Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structural characterization of a modification subunit of a putative type I restriction enzyme from *Vibrio vulnificus* YJ016

**Suk-Youl Park, Hyun-Ju Lee, Jung-Mi Song, Jiali Sun, Hyo-Jeong Hwang, Kosuke Nishi and Jeong-Sun Kim***

Department of Chemistry, Chonnam National University, Gwangju 500-757, Republic of Korea

Correspondence e-mail: jsunkim@chonnam.ac.kr

Received 22 November 2011
Accepted 10 September 2012

**PDB Reference:** vvHsdM, 3ufb

In multifunctional type I restriction enzymes, active methyltransferases (MTases) are constituted of methylation (HsdM) and specificity (HsdS) subunits. In this study, the crystal structure of a putative HsdM subunit from *Vibrio vulnificus* YJ016 (vvHsdM) was elucidated at a resolution of 1.80 Å. A cofactor-binding site for *S*-adenosyl-L-methionine (SAM, a methyl-group donor) is formed within the C-terminal domain of an $\alpha/\beta$-fold, in which a number of residues are conserved, including the G*x*GG and (N/D)PP(F/Y) motifs, which are likely to interact with several functional moieties of the SAM methyl-group donor. Comparison with the N6 DNA MTase of *Thermus aquaticus* and other HsdM structures suggests that two aromatic rings (Phe199 and Phe312) in the motifs that are conserved among the HsdMs may sandwich both sides of the adenine ring of the recognition sequence so that a conserved Asn residue (Asn309) can interact with the N6 atom of the target adenine base (a methyl-group acceptor) and locate the target adenine base close to the transferred SAM methyl group.

## 1. Introduction

Among the several types of restriction and modification systems (RMs; restriction enzymes), which confer on microorganisms resistance towards imported DNA and are classified depending on their composition, cofactor requirements and target DNA recognition, type I restriction enzymes are distinguished from the others by their DNA translocation and restriction mechanisms, as well as their DNA recognition.

Type I restriction enzymes commonly form a hetero-oligomeric protein expressed from three adjacent genes on the chromosome. A specificity subunit (HsdS or S) is responsible for recognition of two asymmetric sequences that are separated by 6–8 base pairs. When one of the two specific DNA sequences is methylated, a methyl group is transferred from *S*-adenosyl-L-methionine (SAM) to the N6 atom of the non-methylated adenine base by the action of a methylation subunit (HsdM or M) (Dryden *et al.*, 2001; Murray, 2000; Sistla & Rao, 2004). When neither of the two cognate bases is methylated, a restriction subunit (HsdR or R) is recruited and type I RM systems start to translocate onto and cut the DNA upon encountering an obstacle at a site located about 1000 base pairs from the recognized DNA sequences. The HsdM and HsdS subunits are capable of forming an independent active trimeric $M_2S_1$ methyltransferase (MTase; Janscak & Bickle, 1998; Taylor *et al.*, 1992). DNA-cleavage and ATPase-

**Table 1**
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

| | Native | SeMet-substituted, peak |
|---|---|---|
| Synchrotron source | 4A, PAL | 8.3.1, ALS |
| Wavelength (Å) | 1.0000 | 0.98962 |
| Space group | $P4_12_12$ | $P4_12_12$ |
| Unit-cell parameters (Å) | $a = b = 78.9$, $c = 165.8$ | $a = b = 79.2$, $c = 164.6$ |
| Resolution (Å) | 50.0–1.80 (1.86–1.80) | 100.0–2.75 (2.85–2.75) |
| Completeness (%) | 99.2 (98.7) | 99.3 (98.8) |
| $R_{\rm merge}$† (%) | 8.0 (33.8) | 12.9 (44.1) |
| Reflections (total/unique) | 659674/48972 | 82895/14247 |
| Multiplicity | 13.5 | 5.8 |
| Temperature (K) | 100 | 100 |
| $\langle I\rangle/\langle\sigma(I)\rangle$ | 19.8 (6.8) | 6.8 (3.1) |
| FOM‡, *SOLVE*/*RESOLVE* | | 0.36/0.71 |
| $R$ factor§/$R_{\rm free}$¶ | 0.18/0.22 | |
| No. of atoms (protein/water) | 3857/542 | |
| Clashscore††, all atoms | 10.6 | |
| Ramachandran plot | | |
| Favoured region (%) | 96.5 | |
| Allowed region (%) | 3.5 | |
| Outliers (%) | 0 | |
| Poor rotamers (%) | 2.6 | |
| Average $B$ factors (Å$^2$) | | |
| Protein | 28.6 | |
| Solvent | 36.6 | |
| R.m.s.d.s‡‡ | | |
| Bonds (Å) | 0.007 | |
| Angles (°) | 1.00 | |

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement of the reflection with Miller indices $hkl$ and $\langle I(hkl)\rangle$ is the mean intensity of this reflection. ‡ Figure of merit = $|\sum P(\alpha)\exp(i\alpha)/\sum P(\alpha)|$, where $P(\alpha)$ is the phase probability distribution and $\alpha$ is the phase (100.0–2.75 Å). § $R$ factor = $\sum_{hkl}||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl}|F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively, for the reflections $hkl$ included in the refinement. ¶ $R_{\rm free}$ is the same as the $R$ factor but calculated using a randomly selected fraction (10%) of the reflection data not included in the refinement. †† Clashscore is the number of serious steric overlaps (>0.4 Å) per 1000 atoms. ‡‡ Root-mean-square-deviations from ideal values (Engh & Huber, 1991)

related processive translocation activities of the HsdR subunit are exerted when a pentameric $M_2S_1R_2$ oligomer of all three subunits is correctly assembled (Dryden *et al.*, 1993; Firman & Szczelkun, 2000; Seidel *et al.*, 2004).

