Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An additional C-terminal loop in endonuclease IV, an apurinic/apyrimidinic endonuclease, controls binding affinity to DNA

Ryuichi Asano,[a] Hirohito Ishikawa,[a] Shuhei Nakane,[a] Noriko Nakagawa,[a,b] Seiki Kuramitsu[a,b] and Ryoji Masui[a,b]*

[a]Department of Biological Sciences, Graduate School of Science, Osaka University, 1-1 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan, and [b]Photon Science Research Division, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan

Correspondence e-mail: rmasui@bio.sci.osaka-u.ac.jp

Received 1 April 2010
Accepted 14 December 2010

PDB References: *Geobacillus kaustophilus* endonuclease IV, 3aal; *Thermus thermophilus* endonuclease IV, 3aam.

Endonuclease IV (EndoIV) is an endonuclease that acts at apurinic/apyrimidinic (AP) sites and is classified as either long-type or short-type. The crystal structures of representative types of EndoIV from *Geobacillus kaustophilus* and *Thermus thermophilus* HB8 were determined using X-ray crystallography. *G. kaustophilus* EndoIV (the long type) had a higher affinity for double-stranded DNA containing an AP-site analogue than *T. thermophilus* EndoIV (the short type). Structural analysis of the two different EndoIVs suggested that a C-terminal DNA-recognition loop that is only present in the long type contributes to its high affinity for AP sites. A mutation analysis showed that Lys267 in the C-terminal DNA-recognition loop plays an important role in DNA binding.

1. Introduction

The apurinic/apyrimidinic (AP) site is one of the most common DNA lesions *in vivo* and is produced by spontaneous hydrolysis, chemical toxins, radiation or DNA glycosylases that remove abnormal DNA bases (Friedberg *et al.*, 2006; Seeberg *et al.*, 1995; Demple & Sung, 2005). The AP site prevents transcription and translation and thus must be repaired by DNA-repair pathways, including base-excision repair. In base-excision repair, an AP endonuclease recognizes the AP site and cleaves the DNA backbone at the 5′-side of the AP site to generate a 5′-deoxyribose phosphate group and a free 3′-hydroxyl end for DNA polymerase repair synthesis (Levin *et al.*, 1988; Barzilay & Hickson, 1995). AP endonucleases are classified on the basis of their structure as members of either the exonuclease III family or the endonuclease IV (EndoIV) family. In yeast, deletion of the EndoIV homologue increases the rate of transversion (AT to GC) mutations and results in hypersensitivity to alkylating agents and chemical oxidants (Kunz *et al.*, 1994). EndoIV is induced by superoxide for defence against oxygen radicals and nitric oxide (Friedberg *et al.*, 2006). All of the available information highlights the physiological importance of EndoIV in DNA repair. The EndoIV family has been widely identified in archaea, bacteria and some eukaryotes but not in most mammals. This makes EndoIV an attractive target for antibiotics (Kunz *et al.*, 1994).

In bacteria, base-excision repair has been studied extensively in the Gram-negative species *Escherichia coli*. The tertiary structure of EndoIV has only been determined for the *E. coli* homologue (Hosfield *et al.*, 1999). In contrast, little effort has been made to study EndoIV from Gram-positive

bacteria (Firmicutes; Shida *et al.*, 1999). DNA-repair processes in Gram-positive bacteria are known to differ from those in Gram-negative bacteria (Chédin & Kowalczykowski, 2002). Furthermore, EndoIVs from Firmicutes commonly have a longer C-terminal extension (about 30 residues) than homologues from Gram-negative bacteria (see Fig. 1). Here, we designate EndoIV from Firmicutes as the long type and the version without the C-terminal extension as the short type. The presence of an additional short sequence at the C-terminus may affect the molecular function of the enzyme. To date, however, the tertiary structure of long-type EndoIV has not been determined.

In *E. coli* exonuclease III is the major AP endonuclease, while EndoIV generally plays a minor role (Yajko & Weiss, 1975). However, genome sequence data revealed that several bacterial genera, including *Chlamydia*, *Aquifex*, *Thermus* and *Thermotoga*, only have EndoIV as an AP endonuclease. Similarly, EndoIV seems to be the only AP endonuclease in several archaeal genera. In these genera EndoIV may have a different functional significance to that of *E. coli* EndoIV (ecEndoIV). This significance might be reflected by a difference in tertiary structure even if these genera have the short-type EndoIV as in *E. coli*.

