Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of an archaeal homologue of the bacterial Fmu/RsmB/RrmB rRNA cytosine 5-methyltransferase

Yasushi Hikida,[a,b] Mitsuo Kuratani,[b] Yoshitaka Bessho,[b,c] Shun-ichi Sekine[a,b,d] and Shigeyuki Yokoyama[a,b,d]*

[a]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, [b]RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, [c]RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [d]Laboratory of Structural Biology, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Correspondence e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Received 27 April 2010
Accepted 20 September 2010

PDB Reference: PH0851, 2yxl.

One of the modified nucleosides that frequently occurs in rRNAs and tRNAs is 5-methylcytidine (m^5C). *Escherichia coli* Fmu/RsmB/RrmB is an *S*-adenosyl-L-methionine (AdoMet)-dependent methyltransferase that forms m^5C967 in 16S rRNA. Fmu/RsmB/RrmB homologues exist not only in bacteria but also in archaea and eukarya and constitute a large orthologous group in the RNA:m^5C methyltransferase family. In the present study, the crystal structure of a homologue of *E. coli* Fmu/RsmB/RrmB from the archaeon *Pyrococcus horikoshii* (PH0851) complexed with an AdoMet analogue was determined at 2.55 Å resolution. The structure and sequence of the C-terminal catalytic domain are highly conserved compared with those of *E. coli* Fmu/RsmB/RrmB. In contrast, the sequence of the N-terminal domain is negligibly conserved between the bacterial and archaeal subfamilies. Nevertheless, the N-terminal domains of PH0851 and *E. coli* Fmu/RsmB/RrmB are both α-helical and adopt a similar topology. Next to the AdoMet-binding site, a positively charged cleft is formed between the N- and C-terminal domains. This cleft is conserved in the archaeal PH0851 homologues and seems to be suitable for binding the RNA substrate.

1. Introduction

Methylation of position 5 of the cytidine base frequently occurs at multiple sites in many rRNA and tRNA molecules. The 5-methylcytidine (m^5C) residues are formed by site-specific methyltransferases (MTases). The *fmu* gene product from *Escherichia coli* (Borodovsky *et al.*, 1994) was the first identified RNA:m^5C MTase and the gene was renamed *rsmB* (Tscherne *et al.*, 1999) and *rrmB* (Gu *et al.*, 1999). The Fmu/RsmB/RrmB protein specifically methylates position 5 of the C967 base of *E. coli* 16S rRNA using *S*-adenosyl-L-methionine (AdoMet) as the methyl donor. C967 is located within the loop of a stem–loop structure (Fig. 1*a*). The minimum substrate for *E. coli* Fmu/RsmB/RrmB is a stem–loop fragment (residues

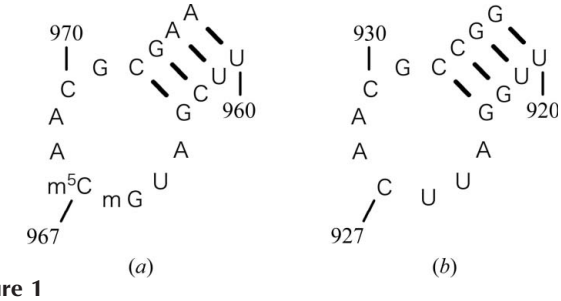

Figure 1
Secondary structures of (*a*) *E. coli* and (*b*) *P. horikoshii* 16S rRNA (960–975). The characters m^5C and mG show the modifications of *E. coli* 16S rRNA. Information about the modified nucleosides in the 16S rRNA from *P. horikoshii* is not available.

960–975) which consists of four base pairs and an eight-residue loop (Gu *et al.*, 1999).

Putative RNA:m^{5}C MTases that share sequence similarity with *E. coli* Fmu/RsmB/RrmB are widespread in eukarya, archaea and bacteria (Reid *et al.*, 1999). The domain that defines the putative RNA:m^{5}C MTases consists of about 300 amino-acid residues and contains six consensus motifs. Motif I (CAAPGGK), motif II (D*xxxxx*R), motif III (DG), motif IV (DAPCS) and motif VI (TC*xxx*E) are required for either methyl-transfer activity or structural maintenance. Several residues within motifs I–III form the AdoMet-binding pocket and the cysteine residue in motif VI serves as the nucleophile in the catalysis (Liu & Santi, 2000). This conserved domain adopts the class I MTase fold, which is composed of a seven-stranded β-sheet (6↑ 7↓ 5↑ 4↑ 1↑ 2↑ 3↑) sandwiched by α-helices to form the $\alpha\beta\alpha$ sandwich fold (Schubert *et al.*, 2003).