The MTases of type I restriction enzymes belong to the N6 adenine-specific DNA MTases (EC 2.1.1.72), along with a DNA MTase of TaqI (mTaqI) from *Thermus aquaticus* (Goedecke *et al.*, 2001) which generates $N^6$-methyladenine. The crystal structures of HsdS from *Methanococcus jannaschii* (Kim *et al.*, 2005), *Mycoplasma genitalium* (Calisto *et al.*, 2005) and *Thermoanaerobacter tengcongensis* (Gao *et al.*, 2011) show three structural units consisting of two target-recognition domains (TRDs) and a couple of long $\alpha$-helices. However, the relative orientations of the three domains of HsdS have suggested two interesting interaction models with HsdM and double-stranded DNA (dsDNA), in which two HsdM subunits are located far away in one model (the open conformation) and two HsdMs come into contact with one another in another model (the closed conformation). Four crystal structures of HsdM subunits from different organisms have also been deposited in the PDB and publicly released: HsdM from *Bacteroides thetaiotaomicron* VPI-5482 (btHsdM; PDB entry 2okc; Joint Center for Structural Genomics); EcoKI from *E. coli* (EcoKI; PDB entry 2ar0; New York SGX Research Center for Structural Genomics); HsdM from *Streptococcus thermophilus* (stHsdM; PDB entry 3lkd; Northeast Structural Genomics Consortium); and HsdM from *Methanosarcina mazei* (mmHsdM; PDB entry 3khk; New York SGX Research Center for Structural Genomics). However, corresponding publications that describe their structures are not available. The electron-microscopy (EM) structure of the EcoKI MTase complexed with the DNA-mimicking protein Ocr provides a structural model for better understanding the assembly of the HsdM and HsdS subunits in the MTases of type I restriction enzymes (Kennaway *et al.*, 2009), in which the C-terminal domain (CTD) of HsdM has been suggested to interact with the TRD of HsdS and the N-terminal domain (NTD) of HsdM might mediate the interaction for HsdM dimerization. Nonetheless, detailed structural information at the atomic level is still lacking and information has not been provided on how the target bases are flipped out for methylation from the paired bases of dsDNA.

To provide further structural information on and a molecular background for a multifunctional type I RM, the present work reports the crystal structure of an HsdM subunit (EC 2.1.1.72) of a putative type I restriction enzyme from *Vibrio vulnificus* YJ016 (vvHsdM; UniProt accession Q7MPU6). The structure was elucidated at a resolution of 1.8 Å by single-wavelength anomalous dispersion (SAD) analysis. Like other deposited HsdM structures, a single elongated structure of the HsdM subunit is formed by the combination of two continuous domains in vvHsdM. A SAM cofactor-binding site is formed within the CTD near the interface with the NTD. The elucidated vvHsdM structure was compared with that of mTaqI, an N6 DNA MTase, and its structural features and possible interactions with target bases are discussed; this will provide a better structural background for deciphering the complicated mechanism of type I RM systems.

## 2. Materials and methods

### 2.1. Cloning, expression and purification of vvHsdM

The cloning and purification have been described in detail elsewhere (Lee *et al.*, 2009). Briefly, cloned full-length vvHsdM (gi_37678450; Met1–Lys530) in pProExHTc expression vector was transformed into *E. coli* B834 (DE3) cells and the expressed native protein was homogeneously purified by sequential nickel-affinity chromatography, cleavage of the N-terminal tag by protease treatment, ion-exchange chromatography and gel-filtration chromatography. Seleno-L-methionine (SeMet) substituted protein was prepared using a similar procedure as that used for the native protein. In both cases, the purified protein was >95% pure as judged by Coomassie-Blue-stained SDS–PAGE (data not shown).

### 2.2. Crystallization, data collection and structure determination

For crystallization, the purified vvHsdM protein was concentrated to 12 mg ml$^{-1}$ in a buffer consisting of 20 m*M*

Tris–HCl pH 7.5, 300 m*M* NaCl. Native crystals suitable for diffraction experiments were obtained within 14 d using the hanging-drop vapour-diffusion method at 295 K by mixing 1 µl protein solution and 1 µl reservoir solution and equilibrating against 200 µl reservoir solution consisting of 22%(*w*/*v*) polyethylene glycol 8000, 0.02 *M* imidazole pH 7.5, 5 m*M* *β*-mercaptoethanol. SeMet-substituted protein crystals were obtained under the same crystallization conditions as used to obtain the native crystals. For data collection, the crystals were briefly immersed in precipitant solution containing 10%(*v*/*v*) *meso*-erythritol and 0.8 *M* *β*-mercaptoethanol and immediately placed in a 100 K nitrogen-gas stream. High-resolution diffraction data for the native vvHsdM crystal were collected on beamline 4A at Pohang Accelerator Laboratory (PAL), Republic of Korea and SAD data were collected from a SeMet-substituted crystal on BL8.3.1 at the Advanced Light Source (ALS), USA. The data were indexed, integrated and scaled using the *HKL*-2000 suite (Otwinowski & Minor, 1997). 17 Se sites of the expected 21 in the asymmetric unit were identified at a resolution of 2.75 Å using *PHENIX* (Adams *et al.*, 2010) combined with *SOLVE* (Terwilliger & Berendzen, 1999). Three selenomethionine residues are present in