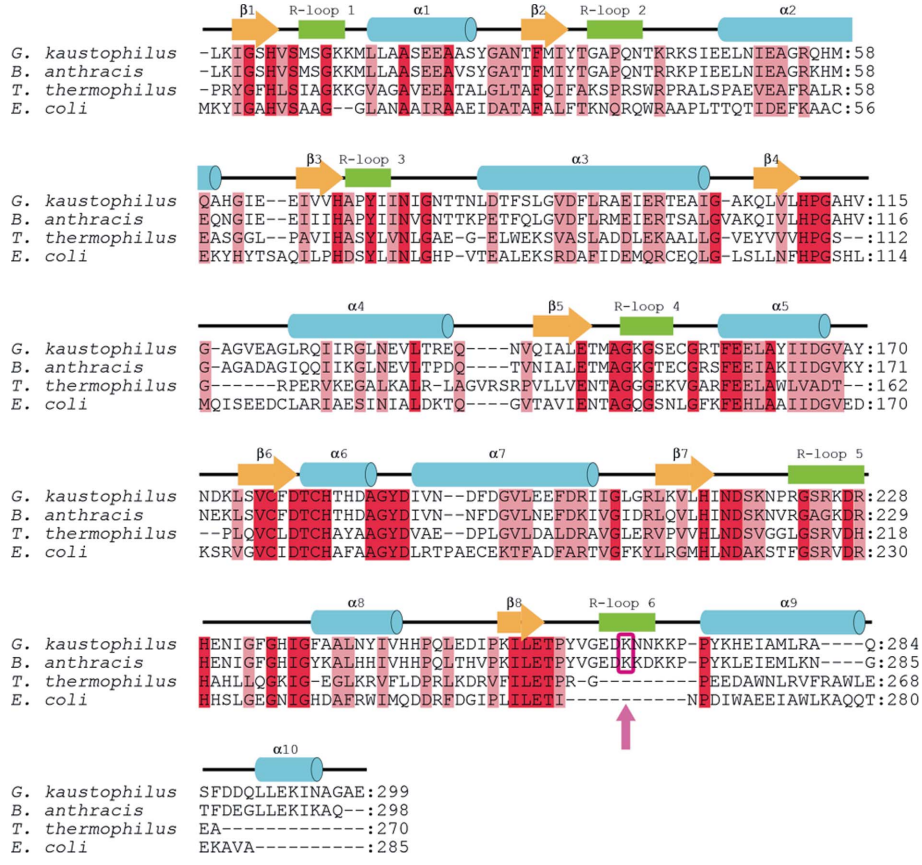

Figure 1
Structure-based sequence alignment of EndoIVs from *G. kaustophilus* (PDB entry 3aal), *B. anthracis* (PDB entry 1xp3), *T. thermophilus* (PDB entry 3aam) and *E. coli* (PDB entry 1qtw; Hosfield *et al.*, 1999). Cylinders indicate α-helices (aqua, $\alpha1$–$\alpha10$) and arrows indicate β-strands (orange, $\beta1$–$\beta8$). Green rectangles indicate DNA-recognition R-loops 1–6. Identical and similar residues are boxed in red and pink, respectively. The magenta arrow indicates the lysyl residue which interacts with the phosphate group of the DNA backbone.

In this report, we chose thermophilic representatives of both types of EndoIV: *Geobacillus kaustophilus* EndoIV (gkEndoIV; NCBI accession No. YP_148327) is of the long-type, while *Thermus thermophilus* HB8 EndoIV (ttEndoIV; NCBI accession No. YP_144100) is of the short-type. gkEndoIV and ttEndoIV are the only AP endonucleases that are present in these species. Proteins from thermophiles are thermostable and often give high-resolution diffraction data in X-ray crystallography. We determined the crystal structures of gkEndoIV and ttEndoIV. The results obtained from structural and DNA-binding analyses indicated that the additional C-terminal residue plays an important role in recognizing DNA. Additionally, the data allowed us to predict the DNA-binding mode of each type of EndoIV in various organisms.

2. Materials and methods

2.1. Expression and purification of gkEndoIV

The gkEndoIV gene was amplified from *G. kaustophilus* genomic DNA by PCR using the primers 5′-GGAATTGTTTAAACCATATGTTAAAAATCGGTTCACATGTTTC-3′ and 5′-GGAATTGGATCCTTATTATTCCGCCCCAGCGTTGATTTTTTC-3′; the amplification product was inserted into the *Pme*I and *Bam*HI restriction sites of the pET-HisTEV3 vector. The gkEndoIV K267A expression plasmid was constructed from pET-HisTEV3/gkEndoIV by the inverse PCR method using the primers 5′-GCGAGGATGCAAACAAC-3′ and 5′-CCACGTATGGGGTTTC-3′. The gkEndoIV expression plasmids were used to transform *E. coli* Rosetta2(DE3). The *E. coli* transformant was cultured at 310 K overnight in 9 l Luria–Bertani medium (Bertani, 1951) containing 50 µg ml^{-1} ampicillin. Protein expression was induced with 50 mg l^{-1} isopropyl β-D-1-thiogalactopyranoside. Cells harbouring the gkEndoIV expression plasmid (35 g) were lysed by ultrasonication in 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 5 m*M* β-mercaptoethanol and 1 m*M* phenylmethanesulfonyl fluoride and the resultant lysate was heat-treated at 333 K for 13 min. After centrifugation, gkEndoIV, which was present in the supernatant, was purified by successive chromatographic steps with HisTrap HP5, HiLoad 16/60 Superdex 200 and HiPrep 26/10 (GE Healthcare). gkEndoIV K267A was prepared using the same method as used for wild-type gkEndoIV.

Table 1
Data-collection and refinement statistics for gkEndoIV and ttEndoIV.

Values in parentheses are for the highest resolution shell.

