Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure of human endonuclease V as an inosine-specific ribonuclease

Zhemin Zhang,[a,b] Zhitai Hao,[a,b] Zhong Wang,[b,c] Qing Li[b,c] and Wei Xie[a,b]*

[a]Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, School of Life Sciences, The Sun Yat-Sen University, Guangzhou 510275, People's Republic of China, [b]Centre for Cellular and Structural Biology, The Sun Yat-Sen University, 132 East Circle Road, University City, Guangzhou 510006, People's Republic of China, and [c]School of Pharmaceutical Sciences, The Sun Yat-Sen University, Guangzhou 510275, People's Republic of China

Correspondence e-mail: xiewei6@mail.sysu.edu.cn

Received 8 March 2014
Accepted 11 June 2014

PDB reference: human endonuclease V, 4nsp

The 6-aminopurine ring of adenosine (A) can be deaminated to form the 6-oxopurine of inosine (I). Endonuclease Vs (EndoVs) are inosine-specific nucleases that cleave at the second phosphodiester bond 3′ to inosine. EndoV proteins are highly conserved in all domains of life, but the bacterial and human enzymes seem to display distinct substrate preferences. While the bacterial enzymes exhibit high cleavage efficiency on various nucleic acid substrates, human EndoV (hEndoV) is most active towards ssRNA but is much less active towards other substrates. However, the structural basis of substrate recognition by hEndoV is not well understood. In this study, the 2.3 Å resolution crystal structure of hEndoV was determined and its unusual RNA-cleaving properties were investigated. The enzyme preserves the general 'RNase H-like' structure, especially in the wedge motif, the metal-binding site and the hypoxanthine-binding pocket. hEndoV also features several extra insertions and a characteristic four-cysteine motif, in which Cys227 and Cys228, two cysteines that are highly conserved in higher eukaryotes, play important roles in catalysis. The structure presented here helps in understanding the substrate preference of hEndoV catalysis.

1. Introduction

An inosine in DNA (dI) is the result of spontaneous or stress-induced deamination of dA (Lindahl, 1979), which converts the 6-aminopurine ring of adenosine to the 6-oxopurine ring (Bass, 2002). dI behaves like dG in terms of base-pairing properties. An A–T base pair will thus be converted to a G–C base pair in DNA-replication processes, producing potential mutations during gene expression. Therefore, dI is regarded as damage to DNA and needs to be excised or repaired. In *Escherichia coli*, the primary enzyme for this repair is endonuclease V (EndoV; Weiss, 2008). EndoV incises DNA without removing the hypoxanthine base, a process that is not well understood. It initiates the process by cleaving the second phosphodiester bond 3′ to inosine (Yao & Kow, 1994; Huang *et al.*, 2001). Additionally, *E. coli* EndoV (EcEndoV) also recognizes apurinic/apyrimidinic (AP) sites, mismatches, insertions/deletions, flaps, pseudo-Y structures, loops and other deaminated DNA bases *in vitro* (Yao *et al.*, 1994; Yao & Kow, 1994, 1996; Gates & Linn, 1977; He *et al.*, 2000; Rosnes *et al.*, 2013). In bacteria, the incision of dI by EndoV is a two-metal-dependent reaction and produces a stable intermediate complex consisting of EndoV and the nicked DNA (Yao & Kow, 1994; Dalhus *et al.*, 2009; Mi *et al.*, 2011; Feng *et al.*, 2006; Lin *et al.*, 2007). A model for the cleavage has been proposed in which Ca^{2+} induces the recognition of dI and Mg^{2+} is required for DNA cleavage (Lin *et al.*, 2007; Feng *et al.*, 2006).

Table 1
The cloning and expression of hEndoV-SF.

Source organism	*Homo sapiens*
DNA source	Human 293T cell cDNA
Forward primer	5′-GATAGGGCCATATGACGCTGTCACTGTGGAA-GCTGTCACTGTGGAAAC-3′
Reverse primer	5′-AATATGGCGGCCGCCGACTTGCGGATGTGCT-ACTTGCGGATGTGCTCTC-3′
Cloning vector	pET-21b(+)
Expression vector	pET-21b(+)
Expression host	*E. coli* BL21 (DE3) strain
Complete amino-acid sequence of the construct produced	MTLSLWKREQARLKAHVVDRDTEAWQRDPAFSGL-QRVGGVDVSFVKGDSVRACASLVVLSFPELEV-VYEESRMVSLTAPYVSGFLAFREVPFLLELVQ-QLREKEPGLMPQVLLVDGNGVLHHRGFGVACH-LGVLTDLPCVGVAKKLLQVDGLENNALHKEKI-RLLQTRGDSFPLLGDSGTVLGMALRSHDRSTR-PLYISVGHRMSLEAAVRLTCCCCRFRIPEPVR-QADICSREHIRKSAAALEHHHHHH

In RNA, inosines are normal residues specifically introduced by deaminases (Zinshteyn & Nishikura, 2009; Keegan *et al.*, 2004). Unlike in DNA, the deamination of adenosine is the most prevalent type of modification of double-stranded RNA (dsRNA) in higher eukaryotes (Zinshteyn & Nishikura, 2009). For example, up to eight eukaryotic tRNAs contain I at position 34 (Sprinzl *et al.*, 1991). I34 is crucial in decoding genes as a wobble base during protein synthesis and allows base pairing with C, T and A (Lim, 1995; Agris *et al.*, 2007). The A-to-I conversion is essential for viability in yeast (Tsutsumi *et al.*, 2007) and defects in A-to-I conversion are connected to many human disorders (Maas *et al.*, 2006).

Recently, human endonuclease V (hEndoV) has been cloned and characterized (Morita *et al.*, 2013; Vik *et al.*, 2013). In contrast to the bacterial orthologues, hEndoV favours RNA substrates over DNA substrates. hEndoV is almost inactive towards dI-containing substrates, but cleaves mRNA transcripts containing inosines or tRNAs with inosine at the wobble position very efficiently. Additionally, hEndoV is localized in both the cytoplasm and the nucleolus, suggesting RNA to be its substrate *in vivo* (Morita *et al.*, 2013; Fladeby *et al.*, 2012). It was further demonstrated that hEndoV prefers ssRNA to dsRNA/DNA (Morita *et al.*, 2013; Vik *et al.*, 2013). However, the structural basis for the substrate preference of hEndoV is still lacking.