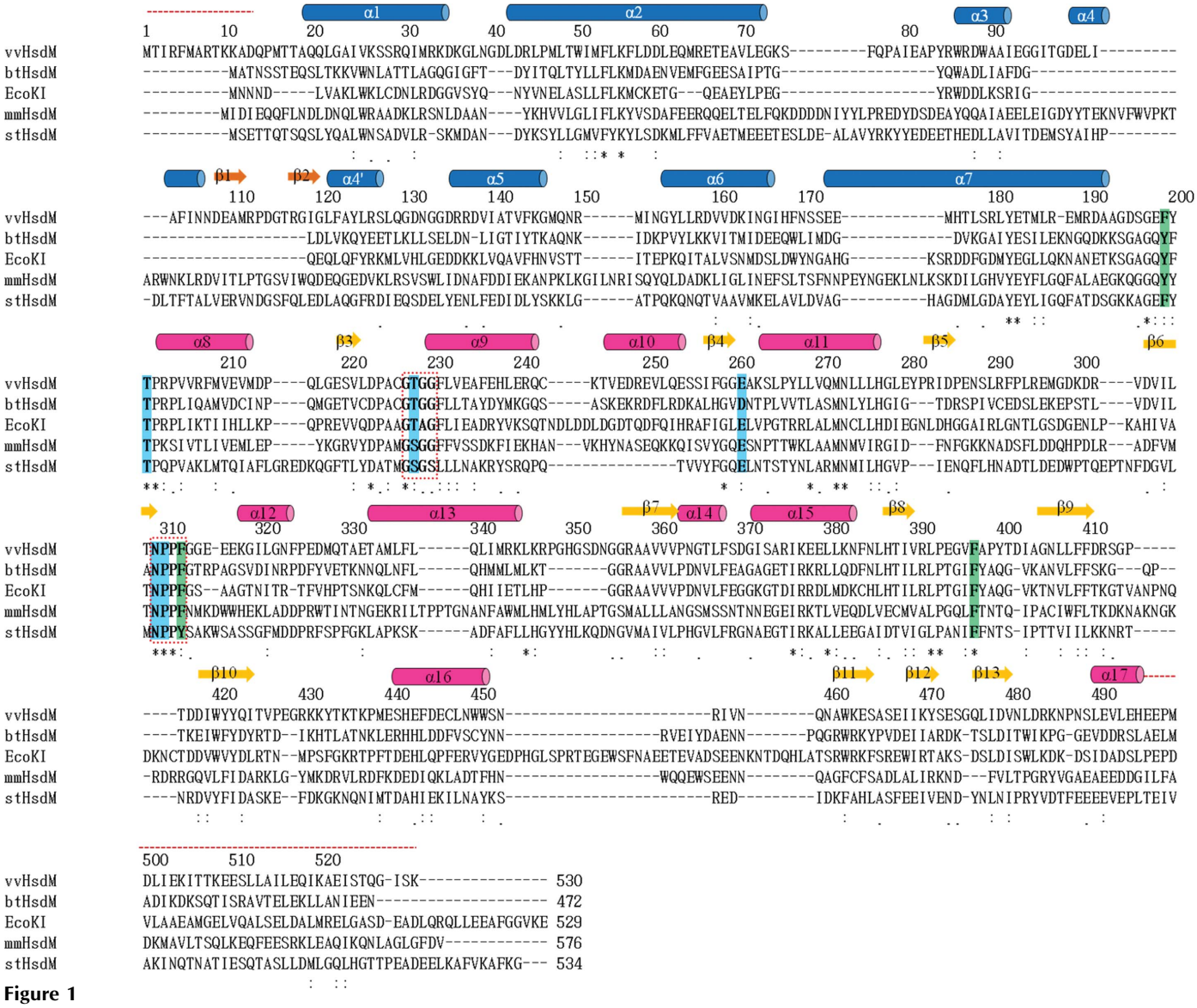


**Figure 1**
Sequence alignment of HsdM subunits of type I restriction enzymes. The amino-acid sequences shown are for HsdM from *V. vulnificus* YJ016 (vvHsdM), StySJI HsdM from *Bacteroides thetaiotaomicron* VPI-5482 (btHsdM; PDB entry 2okc), EcoKI from *E. coli* (EcoKI; PDB entry 2ar0), HsdM from *Methanosarcina mazei* (mmHsdM, PDB entry 3khk) and HsdM from *Streptococcus thermophilus* (stHsdM; PDB entry 3lkd). The cylinders (*α*) and arrows (*β*) above the aligned sequences represent helices and strands, respectively, in vvHsdM. The secondary structures of the NTD are displayed in blue and orange for *α*-helices and *β*-strands, whereas those of the CTD are shown in pink and yellow. The numbering scheme follows the amino-acid sequence of vvHsdM. The $^{227}$G*x*GG$^{230}$ and $^{309}$(N/D)PP(F/Y)$^{312}$ motifs of type I MTases are highlighted by a red dotted box. The conserved residues important for SAM cofactor and target adenine binding are presented on cyan and green backgrounds. The extreme N-terminal and C-terminal regions that were not traced in the present structure are indicated by dashed lines above the aligned sequences. Identical residues are marked by '*'. Strongly conserved and weakly conserved residues are indicated by ':' and '.', respectively. The sequence alignment was prepared using the *ClustalW*2 program from the European Bioinformatics Institute based on the primary amino-acid sequences (Larkin *et al.*, 2007).

**Table 2**
Root-mean-square deviations of compared HsdM structures.

The abbreviations used here are vvHsdM, HsdM from *V. vulnificus* YJ016; btHsdM, StySJI HsdM from *B. thetaiotaomicron* VPI-5482 (PDB entry 2okc); EcoKI, EcoKI from *E. coli* (PDB entry 2ar0); mmHsdM, HsdM from *M. mazei* (PDB entry 3khk); stHsdM, HsdM from *S. thermophilus* (PDB entry 3lkd); DNA MTase of TaqI from *T. aquaticus*, mTaqI. The r.m.s.d. values were obtained using *WinCoot* (Emsley & Cowtan, 2004).

| Domain | vvHsdM sequence range | Other HsdM | No. of aligned residues | Identity (%) | R.m.s.d. of superposed $C^{\alpha}$ atoms (Å) |
|---|---|---|---|---|---|
| NTD | 12–193 | btHsdM | 129 | 11 | 2.8 |
| | | EcoKI | 124 | 17 | 2.3 |
| | | mmHsdM | 123 | 13 | 2.3 |
| | | stHsdM | 119 | 16 | 2.4 |
| CTD | 198–495 | btHsdM | 255 | 39 | 1.8 |
| | | EcoKI | 248 | 33 | 1.6 |
| | | mmHsdM | 238 | 21 | 2.3 |
| | | stHsdM | 240 | 24 | 2.1 |
| | | mTaqI | 210 | 26 | 2.3 |