Data set	gkEndoIV	ttEndoIV
Data collection		
Wavelength (Å)	1.0000	1.0000
Resolution (Å)	50.00–1.60 (1.66–1.60)	50.00–1.58 (1.64–1.58)
Space group	$P6_122$	$P2_1$
Molecules per asymmetric unit	1	1
Unit-cell parameters (Å, °)	$a = b = 85.170$, $c = 140.220$, $\alpha = \beta = \gamma = 120.00$	$a = 36.646$, $b = 85.273$, $c = 36.580$, $\alpha = \gamma = 90.00$, $\beta = 105.370$
Measured reflections	599847	180015
Unique reflections	40267	29574
Completeness (%)	99.7 (97.6)	99.2 (93.9)
Multiplicity	14.9 (6.6)	6.1 (4.0)
$\langle I/\sigma(I)\rangle$	62.8 (6.4)	43.6 (8.4)
$R_{\rm merge}$† (%)	4.9 (21.6)	6.8 (16.1)
Refinement		
Resolution (Å)	23.42–1.60	29.12–1.58
$R_{\rm work}$‡/$R_{\rm free}$§ (%)	19.7/22.2	19.9/22.3
No. of protein atoms	2316	2048
No. of water atoms	110	223
R.m.s.d. bond lengths (Å)	0.013	0.004
R.m.s.d. angles (°)	1.600	1.200
Average B factor (Å^2)	17.6	13.8
Ramachandran plot (%)		
Most favoured	93.4	92.0
Additional	6.2	8.0
Disallowed	0.4	0.0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the average of individual measurements of $I_i(hkl)$. ‡ $R_{\rm work} = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $|F_{\rm obs}|$ and $|F_{\rm calc}|$ are the observed and calculated structure-factor amplitudes, respectively. § $R_{\rm free}$ is calculated against a 10% random sampling of reflections that were removed before structure refinement.

2.2. Expression and purification of ttEndoIV

The ttEndoIV gene was amplified from *T. thermophilus* genomic DNA by PCR using the primers 5′-ATATCATATGCCACGCTACGGGTTCCACCTTTCC-3′ and 5′-ATATGGATCCTTATTAGGCCTCCTCGAGCCAGGCCCT-3′; the amplification product was inserted into the pT7Blue vector using the *Eco*RV restriction site. The gene encoding ttEndoIV was subcloned into the *Nde*I and *Bam*HI sites of pET-15b (Novagen) from pT7Blue/ttEndoIV and was then subcloned into the *Nde*I and *Bam*HI sites of pET-11a (Novagen) from pET-15b/ttEndoIV. The ttEndoIV expression plasmid was used to transform *E. coli* BL21(DE3). The *E. coli* transformant was cultured at 310 K overnight in 12 l Luria–Bertani medium (Bertani, 1951) containing 50 µg ml^{-1} ampicillin. Protein expression was induced with 50 mg l^{-1} isopropyl β-D-1-thiogalactopyranoside. Cells harbouring the ttEndoIV expression plasmid (15 g) were lysed by ultrasonication in 50 m*M* Tris–HCl pH 8.0 and 5 m*M* β-mercaptoethanol and the resultant lysate was heat-treated at 338 K for 10 min. After centrifugation, ttEndoIV, which was present in the supernatant, was purified by successive chromatographic steps with Toyopearl Phenyl, Toyopearl SuperQ (Tosoh), hydroxyapatite (Nacalai) and Superdex75 10/300 GL (GE Healthcare).

2.3. DNA-binding assay

A 21 bp double-stranded DNA containing an AP-site analogue (5′-GGGTGTTGCFTTAGTTGTCAT-3′ and 5′-ATGACAACTAAGGCAACACCC-3′, where F is tetrahydrofuran, an AP-site analogue) was used for the binding assay. The reaction mixture contained 20 m*M* Tris–HCl pH 8.0, 10 m*M* KCl, the 21 bp DNA and EndoIV. 0.5 µ*M* EndoIV solution (200 µl) containing 50 m*M* EDTA was titrated against 10 µ*M* DNA and the fluorescence of the mixture was measured at 340 nm with excitation at 295 nm at 298 K using an F-4500 fluorescence spectrophotometer (Hitachi). The binding curve was analyzed by a similar method as described previously (Watanabe *et al.*, 1994).

2.4. Crystallization, data collection and structure determination

gkEndoIV crystals were obtained using condition No. 6 (0.1 *M* cacodylate pH 6.5, 30% polyethylene glycol 600, 1 *M* NaCl and 10% glycerol) of the Cryo II screen kit (Emerald BioSystems) using the 96-well sitting-drop vapour-diffusion method with drops consisting of 1 µl 20 mg ml^{-1} protein solution and 1 µl reservoir solution equilibrated against 500 µl well solution at 293 K. ttEndoIV crystals were obtained using condition No. 43 (0.2 *M* ammonium phosphate monobasic and 20% polyethylene glycol 3350) of the PEG/Ion screen kit (Hampton Research) with an additional 20 m*M* $MnCl_2$ using the hanging-drop method with drops consisting of 0.5 µl protein solution at 19.7 mg ml^{-1} and 0.5 µl reservoir solution equilibrated against 200 µl well solution at 293 K (Iino *et al.*, 2008). The dimensions of the harvested crystals were 0.05 × 0.05 × 0.2 mm. Data were collected under a liquid-nitrogen stream on the RIKEN Structural Genomics Beamline II (BL26B2; Ueno *et al.*, 2006) at SPring-8 (Hyogo, Japan). The data were processed using the *HKL*-2000 program suite