The unrooted phylogenetic tree of the RNA:m^{5}C MTase family contains four orthologous groups (Bujnicki *et al.*, 2004). The orthologous group containing *E. coli* Fmu/RsmB/RrmB (designated here as the Fmu family) is further divided into bacterial, archaeal and eukaryal subfamilies. The archaeal and eukaryal subfamilies are more closely related to each other than to the bacterial subfamily. In the Fmu family, the crystal structure of the *E. coli* Fmu protein complexed with AdoMet has been determined (Foster *et al.*, 2003). The *E. coli* Fmu structure consists of an N-terminal domain and a C-terminal MTase-fold domain. The characteristic N-terminal domain is specific to Fmu and distinguishes the Fmu family from the

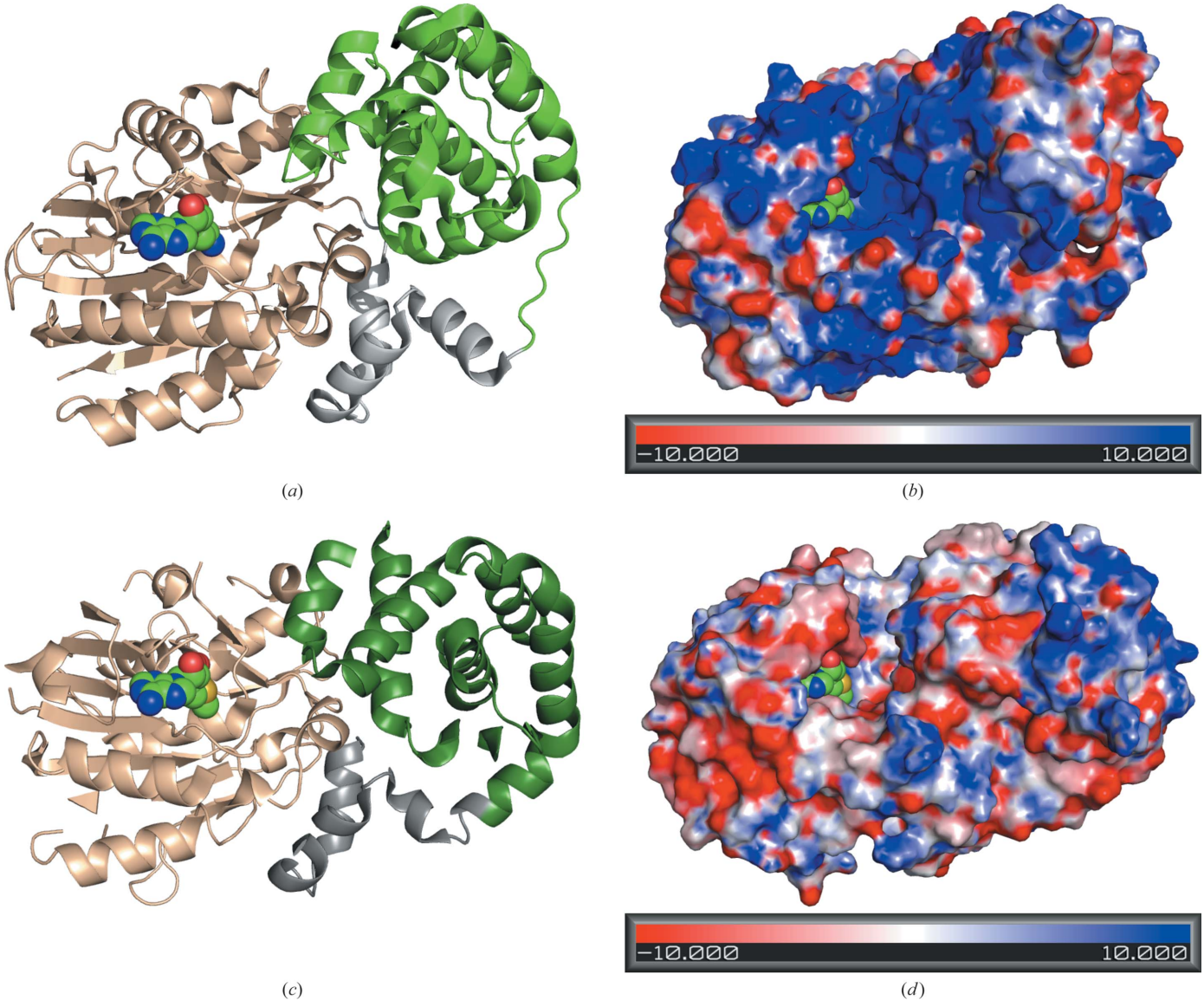

Figure 2
Overall structure of PH0851. (*a*) PH0851 is shown as a ribbon diagram. The C-terminal domain, the linker and the N-terminal domain are coloured green, grey and cream, respectively. The green CPK model depicts the bound sinefungin molecule (an AdoMet analogue). *PyMOL* (DeLano, 2002) was used to create the graphical figures. (*b*) A surface model of PH0851, coloured according to the electrostatic potential distribution (red, −10 *kT*/*e*; blue, +10 *kT*/*e*), as calculated with *APBS* (Baker *et al.*, 2001). (*c*) A ribbon diagram of *E. coli* Fmu coloured in the same manner as PH0851. The CPK model depicts the bound AdoMet molecule. (*d*) Surface model of *E. coli* Fmu showing the electrostatic potential distribution.

other orthologous groups of the RNA:m^{5}C MTase family. The Fmu-specific N-terminal domain is considered to be important for RNA binding (Foster *et al.*, 2003). Human WBSCR20, a gene that is commonly deleted in Williams–Beuren syndrome (Merla *et al.*, 2002), a developmental disorder, also belongs to the Fmu family and its C-terminal MTase-fold domain structure (residues 124–429) has been determined by the Structural Genomics Consortium (PDB code 2b9e; J. R. Min, H. Wu, H. Zeng, P. Loppnau, M. Sundstrom, C. H. Arrowsmith, A. M. Edwards, A. Bochkarev & A. N. Plotnikov, unpublished work).