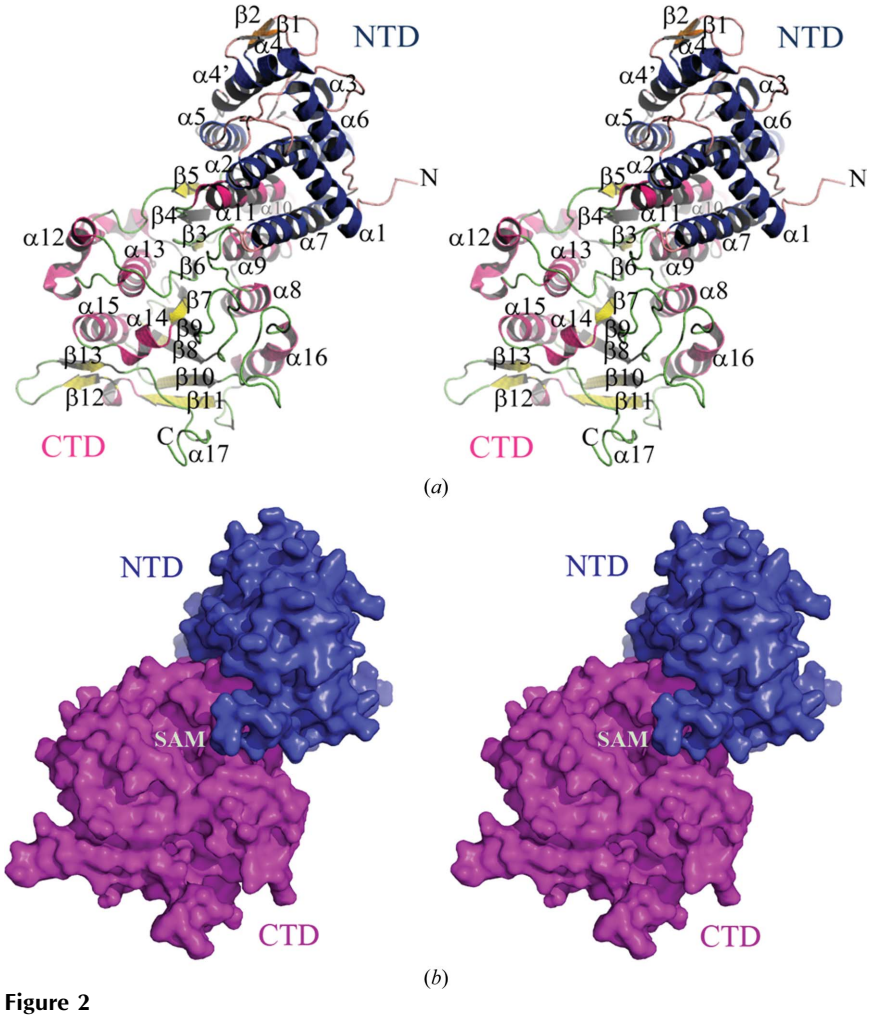


**Figure 2**
Overall structure of vvHsdM. (*a*) The monomeric structure of vvHsdM is displayed as a stereo diagram. The elongated monomeric HsdM structure consists of two domains, which are differentiated using the same colours as in Fig. 1. (*b*) The elongated surface comprising two domains in vvHsdM is also displayed in stereo; the NTD is coloured blue and the CTD orange. The putative SAM-binding pocket is indicated near the interface of the two domains. The figures (with the exception of Fig. 1) were prepared using the *PyMOL* molecular-graphics program (DeLano Scientific).

disordered regions of the structure and the last corresponded to Met16 at the N-terminus of the refined structure. The electron density was further improved by density modification using *PHENIX* (Adams *et al.*, 2010) combined with *RESOLV*E (Terwilliger, 2003), resulting in the automatic modelling of ∼60% of the residues. Further model building was performed manually using the programs *WinCoot* (Emsley & Cowtan, 2004) and *O* (Jones *et al.*, 1991). When ∼80% of the residues had been built, the initial phasing was stopped. Subsequent refinement was performed with *PHENIX* (Adams *et al.*, 2010) including TLS restraints and *CNS* (Brünger *et al.*, 1998) using the native diffraction data by increasing the resolution from 3.5 Å to a final resolution of 1.8 Å *via* resolutions of 3.2, 2.8, 2.4 and 2.0 Å. During refinement, additional amino acids and water molecules were placed, resulting in the final model. Statistics for the collected data and refinement are summarized in Table 1. The quality of the model was analyzed with *WinCoot* (Emsley & Cowtan, 2004) and *MolProbity* (Chen *et al.*, 2010). Coordinates and structure factors for the vvHsdM structure have been deposited in the PDB as entry 3ufb.

## 3. Results and discussion

### 3.1. Overall structure

The putative vvHsdM subunit comprises 530 amino acids with a calculated theoretical molecular weight of ∼60 kDa. Whereas both terminal regions show little primary-sequence conservation among the HsdMs, several sequence motifs, including the $^{227}$G*x*GG$^{230}$ and $^{309}$(N/D)PP(F/Y)$^{312}$ motifs, are conserved in the middle of the aligned HsdM sequences (Fig. 1) and are commonly found in MTases (Timinskas *et al.*, 1995).

The crystal structure of vvHsdM was determined by SAD analysis using an SeMet-substituted protein crystal and was refined to a resolution of 1.8 Å. A single protein molecule comprising two continuous domains is present in the asymmetric unit. In the refined structure, 16 amino acids including residues Met1–Lys11 at the N-terminus, five residues from the expression vector (YFGGA) and 35 residues (Glu496–Lys530) at the C-terminus were disordered (dashed red lines in Fig. 1). All other HsdM structures in the PDB had disordered termini. Analyses of the modelled residues using *WinCoot* (Emsley & Cowtan, 2004) and *MolProbity* (Chen *et al.*, 2010) revealed that all of the modelled residues were located in valid regions of the Ramachandran plot (Table 1).