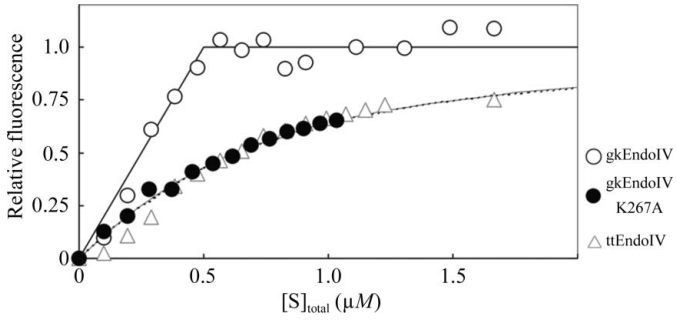

Figure 2
Binding kinetics of EndoIVs to DNA containing an AP-site analogue. The reaction mixture (200 µl) contained 0.5 µ*M* EndoIV, 50 m*M* EDTA and various concentrations of substrate DNA. Fluorescence measurements were performed at 298 K. Changes in fluorescence emission intensity at 340 nm using an excitation wavelength of 295 nm were plotted against the total concentration of DNA. Complete binding of enzyme to substrate was designated as 1.0 on the ordinate axis (relative fluorescence). Open circles, filled circles and open triangles indicate gkEndoIV, gkEndoIV K267A mutant and ttEndoIV, respectively. The black solid line, broken line and grey solid line are the theoretical lines for gkEndoIV, K267A gkEndoIV mutant and ttEndoIV, respectively.

(Otwinowski & Minor, 1997). The structures were solved using *Bacillus anthracis* EndoIV (PDB code 1xp3; M. J. Fogg, V. M. Levdikov, E. V. Blagova, J. A. Brannigan, A. J. Wilkinson & K. S. Wilson, unpublished work) as the model for molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010). The automatic tracing procedures in the program *ARP/wARP* (Morris *et al.*, 2003) were utilized to build the initial models. Model refinements were carried out using the programs *XtalView* (McRee, 1992) and *CNS* (Brünger *et al.*, 1998). According to the *PROCHECK* software (Laskowski *et al.*, 1993), 93.4 and 92.0% of the residues in the final models of gkEndoIV and ttEndoIV, respectively, are located in the most favoured region of the Ramachandran plot and 0.4 and 0% of the residues, respectively, are located in disallowed regions. Data-collection and data-processing statistics are presented in Table 1. The metal ions were determined using X-ray absorption fine structure (XAFS), measurement of inductively coupled plasma (ICP), coordination number and an anomalous difference Fourier map. The coordinates are available in the Protein Data Bank under accession codes 3aal and 3aam. The nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under accession Nos. AP006508 and AP008226. DNA-binding models were constructed based on the structure of ecEndoIV complexed with a product DNA (PDB code 1qum; Hosfield *et al.*, 1999) using *PyMOL* (DeLano, 2002) and *CNS*.