Structures of *Thermotoga maritima* EndoV (TmEndoV) in complex with a hypoxanthine lesion substrate and product DNA have been solved (Dalhus *et al.*, 2009). TmEndoV features an $\alpha + \beta$ structure with a central eight-stranded β-sheet flanked by α-helices on either side. This 'RNase H-like motif' resembles that of *E. coli* RNase H (Katayanagi *et al.*, 1990; Yang *et al.*, 1990) and other nucleic acid-binding proteins (Song *et al.*, 2004; Karakas *et al.*, 2007; Ariyoshi *et al.*, 1994; Ceschini *et al.*, 2001; Davies *et al.*, 2000). A conserved PYIP wedge motif works as the lesion sensor and breaks the base pairing at the lesion site. TmEndoV incises DNA one nucleotide 3′ to dI, while the hypoxanthine base rotates $\sim$90° into the recognition pocket, which is $\sim$8 Å from the catalytic site. Another crystal structure of TmEndoV in complex with DNA containing a one-nucleotide loop (A:TT) has also been determined (Rosnes *et al.*, 2013). The structure of the complex reveals that the enzyme utilizes a base-flipping recognition strategy for DNA loop substrates. The PYIP wedge motif partially splits the two DNA strands by projecting into the minor groove at the loop distortion site. The two thymine bases in the loop are splayed out of the duplex to some extent, while the adenine base is inserted into the hydrophobic recognition pocket. Interestingly, adenine is bound in the same position and orientation as hypoxanthine, and the hydrogen-bonding interactions are mainly formed with the backbone atoms of the residues that make the pocket. Besides TmEndoV, several structures of bacterial EndoVs have been deposited in the PDB without corresponding publications (PDB entries 3hd0, 3goc, 3ga2 and 2qh9; Integrated Center for Structure and Function Innovation, unpublished work; Northeast Structural Genomics, unpublished work; Consortium Midwest Center for Structural Genomics, unpublished work).

To investigate the substrate-recognition mechanism of hEndoV, we solved the crystal structure of apo hEndoV. The 2.3 Å resolution crystal structure of hEndoV provides us with insights into the molecular mechanism of hEndoV catalysis.

2. Experimental procedures

2.1. Cloning, protein expression and purification of hEndoV-SF and mutants

The full-length human *ENDOV* gene (EC 3.1.26.–; GenBank accession No. NM_173627) was amplified from human 293T cell cDNA and cloned into the expression vector pET-21b(+) (Novagen) using the *Nde*I and *Not*I restriction sites. 11 amino acids (AAALEHHHHHH), including a hexahistidine tag, were added to the C-terminus. For crystallization purposes, a truncated fragment representing residues Thr13–Ser250 (hEndoV-SF) was subcloned into pET-21b(+) using the same sites (Table 1). The wild-type full-length *E. coli ENDOV* gene (GenBank accession No. YP_491462.1) was amplified from *E. coli* genomic DNA and cloned into pET-21b(+) using the *Nde*I and *Xho*I restriction sites. The mutants used in this study were generated by the QuikChange method (Stratagene) using hEndoV-SF as the template. The amplifying and mutagenic primers are listed in Supplementary Table S1[1].

The plasmids encoding wild-type or mutant EndoV were transformed into *E. coli* strain BL21 (DE3) cells. The cells were cultured overnight in Luria–Bertani broth containing 50 µg ml^{-1} ampicillin. 2 l fresh culture medium was then inoculated with 5 ml overnight culture. When the OD_{600} reached 0.6–0.8, expression of hEndoV-SF was induced by 0.3 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was shaken at 25°C overnight. The *E. coli* cells were then harvested by centrifugation at 5000 rev min^{-1} for 15 min and resuspended in pre-chilled nickel–nitrilotriacetic acid (Ni–NTA) buffer *A* [20 m*M* Tris–HCl pH 8.0, 250 m*M*

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: QH5006).

NaCl, 10 m*M* imidazole, 1 m*M* β-mercaptoethanol, 1 m*M* phenylmethylsulfonyl fluoride (PMSF)]. The cells were disrupted by ultrasonication and the supernatant was obtained by centrifugation at 14 000 rev min^{-1} for 1 h at 4°C. The supernatant was then applied onto Ni–NTA affinity resin (Qiagen) which had previously been equilibrated with Ni–NTA buffer *A*. The target protein was eluted with Ni–NTA buffer *B* (20 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl, 250 m*M* imidazole, 1 m*M* β-mercaptoethanol, 1 m*M* PMSF). The hEndoV-SF fractions were pooled and dialyzed in a buffer consisting of 20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 1 m*M* DTT. The dialyzed protein was applied onto a HiTrap Heparin HP column (GE Healthcare) equilibrated with Heparin HP buffer *A* (20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 1 m*M* DTT). hEndoV-SF protein was eluted with Heparin HP buffer *B* consisting of 20 m*M* Tris–HCl pH 8.0, 1 *M* NaCl, 1 m*M* DTT. Fractions were concentrated and loaded onto a Superdex 200 size-exclusion column (GE Healthcare) equilibrated with 20 m*M* Tris–HCl pH 8.5, 300 m*M* NaCl. The final fractions of hEndoV-SF were concentrated to 10 mg ml^{-1} using a Millipore centrifugal filter (molecular-weight cutoff 10 kDa) and stored at −80°C.

The same expression and purification strategy was applied for the hEndoV-SF mutants and EcEndoV, except that the expression of EcEndoV and the human mutant DelIns4 were induced overnight with 0.3 m*M* IPTG at 37 and 18°C, respectively. Owing to the instability of DelIns4, it was supplemented with 10% glycerol throughout the purification process and cleavage assays.

2.2. Cleavage assay

Single-stranded deoxyinosine-containing DNA (ssDNA; 5′-GCTCGGCTICGGACCGAG-3′) and inosine-containing RNA (ssRNA; 5′-CUGACUICGGAUCAGGGCC-3′) oligonucleotides were chemically synthesized by Life Technologies and Thermo Fisher Scientific. The ssDNA and ssRNA substrates were dissolved in a buffer consisting of 20 m*M* Tris–HCl pH 8.0 and in TE buffer pH 8.0, respectively. 20 pmol nucleic acid was incubated in a 20 µl mixture consisting of 1× T4 polynucleotide kinase buffer, 1 µl [γ-^{32}P]-ATP (3000 Ci mmol^{-1}, 33 pmol µl^{-1}) and 1 µl T4 polynucleotide kinase. The mixture was incubated at 37°C for 30 min before heat-inactivation at 65°C for 30 min. The labelled ssRNA was cleaned with a Zymo-Spin IC Column (Zymo Research, Irvine, California, USA), eluted with 25 µl TE buffer pH 8.0 and stored at −80°C. Double-stranded DNA substrate containing the I–T base pair (dsDNA/I-T) was generated by mixing the complementary-strand DNA (5′-CTCGGTCC-GTAGCCGAGC-3′) in a equimolar ratio with the ssDNA oligonucleotide described above. dsDNA was heated to 95°C followed by slow cooling to room temperature in the presence of 5 m*M* Mg^{2+}. A standard cleavage reaction was carried out in a buffer consisting of 40 m*M* Tris–HCl pH 7.5, 50 m*M* KCl, 4 m*M* $MnSO_4$, 1 m*M* DTT, 5% glycerol and 1 n*M* of the indicated ssDNA, RNA or dsDNA substrate in the presence of 80 n*M* enzyme. The reaction mixtures were incubated at 37°C for 1 h and 2× sample-loading buffer (95% formamide, 25 µ*M* EDTA) was added to terminate the reactions. A 10 µl sample from each reaction was heated to 50°C for 5 min before separation on a 15% denaturing urea–PAGE gel. After electrophoresis, the gel was removed from the apparatus, wrapped in Saran Wrap and exposed to a phosphorimager plate overnight. The plate was read using a Typhoon scanner (GE Healthcare) and the resulting band intensities were quantified using *ImageJ*.