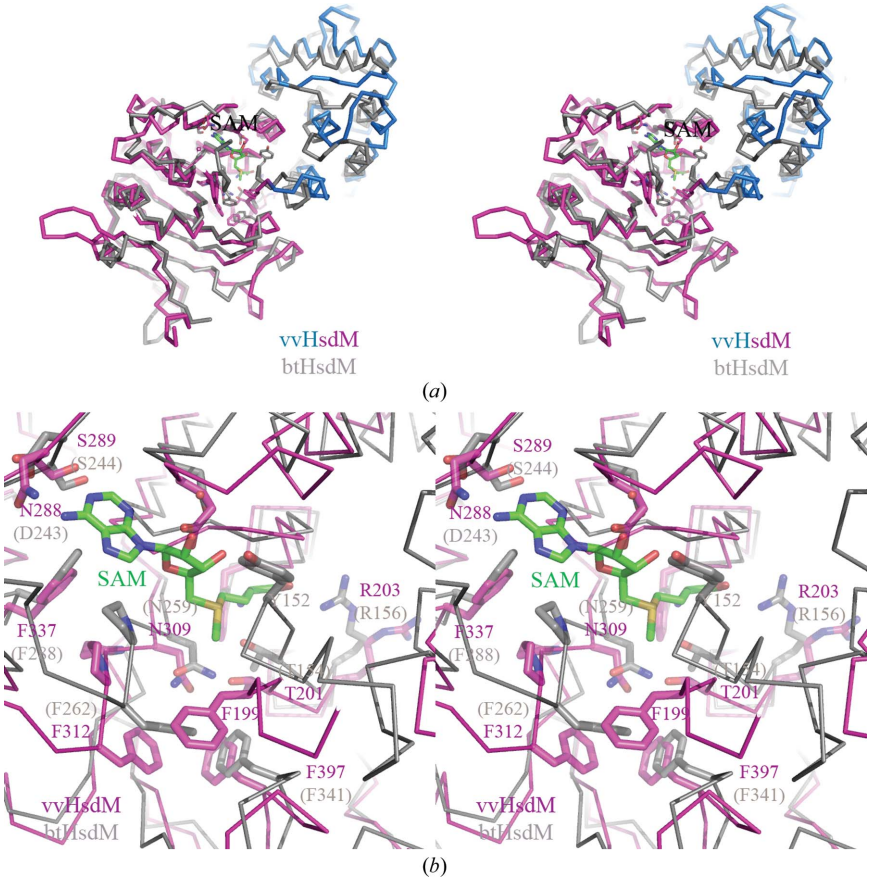


**Figure 3**
Comparison with btHsdM and the SAM-binding pocket. (*a*) Stereo representation of two overall HsdM structures. The SAM-binding pocket of vvHsdM was revealed by comparison of vvHsdM (blue/magenta) with btHsdM (grey). The superposed btHsdM and vvHsdM structures are shown as $C^{\alpha}$ traces and displayed as coils of different colours. The conserved residues around the SAM-binding pocket between the two HsdMs are drawn as stick models. (*b*) A close-up view of the SAM-binding pocket. The SAM (green) bound in btHsdM, the key residues involved in SAM binding in btHsdM and the corresponding residues in btHsdM are drawn as stick models. The two molecules are differentiated by their C-atom colours; N and O atoms are coloured blue and red, respectively.

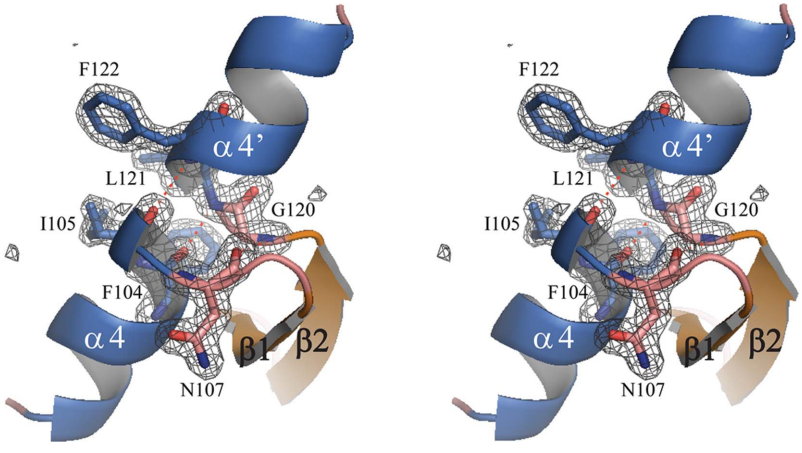


**Figure 4**
Stereo representation of the head-to-tail alignment of the $\alpha4$ and $\alpha4'$ helices observed in vvHsdM. $2F_o - F_c$ density (blue) for these $\alpha$-helices is contoured at $1.0\sigma$. The $\alpha$-helices are shown as ribbons and the residues at the contacting edges of the $\alpha$-helices are represented as stick models. N and O atoms are coloured blue and red, respectively. Hydrogen bonds at the contacting edges of the two $\alpha$-helices in vvHsdM are displayed as red dotted lines between the atoms.

Overall, the vvHsdM structure has an elongated shape formed by the interaction of two globular domains (Fig. 2): an $\alpha$-helical NTD (Ala12–Ala193) and an $\alpha/\beta$ CTD (Glu198–His495). In the present structure, these two domains are connected by a loop (Gly194–Gly197) which has weak electron density. The cofactor-binding site and the catalytic site are formed within the CTD near the NTD.

### 3.2. Structural features of the NTD

The NTD of vvHsdM comprises eight $\alpha$-helices ($\alpha1$, $\alpha2$, $\alpha3$, $\alpha4$, $\alpha4'$, $\alpha5$, $\alpha6$ and $\alpha7$) and two short $\beta$-strands ($\beta1$ and $\beta2$). It forms a globular domain with a long cleft at one side in which the $\alpha11$ helix of the CTD is located and can directly interact with three helices of the NTD ($\alpha2$, $\alpha5$ and $\alpha7$; Fig. 2). The other side of the $\alpha2$ helix is surrounded by five $\alpha$-helices of the NTD ($\alpha1$, $\alpha3$, $\alpha4$, $\alpha4'$ and $\alpha6$), in which many nonpolar residues are present, indicating that hydrophobic residues from the $\alpha$-helices provide the major interactions to form the globular structure of the NTD of vvHsdM.

The NTDs of HsdMs vary in length and sequence (Fig. 1). Nonetheless, the overall structural features of the $\alpha$-helical NTDs are well conserved in the HsdMs (Table 2), with some variations in the spatial positions of the $\alpha$-helices (Fig. 3). Seven $\alpha$-helices constitute a common globular NTD in the deposited HsdM structures, with several insertions between the $\alpha$-helices in some MTases. However, these insertions do not appear to be involved in the assembly of the seven common $\alpha$-helices or in the interactions of the NTDs with the CTDs (Figs. 1 and 3).