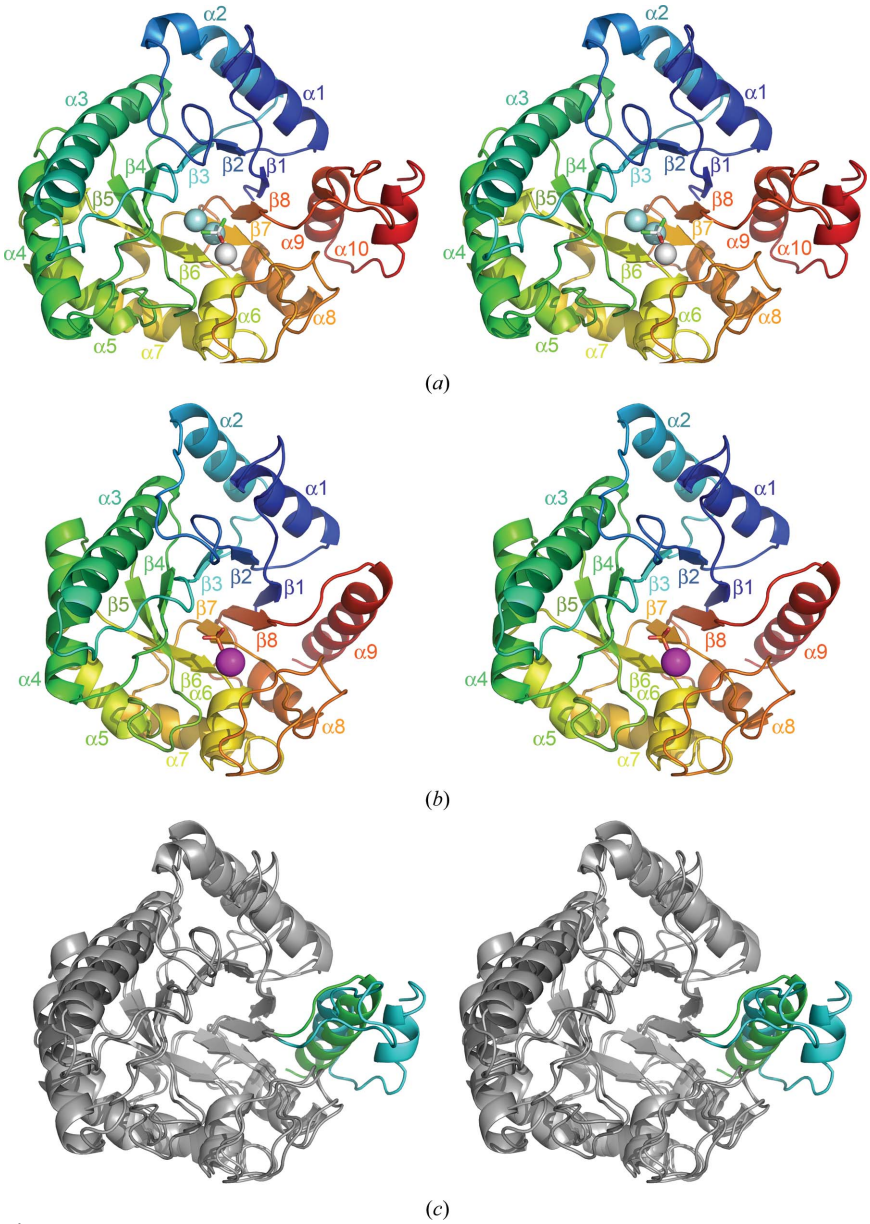

Figure 3
Superposed crystal structures of gkEndoIV and ttEndoIV. Rainbow colouring from blue to green to red indicates the position of the residues from the N-terminus to the C-terminus in the models. (*a*) The overall structure of gkEndoIV. Aqua and white spheres indicate Fe^{2+} and Zn^{2+} ions, respectively. Cacodylate is shown in stick form. (*b*) The overall structure of ttEndoIV. The magenta sphere indicates the Mn^{2+} ion. Phosphate is shown in stick form. (*c*) Superposition of gkEndoIV and ttEndoIV. The C-terminal regions of gkEndoIV and ttEndoIV are coloured green and aqua, respectively.

3. Results and discussion

3.1. Binding affinity of the AP site in double-stranded DNA

We measured the binding affinity of gkEndoIV and ttEndoIV for the substrate analogue (a 21 bp double-stranded DNA containing an AP-site analogue) using intrinsic fluorescence of the protein from tryptophan or tyrosine residues (Fig. 2). When the ttEndoIV solution was titrated against the DNA the fluorescence with an emission maximum at 345 nm, probably from a tryptophan, was quenched on DNA binding. The fluorescence emission spectrum of gkEndoIV had an emission maximum at 320 nm, but it also changed when titrated against the DNA. The titration experiments gave the binding profiles shown in Fig. 2. The affinity of gkEndoIV for the substrate analogue was found to be very high and its dissociation constant was less than 0.01 μ*M*. The break point of total concentration in the presence of 0.5 μ*M* gkEndoIV was 0.5 μ*M* (open circles in Fig. 2). This indicates that one molecule of gkEndoIV binds to one molecule of double-stranded DNA. The affinity of ttEndoIV for the substrate analogue was lower than that of gkEndoIV (open triangles in Fig. 2). Assuming a bimolecular reaction, its dissociation constant was calculated to be 0.37 μ*M*. These results indicate that gkEndoIV has a higher affinity for the AP site in double-stranded DNA than ttEndoIV. It should be noted that the DNA-binding assay was performed at 298 K. The absolute values of the dissociation constants

may be changed at optimum temperatures for both thermophiles.

3.2. Structural determination

To investigate the differences between gkEndoIV and ttEndoIV in their relative affinities for the substrate, we determined their tertiary structures by X-ray crystallography. The overall structures of gkEndoIV (Fig. 3*a*) and ttEndoIV (Fig. 3*b*) were determined at resolutions of 1.60 and 1.58 Å, respectively. In both EndoIVs the first methionine in the amino-acid sequence encoded by their genes was removed in the determined structures. The crystallographic data are summarized in Table 1. Each asymmetric unit contained one molecule that adopted an eight-stranded α/β-barrel fold (TIM barrel). The two structures were very similar, with a root-mean-square deviation (r.m.s.d.) of only 1.65 Å for the C^{α} atoms. gkEndoIV, with 299 amino-acid residues, is longer than ttEndoIV by 29 residues (Fig. 1). The most significant difference between these structures was observed in the C-terminal region, which contained an additional loop and an α-helix in gkEndoIV (Fig. 3*c*). This structural difference may contribute to the increased affinity of gkEndoIV for the substrate (discussed later).

Another difference was observed with respect to the metal ions bound to the proteins. In the determined structures gkEndoIV contained two Fe^{2+} ions and one Zn^{2+} ion (Fig. 4*a*), whereas ttEndoIV contained one Mn^{2+} ion (Fig. 4*b*). The Mn^{2+} ion was probably derived from the crystallization mother liquor. The Mn^{2+}-ion site in ttEndoIV was verified using an anomalous difference Fourier map for data collected at the Zn peak wavelength (1.7000 Å). The assignment of Fe^{2+} and Zn^{2+} ions in gkEndoIV was verified by XAFS and ICP measurements. The numbers of these metal ions were also determined by ICP measurements. A cacodylate ion was bound to the active site of gkEndoIV and a phosphate ion was bound to that of ttEndoIV. The cacodylate and phosphate ions were components of the crystallization mother liquor. The sites to which they bound were expected to be the site of the DNA phosphate group.