Table 2
Diffraction data-collection and refinement statistics.

Values in parentheses are for the outer shell.

Diffraction source	Oxford Diffraction Xcalibur Nova diffractometer
Wavelength (Å)	1.54
Temperature (K)	153
Detector	Oxford Onyx CCD
Crystal-to-detector distance (mm)	65
Rotation range per image (°)	0.75
Total rotation range (°)	74.25
Exposure time per image (s)	120
Space group	$P2_12_12_1$
a, *b*, *c* (Å)	39.75, 48.71, 108.53
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.82
Resolution range (Å)	26.76–2.30 (2.42–2.30)
Total No. of reflections	36828
No. of unique reflections	9877 (1404)
Completeness (%)	99.5 (99.7)
Multiplicity	3.7 (3.8)
$\langle I/\sigma(I)\rangle$	9.6 (3.5)
$R_{\rm r.i.m.}$†	0.125 (0.407)
Overall *B* factor from Wilson plot (Å^2)	25.2
Resolution range (Å)	26.755–2.300 (2.633–2.300)
Completeness (%)	99.3
σ Cutoff	$F > 1.5\sigma(F)$
No. of reflections, working set	9803 (3063)
No. of reflections, test set	471 (154)
Final $R_{\rm cryst}$	0.210 (0.235)
Final $R_{\rm free}$	0.268 (0.310)
No. of non-H atoms	
Protein	1848
Water	84
Total	1932
R.m.s. deviations	
Bonds (Å)	0.004
Angles (°)	0.889
Average *B* factors (Å^2)	
Protein	17.9
Water	21.4
Ramachandran plot	
Most favoured (%)	96.58
Allowed (%)	3.42

† Estimated $R_{\rm r.i.m.} = R_{\rm merge}[N/(N-1)]^{1/2}$, where *N* is the data multiplicity.

2.3. Crystallization

Initial crystallization screening was set up by a Mosquito crystallization robot (TTP Labtech) using the sitting-drop vapour-diffusion method with the commercial screens Index, Crystal Screen and Crystal Screen 2 (Hampton Research), as well as a homemade PEG/ammonium sulfate-based screen, at 25°C in 96-well plates. Hits were observed in 2–3 d and two crystal forms were obtained: needle-shaped crystals and diamond-shaped crystals. The former were found in a condi-

tion consisting of 0.2 *M* sodium formate, 25% PEG 3350, while the latter were found in a condition consisting of 0.1 *M* sodium chloride, 0.1 *M* bis-tris pH 6.5, 1.5 *M* ammonium sulfate. After optimization, thick rod-shaped crystals were obtained from a condition consisting of 19% PEG 3350, 0.2 *M* sodium formate, 0.1 *M* Tris–HCl pH 7.5, while diamond-shaped crystals were produced from a condition consisting of 1.4 *M* ammonium sulfate, 0.2 *M* sodium acetate, 0.1 *M* Tris–HCl pH 7.5. Crystals were soaked in a freshly made cryoprotective solution containing all of the components of the reservoir solution plus 20%(*v*/*v*) glycerol. The soaked crystals were mounted on nylon loops and flash-cooled in liquid nitrogen. Using an in-house Oxford Diffraction Xcalibur Nova diffractometer operating at 50 kV and 0.8 mA, a full native data set (a total of 99 frames) was collected with a rotation of 0.75° per frame at −120°C. The crystal-to-detector distance was 65 mm and the exposure time was 120 s, and the data were recorded with a 165 mm Onyx CCD detector. The data were processed and scaled using *CrysAlis*^{Pro} (v.1.171.33.49; Oxford Diffraction) and *SCALA* from the *CCP*4 suite (Winn *et al.*, 2011; Table 2).

2.4. Data collection and structure determination

The space group of the crystal was $P2_12_12_1$ and the crystal diffracted to 2.3 Å resolution with a completeness of 99.7% (Table 2). The structure was solved by molecular replacement using *PHENIX* (Adams *et al.*, 2010) with the coordinates of the TmEndoV structure (PDB entry 2w35; Dalhus *et al.*, 2009) as the search model. The initial model was extended by *PHENIX* using the autobuild option and the resulting model was further built manually according to the electron-density map with *Coot* (Emsley *et al.*, 2010). Multiple cycles of