Unlike the seven $\alpha$-helices in the NTDs of other HsdMs, the globular NTD of vvHsdM basically comprises eight $\alpha$-helices. In other HsdM structures, the $\alpha4$ helix in the NTD comprises 16–20 amino acids representing four or five helical turns. In vvHsdM, however, helix $\alpha4$ is broken in the middle by a 14-amino-acid insertion between Asn106 and Leu121. The two subhelices, termed $\alpha4$ and $\alpha4'$, nevertheless share a common axis (Fig. 1). The inserted residues (Asn107–Gly120) at this position form a $\beta$-sheet consisting of two short $\beta$-strands ($\beta1$ and $\beta2$ in Figs. 1 and 4). Four hydrogen bonds are frequently observed among the main-chain atoms that form each turn of a stable

common $\alpha$-helix. The continuous $\alpha$-helices $\alpha4$ and $\alpha4'$ have only two hydrogen bonds at the interruption point (Fig. 4). Further hydrogen bonding near the $\alpha4$ helix and the $\alpha4'$ helix is provided by the inserted residues, for example the carbonyl O atom of Asn107 and the N atom of Gly120 on the extruded $\beta$-hairpin (Fig. 4), which may stabilize this interrupted $\alpha4$ helix and result in the formation of similar single $\alpha4$ helices in other HsdMs.

### 3.3. Structural features of the CTD

Unlike the helical NTD, which has poor sequence conservation across the family, the globular CTD structure is formed from mixed $\alpha$-helix and $\beta$-strand secondary structures and includes several conserved regions of primary sequence (Fig. 1). Nine $\beta$-strands ($\beta3$–$\beta11$) form a single pleated $\beta$-sheet in the centre of the globular CTD which is sandwiched by four $\alpha$-helices on one side and three $\alpha$-helices on the other (Fig. 2).

Comparison of the CTD of vvHsdM composed of ~300 amino acids with other HsdMs shows good $C^{\alpha}$ superposition for most of the $\alpha$-helices and the central $\beta$-sheet (Fig. 3), with root-mean-square-deviations (r.m.s.d.s) of less than 3.0 Å (Table 2). Deviations between the CTD structures appear in the loops, which are far from the putative methylation site and include a SAM-binding pocket, which is described in detail below.

The interaction of the CTD with the NTD forms an overall elongated shape of vvHsdM. The $\alpha11$ helix of the CTD of vvHsdM is located in the open cleft formed on the surface of the NTD and subsequently forms an $\alpha$-helical bundle structure with the $\alpha2$, $\alpha5$ and $\alpha7$ helices of the NTD (Fig. 2). A hydrophobic core is formed by several nonpolar residues (Leu265, Leu268, Leu269, Met272, Leu275 and Leu276) and includes a conserved sequence motif ($^{269}$L*xx*MN$^{273}$) along one side of the $\alpha11$ helix which mediates interactions with the NTD. Additional interactions are observed at the interface between the two domains: there are three polar interactions (Thr228 of the CTD with Glu184 of the NTD through their side-chain atoms, Glu280 of the CTD with the amide of the main-chain atom of Arg136, and Arg65 with the carbonyl O atom of Leu275) and one ionic interaction (Glu237 of the CTD with Arg180 of the NTD). However, these polar interactions are not conserved in other HsdMs.

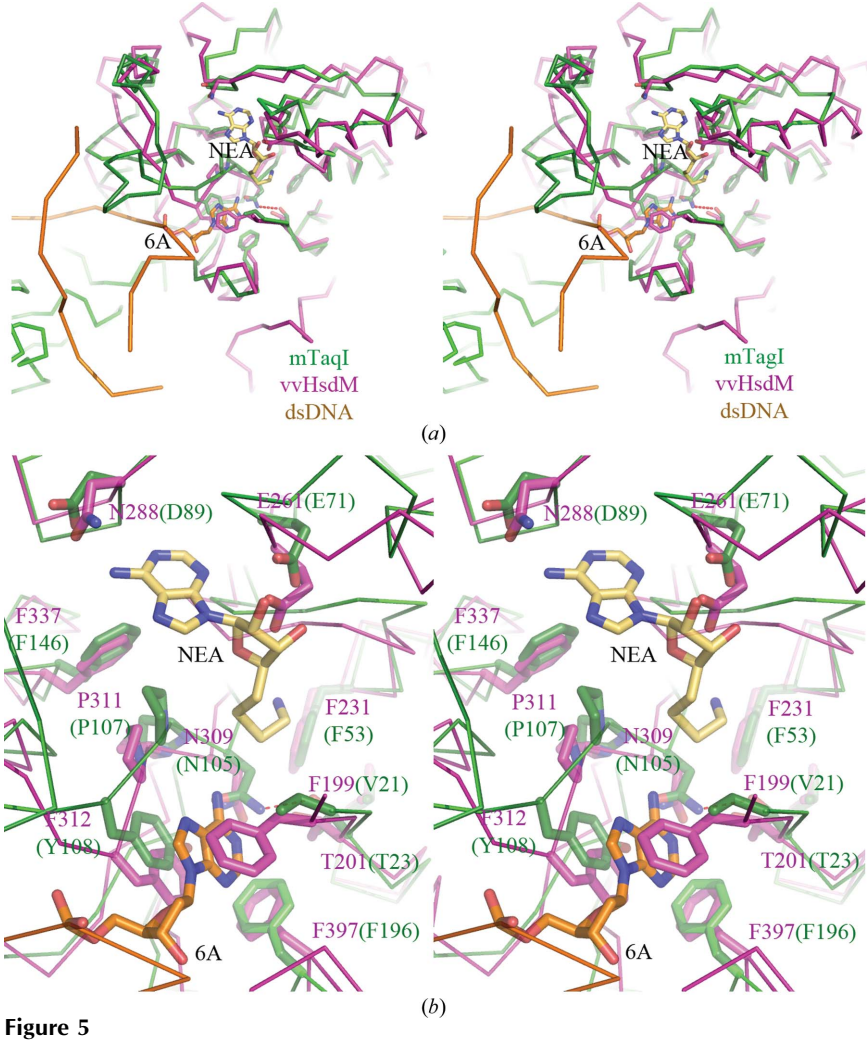


**Figure 5**
Implications for methylation in type I RM MTases. (*a*) Comparison of N6 DNA MTases; the backbone diagram of the CTD of vvHsdM (magenta) superposed on the tertiary complex of mTaqI (green), NEA (yellow) and dsDNA (orange). The conserved residues around the SAM-binding pocket are displayed as stick models. The flipped adenine base at the sixth nucleotide (6A) of dsDNA bound to mTaqI is drawn in stick representation. The two molecules are differentiated based on their C-atom colours; N and O atoms are coloured blue and red, respectively. (*b*) A close-up view of the methylation site of the N6 DNA MTases. The conserved residues for forming the SAM-binding pocket, residues important for interaction with the target base and the methyl-group acceptor (the N6 atom of the target adenine ring in the bound dsDNA; 6A) are displayed as stick models.