In ttEndoIV, the Mn^{2+} ion was coordinated by His172, Asp217, His219, phosphate and one water molecule. In gkEndoIV, the Zn^{2+} ion was coordinated by His182, Asp227, His229 and cacodylate; Fe^{2+}(1) was coordinated by Glu145, Asp179, His214, Glu259 and cacodylate; and Fe^{2+}(2) was coordinated by His69, His110 and Glu145 (Fig. 4). The Zn^{2+} ion in gkEndoIV and the Mn^{2+} ion in ttEndoIV occupied the same binding site. In ttEndoIV, a water molecule occupied the same position as the Fe^{2+}(2) ion in gkEndoIV. The zinc ion had nearly tetrahedral coordination geometry, whereas the Mn^{2+} and Fe^{2+}(1) ions seemed to have octahedral geometry. The positions of the three metal ions in gkEndoIV were almost the same as those of the three Zn^{2+} ions in ecEndoIV. The metal-chelating residues and their positions were conserved among ttEndoIV, gkEndoIV and ecEndoIV (Hosfield *et al.*, 1999), with the exception of His108 of ttEndoIV. His108 of ttEndoIV corresponds to His110 of gkEndoIV and His109 of ecEndoIV. In gkEndoIV and ecEndoIV the side chains of these histidine residues coordinated to a metal ion in the active site (Fig. 4*a*). In ttEndoIV, however, the side chain of His108 was directed in the opposite direction (Fig. 4*b*) and made a hydrogen bond to the carbonyl O atom of Val74.

The difference in the orientation of the histidine side chain might be derived from differences in the solvent exposure of the active sites of these proteins. In gkEndoIV the location of residues 149–153 (GKGSE) partially shields the active site from solvent. A similar pattern of location also occurs in ecEndoIV (residues 149–153; GQGSN). In these structures, bulky residues such as Lys and Gln seem to form an effective shield. In ttEndoIV residues 143–147 (GGGEK) are located in a similar manner to those in gkEndoIV and ecEndoIV. However, the bulky residues Glu and Lys are positioned away from the active site, making it more exposed to solvent. Metal ions in the active site that have greater exposure to solvent should dissociate more easily. Consequently, the loss of the metal ion may have allowed the side chain of His108 to adopt a different conformation.

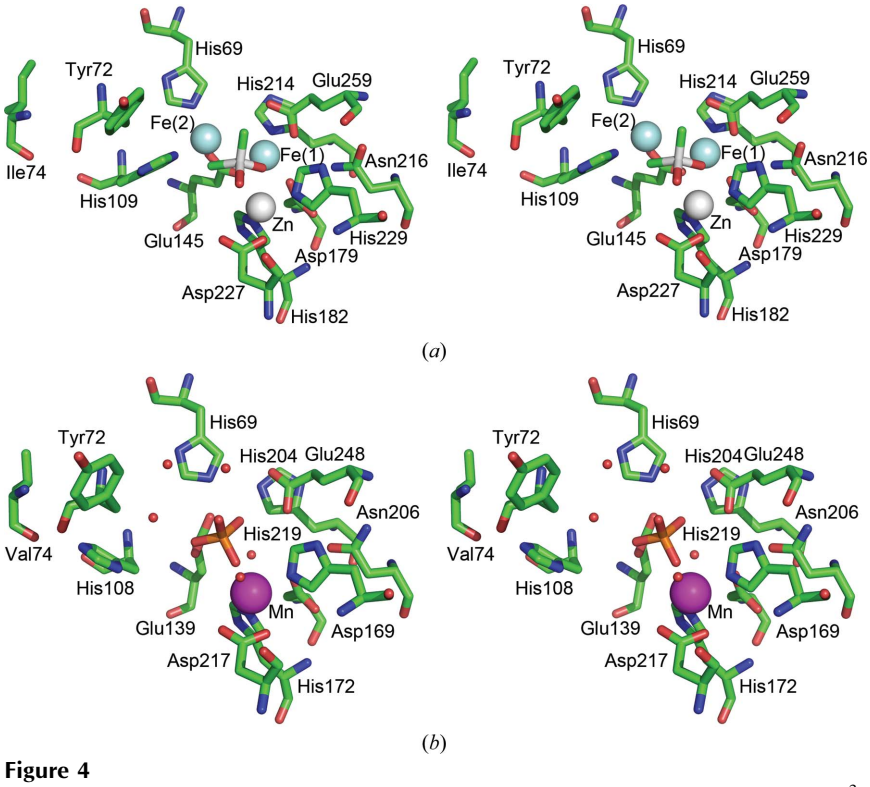

Figure 4
Active sites of gkEndoIV (*a*) and ttEndoIV (*b*). In (*a*), aqua and white spheres indicate Fe^{2+} and Zn^{2+} ions, respectively. Cacodylate is shown in stick form. In (*b*), the magenta sphere indicates the Mn^{2+} ion. Red spheres indicate water molecules.