In this study, we determined the crystal structure of an archaeal member of the Fmu family, PH0851 from *Pyrococcus horikoshii* (Kawarabayasi *et al.*, 1998). *P. horikoshii* contains five putative RNA:m^{5}C MTases (Reid *et al.*, 1999), the substrates of which have not yet been identified. PH0851 consists of 450 amino-acid residues and its C-terminal 270 residues constitute the core MTase domain homologous to that of *E. coli* Fmu. In contrast, the N-terminal 180 residues share very low similarity to those of *E. coli* Fmu. We now report the crystal structure of PH0851 complexed with an AdoMet analogue and its structural comparison with that of *E. coli* Fmu.

2. Materials and methods

2.1. Cloning, expression and purification of PH0851

ORF PH0851 (Kawarabayasi *et al.*, 1998) was amplified by PCR from *P. horikoshii* genomic DNA and was cloned between the *Nde*I and *Bam*HI sites of the expression vector pET-11a (Novagen). Wild-type PH0851 (native) protein and its selenomethionine (SeMet) derivative were overexpressed in *E. coli* strains Rosetta2 (Novagen) and BL21 (DE3) RIL, respectively. Cells were grown at 310 K in LB medium containing 50 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol until the A_{600} reached 0.5; isopropyl β-D-1-thiogalactopyranoside was then added to 0.1 m*M* concentration. Cells were grown at 310 K for an additional 19 h and were harvested by centrifugation. The pellet was resuspended in buffer *A* (20 m*M* Tris–HCl buffer pH 8.0 containing 50 m*M* NaCl, 5 m*M* β-mercaptoethanol and 1 m*M* phenylmethylsulfonyl fluoride) and homogenized by sonication. The extract was heated for 30 min at 353 K to denature most of the *E. coli* proteins. The heat-treated fraction was centrifuged and the supernatant fraction was loaded onto a Q-Sepharose Fast Flow column (Amersham Biosciences) equilibrated with buffer *A*. The protein was eluted with a linear gradient of 0.05–1 *M* NaCl. The PH0851 fractions were dialyzed against buffer *B* (20 m*M* MES–NaOH buffer pH 6.0 containing 50 m*M* NaCl and 5 m*M* β-mercaptoethanol) and applied onto a HiTrap Heparin column (Amersham Biosciences) equilibrated with buffer *B*. The protein was eluted with a linear gradient of 0.05–2 *M* NaCl. The fractions containing PH0851 were collected and dialyzed against buffer *C* (20 m*M* Tris–HCl buffer pH 8.0 containing 50 m*M* NaCl, 5 m*M* β-mercaptoethanol and 1 m*M* phenylmethylsulfonyl fluoride) and applied onto a UnoS column (Bio-Rad) equilibrated with buffer *C*. The protein was eluted with a linear gradient of 0.05–2 *M* NaCl. The final preparation was over 95% homogeneous as judged by SDS–PAGE.

2.2. Crystallization of PH0851

Prior to crystallization, the protein was dialyzed against buffer *D* (20 m*M* Tris–HCl buffer pH 8.0 containing 400 m*M* NaCl) and was concentrated to 1.7 mg ml^{-1} by ultrafiltration (Vivaspin, GE Healthcare). The AdoMet analogue sinefungin was added to a final concentration of 1 m*M* before crystallization. Crystallization was performed by the hanging-drop vapour-diffusion method. Crystals were obtained at 293 K by mixing protein solution and reservoir solution [80 m*M* HEPES–Na buffer pH 7.0 containing 1.15 *M* sodium citrate and 7–15%(*w*/*v*) 1,6-hexanediol] in a 4:1 volume ratio and equilibrating the mixture against the reservoir solution. The crystals grew in a week to maximum dimensions of 0.5 × 0.5 × 1.0 mm. Crystals of the SeMet derivative were prepared in the same manner as the native protein. Before data collection, the crystals were flash-cooled in liquid nitrogen. As the drop solution containing the crystals was effective as a cryoprotectant, no additional protectant was needed for the cryocooling.

2.3. Data collection and structure determination

The native data set was collected on beamline BL41XU at SPring-8 (Harima, Japan) and the MAD data sets for the SeMet derivative were collected on beamline NW12A at the Photon Factory (KEK, Tsukuba, Japan). All data were processed and scaled with *HKL*-2000 (Otwinowski & Minor, 1997).

