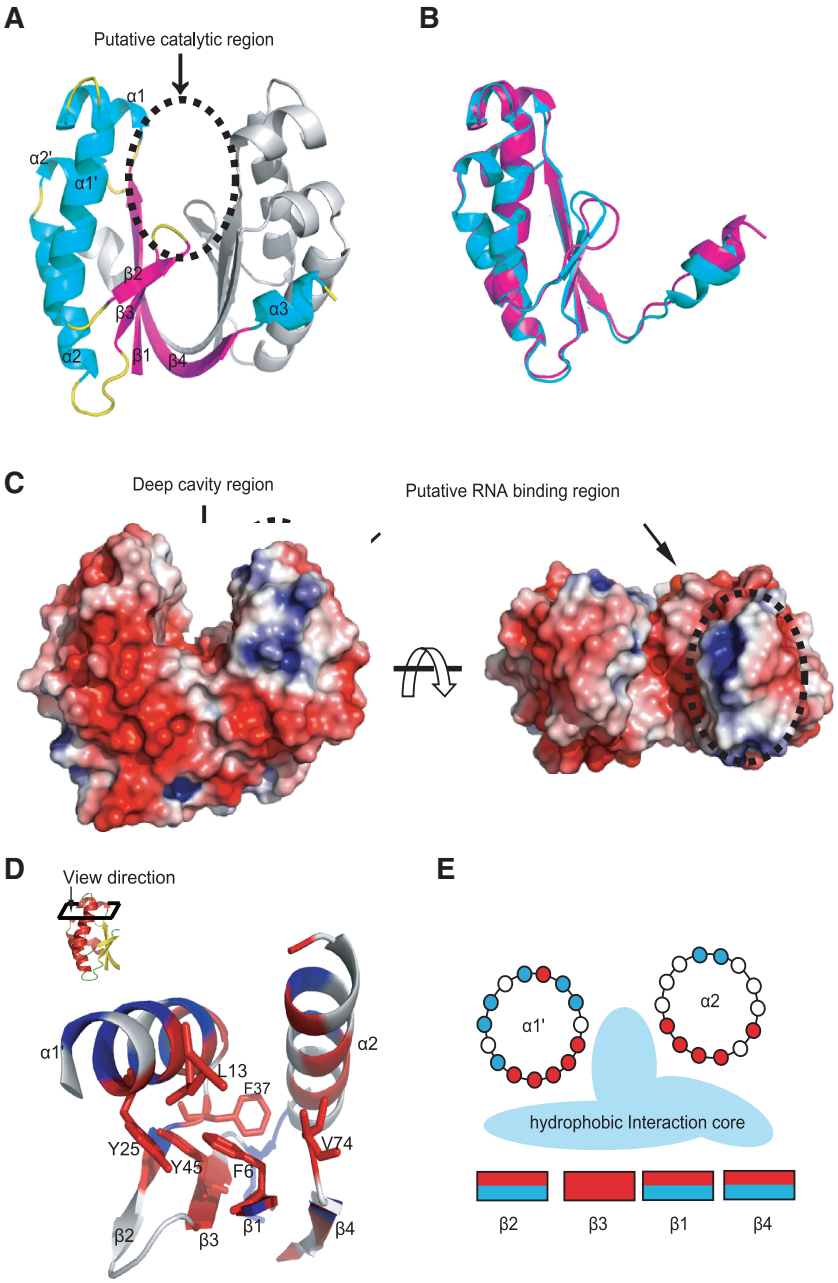

Figure 1. Structure of HP0315 from *H. pylori*. (**A**) Cartoon representation of dimer of HP0315 (α-helices, β-strands and loops are cyan, magenta and yellow, respectively). Dotted circle represents the putative catalytic region located at deep cavity region. (**B**) Superposition of molecules A and B from HP0315 with an average r.m.s.d. of 0.362 Å. Two molecules are slightly different from each other in the loop between β2 and β3 and C-terminal region. (**C**) Surface representation of HP0315 showing positive electrostatic potential in blue and negative in red. Dotted circle represents the putative RNA-binding region. This region would be related to initial binding with RNA, and then second catalytic reaction would be happen around deep cavity region. (**D**) The image seeing down from top indicates that highly conserved residues (F6, L13, Y25, F37, Y45 and V74) from the result of structure based sequence alignment are concerned with hydrophobic interaction. Hydrophobic residues are colored in red and hydrophilic residues are in blue, respectively. (**E**) This figure shows that clustered hydrophobic residues into inside form the hydrophobic interaction core. All figures were prepared using PyMOL (52).

A deep cavity is situated in the middle of the dimer structure of HP0315. The solvent-accessible residues on the inside of the deep channel can be considered as a putative active site. The two pinnacles surrounding the channel are positively charged (Figure 1C), indicating that they may be putative RNA-binding sites. In the light of these observations, the RNA-binding region and the putative active site region appear to be slightly separated from each other.

Figure 2A shows the distribution of the α carbon B-factor of the A molecule in the HP0315. High B-factor regions are located on α1 helix, the loop between β2 and β3 strands, the loop between β3 strand and α2 helix, and the C terminal tail (Figure 2B).