3.3. DNA-binding site

To obtain further insights into the possible reasons for the differences in affinity for the substrate shown by the two types of enzyme, we constructed DNA-binding models for ttEndoIV and gkEndoIV based on the structure of ecEndoIV complexed with product DNA (Hosfield *et al.*, 1999; Fig. 5). The ecEndoIV structure in the DNA complex could be superimposed on those of gkEndoIV and ttEndoIV with r.m.s.d.s (for C^{α} atoms) of 1.65 and 1.69 Å, respectively. The five DNA-recognition loops (R-loops) that have been defined previously (Garcin *et al.*, 2008) were identified in the model structures of both gkEndoIV and ttEndoIV (Figs. 5*a* and 5*b*, respectively). The presence of these loops suggests that the interactions with DNA are conserved in these proteins. In the ttEndoIV–DNA model structure Trp41 was located near the DNA, suggesting that the fluorescence of this Trp residue was quenched upon DNA binding (Fig. 5*b*). This residue was not conserved in gkEndoIV (Fig. 1). Instead, Tyr34 and Tyr72 were found near the DNA in the gkEndoIV–DNA model structure (Fig. 5*a*).

Unexpectedly, in the model structure of gkEndoIV complexed with DNA an additional loop (265–269) between the last two helices was identified and was presumed to interact with the bound DNA (Fig. 5*a*). The gkEndoIV–DNA binding model suggested a possible interaction between the phosphate group of the DNA backbone and Lys267 in this loop (Fig. 5*c*). In contrast, it was not present in ttEndoIV (Fig. 5*d*). The amino-acid sequence of this loop was not restricted to gkEndoIV but was also conserved in EndoIVs from other Firmicutes, including *Bacillus*, *Listeria*, *Staphylococcus* and *Enterococcus*. Thus, we designated this loop the sixth R-loop (R-loop 6). We further hypothesized that R-loop 6 in gkEndoIV contributes to the high binding affinity for DNA containing an AP site.