The native crystals belonged to space group $P3_221$, with unit-cell parameters $a = b = 116$, $c = 93$ Å, whereas the SeMet crystals belonged to space group $P3_121$, with unit-cell parameters $a = b = 116$, $c = 186$ Å. Using the program *autoSHARP* (Vonrhein *et al.*, 2006), 14 selenium sites out of 20 expected sites were identified and the initial phases were calculated. The phases were extended to 3.0 Å by density modification with the program *RESOLVE* (Terwilliger, 2000). The atomic model was built using the program *Coot* (Emsley & Cowtan, 2004) and was refined using *CNS* (Adams *et al.*, 1997) and *REFMAC*5 (Winn *et al.*, 2001). At the initial stage of model building, the model was built into the electron density by step-by-step comparisons with the N-terminal domain structure of *E. coli* Fmu (PDB code 1sqf; Foster *et al.*, 2003). The asymmetric unit contains two PH0851 molecules, with a calculated solvent content of 65%.

The native protein crystal diffracted to 2.55 Å resolution. We determined the native structure by molecular replacement using the *MOLREP* program (Vagin & Teplyakov, 1997) with the SeMet-derivative model as the search model. The model of the native crystal was built and refined in a similar manner as the SeMet-derivative structure and was refined using *REFMAC*5. The final refinement protocol included TLS refinement (Painter & Merritt, 2006*a*,*b*) with four TLS groups

per molecule. The phasing and refinement statistics are shown in Table 1.

The final model consists of 443 residues (residues 6–448), with R_{work} and R_{free} factors of 21.2% and 24.6%, respectively (Table 1). A Ramachandran plot analysis (Laskowski *et al.*, 1993) revealed good stereochemistry, with 99% of the residues in the most favoured and allowed regions.

3. Results and discussion

3.1. Overall structure

The 2.55 Å resolution structure of PH0851 complexed with an AdoMet analogue (PDB code 2yxl) is monomeric and consists of two domains and one linker: an N-terminal domain (residues 1–145), a C-terminal domain (residues 185–448) and an α-helical linker (residues 146–184) (Fig. 2*a*).

The N-terminal domains of PH0851 and *E. coli* Fmu superpose well (Fig. 3*d*), but the overall superposition (with respect to the catalytic C-terminal domain and the AdoMet analogue) of PH0851 and *E. coli* Fmu (Fig. 3*c*) revealed that the N-terminal domains are poorly superposable. The relative orientation between the N- and C-terminal domains of PH0851 differs by more than 20° from that of *E. coli* Fmu, although each domain has almost the same fold in PH0851 and *E. coli* Fmu. The surface area of PH0851 is 18 500 Å^2 (N-terminal domain, 8300 Å^2; linker, 2700 Å^2; C-terminal domain, 12 400 Å^2).

The contact areas between the N-terminal domain and the linker, between the N-terminal domain and the C-terminal domain, and between the linker and the C-terminal domain are 550, 1100 and 960 Å^2, respectively. The α-helical linker