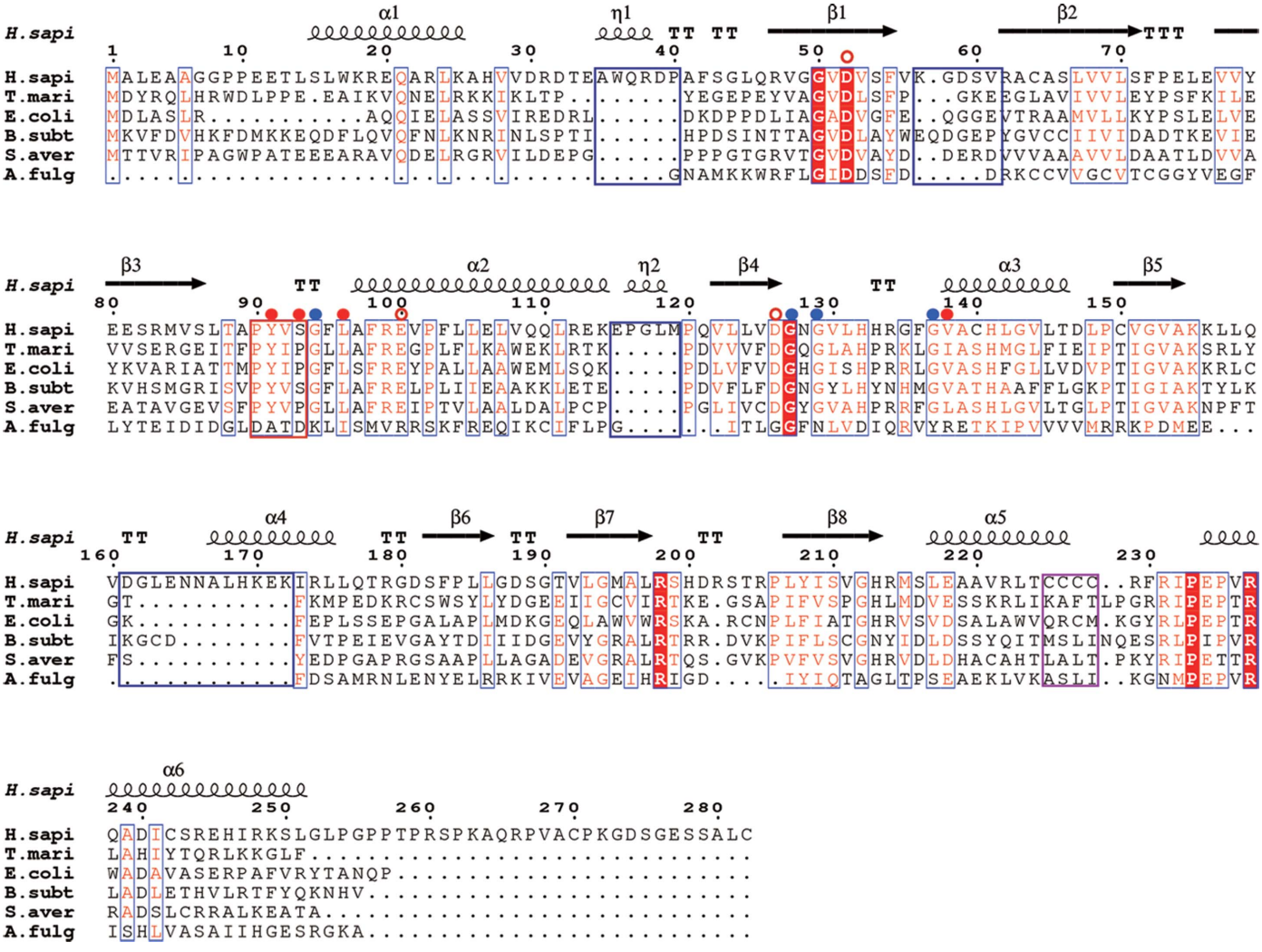

Figure 1
Multiple sequence alignment of human (H.sapi), *T. maritima* (T.mari), *E. coli* (E.coli), *B. subtilis* (B. subt), *S. avermitilis* (S.aver) and *A. fulgidus* (A.fulg) EndoV sequences using *ESPript*. The full-length hEndoV is shown and its secondary structure is shown at the top. Several regions are boxed: the wedge motif (Pro90–Ser93) in red, insertion 1 (Ala35–Pro40), insertion 2 (Lys57–Ser60), insertion 3 (Glu115–Met119) and insertion 4 (Asp161–Asn166) in dark blue and the four-cysteine motif (Cys225–Cys228) in magenta. The red solid dots represent the hydrophobic residues that form the hypoxanthine-binding pocket, while the blue dots represent the invariant glycine residues that form backbone hydrogen bonds to the deaminated base. The open dots are the Mg^{2+} metal-binding residues. Residues that are identical in all four sequences are highlighted in red, while similar residues are shown in light blue boxes.

refinement alternating with model rebuilding were carried out by *phenix.refine* (Afonine *et al.*, 2012). The final *R* factor was 20.97% (R_{free} = 26.83%). The Ramachandran plot of the final model has 96.58, 3.42 and 0% of the residues in the most favoured, generously allowed and disallowed regions (Table 2). The final model was validated by *SFCHECK* and *PROCHECK* (Vaguine *et al.*, 1999; Laskowski *et al.*, 1993). All figures were produced with *PyMOL* (http://www.pymol.org) and the charge distribution on the protein surface was calculated by *APBS* (Baker *et al.*, 2001). The secondary structure of hEndoV-SF was prepared by *ESPript* (http://espript.ibcp.fr; Gouet *et al.*, 2003). The atomic coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org/) as entry 4nsp.

3. Results

3.1. Structure overview and comparison to TmEndoV

hEndoV has a proline-rich C-terminal extension of ∼30 residues compared with its bacterial orthologues (Fig. 1). This C-terminal extension, along with several residues at the N-terminus, is predicted to form an unstructured region which interferes with crystallization (data not shown). Consequently, we removed both regions for crystallization purposes. Such a construct (Thr13–Ser250), named hEndoV-SF, retains around one third of the activity of the full-length enzyme on ssRNA substrate (Supplementary Fig. S1). More importantly, the hEndoV-SF fragment readily crystallizes in two crystal forms: a rod-shaped form obtained in the presence of PEG 3350 and a diamond-shaped form obtained in the presence of ammonium sulfate. We collected data from the former crystals as they gave better quality diffraction.

The rod-shaped crystals belonged to an orthorhombic system, with space group $P2_12_12_1$, and each asymmetric unit contains one molecule with an estimated solvent content of 38.8%. Gel-filtration chromatography indicates that hEndoV-SF exists as a single species with an approximate mass of 28.8 kDa in solution, consistent with the molecular weight of a monomer (the theoretical molecular weight of hEndoV-SF is 27.8 kDa). Continuous and well defined electron density was observed for the entire polypeptide except for the disordered Lys57-Gly58 dipeptide, which is located on a loop connecting β_1 and β_2 (Fig. 2). The electron density for the C-terminal His_6 tag is also absent apart from the very first histidine residue. The refined model contains a total of 240 amino acids and 84 water molecules. The refinement statistics are summarized in Table 2.

hEndoV-SF forms a single $\alpha + \beta$ domain with a spherical shape composed of a β-sheet of eight mixed strands flanked by α-helices on both sides. This three-layered sandwich fold, or 'RNase H-like motif', resembles that of TmEndoV, which shares 31% sequence identity and 50% sequence similarity with hEndoV-SF. The two proteins are superimposable with an r.m.s. deviation of 2.1 Å over 223 C^{α} atoms.

A highly conserved PYIP motif (Pro79–Pro82) of TmEndoV in the DNA-bound form acts as the essential wedge-like recognition element. It inserts into the lesion region and distorts the helix. The equivalent fragment in hEndoV-SF is 90PYVS93 and the side chains of these residues are in the same orientation (Fig. 3*a*). However, the PYIP residues in TmEndoV push further into the DNA helix: the motif moves 3.2–4.1 Å towards the complementary DNA strand, with the largest shift of 4.1 Å observed for the hydroxyl group of tyrosine. This translation is most likely to be induced by the binding of the dsDNA substrate, as the same motif in apo TmEndoV does not display such a dramatic movement, assuming an intermediate position between the TmEndoV–DNA complex and apo hEndoV-SF (Fig. 3*a*).