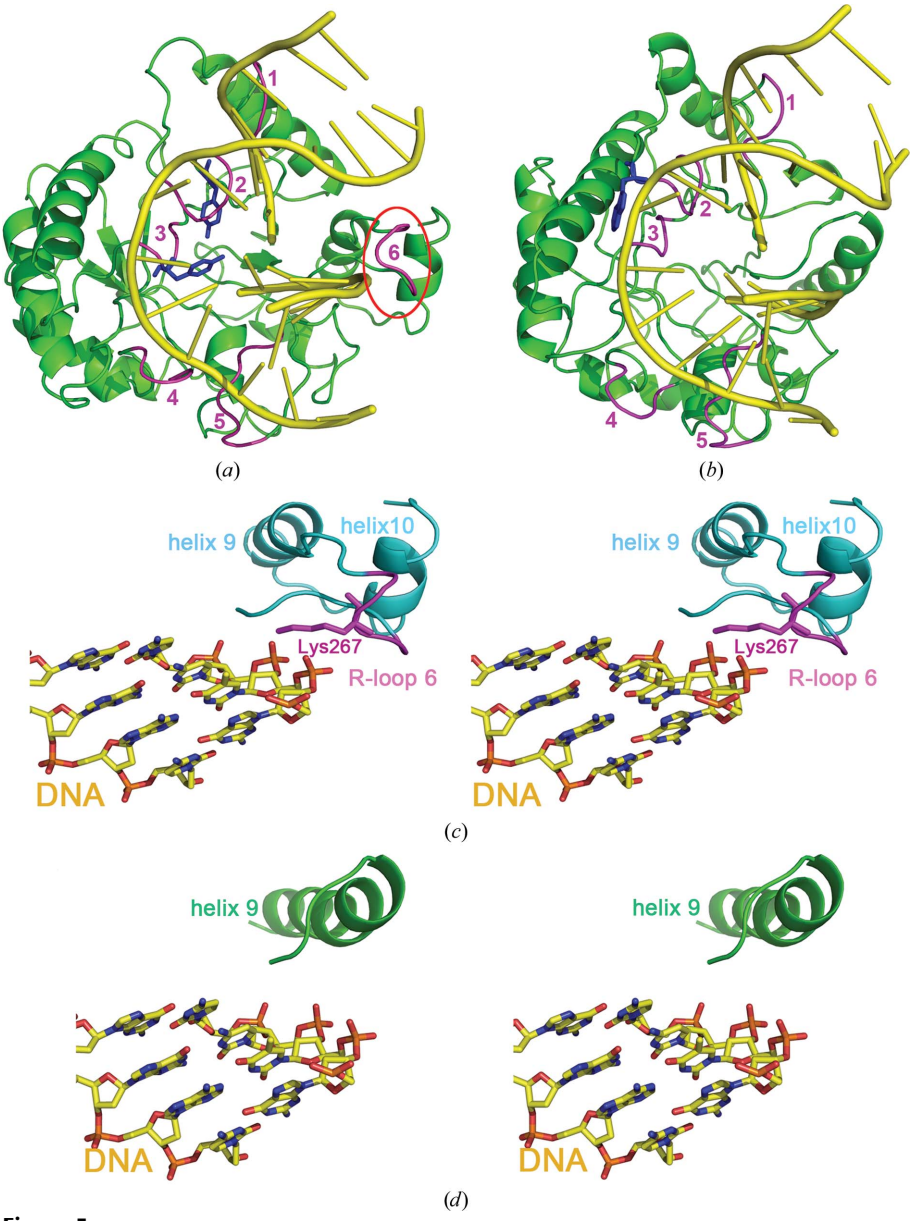

Figure 5
Model structures of the gkEndoIV–DNA complex (*a*) and the ttEndoIV–DNA complex (*b*). DNA is shown in yellow. R-loops are indicated by magenta loops with numbers. The red circle indicates R-loop 6. In (*a*) Tyr34 and Tyr72 are indicated by the upper and lower blue sticks, respectively. In (*b*) Trp41 is shown as blue sticks. (*c*, *d*) Expanded stereoviews around the C-terminus including R-loop 6 in (*a*) and (*b*), respectively. Lys267 on R-loop 6 is shown as magenta sticks.

3.4. Lys267 in R-loop 6

In order to verify our hypothesis that R-loop 6 is involved in the tight binding of gkEndoIV to DNA, we constructed the K267A gkEndoIV mutant and measured its affinity for double-stranded DNA containing an AP-site analogue. Compared with wild-type gkEndoIV, the affinity of K267A gkEndoIV for the substrate analogue was drastically decreased: the dissociation constant of the mutant was determined to be 0.38 μM (filled circles in Fig. 2). This value was at least 40-fold larger than that of the wild type. This result suggested that Lys267 in the R-loop 6 of gkEndoIV directly interacts with DNA. Since the affinity of K267A gkEndoIV for DNA was comparable with that of ttEndoIV, it is possible that this residue contributes to the higher DNA-binding affinity of gkEndoIV. Verification of this hypothesis requires the crystal structure of the gkEndoIV–DNA complex. Lysine residues corresponding to Lys267 of gkEndoIV are conserved in the sequences of EndoIVs from Firmicutes.

In conclusion, we have demonstrated that the C-terminal extension in gkEn-

doIV contributes to its tight binding to substrate DNA. This conclusion can be extended to the long-type EndoIV. It remains unclear why the long-type EndoIV binds more tightly to DNA than the short-type. EndoIV is known to be associated with other proteins in DNA-repair processes (Friedberg *et al.*, 2006). Firmicutes may require different cooperation partners for these processes.

We thank Michiyo Takahara for her help in the purification of gkEndoIV, Toshi Arima for her help in gkEndoIV crystallization, Yuka Nonaka for her help in data collection from gkEndoIV crystals and Ryoichi Arai for providing the pET-HisTEV3 vector. This work was supported by Grants-in-Aid for Scientific Research 17770089 (to NN) and 20570131 (to RM) from the Ministry of Education, Science, Sports and Culture of Japan.

References

Barzilay, G. & Hickson, I. D. (1995). *Bioessays*, **17**, 713–719.
Bertani, G. (1951). *J. Bacteriol.* **62**, 293–300.
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.
Chédin, F. & Kowalczykowski, S. C. (2002). *Mol. Microbiol.* **43**, 823–834.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
Demple, B. & Sung, J.-S. (2005). *DNA Repair*, **4**, 1442–1449.
Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A. & Ellenberger, T. (2006). *DNA Repair and Mutagenesis*, 2nd ed., pp. 9–69. Washington: ASM Press.
Garcin, E. D., Hosfield, D. J., Desai, S. A., Haas, B. J., Björas, M., Cunningham, R. P. & Tainer, J. A. (2008). *Nature Struct. Mol. Biol.* **15**, 515–522.
Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P. & Tainer, J. A. (1999). *Cell*, **98**, 397–408.
Iino, H., Naitow, H., Nakamura, Y., Nakagawa, N., Agari, Y., Kanagawa, M., Ebihara, A., Shinkai, A., Sugahara, M., Miyano, M., Kamiya, N., Yokoyama, S., Hirotsu, K. & Kuramitsu, S. (2008). *Acta Cryst.* F**64**, 487–491.
Kunz, B. A., Henson, E. S., Roche, H., Ramotar, D., Nunoshiba, D. & Demple, B. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 8165–8169.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Levin, J. D., Johnson, A. W. & Demple, B. (1988). *J. Biol. Chem.* **263**, 8066–8071.
McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Seeberg, E., Eide, L. & Bjoras, M. (1995). *Trends Biochem. Sci.* **20**, 391–397.
Shida, T., Ogawa, T., Ogasawara, N. & Sekiguchi, J. (1999). *Biosci. Biotechnol. Biochem.* **63**, 1528–1534.
Ueno, G., Kanda, H., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2006). *J. Struct. Funct. Genomics*, **7**, 15–22.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Watanabe, R., Masui, R., Mikawa, T., Takamatsu, S., Kato, R. & Kuramitsu, S. (1994). *J. Biochem.* **116**, 960–966.
Yajko, D. M. & Weiss, B. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 688–692.