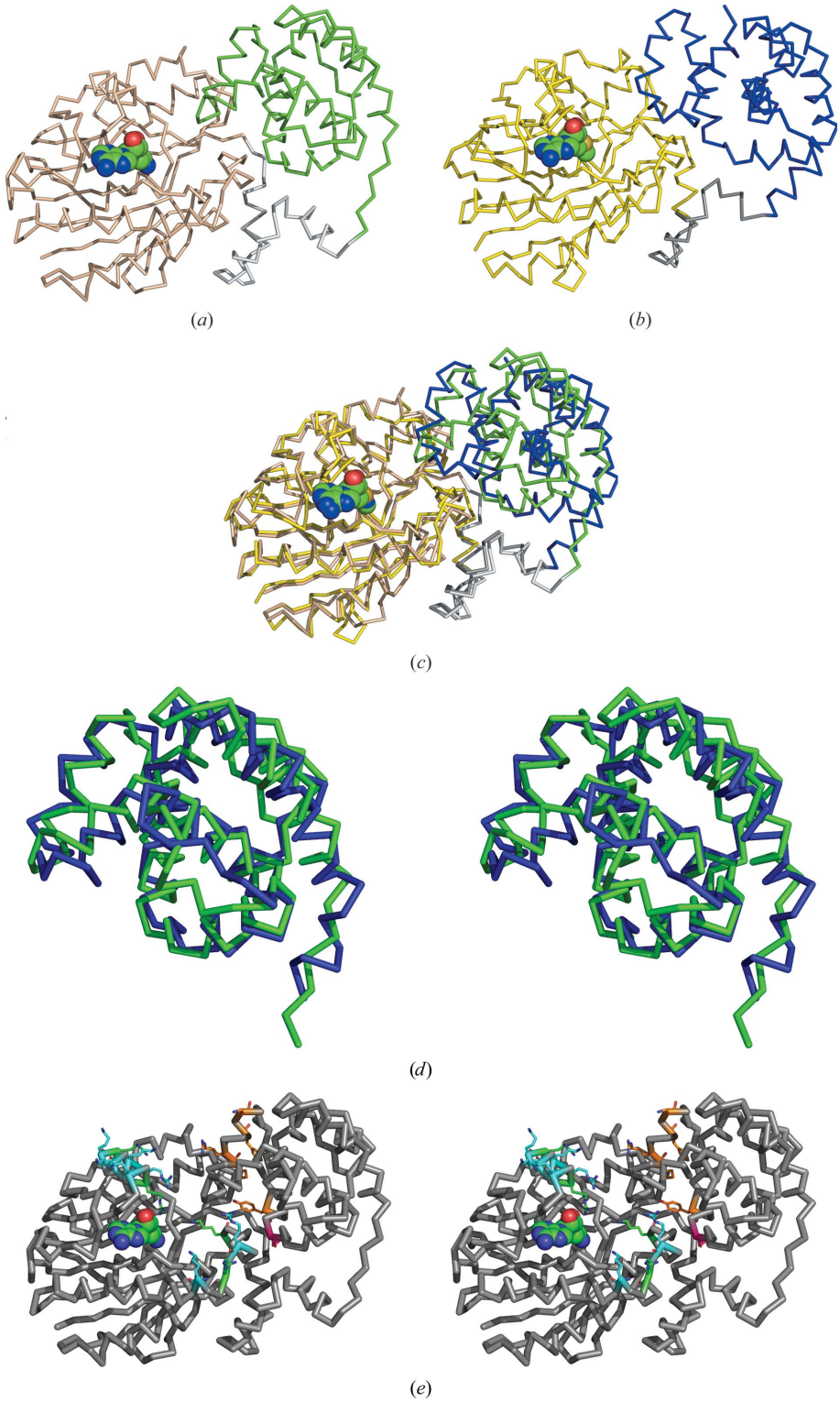

Figure 3
Comparison between the PH0851 and *E. coli* Fmu structures. The C$^\alpha$ traces of (*a*) PH0851 and (*b*) *E. coli* Fmu are shown. The PH0851 model is coloured as in Fig. 2(*a*), while the N-terminal domain, the C-terminal domain and the linker of Fmu are coloured blue, yellow and grey, respectively. The enzyme-bound AdoMet or sinefungin is depicted by green CPK models. (*c*) Superposition of the PH0851 and *E. coli* Fmu structures. (*d*) Superposition of the N-terminal domain of PH0851 and that of *E. coli* Fmu (stereoview, walleye). (*e*) Conserved basic residues around the putative RNA-binding site are shown as stick models on the PH0851 molecule. For the C-terminal domain, the residues conserved in PH0851 and *E. coli* Fmu (Arg296, Arg306, Lys348 and Arg357) are coloured green, while the residues conserved in species close to *P. horikoshii* (Lys294, Lys298, Arg299, Lys301, Lys305, Arg355 and Lys360) are coloured cyan. For the N-terminal domain, the residue (Lys62) conserved between PH0851 and *E. coli* Fmu is coloured pink, while those conserved in species close to *P. horikoshii* (Lys35, Arg36, Phe39, Lys41, Tyr58) are coloured orange (stereoview, walleye).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell. We collected and indexed data in the 50–2.2 Å resolution range, but used the data in the 50–2.55 Å range for refinement owing to the poorer statistics at higher resolution.

		SeMet		
	Native	Peak	Edge	Remote
Data collection				
Wavelength (Å)	1.0000	0.97912	0.97898	0.96390
Resolution (Å)	50–2.2 (2.28–2.2)	50–3.0 (3.11–3.0)	50–3.0 (3.11–3.0)	50–3.0 (3.11–3.0)
Space group	$P3_221$	$P3_121$	$P3_121$	$P3_121$
Unit-cell parameters				
$a = b$ (Å)	116.6	116.3	116.6	116.7
c (Å)	93.6	186.1	186.4	186.5
Unique reflections	37661 (3673)	29745 (2886)	29617 (2716)	29793 (2380)
Multiplicity	10.4 (6.4)	21.5 (16.3)	10.5 (6.7)	10.2 (4.5)
Completeness (%)	99.4 (98.7)	99.9 (99.5)	99.3 (93.3)	97.9 (80.0)
$I/\sigma(I)$	32.4 (2.1)	23.1 (2.0)	15.3 (0.9)	12.3 (0.5)
R_{merge}† (%)	7.1 (80.3)	14.9 (—)	13.3 (—)	13.3 (—)
Phasing				
Phasing power				
Iso (centric)			0.020	0.349
Ano (50–2.98 Å)		0.514	0.647	0.465
FOM (acentric/centric)		0.28/0.19		
Refinement				
Resolution (Å)	50–2.55			
R_{work} (%)	21.2			
R_{free}‡ (%)	24.6			
R.m.s.d. bond lengths (Å)	0.007			
R.m.s.d. bond angles (°)	1.1			

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ R_{free} was calculated for a subset of reflections (5%) that were excluded from all stages of refinement.

(residues 146–184), which connects the N- and C-terminal domains, consists of three α-helices. It interacts with the N- and C-terminal domains to form hydrophobic interfaces that stabilize the relative positions of the N- and C-terminal domains.