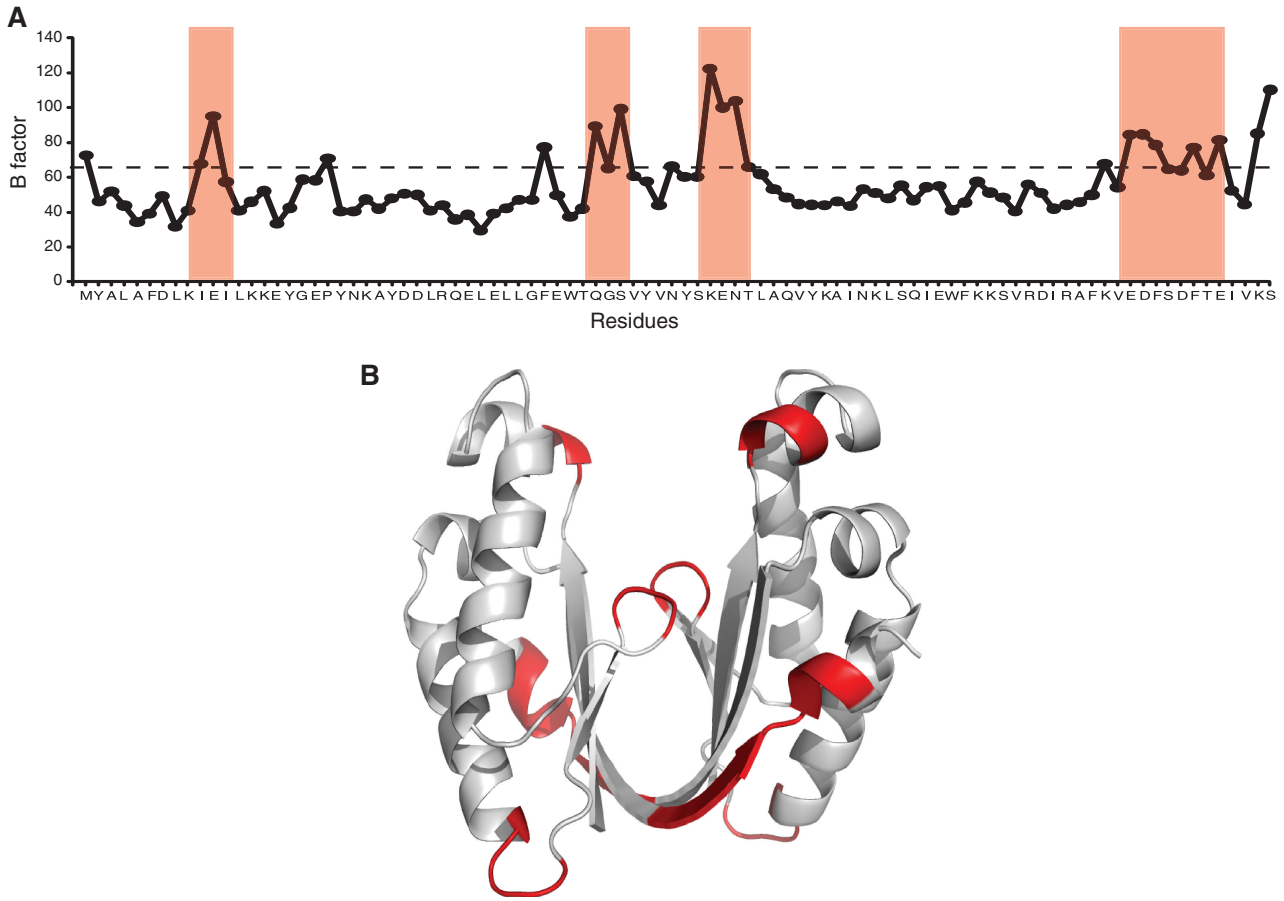

Figure 2. Structural analysis according to α carbon *B*-factor. (**A**) The distribution of α carbon *B*-factor of each residues. The regions with high α carbon *B*-factor are colored under red. (**B**) High distribution of α carbon B-factor is represented with red color on the cartoon structure of HP0315.

Sequence analysis

The molecular function of VapD is unknown. As a first step toward a functional evaluation, we performed a sequence-based homology search (Figure 3A). There are many sequence homologs among the proteins of the VapD family. In particular, residues in the first β-strand (β1), the loops between β2 and β3, and C-terminal region are highly conserved (Figure 3A). By considering that the C-terminal conserved region resulting from sequence alignments is related to dimerization, it is thought that the other region may contain active-site residues.

According to the Pfam database, HP0315 is related to the CRISPR-associated protein Cas2, with an *E*-value 4.1e-08. In addition, Makarova *et al.* (6) investigated the interrelationships between VapD and Cas2 proteins. According to the aspects of the sequence and/or structure (see the structural homologs part), VapD was determined to be a protein that was related to the Cas2 proteins. However, a CRISPRs sequence composed of 25–50 bp, which is characteristic of Cas2 proteins, was not found near the *hp0315* gene.

The result of the *H. pylori* genome analysis indicates that there is another interesting feature. The two genes *hp0315* and *hp0316* exist as an operon, which is a functional unit of genomic DNA containing partially overlapped genes under the control of a single regulatory signal or promoter (gene coordinates: *hp0315* 330872–330588, *hp0316* 331245–330853) (Supplementary Figure S3A) (36). HP0316 is an unknown protein and has a sequence similarity of 88.9% with HP0895 (Supplementary Figure S3B). Recently, HP0895 provided a clue about the function of HP0316 (37). *hp0895* forms an operon with *hp0894,* and this operon was found to be a TA system. HP0894 is known as a toxin possessing RNase activity. Based on this similarity, despite the fact that HP0315 has no sequence and structural homology with HP0894, we postulated that HP0315 may have the same role as HP0894. To test whether HP0315 is a toxin or not, we undertook a HP0315–HP0316 binding assay. The result of a His-tag pull-down assay indicated that HP0315 did not bind to HP0316 (Supplementary Figure S3C). Furthermore, *in vivo* toxicity experiments revealed that HP0315 and HP0316 do not form a TA system (Supplementary Figure S3D). Therefore, the biological roles of HP0315 and HP0316 should be different from those of HP0894 and HP0895.

Structural homologs

A DALI search (38) using the dimer structure may supply a possible function of this molecule. Nine structures were