### 3.4. SAM-binding pocket

The HsdM and HsdS subunits of type I restriction enzymes can recognize methylation statuses at two target-recognition sequences; if one of the two cognate bases is unmethylated, the HsdM subunit transfers a methyl group to the N6 atom of the adenine ring using SAM as a methyl-group donor (Wilson & Murray, 1991).

The crystal structures of btHsdM and mTaqI are of complexes with the SAM cofactor and with its analogue 5′-deoxy-5′-[2-(amino)ethylthio]adenosine (NEA; Figs. 3 and 5). The CTD of vvHsdM (Glu198–His495) superposes well on that of btHsdM (Ala149–Pro432) and the MTase domain (Val21–Trp238) of mTaqI, with r.m.s.d.s of 1.8 and 2.3 Å, respectively, calculated on their $C^{\alpha}$ atoms (Table 2). Therefore, the SAM-binding site of vvHsdM could easily be found by comparing vvHsdM with these

two structures and is located near the interface of the two domains of the overall elongated structure (Fig. 2). As in other HsdMs, as well as mTaqI, the SAM-binding pocket of vvHsdM is exposed to solvent.

The methyl-group donor SAM has several functional moieties that could interact with various protein atoms. The binding site of its adenine-ring moiety is located in the deep cleft formed by several conserved residues, specifically Ser289, Leu290, Pro311 and Phe337 (Fig. 3). The N1 atom of the SAM adenine ring is close enough to form a polar interaction with the hydroxyl group of the conserved Ser/Thr residue (Ser244 in btHsdM, Ser289 in vvHsdM, Ser318 in mmHsdM, Thr251 in EcoKI and Thr285 in stHsdM). The hydrophobic interaction between the adenine ring of SAM and the conserved Phe/Trp residue (Phe288 in btHsdM, Trp376 in mmHsdM, Phe337 in vvHsdM and Phe146 in mTagI) appears to be involved in additionally fixing the SAM adenine ring.

The ribose sugar ring moiety of SAM also interacts with the conserved residues of HsdMs. The 2′-hydroxyl of the ribose forms a polar interaction with a carboxyl group of Glu261 in the conserved Glu/Asp sequence (Asp214 in btHsdM, Glu261 in vvHsdM and Glu71 in mTaqI; Figs. 1, 3 and 5). The SAM sugar ring faces Pro310 of the conserved sequence motif $^{309}$(D/N)PP(F/Y)$^{312}$ on the opposite side of this conserved Glu261 in a manner similar to the sugar ring of NEA, which faces Pro107 of the $^{105}$NPPY$^{108}$ motif in the NEA–mTaqI complex structure (Fig. 5).

The local binding pocket for the amine and carboxyl moieties of SAM is formed opposite the site of the adenine-ring moiety in the overall SAM binding pocket. In the SAM–btHsdM structure (Fig. 3) the amine group of SAM is located close to the hydroxyl group of the conserved Thr201, implying a possible hydrogen bond between them. Thr228 in the conserved $^{227}$G(S/T)(A/G)G$^{230}$ motif of vvHsdM (Fig. 1) is located close to the carboxyl group of SAM, allowing a possible polar interaction. Another positively charged Lys/Arg residue (Arg156 in btHsdM and Arg203 in vvHsdM), which is replaced by a Gln residue in stHsdM (Fig. 1), fixes the carboxyl group of SAM in the SAM–btHsdM complex structure (Fig. 3), implying that similar polar interactions are additionally involved in fixing the carboxyl group of SAM in vvHsdM.

Collectively, the SAM binding pocket is also structurally conserved among HsdMs and mTagI, with similar secondary structures and highly conserved residues around this pocket.

### 3.5. Structural view of methylation by type I RM MTases

During the methylation of a target molecule, the transferred methyl group should be near the acceptor molecule. In the type I RM system the methyl-group donor and acceptor are SAM and the N6 atom of the adenine base of the target nucleotide sequence, respectively. SAM can interact *via* several functional moieties with a number of conserved residues, and these interactions may place the leaving methyl group of SAM close to the accepting N6 atom of the target adenine base.

As mentioned above, the MTase domain of mTaqI in complex with its target DNA superposes well with the CTD of the elucidated vvHsdM structure, and several sequence motifs are highly conserved around the SAM-binding pocket (Fig. 5). Therefore, the interaction of HsdM with DNA and recognition of the target adenine nucleotide can be inferred based on a comparison of these two structures. The $^{105}$NPPY$^{108}$ motif of mTaqI draws attention because Tyr108 faces the adenine ring of the target nucleotide sequence (Fig. 5). Asn105 of mTaqI, the position of which is fixed by Thr23 through a hydrogen bond (Fig. 5), interacts with the N6 atom of the adenine base of the target nucleotide sequence. The environment of this region is also conserved in vvHsdM and the other HsdMs. Based on comparison of vvHsdM with the mTaqI structure, the target adenine base to be methylated in HsdMs is probably located in the narrow cleft formed between the two loops $\alpha$7–$\alpha$8 and $\beta$5–$\alpha$12, leading to formation of the SAM binding pocket within the CTD of vvHsdM (Fig. 5). These two loops contain two highly conserved sequence motifs among HsdMs: $^{197}$G(E/Q)(F/Y)(F/Y)TP$^{202}$ and $^{309}$(N/D)PP(F/Y)$^{312}$ (Fig. 1). Interestingly, two aromatic residues from these conserved motifs, Phe199 and Phe312, and the polar residue Asn309 are positioned within this cleft. The aromatic ring of Phe312 (Phe262 in btHsdM, Phe271 in EcoKI, Tyr307 in stHsdM and Phe338 in mmHsdM) is spatially conserved in the HsdM structures and corresponds to Tyr108 of mTaqI; that of Phe199 in vvHsdM and its equivalent residues in other HsdMs show large structural displacements among the superposed HsdM structures and correspond to Val21, which provides a van der Waals interaction on the opposite side of Tyr108, in mTaqI (Fig. 5). Based on these structural features of HsdMs, the aromatic rings of Phe199 and Phe312 are likely to sandwich the adenine ring of the cognate nucleic sequences from both sides. Furthermore, the flexible $^{197}$G(E/Q)(F/Y)(F/Y)TP$^{202}$ motif present in the reported HsdM structures, including btHsdM (Fig. 5), might be ordered upon DNA binding. Like Asn105 of mTaqI, which interacts with Thr23 and the N6 atom of the adenine base of the target nucleotide sequence, the conserved Asn309 of vvHsdM, the spatial position of which is also fixed by the hydroxyl group of Thr201 (Fig. 5), appears to function similarly to Asn105 in mTaqI, which helps in suitably positioning the methyl-group acceptor close to the transferring methyl-group moiety of the methyl-group donor SAM in vvHsdM. The hydroxyl group of Thr201 is conserved among HsdMs, even though it exists as a Ser or Thr residue (Fig. 1).