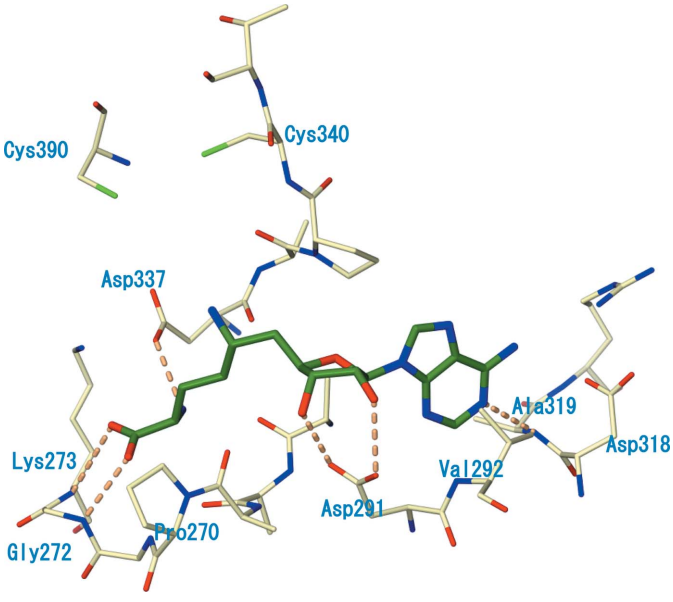

Figure 4
The AdoMet-binding pocket. Sinefungin (green) and the residues around the AdoMet-binding pocket are shown as stick models. Hydrogen bonds are depicted by dashed lines.

3.2. The C-terminal domain

The C-terminal domain of PH0851 (residues 185–450) adopts the typical fold of the class I MTases (Fig. 2). The C-terminal domains of PH0851 and *E. coli* Fmu superpose well (Fig. 3), with an average root-mean-square deviation (r.m.s.d.) of 1.53 Å over 245 C^{α} atoms, as calculated by *SSM* (Krissinel & Henrick, 2004). Human WBSCR20 also belongs to the Fmu family (the orthologous group to which *E. coli* Fmu and PH0851 belong; Bujnicki *et al.*, 2004). The C-terminal domain structure of WBSCR20 also superposes well on that of PH0851, with an average r.m.s.d. of 1.58 Å over 233 C^{α} atoms. Therefore, these three MTase-domain structures, which represent the bacterial, archaeal and eukaryal subfamilies of the Fmu family, are highly conserved.

The PH0851 C-terminal domain binds the AdoMet analogue, as shown in Fig. 4. The AdoMet analogue molecule adopts an extended conformation and buries 399 $Å^2$ (68% of its entire surface of 585 $Å^2$). The adenine base is stacked with the hydrophobic side chain of Val292 and hydrogen bonds are formed between Ado N1 and Ala319 NH (3.06 Å) and between Met NH and Asp337 OD2 (2.71 Å). The carboxyl-terminal O atoms of the AdoMet analogue form two hydrogen bonds to Gly272 NH (2.90 Å) and Lys273 NH (3.24 Å). The ribose moiety is fixed by two hydrogen bonds between Ado O2′ and Asp291 OD2 (2.83 Å) and between Ado O3′ and Asp291 OD1 (2.79 Å). In addition, several conserved residues, including Pro270, Asp291, Asp318, Cys340 and Cys390, participate in forming the pocket that binds the AdoMet analogue (Fig. 4). The positions of these residues relative to the AdoMet analogue are almost the same as those in the *E. coli* Fmu structure.

3.3. The N-terminal domain

The N-terminal domain of PH0851 comprises ten α-helices (Fig. 2*a*). Its overall architecture is similar to that of *E. coli* Fmu, which also consists of α-helices connected with a similar topology (Fig. 3). On the basis of the correspondence of the helices between the two structures, the average r.m.s.d. over 124 C^{α} atoms was estimated to be 2.62 Å. According to the *DALI* server (Holm *et al.*, 2008), the closest structural homologues of the N-terminal domain of PH0851 are those of *E. coli* Fmu (Z score = 13.3) and *Thermotoga maritima* NusB (Z score = 16.2; PDB codes 1tzt, 1tzu, 1tzv, 1tzw and 1tzx; Bonin *et al.*, 2004). NusB is a prokaryotic transcription factor that interacts with mRNA. The N-terminal domain of PH0851 shares a low sequence homology with NusB (18% identity and 43% homology). However, the N-terminal domain of PH0851

shares a similar fold with the NusB protein and also exhibits structural homology with that of *E. coli* Fmu (Fig. 3). The sequence homology between the N-terminal domains of PH0851 and *E. coli* Fmu is quite low and thus only a structure-based alignment facilitated identification of the amino-acid correspondences (Figs. 3*e* and 5). The structure-based alignment revealed that the two domains share 7% amino-acid identity. Thus, in the present study the N-terminal domain of PH0851 was found to assume an all-α-helical fold homologous to that of *E. coli* Fmu, despite the very low sequence homology. We speculate that this α-helical fold is conserved in the Fmu-family proteins, including WBSCR20 (the N-terminal domain of which shares 6% amino-acid identity to that of PH0851).