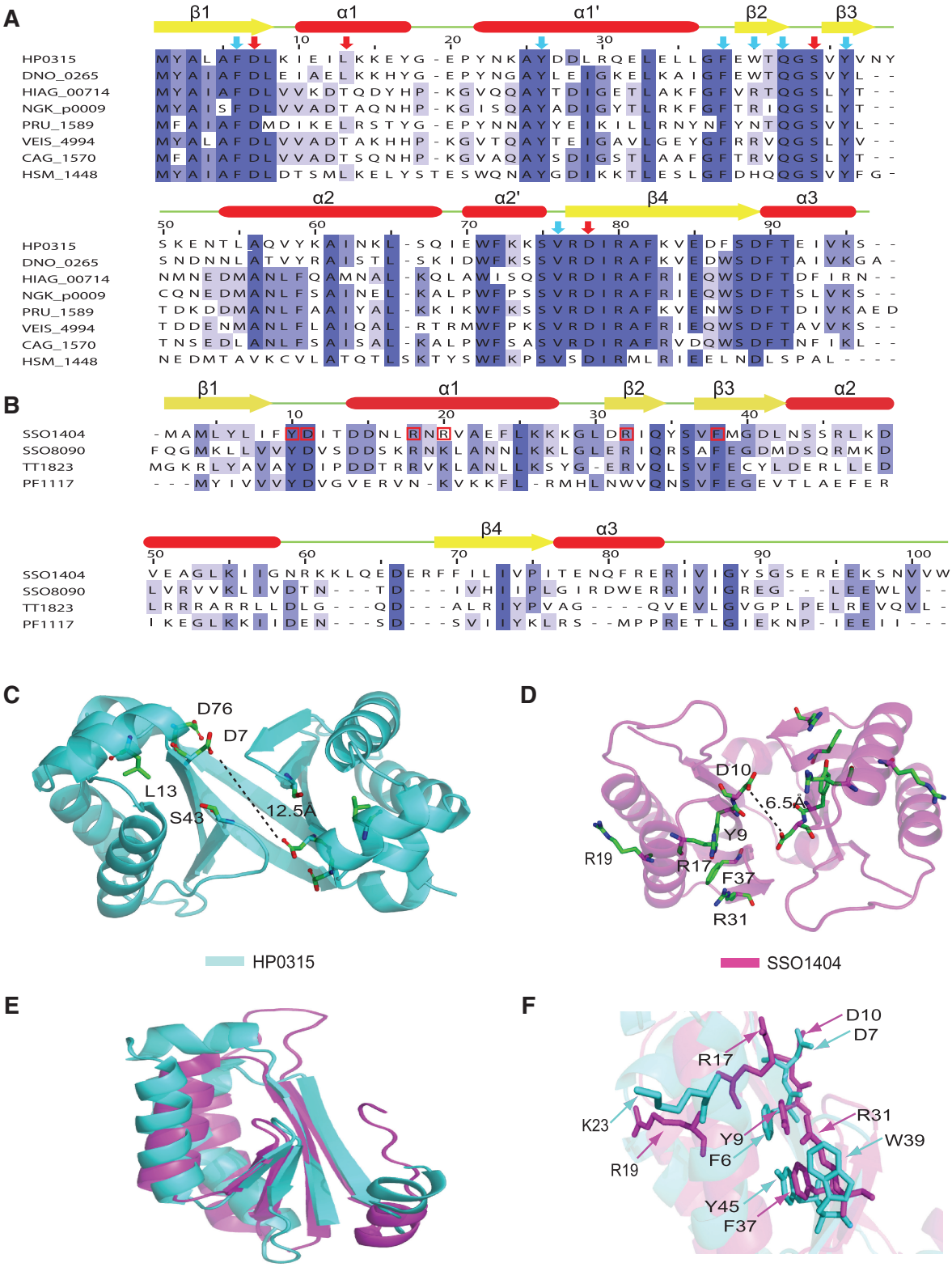

Figure 3. Structure-based sequence analysis of HP0315 with other proteins belonging to VapD family and Cas2 proteins. (**A**) Residues conserved in all VapD family are colored under dark blue, highly conserved residues under medium blue and somewhat conserved residues under light blue. The secondary structure elements derived from the structure of HP0315 are displayed above the alignment. Residues marked with arrow were highly conserved among VapD family and selected as candidates for mutational studies. Residues with red arrow are related to catalytic activity based on mutational study. The compared proteins are as follows: HP0315 (NCBI accession ID: NP_207113), DNO_0265 (YP_001209119) from *D. nodosus*, HIAG_00714 (ZP_05850077) from *H. influenza*, NGK_p0009 (YP_002000622) from *N. gonorrhoeae*, PRU_1589 (YP_003574888) from *P. ruminicola*, VEIS_4994 (YP_999698) from *V. diseniae*, CAG_1570 (YP_379868) from *C. chlorochromatii*, HSM_1448 (YP_001784768) from *H. somnus*. (**B**) Structure-based sequence analysis of Cas2 proteins. Residues are colored like above sequence alignment. Residues marked with red box of SSO1404 are important for the catalytic activity of SSO1404 (26). The compared proteins are as follows: SSO1404 from *S. solfataricus*, SSO8090,

(continued)

identified with a *Z*-score between 6.0 and 6.8. All of them have a ferredoxin-like fold. With only *Z*-scores that exceeded 6.4, the DALI program filtered the top three hypothetical unknown proteins (Supplementary Figure S4), TT1823 from *Thermus thermophilus* ($Z = 6.8$, PDB ID 1zpw r.m.s.d. 2.2 Å), SSO8090 from *Sulfolobus solfataricus* ($Z = 6.8$, PDB ID 3exc r.m.s.d. 2.4 Å) and PF1117 from *Pyrococcus furiosus* ($Z = 6.4$, PDB ID 2i0x r.m.s.d. 2.5 Å). The molecules with the next highest *Z*-scores with known functions were the adenylyl cyclase fragment ($Z = 5.9$, PDB ID 1y11) and various ribosomal proteins.

According to the Pfam database, the top three proteins above belong to Cas2 family. A function was reported on one of the Cas2 proteins ($Z = 6.0$ PDB ID 2ivy) (26), although it did not belong to top three proteins in our DALI search. The reported structure of SSO1404 (PDB ID 2ivy r.m.s.d 3.1 Å) from *S. solfataricus* can be a representative structure of the proteins associated with the CRISPRs (6). Cas2 proteins are a product of the *cas2* gene, one of the two *cas* (CRISPR-associated) genes, *cas1* and *cas2*, belonging to the COG1343 and COG3512 groups of the clusters of orthologous groups (COG) classification system (6, 39). The overall structures of VapD and SSO1404 (26) are very similar to each other, although there are some differences in the distribution of the secondary structure elements. The major difference is in the β4 strand region of VapD, which was replaced by a helix in SSO1404. Another difference lies at the α1 region of VapD which was replaced by a loop in SSO1404. The homodimer interface part for both proteins shows a high level of structural similarity. This interface part corresponds to the N-terminal region of VapD. Those residues that are highly conserved in VapD shown as by the Consurf program (40–42) appear to correspond well to the active site region of SSO1404 and appear to be mainly distributed on the interface region of the dimer structure (Figure 3C, D and Supplementary Figure S5). A comparison of the key residues of SSO1404 with the corresponding residues of HP0315 shows that Y9, D10 and F37 of SSO1404 are remarkably conserved as F6, D7 and Y45 of HP0315, respectively (Figure 3B, E and F). From the structure and structure-based alignment of key residues, we speculate that HP0315 may show ribonuclease activity.