The wedge in TmEndoV disrupts the I–C base pair and also flips the hypoxanthine base into a hydrophobic rigid binding pocket. The pocket is partially formed by two residues from the wedge motif (Pro82 and Tyr80) as well as two additional residues: Ile122 and Leu85. The equivalent residue to Ile122 is Val138 in hEndoV-SF and it remains in nearly the same position. In contrast, the Leu96 (corresponding to Leu85) side chain reorients and points towards the wedge motif (Fig. 3*b*), the same direction as in apo TmEndoV. This change is therefore a consequence of the movement of the adjacent wedge, which forces Leu85 into the opposite direction owing to steric hindrance. In TmEndoV, Leu85 serves as a physical barrier between the catalytic site and the lesion-recognition pocket. The side-chain orientation of Leu85 may represent a substrate-bound state, while that of Leu96 represents the initial state of the free enzyme. Aside from the hydrophobic interactions, the hypoxanthine base is recognized by

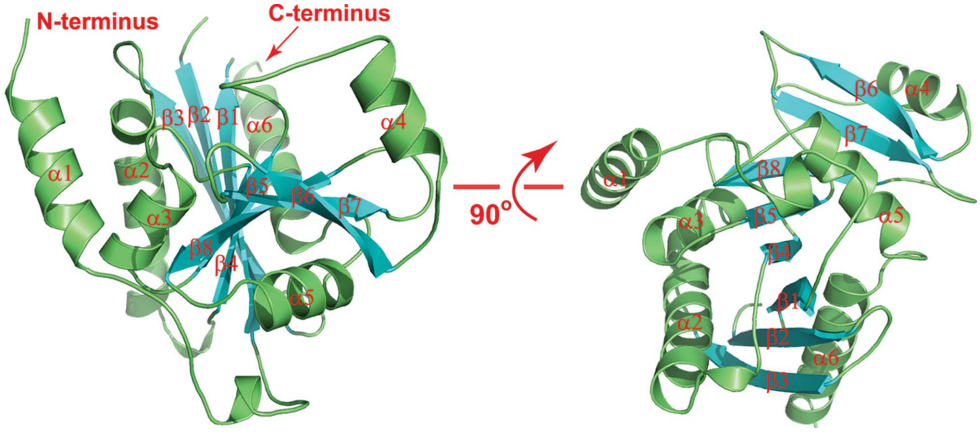

Figure 2
The overall structure of hEndoV-SF in ribbon representation. Side and top views are illustrated. The central β-sheet is coloured cyan and the rest is in green. The major secondary-structure elements and the N- and C-termini are labelled.

the polar backbone interactions of four conserved glycine residues in TmEndoV: Gly83, Gly111, Gly113 and Gly121. The glycines in hEndoV-SF that make backbone interactions with the deaminated base (Gly94, Gly127, Gly129 and Gly137) are absolutely conserved not only in primary sequence but also in three-dimensional space, apart from Gly94, which may also be affected by the wedge movement (Fig. 3*b*). The conservation of these glycine residues ensures a rigid binding pocket and high selectivity for the hypoxanthine base.

In TmEndoV, Asp43, Glu89 and Asp110 coordinate to the active-site magnesium ion. Mutations of these residues remarkably reduce the cleavage efficiency (Huang *et al.*, 2002). In the structure of hEndoV-SF no electron density for a magnesium ion can be found at the this site, but Asp52, Glu100 and Asp126 in hEndoV-SF still align well in the superimposed structures (Fig. 3*b*).

3.2. Sequence and structure alignment with bacterial orthologues

EndoVs are conserved in all three kingdoms. A sequence alignment with several previously characterized EndoVs is shown in Fig. 1 (Dalhus *et al.*, 2009; Liu *et al.*, 2000). Not surprisingly, the most conserved residues primarily cluster in the wedge motif, the metal-binding site and the hypoxanthine-binding pocket. The residues that form the base-binding

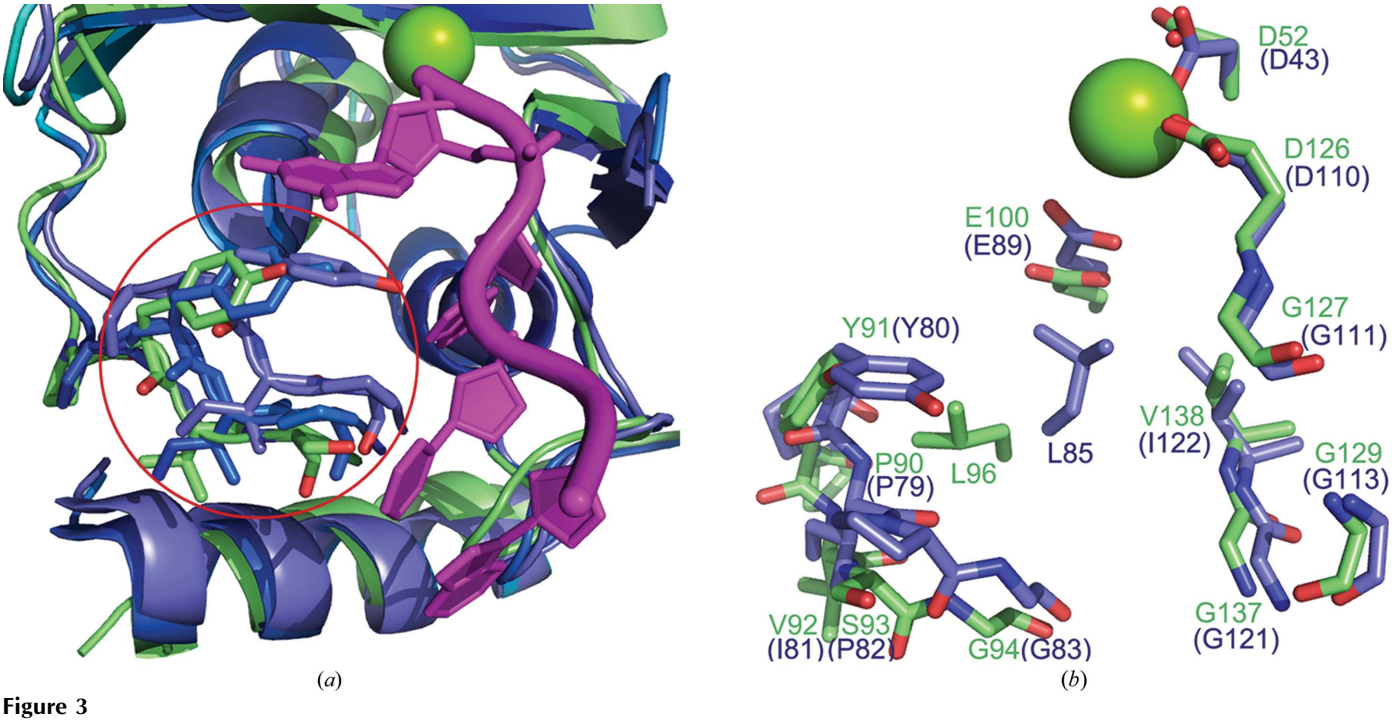

Figure 3
Comparison of key residues for substrate recognition and catalysis. (*a*) Comparison of the wedge motif (circled and shown in stick representation) of apo TmEndoV (PDB entry 3hd0, marine), the TmEndoV–DNA complex (PDB entry 2w35; coloured slate for the protein and magenta for the DNA) and apo hEndoV-SF (green). The magnesium ion is indicated by a green sphere. (*b*) Comparison of the residue positions (shown as stick representations) that form the substrate-binding pocket and the Mg^{2+}-coordinating site in the TmEndoV complex (slate) and hEndoV-SF (green).