3.4. Putative RNA-binding site of PH0851

PH0851 contains a positively charged cleft next to the AdoMet-binding site (Fig. 2*b*). The cleft is formed between the N- and C-terminal domains and could be the potential RNA-binding site if the PH0851 protein binds RNA. The residues in this cleft are more conserved than the other surface residues of the N- and C-terminal domains (Fig. 6). In the C-terminal domain there are many basic amino-acid residues in the cleft, including Lys294, Arg296, Lys298, Arg299, Lys301, Lys305, Arg306, Lys348, Arg355, Arg357 and Lys360 (Fig. 3*e*). Of these, *E. coli* Fmu conserves basic amino-acid residues, Arg282, Arg292, Arg342 and Arg333, at the positions corresponding to Arg296, Arg306, Lys348 and Arg357, respectively, of PH0851. In contrast, in the N-terminal domain of PH0851 Lys35, Arg36, Phe39, Lys41 and Tyr58 are in the cleft. These residues are only conserved in closely related species and are not conserved in *E. coli* Fmu. *E. coli* Fmu has a patch of positively charged residues in the N-terminal domain (Fig. 2), but they do not correspond to the basic residues of PH0851. Only one basic residue, Lys62 (Arg53 in *E. coli* Fmu), is commonly conserved between PH0851 and *E. coli* Fmu (Fig. 3*e*). Arg53 of *E. coli* Fmu is exposed on the surface of the N-terminal domain. In contrast, Lys62 of PH0851 has no space for rRNA to access, suggesting a different function to that of Arg53 of *E. coli* Fmu. Nevertheless, if Lys62 interacts with rRNA then the N-terminal domain must change its position relative to the catalytic C-terminal domain.

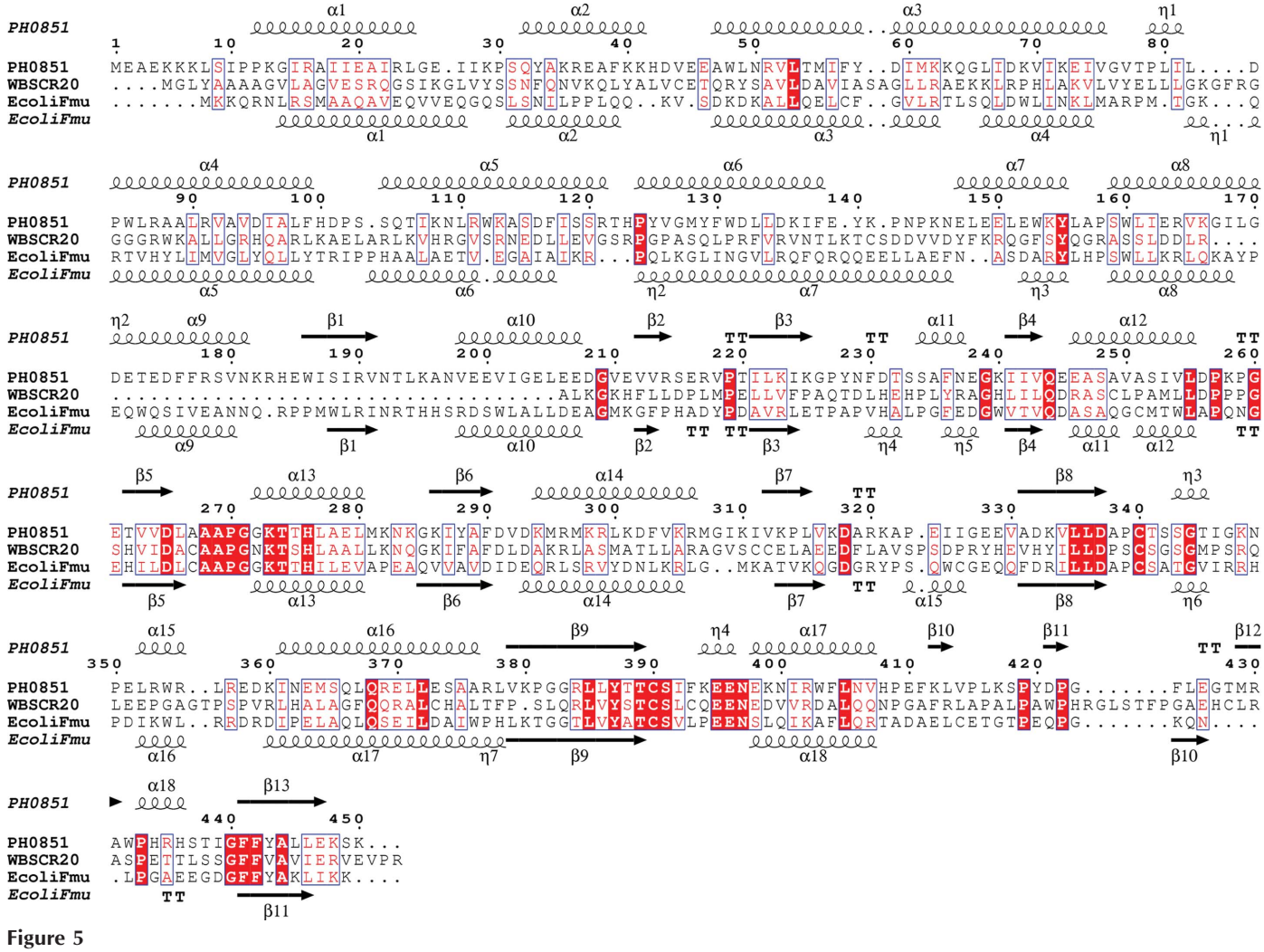

Figure 5
Alignment of PH0851, WBSCR20 and *E. coli* Fmu. The secondary structures of P0851 and *E. coli* Fmu are shown at the top and bottom, respectively.