Enzymatic activity of HP0315

The results presented in the previous section showed that HP0315 may display ribonuclease activity, similar to Cas2 proteins. To identify the biochemical activity of HP0315, we performed a primer extension experiment as well as agarose-based nuclease experiments. The purified protein and various nucleotide substrates including linear double-stranded DNA (dsDNA: PCR amplified of *hp0315* gene, 308 nt), linear single-stranded DNA (ssDNA: C-terminal oligonucleotide of *hp0315* gene, 42 nt) and single-stranded RNA (ssRNA: *in vitro* transcript of *hp0315* gene, 308 nt) were used to test for nuclease activity (Figure 4A). Interestingly, GB1_HP0315 cleaved only ssRNA among them. In addition, GB1_HP0315 showed RNase activity over a broad range of pH with the maximum activity at pH 7.0–7.5 (Figure 4B). Although some nucleases, including Cas2 proteins, show the highest level of activity in the presence of divalent cations, e.g. Mg^{2+} and Mn^{2+} (26,43,44), the ribonuclease activity of GB1_HP0315 was not dependent on metal ions with ranges of Mg^{2+} and Mn^{2+} concentrations of 1, 10, 50, 100 μM, 1 and 2.5 mM (Figure 4C) (data are not shown). In addition to divalent ions, the RNase activity of GB1_HP0315 was not influenced by monovalent ions. We performed agarose-based electrophoretic mobility shift assays according to the concentration of the monovalent ions of Na^+ or K^+ ions from 0 to 400 mM. However, in the case of GB1_HP0315, the RNase activity was not significantly influenced by the Na^+ or the K^+ ions (Figure 4D).

Cleavage of the long mRNA

To confirm the RNase activity of HP0315 and to identify the cleavage site of the mRNAs of HP0315, primer extension inhibition experiments were performed. As a result, GB1_HP0315 predominantly cleaved mRNAs before adenine (A) and guanine (G) (Figure 5). Therefore, HP0315 is expected to be a purine-specific endoribonuclease.

Cleavage of short RNAs

To identify the limitation of the mRNA length which can be cleaved by GB1_HP0315, various short-length RNA templates (from 14 to 6 nt) were used for an RNase activity test (Figure 6). Initially, we tested the sequence (5′-GACUUUAGCGAUUU-3′: 14 nt) including a cleavable site, based on the primer extension results. Other small RNA templates were prepared by cutting 2 nt (1 nt at each terminal site). As shown in Figure 6, GB1_HP0315 could cleave at least 6 nt RNA. RNAs shorter than 6 nt were not tested due to the technical difficulty associated with this type of synthesis.

Putative active site and mutational study of HP0315

It is expected that the residues which are important for RNase activity meet the following conditions: (i) residues

Figure 3. Continued
TT1823 from *T. thermophilus* and PF1117 from *P. furiosus*. (**C**) Important residues for its activity of HP0315. According to mutational study, D7, L13, S43 and D76 are important residues for its activity. D7 and D76 are supposed to be a nucleophile which can attack and cleave phosphodiester bond of ssRNAs. L13 located on the α1 helix region can affect the local environment of α1 helix which is a candidate for first RNA-binding region using positive residues. S43 located between β2 and β3 might bind with uracil base, which might help the specific cleavage of RNA. (**D**) Important residues for its activity of SSO1404. D10 is known as a catalytic active site (26). (**E**) Superposition of HP0315 and SSO1404. HP0315 is colored by cyan and SSO1404 is colored by magenta. (**F**) Superposition of key residues of SSO1404 with the corresponding residues of HP0315. The sequences were aligned by CLUSTALW2 (53), and the figure was generated using Jalview (54).