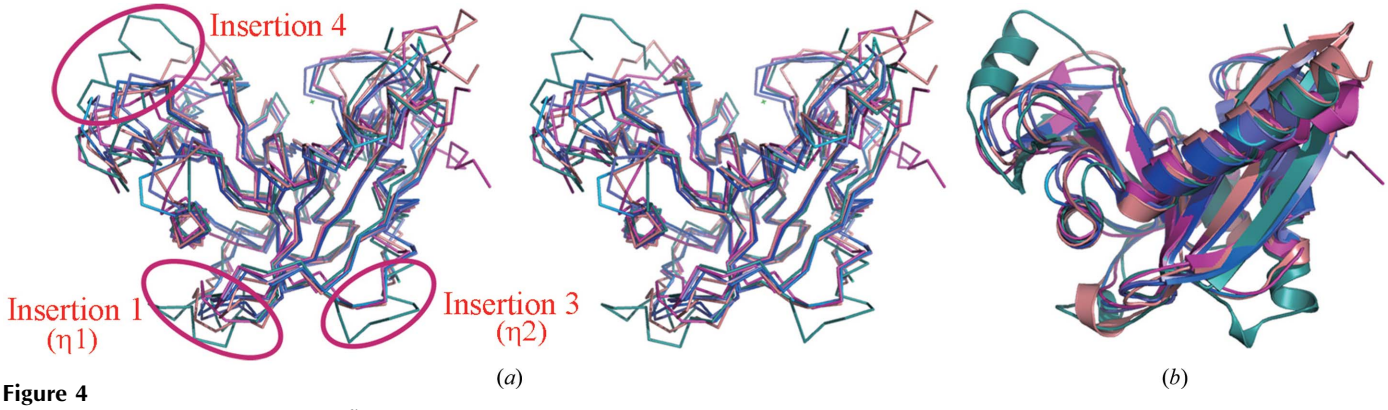

Figure 4
Structure overlay of EndoVs. (*a*) C^{α}-trace representation in stereoview of human EndoV (teal), *B. subtilis* EndoV (salmon; PDB entry 3ga2), *S. avermitilis* EndoV (magenta; PDB entry 3goc), apo *T. maritima* EndoV (cyan; PDB entry 3hd0) and *T. maritima* EndoV in the DNA-bound form (blue; DNA not shown; PDB entry 2w35). Three regions of hEndoV-SF significantly deviate from the other orthologues in structure and are circled by red ovals: insertion 1(η1), insertion 3 (η2) and insertion 4. (*b*) The superimposition in ribbon representation.

pocket are the most conserved sites. In contrast to bacterial orthologues, hEndoV-SF possesses several extra insertions in the middle of the sequence: insertion 1 (η1; residues Ala35–Pro40), insertion 2 (residues Lys57–Ser60), insertion 3 (η2; residues Glu115–Met119), insertion 4 (α4; residues Asp161–Asn166) and the proline-rich loop at the C-terminus (Fig. 1). Notably, insertions 1, 3 and 4 form helical structures in hEndoV-SF, whereas in the bacterial counterparts these regions are short flexible loops. These structural alterations suggest that functional roles for these regions are gained in hEndoV-SF. Another intriguing discovery is the presence of a 225CCCC228 motif at the end of α5 of hEndoV-SF. This motif is not found in bacterial EndoVs but is relatively conserved in eukaryotic EndoVs, especially at the last two cysteine residues (Fig. 1 and Supplementary Fig. S2). The electron-density map clearly demonstrates that these cysteines do not participate in the formation of any disulfide bonds or metal-binding sites. Further functional investigation is needed.

A *DALI* search for structural homologues retrieved several closely related structures: *T. maritima* EndoV in the apo form (PDB entry 3hd0), *Bacillus subtilis* EndoV (PDB entry 3ga2), *Streptomyces avermitilis* EndoV (PDB entry 3goc) and *T. maritima* EndoV in the DNA-bound form (PDB entry 2w35) (Fig. 4). Despite the different substrate specificity, the 'RNase H-like motif' is generally preserved in the catalytic domain. The largest local variations are observed in insertions 1, 3 and 4, which is in agreement with the sequence alignment (Figs. 1 and 4).

3.3. Substrate-recognition specificity

Prokaryotic and human EndoVs show distinct substrate preferences. EcEndoV has a broad substrate spectrum with equal efficacies towards DNA and RNA. Similarly, TmEndoV is also highly active towards DNA substrates. In spite of high sequence homology with TmEndoV, hEndoV only exhibits high cleavage efficiency on ssRNA, with intermediate efficiency on dsRNA or ssDNA containing a ribonucleotide 3′ to dI, but low or no activity on dsDNA. The theoretical pIs for both proteins were similar (∼8.4); consequently, the high activity of TmEndoV on dsDNA is not owing to an overall higher affinity for nucleic acid substrates. We calculated the electrostatic charges on the protein surfaces and found that the charge distribution of the two proteins was different. The DNA-binding surface of TmEndoV comprises a central cleft of conserved residues and the DNA substrate fits deeply into the positive groove (Fig. 5). The positive patches are also located in different areas of the two proteins. Whether these differences contribute to the different substrate-recognition specificities remains to be seen.

Both sequence and structure alignments suggest that the several insertions and the conserved CCCC motif may be functionally important. Thus, we introduced several mutations in these regions to determine whether structure-based engineering could lead to altered substrate specificities.

We first created the following deletion/hybrid mutants of hEndoV-SF: DelIns 1 (deletion of η1, residues Ala35–Pro40), DelIns 3 (deletion of η2, residues Glu115–Met119), DelIns 4 (deletion of α4, residues Asp161–Asn166) and RepIns2 [replacement of insertion 2 (residues Lys57–Ser60) with that of EcEndoV (residues Gln40–Glu43)]. RepIns2 had an expression level comparable to that of the wild type. In contrast, DelIns 1 was expressed in inclusion bodies, while DelIns 3 and DelIns 4 were poorly expressed and were only partially purified (Fig. 6*a*). We then tested the cleavage efficiencies of these mutants on the ssDNA and dsDNA/I-T substrates. The effective concentration of the mutants was kept at 80 n*M* after adjustment of the concentration for contaminant proteins, and all assays were performed in the presence of 4 m*M* Mn^{2+}. However, these mutants with simple deletions or substitutions were still inactive towards either the ssDNA or the dsDNA/I-T substrate (Supplementary Fig. S3).