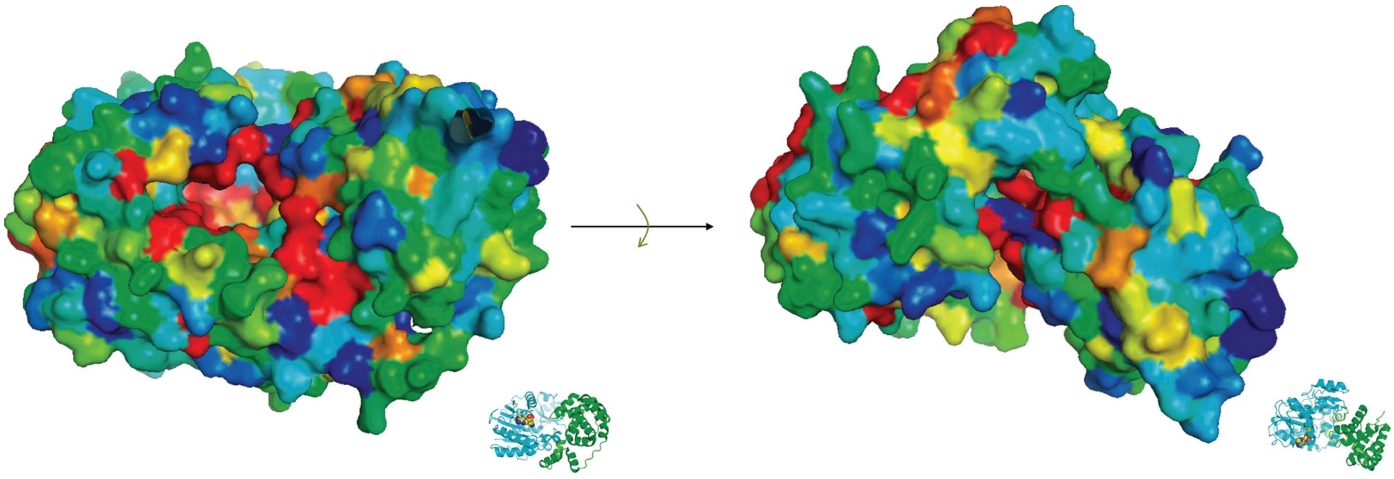

Figure 6
PH0851 ribbon model coloured by domain and surface coloured by conserved ratio. The conserved ratio was calculated using eight proteins with high homology to PH0851. Orange-coloured residues are more conserved and blue-coloured residues are less conserved. The AdoMet analogue is depicted by a CPK model in the ribbon model.

Although information about the modified nucleosides in the 16S rRNA from *P. horikoshii* is not available (MODOMIC database), it contains C at the position corresponding to m^5C967 in *E. coli* 16S rRNA (C927; Fig. 1*b*). The sequence similarity suggests that the region around C927 could adopt a stem–loop structure resembling that of the Fmu recognition site in *E. coli* 16S rRNA (Fig. 1). Considering the structural similarity between PH0851 and Fmu, we propose that the putative substrate of PH0851 would be C927 within 16S rRNA.

We are grateful to Dr Ihsanawati for assisting with the crystallization, Drs M. Kawamoto and N. Shimizu for technical assistance with the native data collection on BL41XU at SPring-8 (Harima, Japan) and Dr N. Igarashi for technical assistance with SeMet data collection on beamline NW12A at the Photon Factory (Tsukuba, Japan). We thank T. Nakayama and A. Ishii for clerical help. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses and in part by a Grant-in-Aid for Scientific Research 21603018 (to YB) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Bonin, I., Robelek, R., Benecke, H., Urlaub, H., Bacher, A., Richter, G. & Wahl, M. C. (2004). *Biochem. J.* **383**, 419–428.
Borodovsky, M., Rudd, K. E. & Koonin, E. V. (1994). *Nucleic Acids Res.* **22**, 4756–4767.
Bujnicki, J. M., Feder, M., Ayres, C. L. & Redman, K. L. (2004). *Nucleic Acids Res.* **32**, 2453–2463.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Foster, P. G., Nunes, C. R., Greene, P., Moustakas, D. & Stroud, R. M. (2003). *Structure*, **11**, 1609–1620.
Gu, X. R., Gustafsson, C., Ku, J., Yu, M. & Santi, D. V. (1999). *Biochemistry*, **38**, 4053–4057.
Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
Kawarabayasi, Y. *et al.* (1998). *DNA Res.* **5**, 55–76.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Liu, Y. & Santi, D. V. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8263–8265.
Merla, G., Ucla, C., Guipponi, M. & Reymond, A. (2002). *Hum. Genet.* **110**, 429–438.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.
Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.
Reid, R., Greene, P. J. & Santi, D. V. (1999). *Nucleic Acids Res.* **27**, 3138–3145.
Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). *Trends Biochem. Sci.* **28**, 329–335.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Tscherne, J. S., Nurse, K., Popienick, P., Michel, H., Sochacki, M. & Ofengand, J. (1999). *Biochemistry*, **38**, 1884–1892.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2006). *Methods Mol. Biol.* **364**, 215–230.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.