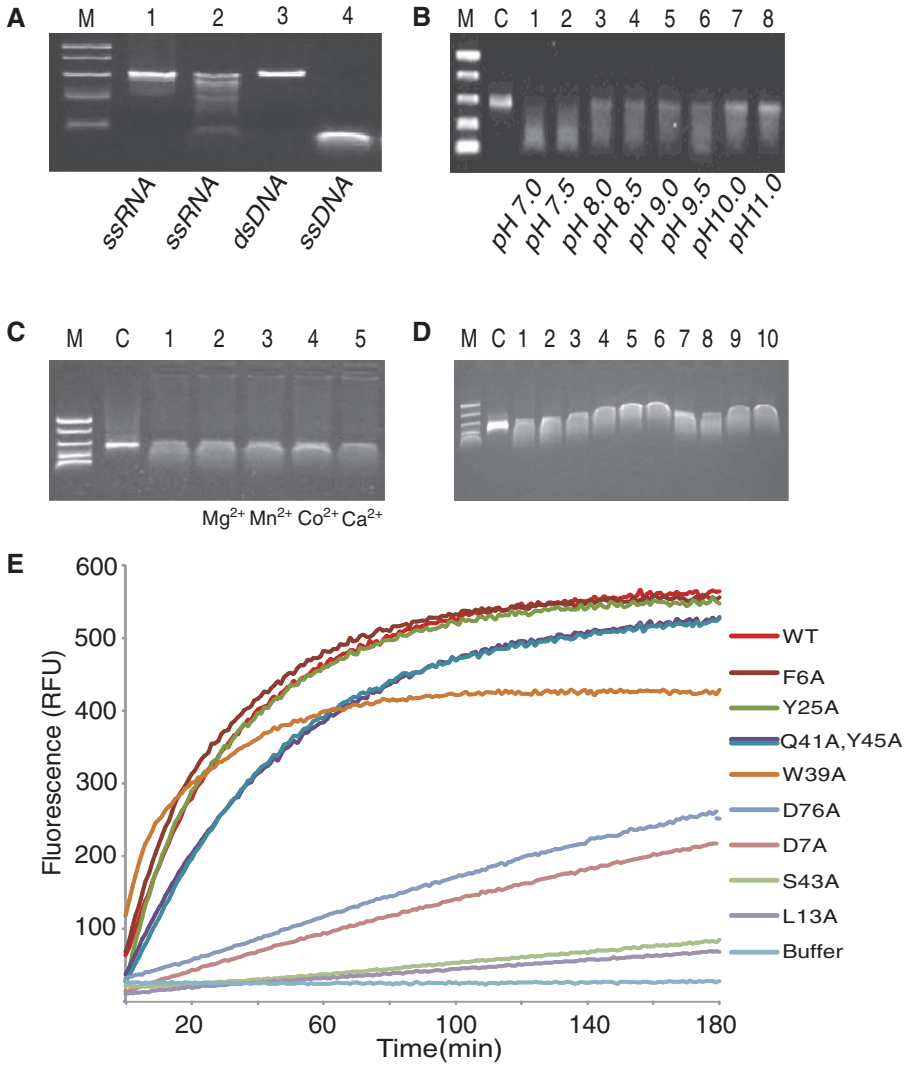

Figure 4. Enzymatic activity experiments. (**A**) Lane 1 is ssRNA alone. Lane 2 is ssRNA with GB1_HP0315. Lane 3 is dsDNA with GB1_HP0315. Lane 4 is ssDNA with GB1_HP0315. (**B**) Lane C is RNA alone, lane 1 is at pH 7.0, lane 2 at pH 7.5, lane 3 at pH 8.0, lane 4 at pH 8.5, lane 5 at pH 9.0, lane 6 at pH 9.5, lane 7 at pH 10.0 and lane 8 at pH 11.0. Some precipitations are occurred at and below pH 6.0. (**C**) Lane C is RNA alone and lane 1 is RNA and GB1_HP0315 without divalent cation. Lane 2 is RNA and GB1_HP0315 with 2.5 mM Mg^{2+} ion, lane 3 with 2.5 mM Mn^{2+}, lane 4 with 2.5 mM Co^{2+} and lane 5 with 2.5 mM Ca^{2+} (**D**) Lane C is RNA alone. Lane 1 is RNA and GB1_HP0315 with no monovalent ion. Lane 2 is RNA and HP0315 with 100 mM NaCl, lane 3 is with 200 mM NaCl, lane 4 is with 300 mM NaCl, lane 5 is with 400 mM NaCl and lane 6 is with 500 mM NaCl. Lane 7 is RNA and HP0315 with 100 mM KCl, lane 8 is with 200 mM KCl, lane 9 is 300 mM KCl and lane 10 is with 400 mM KCl. (**E**) Ribonuclease activity of mutated GB1_HP0315 using fluorescence. A fluorescent substrate is incubated with various mutated GB1_HP0315 in same condition. When fluorescence-quenched oligonucleotide probes are cleaved, fluorescence can be appeared in the process of time. Wild-type protein of HP0315 cleaved the substrate most quickly. F6A, Y25A, Q41A, Y45A, W39A, D76A, D7A, S43A and L13A showed highest RNase activity in order. This result clearly indicates that D7, L13, S43 and D76 are most important residues for ribonuclease activity.

which are conserved by the multiple sequence alignment of VapD proteins and (ii) residues that are conserved by the structure-based sequence alignment of HP0315 and selected Cas2 proteins.

Referring to the two criteria, several residues were selected: Phe-6, Asp-7, Leu-13, Tyr-25, Phe-37, Trp-39, Gln-41, Ser-43, Tyr-45, Val-74 and Asp76. To identify the residues required for its activity, we performed alanine mutagenesis screening of the above residues. All of the selected residues did not have direct relevance to the catalytic active site. Among the selected residues, some residues appear to be connected to the folding and/or stability of the structure, as F37A and V74A seriously reduced the stability of HP0315 and thus could not be purified. D7A, L13A, S43A and D76A significantly reduced the ribonuclease activity of GB1_HP0315 (Figure 4E).

The F6, L13, Y25, F37, Y45 and V74 residues appeared to be more involved with maintaining the structure of the key catalytic region than with the catalytic activity itself. As stated earlier, the above residues face into the inside of the structure, forming a hydrophobic core (Figure 1D). Especially, L13 contributes to the stability of α1 helix. As the α1 helix is located at the top of the dimer structure (Figure 1D), it may be a candidate for the initial RNA binding.

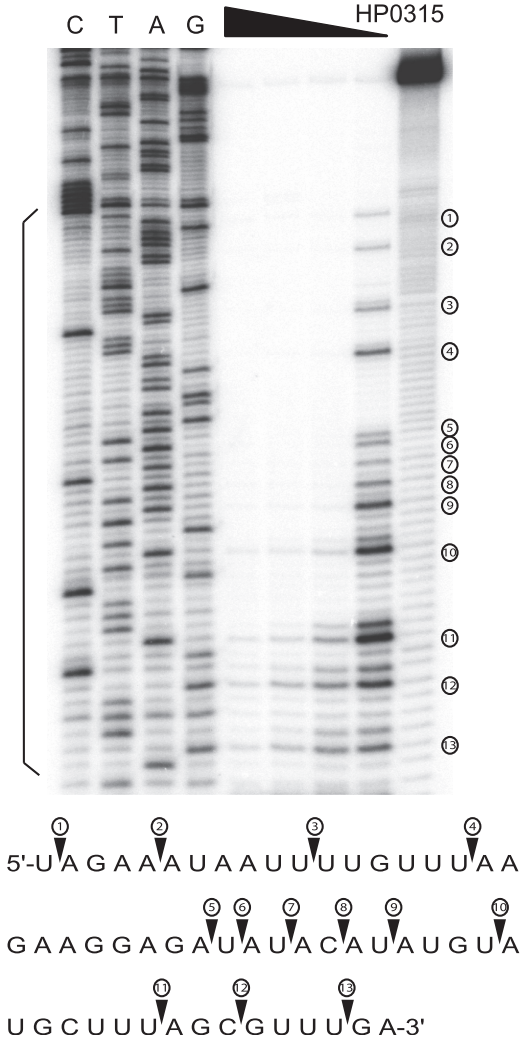

Figure 5. Primer extension inhibition analyses of mRNA cleavage by GB1_HP0315. Lanes 1–4 are DNA sequence ladders. Lane 5–9 is primer extension products of mRNA of HP0315 after digestion with various amounts of GB1_HP0315. Cleavage sites are indicated by sequential numbers on right side of the images. The RNA recognition sequences analyzed by the DNA sequencing ladder are on bottom side of the images. Preferential cleavage sites are mainly before the bases A and G.

DISCUSSION

VapD, the product of a virulence-associated gene with about 90 amino residues, is present in most bacterial species. Thus far, little information on VapD has been available but, in this article, we describe the first structural and biochemical characterization of VapD.