## 4. Concluding remarks

Comparison of the elucidated vvHsdM structure with that of mTaqI shows that the overall structure and several key residues of the N6 DNA MTases are highly conserved between these two enzymes, although some small differences were observed, suggesting similar enzymatic mechanisms in type I RMs and mTaqI. As the elucidated HsdS structure itself was not sufficient to explain the methylation of type I MTases, the crystal structure of vvHsdM alone cannot satisfactorily explain certain observations such as the binding of dsDNA by type I

MTases for base flipping, the transfer of information on the methylation of one of the cognate bases recognized by one HsdM subunit to the other HsdM subunit and the recruitment of the HsdR subunit depending on the methylation of the cognate bases.

The vvHsdM structure also provided a concrete structural background for suggesting domain assembly and possible interaction with other subunits and with the DNA molecule. The CTD, the extreme C-terminal region of which was disordered in the present structure as in the other deposited HsdM structures, was suggested to possibly be responsible for complex formation with the TRDs of HsdS (Powell *et al.*, 2003), resulting in interdomain arrangements of two subunits similar to the two intradomain arrangements of a single-chain TaqI MTase. Two putative roles have been suggested for the NTD. It has previously been shown that the HsdR and HsdM subunits of EcoKI form a complex (Dryden *et al.*, 1997). Since the other three domains at the N-terminus of HsdR are responsible for the restriction and translocation activities, the disordered extreme C-terminal region of ∼200 residues in HsdR was suggested to interact with the NTD of HsdM (Lapkouski *et al.*, 2009; Uyen *et al.*, 2009). However, there is no evidence to indicate the existence of a direct interaction between these two regions. Another putative role of the NTD of HsdM deduced from the previously reported EM structure (Kennaway *et al.*, 2009) was the mediation of dimerization of HsdM in the closed state, in which the single $\alpha$-helix ($\alpha$4) provides the major contact area. However, the sequence around the $\alpha$4 helix is not well conserved (Fig. 1) and the secondary structure also does not superpose well. Therefore, such detailed information may be further clarified from the complex structures of HsdM, HsdR and HsdS with or without DNAs with cognate bases at high resolution.

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology of Korea (2010-0008560), by the World Class Institute (WCI) Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology of Korea (MEST; NRF Grant No. WCI 2009-002) and partially by the Ministry of Science and Technology, Republic of Korea (21C Frontier Microbial Genomics and Application Center Program). The authors wish to thank the staff scientists at PAL and ALS for their assistance with data collection.

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Brünger, 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. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Calisto, B. M., Pich, O. Q., Piñol, J., Fita, I., Querol, E. & Carpena, X. (2005). *J. Mol. Biol.* **351**, 749–762.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Dryden, D. T., Cooper, L. P. & Murray, N. E. (1993). *J. Biol. Chem.* **268**, 13228–13236.
Dryden, D. T., Cooper, L. P., Thorpe, P. H. & Byron, O. (1997). *Biochemistry*, **36**, 1065–1076.
Dryden, D. T., Murray, N. E. & Rao, D. N. (2001). *Nucleic Acids Res.* **29**, 3728–3741.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Firman, K. & Szczelkun, M. D. (2000). *EMBO J.* **19**, 2094–2102.
Gao, P., Tang, Q., An, X., Yan, X. & Liang, D. (2011). *PLoS One*, **6**, e17346.
Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. (2001). *Nature Struct. Biol.* **8**, 121–125.
Janscak, P. & Bickle, T. A. (1998). *J. Mol. Biol.* **284**, 937–948.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kennaway, C. K., Obarska-Kosinska, A., White, J. H., Tuszynska, I., Cooper, L. P., Bujnicki, J. M., Trinick, J. & Dryden, D. T. (2009). *Nucleic Acids Res.* **37**, 762–770.
Kim, J.-S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R. & Kim, S.-H. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 3248–3253.
Lapkouski, M., Panjikar, S., Janscak, P., Smatanova, I. K., Carey, J., Ettrich, R. & Csefalvay, E. (2009). *Nature Struct. Mol. Biol.* **16**, 94–95.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Lee, H.-J., Nishi, K., Song, J.-M. & Kim, J.-S. (2009). *Acta Cryst.* F**65**, 1271–1273.
Murray, N. E. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 412–434.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Powell, L. M., Lejeune, E., Hussain, F. S., Cronshaw, A. D., Kelly, S. M., Price, N. C. & Dryden, D. T. (2003). *Biophys. Chem.* **103**, 129–137.
Seidel, R., van Noort, J., van der Scheer, C., Bloom, J. G., Dekker, N. H., Dutta, C. F., Blundell, A., Robinson, T., Firman, K. & Dekker, C. (2004). *Nature Struct. Mol. Biol.* **11**, 838–843.
Sistla, S. & Rao, D. N. (2004). *Crit. Rev. Biochem. Mol. Biol.* **39**, 1–19.
Taylor, I., Patel, J., Firman, K. & Kneale, G. (1992). *Nucleic Acids Res.* **20**, 179–186.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Timinskas, A., Butkus, V. & Janulaitis, A. (1995). *Gene*, **157**, 3–11.
Uyen, N. T., Park, S.-Y., Choi, J.-W., Lee, H.-J., Nishi, K. & Kim, J.-S. (2009). *Nucleic Acids Res.* **37**, 6960–6969.
Wilson, G. G. & Murray, N. E. (1991). *Annu. Rev. Genet.* **25**, 585–627.