To test the involvement of disulfide-bond formation by the four-cysteine motif, we incubated hEndoV-SF with 2 m*M* oxidized glutathione. The oxidation of hEndoV-SF by oxidized glutathione for 3 h at 4°C had a minimal effect on enzymatic performance (Figs. 6*b* and 6*c*). We also generated a series of cysteine-to-serine mutants for each cysteine residue in the motif 225CCCC228 and compared their activities with that of wild-type enzyme towards the ssRNA substrate. The C225S and C226S mutations did not cause significant activity loss. In contrast, a single mutation of the highly conserved Cys227 or Cys228 decreased the cleavage activities of hEndoV-SF by 68 and 46%, respectively (Figs. 6*b* and 6*c*).

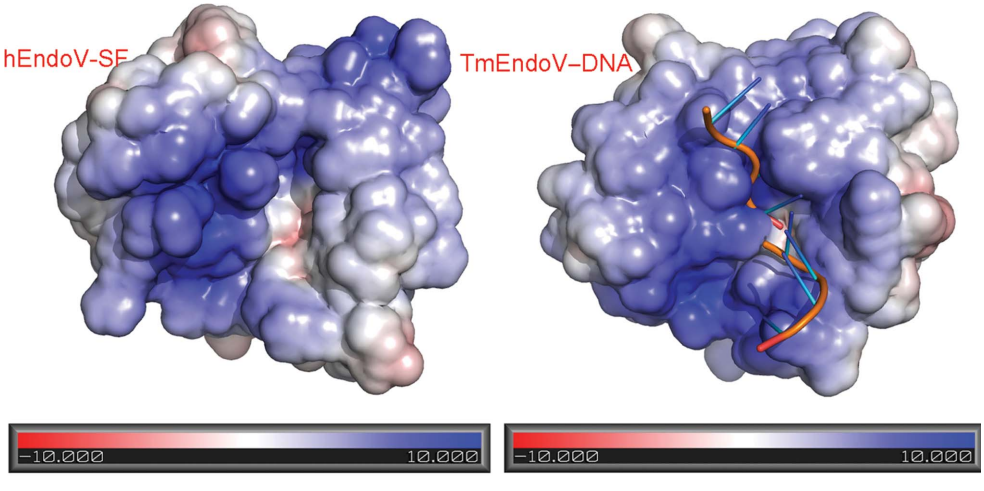

Figure 5
The comparison of the surface charge distribution of apo hEndoV-SF (left) and TmEndoV in the DNA-bound form (right). Surface charge potential was calculated by *APBS*.

4. Discussion

EndoV is a unique enzyme that recognizes I-containing nucleic acids and generates a nick in the second phosphodiester bond 3′ to the inosine in the damaged strand. The cleavage initiates the repair pathway and recruits more enzymes to the lesion site for downstream repair.

EndoVs are conserved in all three kingdoms. Based on our multiple sequence alignment, *Archaeoglobus fulgidus* EndoV (AfEndoV) appears to be the most distant EndoV among all of the orthologues. It lacks the starting helix at the N-terminus and exhibits poor conservation even in the most conserved regions, suggesting that this enzyme adopts a totally different substrate-binding strategy (Fig. 1). AfEndoV has been reported to cleave DNA (Liu *et al.*, 2000) and it would be of interest to explore how it carries out the cleavage reaction.

Despite relative high sequence and structural similarities, bacterial and human EndoVs seem to possess different substrate preferences. The 2.3 Å resolution crystal structure of hEndoV demonstrates that the protein retains the same 'RNase H-like motif' while possessing several insertions and a characteristic four-cysteine motif. Although enzymatic assays do not reveal the immediate functions of these insertion domains, mutations of the highly conserved cysteine residues Cys227 and Cys228 that are far away from the active site dramatically reduced the activity. We could rule out the possibility of the involvement of disulfide bonds or metal coordination from the oxidation experiment. The loss of cleavage efficiency of the cysteine-to-serine mutants suggests an important role for the four-cysteine motif, which is somehow connected to the activity of the enzyme.

hEndoV has high cleavage efficiency towards RNA but a much lower activity towards DNA. Modelling studies indicated that the 2′-hydroxyl group of the adjacent ribose interacts with the highly conserved residue Glu100 (Vik *et al.*, 2013). The lack of activity towards the ssDNA substrate is probably a consequence of the reduced substrate-binding affinity owing to the loss of the original interaction with the 2′-hydroxyl group. This is consistent with the observation that hEndoV still retains most of its activity on dIrG-containing ssDNA. Nevertheless, the usual substrate preference of hEndoV requires the cocrystal structure of hEndoV bound to its authentic substrate in order to reveal the structural basis underlying this interesting enzyme.