Comparisons of the structure of HP0315 with those of other homologs belonging to the Cas2 family as well as the result of a sequence blast strengthened the functional relationship between HP0315 and Cas2 proteins. For this reason, HP0315 was assumed to display ribonuclease activity, similar to Cas2 proteins. As expected, our study shows that HP0315 displays ribonuclease activity specific to ssRNA substrates. For an accurate assessment of the ribonuclease activity, in our experiments, we excluded the presence of any trace of imidazole and guanidinium derivatives, which also have weak ribonuclease activity (45).

The appearance of the overall monomeric structure of HP0315 is similar to those of the Cas2 proteins, particularly the β-strands (Figure 3E). Furthermore, the key residues on the β strands, Y9, D10 and F37, from SSO1404, a recently reported Cas2 protein, correspond very well to F6, D7 and Y45, respectively, of HP0315 (Figure 3F). However, the mechanism of SSO1404 activity is somewhat different from that of HP0315. The basic principle of the mechanism suggested for the activity of SSO1404 begins with coordination between Mg^{2+} and two D10 residues from dimer molecules. These two Asp residues from SSO1404 dimer molecules can create coordination with a Mg^{2+} ion, as the distance between the side chains from the two Asp residues is only 6.5 Å (Figure 3D). However, in the case of HP0315, the distance between the side chains of two instances of D7, which is the corresponding residue of D10 of SSO1404, is around 12.5 Å (Figure 3C), which complicates the coordination between Mg^{2+} (or any other metal ion) and the two Asp side chains. In fact, HP0315 showed ribonuclease activity without the addition of metal ions. In our mutational studies, D7 of HP0315 was also found to be an important residue for its ribonuclease activity, like the corresponding residue D10 of SSO1404. However, it may have a different role from the previously mentioned mechanism of SSO1404. Besides D7, our mutational screen results indicated that L13, S43 and D76 could also be considered as important residues for ribonuclease activity. The D7, L13 and D76 residues are located in β1, α1 and β4, respectively, and the S43 residue is located on the short-turn region between β2 and β3 (Figure 3C). It was expected that D7 and D76 were catalytic residues for ribonuclease activity because these are most highly conserved from the sequence blast as well as acidic amino acid, which must be a catalytic nucleophile for RNA cleavage. We do not know the exact mechanism of these two aspartic residues as a nucleophile due to the deficiency of complex structure. However, considering the results of mutational studies and the comparison with Cas2, two aspartic residues are strong candidates as a catalytic site.

As mentioned previously, it has been predicted that α1 is an important region for RNA binding, considering the distribution of positive charges on the electropotential map. Interestingly, an analysis of the *B* factor indicates that the α1 helix has relatively high α carbon *B*-factor values (Figure 2). That is, some residues on these regions have atomic movements containing residue flexibility and/or conformational changes (46). These high *B*-factor values and the location of the top position strongly suggest that the residues on the α1 helix can be candidates for interacting with RNA. In a protein–protein or protein–nucleic acid interaction system, protein flexibility and/or intrinsic disorder can be an important characteristic for interacting (47). The α1 helix is stabilized by the hydrophobic interaction between the hydrophobic core and L13. Consequentially, L13A could give rise to serious changes in the environment around the α1 helix,

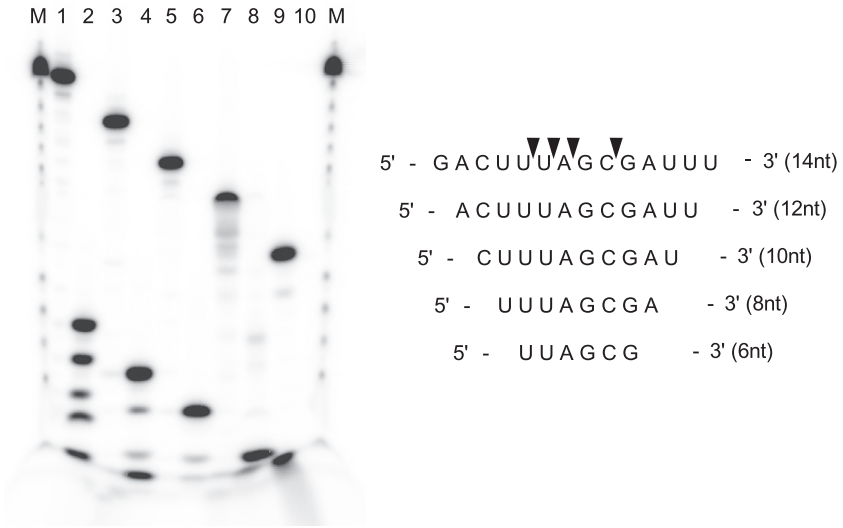

Figure 6. Cleavage experiments of small RNAs. Lanes M are imidazole ladder marker, which was prepared by partial RNA cleavage by imidazole. Lane 1 is 14 nt ssRNA without GB1_HP0315, and Lane 2 is 14 nt ssRNA with GB1_HP0315. Lane 3 is 12 nt ssRNA alone and Lane 4 is 12 nt ssRNA with GB1_HP0315. Lane 5 is 10 nt ssRNA alone and Lane 6 is 10 nt ssRNA with GB1_HP0315. Lane 7 is 8 nt ssRNA alone and Lane 8 is 8 nt ssRNA with GB1_HP0315. Lane 9 is 6 nt ssRNA alone and Lane 10 is 6 nt ssRNA with GB1_HP0315. Cleavage sites are indicated by closed triangles.

which may cause the disturbance in the first interaction between the protein and the RNA.

Our result revealed that RNA cleavage occurs before the purine residues, which implies the X:A or X:G sites (X: any nucleotide), but the preferential site cleaved most efficiently is U:A (Figure 5 and Supplementary Figure S6). In the case of ribonuclease A (48), a side chain of serine in the catalytic region is known to form a hydrogen bond with uracil. Thus it is associated with the specificity of ribonuclease. HP0315 mostly cleaved just before the purine residues, but it frequently cleaved between U and A, which may be mediated by S43 in a similar way. Moreover, the loop region containing S43, between the β2- and β3-strands, also shows a high *B*-factor value, which can be evidence of interaction between the S43 residue and uracil.