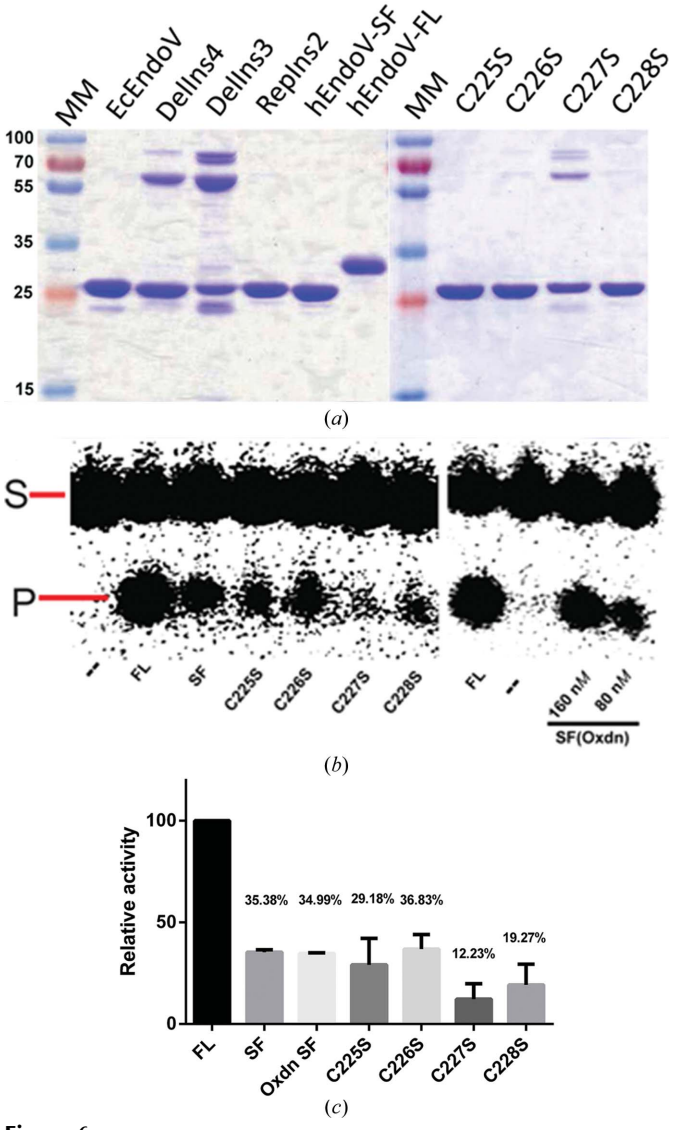

Figure 6
The functions of the insertions and the four-cysteine motif. (*a*) Mutants were created using the QuikChange method and purified mutants were analysed by SDS–PAGE. DelIns3 and DelIns4 were only partially purified. Lane MM, molecular-weight markers (labelled in kDa). (*b*) Cleavage assay of the cysteine-to-serine mutants and the oxidized hEndoV-SF on the ssRNA substrate. S and P represent the substrate and product, respectively. (*c*) Summary of the activities in (*b*) as a bar graph. The activity of the full-length wild type was regarded as 100% and the activities of the mutants were normalized with respect to that of the wild type. Error bars represent SDs calculated from two measurements. The activity of each mutant (as a percentage) is indicated by the number above the bar. The effective concentration of the proteins was kept at 80 n*M* after adjustment of the concentration for contaminant proteins, with the exception of the oxidation experiment. The oxidation experiment of hEndoV-SF was performed at enzyme concentrations of 80 and 160 n*M*, but only the activity at 80 n*M* was used for quantitation. –, blank; FL, full-length hEndoV; SF, hEndoV-SF; oxdn SF, oxidized hEndoV-SF.

The authors thank Zhang Hong and Zhang Liping for technical assistance. This work was supported by the National Sciences Foundation of China (4103041), the Fundamental Research Funds for the Central Universities (SYSU: 11lgjc09) and the Innovative R&D Team Leadership of Guangdong Province (People's Republic of China) (2011Y038).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Agris, P. F., Vendeix, F. A. & Graham, W. D. (2007). *J. Mol. Biol.* **366**, 1–13.
Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H. & Morikawa, K. (1994). *Cell*, **78**, 1063–1072.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Bass, B. L. (2002). *Annu. Rev. Biochem.* **71**, 817–846.

Ceschini, S., Keeley, A., McAlister, M. S., Oram, M., Phelan, J., Pearl, L. H., Tsaneva, I. R. & Barrett, T. E. (2001). *EMBO J.* **20**, 6601–6611.
Dalhus, B., Arvai, A. S., Rosnes, I., Olsen, Ø. E., Backe, P. H., Alseth, I., Gao, H., Cao, W., Tainer, J. A. & Bjørås, M. (2009). *Nature Struct. Mol. Biol.* **16**, 138–143.
Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. & Rayment, I. (2000). *Science*, **289**, 77–85.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Feng, H., Dong, L. & Cao, W. (2006). *Biochemistry*, **45**, 10251–10259.
Fladeby, C., Vik, E. S., Laerdahl, J. K., Gran Neurauter, C., Heggelund, J. E., Thorgaard, E., Strøm-Andersen, P., Bjørås, M., Dalhus, B. & Alseth, I. (2012). *PLoS One*, **7**, e47466.
Gates, F. T. III & Linn, S. (1977). *J. Biol. Chem.* **252**, 1647–1653.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
He, B., Qing, H. & Kow, Y. W. (2000). *Mutat. Res.* **459**, 109–114.
Huang, J., Lu, J., Barany, F. & Cao, W. (2001). *Biochemistry*, **40**, 8738–8748.
Huang, J., Lu, J., Barany, F. & Cao, W. (2002). *Biochemistry*, **41**, 8342–8350.
Karakas, E., Truglio, J. J., Croteau, D., Rhau, B., Wang, L., Van Houten, B. & Kisker, C. (2007). *EMBO J.* **26**, 613–622.
Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990). *Nature (London)*, **347**, 306–309.
Keegan, L. P., Leroy, A., Sproul, D. & O'Connell, M. A. (2004). *Genome Biol.* **5**, 209.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lim, V. I. (1995). *J. Mol. Biol.* **252**, 277–282.
Lin, J., Gao, H., Schallhorn, K. A., Harris, R. M., Cao, W. & Ke, P. C. (2007). *Biochemistry*, **46**, 7132–7137.
Lindahl, T. (1979). *Prog. Nucleic Acid Res. Mol. Biol.* **22**, 135–192.
Liu, J., He, B., Qing, H. & Kow, Y. W. (2000). *Mutat. Res.* **461**, 169–177.
Maas, S., Kawahara, Y., Tamburro, K. M. & Nishikura, K. (2006). *RNA Biol.* **3**, 1–9.
Mi, R., Abole, A. K. & Cao, W. (2011). *Nucleic Acids Res.* **39**, 536–544.
Morita, Y., Shibutani, T., Nakanishi, N., Nishikura, K., Iwai, S. & Kuraoka, I. (2013). *Nature Commun.* **4**, 2273.
Rosnes, I., Rowe, A. D., Vik, E. S., Forstrom, R. J., Alseth, I., Bjoras, M. & Dalhus, B. (2013). *Structure*, **21**, 257–265.
Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. (2004). *Science*, **305**, 1434–1437.
Sprinzl, M., Dank, N., Nock, S. & Schon, A. (1991). *Nucleic Acids Res.* **19**, 2127–2171.
Tsutsumi, S., Sugiura, R., Ma, Y., Tokuoka, H., Ohta, K., Ohte, R., Noma, A., Suzuki, T. & Kuno, T. (2007). *J. Biol. Chem.* **282**, 33459–33465.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Vik, E. S., Nawaz, M. S., Strøm Andersen, P., Fladeby, C., Bjørås, M., Dalhus, B. & Alseth, I. (2013). *Nature Commun.* **4**, 2271.
Weiss, B. (2008). *DNA Repair*, **7**, 205–212.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). *Science*, **249**, 1398–1405.
Yao, M., Hatahet, Z., Melamede, R. J. & Kow, Y. W. (1994). *J. Biol. Chem.* **269**, 16260–16268.
Yao, M. & Kow, Y. W. (1994). *J. Biol. Chem.* **269**, 31390–31396.
Yao, M. & Kow, Y. W. (1996). *J. Biol. Chem.* **271**, 30672–30676.
Zinshteyn, B. & Nishikura, K. (2009). *Wiley Interdiscip. Rev. Syst. Biol. Med.* **1**, 202–209.