Therefore, we speculate that D7 and D76 are related to the catalytic activity, while L13 and S43 seem to be associated with the related specificity. This indicates that HP0315 has a similar structure but not an identical mechanism to cleave ssRNAs compared to SSO1404. Therefore, the present results strongly suggest that HP0315 is a member of a novel ribonuclease family other than the Cas2 family.

It was previously reported that VapD is functionally linked to the VapBC TA system (6). The genes of *hp0315* and *hp0316* are arranged as an operon. Interestingly, HP0316 has a high sequence homology with HP0895, antitoxin from *H. pylori*, which was reported recently by our group (37). This type of gene arrangement of *hp0315* raises the possibility that HP0315 belongs to the TA system as a toxin. However, HP0315 did not bind to HP0316 based on the results of a His-tag pull-down assay. HP0315 and HP0316 did not show TA activity in an *E. coli* viability test, either. This suggests that HP0315 and HP0316 do not belong to a TA system. Moreover, not only HP0967, another VapD protein from *H. pylori*, but also most VapD proteins from various organisms do not exist as an operon. This gene arrangement of HP0315 is very unique in this respect among the VapD family.

It was also reported that VapD is distantly related to Cas2 proteins and linked to the TA system (49). However, little evidence of this has been published. In this work, we revealed that HP0315, a member of the VapD family, has structural similarity with the Cas2 family and has gene arrangement similarity to the TA system, whereas it does not belong to any of them, like an evolutionary intermediate. Our result can serve as direct evidence that proves the previously reported suggestions (49). The facts that HP0315 is related to Cas2 proteins and TA system and has ribonuclease activity provide some information about the function of HP0315. The global inhibition of translation by mRNA cleavage would be a basic principle pertaining to the biological role of both Cas2 proteins and TA systems, including RelBE, MazEF, PemIK and ChpBIK (19,26,50,51). The exact function of HP0315 has not been determined. However, considering its relationship with Cas2 and the TA system, the role of HP0315 based on endoribonuclease activity may be related to either cell maintenance or a defense mechanism against invasion or possibly both like Cas2 and/or the TA system.

This structural study is the first step toward the understanding of the biological meaning of this protein and related predictions involving it in *H. pylori* virulence.

ACCESSION NUMBER

PDB ID 3UI3.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online: Supplementary Figures 1–6.

ACKNOWLEDGEMENTS

The authors thank Dr David B. Mckay at the University of Colorado for the helpful discussion and technical assistance during the phasing process and Dr Gerhard Wagner at the Harvard Medical School for providing of GB1 fused vector. They also thank the staff members of BL-4 A of Pohang Light Source for assistance with data collection.

FUNDING

National Research Foundation (NRF) of Korea grant funded by Korean government (MEST) (Grant number 20110001207); Korea Healthcare Technology R&D Project, Ministry for Health & Welfare Affairs, Republic of Korea (Grant number: A084420 and A092006); 2011 Brain Korea 21 Project for Medicine, Dentistry and Pharmacy.

Conflict of interest statement. None declared.

REFERENCES

1. Wassenaar,T.M. and Gaastra,W. (2001) Bacterial virulence: can we draw the line? *FEMS Microbiol. Lett.*, **201**, 1–7.
2. Zive,D., Koprowicz,K., Schmidt,T., Stiell,I., Sears,G., Van Ottingham,L., Idris,A., Stephens,S. and Daya,M. (2011) Variation in out-of-hospital cardiac arrest resuscitation and transport practices in the Resuscitation Outcomes Consortium: ROC Epistry-Cardiac Arrest. *Resuscitation*, **82**, 277–284.
3. Cheetham,B.F., Tattersall,D.B., Bloomfield,G.A., Rood,J.I. and Katz,M.E. (1995) Identification of a gene encoding a bacteriophage-related integrase in a vap region of the Dichelobacter nodosus genome. *Gene*, **162**, 53–58.
4. Catani,C.F., Azzoni,A.R., Paula,D.P., Tada,S.F., Rosselli,L.K., de Souza,A.P. and Yano,T. (2004) Cloning, expression, and purification of the virulence-associated protein D from Xylella fastidiosa. *Protein Expr. Purif.*, **37**, 320–326.
5. Katz,M.E., Strugnell,R.A. and Rood,J.I. (1992) Molecular characterization of a genomic region associated with virulence in Dichelobacter nodosus. *Infect. Immun.*, **60**, 4586–4592.
6. Makarova,K.S., Grishin,N.V., Shabalina,S.A., Wolf,Y.I. and Koonin,E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.*, **1**, 7.
7. Takai,S., Hines,S.A., Sekizaki,T., Nicholson,V.M., Alperin,D.A., Osaki,M., Takamatsu,D., Nakamura,M., Suzuki,K., Ogino,N. *et al.* (2000) DNA sequence and comparison of virulence plasmids from Rhodococcus equi ATCC 33701 and 103. *Infect. Immun.*, **68**, 6840–6847.
8. Polidori,M. and Haas,A. (2006) VapI, a new member of the Rhodococcus equi Vap family. *Antonie Van Leeuwenhoek*, **90**, 299–304.
9. Cheetham,B.F. and Katz,M.E. (1995) A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.*, **18**, 201–208.
10. Katz,M.E., Wright,C.L., Gartside,T.S., Cheetham,B.F., Doidge,C.V., Moses,E.K. and Rood,J.I. (1994) Genetic organization of the duplicated vap region of the Dichelobacter nodosus genome. *J. Bacteriol.*, **176**, 2663–2669.
11. Daines,D.A., Jarisch,J. and Smith,A.L. (2004) Identification and characterization of a nontypeable *Haemophilus influenzae* putative toxin-antitoxin locus. *BMC Microbiol.*, **4**, 30.
12. Benoit,S., Benachour,A., Taouji,S., Auffray,Y. and Hartke,A. (2001) Induction of vap genes encoded by the virulence plasmid of Rhodococcus equi during acid tolerance response. *Res. Microbiol.*, **152**, 439–449.
13. Galli,D.M. and LeBlanc,D.J. (1994) Characterization of pVT736-1, a rolling-circle plasmid from the gram-negative bacterium *Actinobacillus actinomycetemcomitans*. *Plasmid*, **31**, 148–157.
14. Jensen,R.B., Grohmann,E., Schwab,H., Diaz-Orejas,R. and Gerdes,K. (1995) Comparison of ccd of F, parDE of RP4, and parD of R1 using a novel conditional replication control system of plasmid R1. *Mol. Microbiol.*, **17**, 211–220.
15. Roberts,R.C., Strom,A.R. and Helinski,D.R. (1994) The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *J. Mol. Biol.*, **237**, 35–51.
16. Gotfredsen,M. and Gerdes,K. (1998) The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family. *Mol. Microbiol.*, **29**, 1065–1076.
17. Galvani,C., Terry,J. and Ishiguro,E.E. (2001) Purification of the RelB and RelE proteins of Escherichia coli: RelE binds to RelB and to ribosomes. *J. Bacteriol.*, **183**, 2700–2703.
18. Masuda,Y., Miyakawa,K., Nishimura,Y. and Ohtsubo,E. (1993) chpA and chpB, Escherichia coli chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100. *J. Bacteriol.*, **175**, 6850–6856.
19. Gerdes,K., Christensen,S.K. and Lobner-Olesen,A. (2005) Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.*, **3**, 371–382.
20. Malfertheiner,P., Bornschein,J. and Selgrad,M. (2010) Role of *Helicobacter pylori* infection in gastric cancer pathogenesis: a chance for prevention. *J. Dig. Dis.*, **11**, 2–11.
21. McNamara,D. and El-Omar,E. (2008) *Helicobacter pylori* infection and the pathogenesis of gastric cancer: a paradigm for host-bacterial interactions. *Dig. Liver Dis.*, **40**, 504–509.
22. Pounder,R.E. and Ng,D. (1995) The prevalence of *Helicobacter pylori* infection in different countries. *Aliment Pharmacol. Ther.*, **9(Suppl. 2)**, 33–39.
23. Cao,P. and Cover,T.L. (1997) High-level genetic diversity in the vapD chromosomal region of *Helicobacter pylori*. *J. Bacteriol.*, **179**, 2852–2856.
24. 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. *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, **388**, 539–547.
25. Haft,D.H., Selengut,J., Mongodin,E.F. and Nelson,K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.*, **1**, e60.
26. Beloglazova,N., Brown,G., Zimmerman,M.D., Proudfoot,M., Makarova,K.S., Kudritska,M., Kochinyan,S., Wang,S., Chruszcz,M., Minor,W. *et al.* (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.*, **283**, 20361–20371.
27. Van Duyne,G.D., Standaert,R.F., Karplus,P.A., Schreiber,S.L. and Clardy,J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.*, **229**, 105–124.
28. Otwinowski,Z. and Minor,W. (1997) In: Charles,W. and Carter.Jr (eds), *Methods in Enzymology*, Vol. 276. Academic Press, New York, pp. 307–326.
29. Miller,R., Shah,N., Green,M.L., Furey,W. and Weeks,C. (2007) Shake-and-Bake on the grid. *J. Appl. Cryst.*, **40**, 938–944.
30. Terwilliger,T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 965–972.
31. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
32. Brunger,A.T. (2007) Version 1.2 of the crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.

33. 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. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
34. 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. Cryst.*, **26**, 283–291.
35. Cheng,Y. and Patel,D.J. (2004) An efficient system for small protein expression and refolding. *Biochem. Biophys. Res. Commun.*, **317**, 401–405.
36. Jacob,F., Perrin,D., Sanchez,C. and Monod,J. (1960) [Operon: a group of genes with the expression coordinated by an operator]. *C R Hebd. Seances Acad. Sci.*, **250**, 1727–1729.
37. Han,K.D., Matsuura,A., Ahn,H.C., Kwon,A.R., Min,Y.H., Park,H.J., Won,H.S., Park,S.J., Kim,D.Y. and Lee,B.J. (2011) Functional identification of toxin-antitoxin molecules from *Helicobacter pylori* 26695 and structural elucidation of the molecular interactions. *J. Biol. Chem.*, **286**, 4842–4853.
38. Holm,L. and Sander,C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**, 478–480.
39. Tatusov,R.L., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Kiryutin,B., Koonin,E.V., Krylov,D.M., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N. *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**, 41.
40. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38(Suppl.)**, W529–W533.
41. Glaser,F., Pupko,T., Paz,I., Bell,R.E., Bechor-Shental,D., Martz,E. and Ben-Tal,N. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
42. Landau,M., Mayrose,I., Rosenberg,Y., Glaser,F., Martz,E., Pupko,T. and Ben-Tal,N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.*, **33**, W299–W302.
43. Miallau,L., Faller,M., Chiang,J., Arbing,M., Guo,F., Cascio,D. and Eisenberg,D. (2009) Structure and proposed activity of a member of the VapBC family of toxin-antitoxin systems. VapBC-5 from Mycobacterium tuberculosis. *J. Biol. Chem.*, **284**, 276–283.
44. Guy,C.P., Majernik,A.I., Chong,J.P. and Bolt,E.L. (2004) A novel nuclease-ATPase (Nar71) from archaea is part of a proposed thermophilic DNA repair system. *Nucleic Acids Res.*, **32**, 6176–6186.
45. Bibillo,A., Figlerowicz,M. and Kierzek,R. (1999) The non-enzymatic hydrolysis of oligoribonucleotides VI. The role of biogenic polyamines. *Nucleic Acids Res.*, **27**, 3931–3937.
46. Rhodes,G. (1993) *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. Academic, New York.
47. Grunberg,R., Nilges,M. and Leckner,J. (2006) Flexibility and conformational entropy in protein-protein binding. *Structure*, **14**, 683–693.
48. Raines,R.T. (1998) Ribonuclease A. *Chem. Rev.*, **98**, 1045–1066.
49. Koonin,E.V. and Makarova,K.S. (2009) CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol. Rep.*, **1**, 95.
50. Zhang,Y., Zhu,L., Zhang,J. and Inouye,M. (2005) Characterization of ChpBK, an mRNA interferase from Escherichia coli. *J. Biol. Chem.*, **280**, 26080–26088.
51. Zhang,Y. and Inouye,M. (2009) The inhibitory mechanism of protein synthesis by YoeB, an Escherichia coli toxin. *J. Biol. Chem.*, **284**, 6627–6638.
52. Schrödinger,L.L.C. *The PyMOL Molecular Graphics System, Version 1.3*. Schrödinger, LLC.
53. 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.
54. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.